БАКАЛАВРИАТ

В.А. Малугин, Л.Н. Фадеева

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

УЧЕБНИК

ВЫСШЕЕ ОБРАЗОВАНИЕ – БАКАЛАВРИАТ

серия основана в 1996 г.

В.А. Малугин
Л.Н. Фадеева

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

УЧЕБНИК

Допущено Учебно-методическим объединением по классическому университетскому образованию в качестве учебника для студентов высших учебных заведений, обучающихся по направлению подготовки 080100 «Экономика»

Электронно-Библиотечная Система
znanium.com

Соответствует Федеральному государственному образовательному стандарту 3-го поколения

Москва
ИНФРА-М
2014

УДК 519.2(075.8)
ББК 22.172я73
М18

ФЗ № 436-ФЗ	Издание не подлежит маркировке в соответствии с п. 1 ч. 1 ст. 11

Рецензенты:

Алехин Е.И. — канд. физ.-мат. наук, доцент, зав. кафедрой математического и информационного анализа экономических процессов ГОУ ВПО «Орловский государственный университет»;

Максимов В.П. — д-р физ.-мат. наук, профессор кафедры информационных систем и математических методов в экономике Пермского государственного университета

Об авторах:

Малугин В.А. — кандидат физико-математических наук, доцент кафедры «Математические методы анализа экономики» экономического факультета МГУ им. М.В. Ломоносова;

Фадеева Л.Н. — кандидат физико-математических наук, доцент кафедры «Математические методы анализа экономики» экономического факультета МГУ им. М.В. Ломоносова

Малугин В.А., Фадеева Л.Н.

М18 Количественный анализ в экономике и менеджменте: Учебник. — М.: ИНФРА-М, 2014. — 615 с. + Доп. материалы [электронный ресурс; режим доступа http://www.znanium.com]. — (Высшее образование: Бакалавриат). — DOI 10.12737/758 (www. doi.org).

ISBN 978-5-16-004832-1

В книге рассмотрены основные разделы математики — как элементарной, так и высшей, включая теорию вероятностей и математическую статистику. Математический инструментарий затем использован в экономических приложениях (эконометрика, микроэкономика). Полученные знания применены в создании моделей экономики и управления и изучении методов их анализа, в решении задач по оптимизации.

Материал снабжен большим количеством примеров и иллюстраций, а также задачами для самостоятельного решения. Теоретический и практический курс количественного анализа рассчитан на два учебных года.

Для студентов и преподавателей вузов и факультетов экономического профиля.

ББК 22.172я73

Материалы, отмеченные знаком, доступны в электронно-библиотечной системе znanium (www.znanium.com)

ISBN 978-5-16-004832-1

ПРЕДИСЛОВИЕ

Антицитата: «высшая математика убивает креативность».

А.А. Фурсенко,
министр образования и науки РФ,
о преподавании математики в школе
(РИА «Новости», 12 февр. 2009)

Математический курс с приложениями в экономике и менеджменте читается студентам по семестрам последовательно в течение двух лет. Он написан преподавателями, ряд лет ведущими занятия по соответствующим дисциплинам на отделении экономики и менеджмента. По нему старательный студент самостоятельно *реально* может научиться основам количественного анализа.

Цель, которую ставит перед собой коллектив преподавателей, не сводится к тому, чтобы дать студенту сумму знаний. Цель в другом. Знания со временем устаревают, технологии меняются. Что же остается у студента, когда-то закончившего университет?

В течение всех лет преподавания в университете идет процесс интеллектуального воспитания и развития студента. Ему прививается научный подход к решению проблем экономики и управления. Достигается это на основе материала книги в процессе живого общения на лекциях и семинарах. Студент обучается логическому мышлению, применению системного подхода, использованию современных достижений науки в своей деятельности. Кроме того, он начинает понимать, как надо учиться, чтобы достичь успеха, и что надо учиться всегда. Научный метод, плюс умение логически мыслить, плюс фундаментальные знания вооружают его на всю жизнь. В этом мы видим достоинство университетского образования в науке экономики и управления.

В последние десятилетия научно обусловленный подход в управлении приобретает все большее значение в жизни людей. Слабое руководство предприятием или фирмой приводит к банкротству. Непродуманные действия по управлению экономикой регионов и государств приводят к провалам в экономической политике и понижают жизненный уровень людей. Отсутствие внимания к научным достижениям и научно обоснованным прогнозам развития человечества в области экологии, политологии, социологии и т.д. со стороны надгосударственных организаций (ООН, ЕС и др.) приводит к развитию «болезней» человечества, лечение которых потом дорого будет обходиться потомкам.

Считается, что «фундаментальность университетской подготовки требует знаний не только специальных, но и общих» в сфере бизнеса и управления, причем к общим относят язык, физкультуру и математику. На наш взгляд, изучение математики с ее приложениями в области экономики и управления для управленческих специальностей играет иную, неизмеримо более важную роль.

Научная мысль — это не поток ощущений, эмоций или аморфного сознания, это логическая цепочка рассуждений и вычислений, в конце которой, как правило, содержится ответ на поставленный вопрос. Умение сформулировать проблему, построить научную модель развития исследуемого объекта, найти и просчитать решение, правильно его интерпретировать — прерогатива высоко научно-образованных людей. Расцвет в большинстве научных отраслей в значительной степени определяется тем, насколько глубоко математика проникла в эти отрасли, насколько серьезно исследователи используют математический аппарат в своей работе.

Ближе других по роду деятельности к решению проблем бизнеса и управления стоят специалисты по политологии, по государственному управлению, по менеджменту. Именно они в первую очередь должны обладать логическим мышлением, применять научный метод к сложным проблемам. Ни одна наука не учит так эффективно умению размышлять, из причин выводить следствия, обосновывать свои выводы, как математика. Четкие законы математики исключают расплывчатость, туманность, многозначность суждений. Если хочешь научиться *логически мыслить*, решай задачи по математике. Еще один важный аргумент в пользу занятий математикой — *известный ответ*. Если исследователь нашел идею, верно использовал формулы, не сделал ошибок в расчетах, он получит правильный ответ. Правильными ли были его действия при решении проблем в экономике, бизнесе, управлении и т.д., можно узнать спустя длительное время. Каждое неверное решение на рабочем месте — это не только чувство неудовлетворенности, это еще и материальные и финансовые потери: как минимум — исследователя, как максимум — целого государства.

Математика не сводится к решению уравнений или нахождению интегралов. На математическом языке строятся модели экономических, социологических, политологических проблем. Разработаны методы их решения. Если аккуратно с учетом основных особенностей построена, например, экономико-математическая модель, то, варьируя ее параметры и анализируя решение, можно разобраться в том, какие действия следует предпринять для решения реальной экономической проблемы. Перед принятием управленческого решения нужно проиграть ситуацию на модели.

Данный учебник дает будущему специалисту в области бизнеса и управления возможность развить в себе навыки логического мышления даже в том случае, если его знания далее четырех действий арифметики не распространяются. Познакомившись с основными математическими идеями и инструментами, он затем научится строить математические модели, анализировать их, а следовательно, с бо́льшим успехом находить правильные решения. Структура книги построена так, что, разобравшись с элементарной математикой и познакомившись с идеями высшей, будущий специалист выходит на экономико-математические и управленческие задачи, размещенные во второй части — «Количественные методы в экономике и управлении». Одним из основных разделов здесь является теоретический курс *исследования операций*, т.е. применение научного подхода к сложным проблемам, возникающим в управлении сообществами людей, совокупностью машин, финансовыми потоками. Исследование операций — раздел на стыке математики, экономики и управления, предназначенный дать в руки менеджеру всю мощь науки для нахождения идей разрешения проблем и реализации этих идей наилучшим образом. Исследование операций позволяет обосновывать и использовать методы математического моделирования различных процессов в экономике, социологии, управлении, психологии, строить экономико-математические модели, подвергнув их анализу, выбирать оптимальные решения.

Материал учебника преподносится в доступной форме, отличается системным подходом к изложению. Лекции сопровождаются достаточно большим количеством задач и упражнений с ответами. Математическая теория и специально подобранные задачи в одной книге позволяют легко сочетать лекции и семинары в процессе обучения. Часть задач решена, но большая часть предложена для самостоятельного решения или разбора на семинарских занятиях. Реализованная в книге идея объединить математическую и экономическую теорию, а также сборник задач и упражнений важна для успешной работы преподавателя и студента в условиях новых стандартов высшего образования третьего поколения, предполагающих наличие в учебном плане как аудиторных часов, обязательно включающих «контактную» часть, так и времени на самостоятельную работу.

Особенностью книги является краткий курс по элементарной математике, предваряющий разделы высшей математики. Он позволяет школьникам подготовиться к изучению высшей математики, а студентам — повторить или вновь освоить некоторые разделы школьной математики, без которой восприятие высшей невозможно. Книга, таким образом, содержит весь математический аппарат, необходимый будущему предпринимателю или менеджеру при получении им высшего образования, начиная со школьной алгебры и заканчивая теорией оптимизации в управлении.

Значительная часть материала учебника посвящена классическим методам теории оптимизации и включает в себя такие важные части, как метод Лагранжа, используемый при нахождении условного экстремума, симплекс-метод и др. На этой основе рассматриваются важные теоретические экономико-математические оптимизационные задачи, ориентирующие обучающегося не только на качественное, но и на количественное обоснование принимаемых решений по оптимизации.

Количественный анализ для экономических и управленческих специальностей — развивающийся предмет. В учебный курс в силу недостаточного количества часов не включены многие разделы, полезные экономисту и управленцу в их практической деятельности. И тем не менее, авторы надеются, что материал, изложенный в книге, будет полезен также обучающимся по родственным специальностям:

080105 «Финансы и кредит»;

080111 «Маркетинг»;

080109 «Бухгалтерский учет и аудит»;

080301 «Коммерция, торговое дело»;

080502 «Экономика и управление на предприятии»;

080504 «Государственное и муниципальное управление»;

080505 «Управление персоналом»;

080507 «Менеджмент организации»;

080700 «Бизнес-информатика»;

080801 «Прикладная информатика в экономике».

Основной единицей деления учебника является раздел, посвященный, как правило, одной дисциплине. В каждом разделе рассматриваются базовые понятия, без которых не возникает понимания дисциплины в ее целостности. Затем изучается материал, менее значимый для самой дисциплины, но существенный для дальнейшего изложения количественных методов в экономике и управлении. Все эти разделы в той или иной степени читались учащимся.

Раздел 1, 2, 3 — канд. физ.-мат. наук, доц. Малугин В.А.; раздел 4, 5 — канд. физ.-мат. наук, доц. Фадеева Л.Н; раздел 6 — канд. физ.-мат. наук, доц. Малугин В.А. (глава 23); канд. физ.-мат. наук, доц. Фадеева Л.Н. (глава 24, п. 1, 2, 3); доц. Лебедев А.В. (глава 24, п. 4); раздел 7 — канд. физ.-мат. наук, доц. Малугин В.А. (главы 26, 27, 28, 30, 31); канд. экон. наук, доц. Павлова Л.С., канд. экон. наук, доц. Слепак Б.Э. (глава 29).

Часть I
МАТЕМАТИЧЕСКИЕ ОСНОВЫ ЭКОНОМИЧЕСКОГО АНАЛИЗА

Раздел I
ОСНОВНЫЕ ПОНЯТИЯ ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

Глава 1
АРИФМЕТИКА И АЛГЕБРА

1.1. ОЧЕРК РАЗВИТИЯ МАТЕМАТИКИ

Источники и движущие силы зарождения и первых шагов математики

Слово «математика» происходит от греческого *mathematike*, что значит знание, наука. В древности это была наука о числах (арифметика) и наука о формах на плоскости (геометрия). Математические знания независимо и параллельно возникали у разных народов.

Как арифметика, так и геометрия выросли из потребностей *обыденной жизни*, были обусловлены запросами хозяйствования, сводившимися к счету предметов, измерению количества продуктов, площадей земельных участков, определению размеров отдельных частей архитектурных сооружений, измерению времени, коммерческим расчетам, навигации. Развивающаяся торговля требовала количественного выражения в торговых операциях. Землемерные работы при продаже или покупке земли с возвращением рек в свои русла после разливов нуждались в геометрических построениях и вычислениях.

Другой движущей силой являлось осмысление действительности, *философские исследования*. Они требовали абстрактных логически обусловленных построений. Происходила формализация философских рассуждений. Математические рассуждения отходили от конкретных объектов. Исследование абстрактных форм приводило к новому математическому знанию, которое затем возвращалось к материальной действительности в виде умения решать практические задачи. Математика выделилась из науки *философии* при ее формализации и соединилась с простым счетом и геометрическими построениями предприимчивых и образованных людей.

Параллельно с развитием элементов математики проводились *астрономические исследования.* Наблюдения за движением звезд требовали количественных отношений. Астрономия являлась первой научной областью, которая предъявила математике свои особые и очень большие требования. Это привело к раннему развитию тригонометрии.

Развитие различных разделов математики в значительной мере подогревалось *любознательностью* образованных представителей общества. Математика, осознавшая себя как научное направление, начала ставить перед собой задачи, которые, казалось, уже не имели непосредственного практического применения. Во времена Пифагора начала развиваться, например, теория чисел. Религиозное мировоззрение Древнего мира наложило свой отпечаток на эти ростки. Числовым соотношениям приписывались магические значения. Одни числа считались расположенными к людям, другие несли негативный мистический оттенок, приносили несчастья, их боялись и старались не употреблять в счете.

Наибольшего развития *элементы математики* достигли в древней Греции к VI–V вв. до н. э. В результате накопления достаточно большого фактического материала, суммирования и осмысления большого количества разрозненных знаний складывалось понимание математики как самостоятельной науки — науки *элементарной математики*, содержащей в себе арифметику, теорию чисел, начала геометрии и алгебры. Само слово «алгебра» появилось много позже. Основы математической науки закладывались как систематическое и логически последовательное построение.

Потребности измерения (количества зерна, длины дороги и т.д.) приводят к появлению названий и обозначений простейших дробных чисел и к разработке приемов выполнения арифметических действий над дробями. Появились первые математические доказательства. Математика, как и все научное и художественное творчество, перестала быть безымянной. Она создается теперь известными по именам математиками, оставившими после себя математические сочинения. К сожалению, многие из них не дошли до нас даже в отрывках.

Геометрия Евклида — первая естественно-научная теория

В V–III вв. до н.э. научная жизнь сначала сосредоточивается в Афинах, через два столетия после завоеваний Александра Македонского перемещается в Александрию. Нам известны имена первых греческих геометров и философов Фалеса Милетского и Пифагора Самосского. Значительный вклад в развитие геометрии сделали Евклид и Архимед.

В Афинах, а особенно в Александрии, куда стекались образованные представители эллинистического общества, где объединялись различные мировые культуры, греческая математика достигла своего высшего расцвета. Этому способствовало невиданное ранее по своей широте государственное покровительство науке. Александрийские библиотеки обладали огромной притягательной силой.

Наиболее значительным сочинением этого времени из дошедших до нас в копиях и комментариях поздних авторов, явилось сочинение «Начала» Евклида. Оно не было энциклопедией математических знаний своей эпохи, поскольку многие математические достижения того времени в нем отсутствовали. Но другие разделы были изложены столь подробно и последовательно, что на многие века стали образцом математического учебника, по которому учились студенты, которым как справочником пользовались ученые и вообще образованные люди. «Начала» Евклида посвящены элементарной геометрии, методам нахождения площадей и объемов, теории чисел, алгебре и некоторым другим разделам. Евклид подвел итог трехсотлетнему развитию греческой математики. «Начала» построены по так называемой дедуктивной системе. Сначала приводятся определения, постулаты и аксиомы, затем теоремы. Постулаты посвящены возможности выполнения некоторых геометрических построений, аксиомы вводят некоторые предложения, касающиеся свойств отношений равенства и неравенства между величинами. Первый постулат, например, утверждает, что «от всякой точки до всякой точки можно провести прямую линию», а первая аксиома говорит: «Равные одному и тому же равны между собой». Затем формулируются теоремы, одна из которых — теорема Пифагора. Приводятся их доказательства, которые служат образцом научной строгости. Метод изложения материала, основательность и строгость в изложении привели к тому, что геометрия Евклида признана первой естественно-научной теорией.

Основные этапы развития математики

После александрийской эпохи расцвета, увлеченности самых широких слоев образованного общества математикой продолжался рост объема научных знаний в этой области. Однако математическая наука уже не достигала прежней цельности и глубины, оставаясь уделом отдельных ученых. Последующие столетия были эпохой возрастающего влияния религии в жизни общества. Вера и научное знание сталкивались. Ученых преследовала инквизиция. Боролись между собой научные школы. Например, неясно было, как строить фигуры циркулем и линейкой, если существуют несоизмеримые отрезки. В связи с этим еще в III в. до н. э. Аристотелем был наложен запрет на применение арифметики к геометрии. Только открытие иррациональных чисел позволило разобраться в этой проблеме.

Центр развития наук, в том числе и математики, переместился на Восток. В Индии в V–XII вв. развитие математики привело к созданию десятичной системы счисления, к употреблению нуля для обозначения отсутствия единиц данного разряда. Была создана алгебра, оперирующая не только дробями, но и иррациональными и отрицательными числами.

В Средней Азии и на Ближнем Востоке арабские завоевания, объединение огромных территорий под властью арабских халифов привели к тому, что в течение IX—XV вв. ученые Средней Азии, Ближнего Востока и Пиренейского п-ова пользовались арабским языком. Наука здесь развивается в мировых торговых городах и при государственной поддержке. В первой половине IX в. арабский математик Хорезми впервые дал изложение алгебры как самостоятельной науки. Термин «алгебра» происходит от начала названия сочинения Хорезми «Аль-джебр», по которому европейские математики раннего Средневековья познакомились, например, с решением квадратных уравнений. В связи с астрономическими и геодезическими работами большое развитие получила тригонометрия. Выходец из Сирии Аль-Баттани ввел в употребление первые тригонометрические функции синус и тангенс. Блестящим завершением этой эпохи явилась в XV в. деятельность Улугбека, который при своем дворе и обсерватории в Самарканде собрал более 100 ученых и организовал долго остававшиеся непревзойденными астрономические наблюдения, вычисление математических таблиц и т.д.

Математический анализ — качественный скачок в развитии

С XVII в. начинается новый период развития математики. Центр развития наук опять перемещается в Европу. Поворотным пунктом в математике была введенная Р. Декартом *переменная величина*. Количественные отношения стали рассматриваться в процессе их изменения, и стали изучаться сами эти изменения. Была создана *декартова система координат*, на которой показывались изменения одной величины в зависимости от другой. На первый план выдвигается понятие *функции*, играющее в дальнейшем такую же роль основного и самостоятельного предмета изучения, как ранее понятия величины или числа. Изучение переменных величин и функциональных зависимостей приводит далее к идеям бесконечно малой и бесконечно большой величин. Возникают понятия *предела, производной, дифференциала и интеграла* — основные понятия новой математической науки — математического анализа. Анализ бесконечно малых величин, изучение свойств новых понятий формируют разделы математического анализа — дифференциальное и интегральное исчисления. Созданием в XVII в. новой математики переменных величин — математическому анализу — в первую очередь мы обязаны И. Ньютону и Г. Лейбницу. К математическим достижениям XVII в. следует отнести также открытие логарифмов Дж. Непером.

В Западной Европе к XIX в. создается современный стиль математических исследований, Полученные результаты не рассылаются с оказией знакомым коллегам, а оформляются в виде статей и публикуются в периодических изданиях, вскоре приобретших вид научных журналов. Новые знания быстро становятся достоянием всего мира и используются в исследованиях других ученых, не знакомых между собой. К изучению математики привлекаются талантливые молодые люди из разных кругов общества. Леонард Эйлер — сын пастора — в 23 года становится профессором Петербургской академии наук, Ж. Лагранж из семьи французского чиновника в 19 лет — профессор в Турине, П. Лаплас, сын французского крестьянина, в 22 года — профессор военной школы в Париже, в 36 лет — член Парижской академии наук.

XIX и XX вв. отмечены взрывным интересом к математике. Она расширяет свои разделы, появляются и развиваются новые математические отрасли: топология, теория функций, функциональный анализ, дифференциальные уравнения, уравнения математической физики, теория вероятностей, математическая статистика, вычислительная математика. Быстрый прогресс вычислительной техники в последние десятилетия приводит к перемещению основных усилий математиков внутри этих сложившихся разделов математики к появлению целого ряда новых математических дисциплин. Среди них теория алгоритмов и программирование, теория информации, теория игр, исследование операций, кибернетика. Вопросы о наилучшем управлении физическими или механическими системами, оптимальном управлении экономическими объектами привели к созданию математической теории оптимального управления. Исследования в области проблем управления в соединении с прогрессом вычислительной техники дают основу для автоматизации и оптимального управления все новыми сферами человеческой деятельности. Эта книга познакомит будущих менеджеров с основными достижениями в области математики, включая элементы теории классической оптимизации, и покажет, как математические достижения используются в различных экономических областях и, в частности, в управлении экономическими процессами.

1.2. ЭЛЕМЕНТЫ АРИФМЕТИКИ

Числа и выражения

Напомним некоторые арифметические суждения или приемы, которые будут активно использоваться в дальнейшем.

Квадраты некоторых чисел. Составим наиболее часто встречающиеся квадраты чисел:

$11^2 = 121,\qquad 16^2 = 256,$

$12^2 = 144,\qquad 17^2 = 289,$

$13^2 = 169,\qquad 18^2 = 324,$

$14^2 = 196,\qquad 19^2 = 361,$

$15^2 = 225,\qquad 21^2 = 441.$

Надо постараться их запомнить.

Квадраты чисел, оканчивающихся на 5

Существует простое правило, позволяющее вычислить подобные квадраты: число, стоящее перед пятью, надо умножить на число, на единицу большее, и справа приписать 25.

Пример 1.1. Возвести в квадрат число 35.

Решение. Для того чтобы возвести в квадрат число 35, следует число 3 умножить на число 4 и приписать справа число 25. Получим: $35^2 = 1225$.

Пример 1.2. Возвести в квадрат число 115.

Решение. Для возведения в квадрат числа 115 надо число 11 умножить на число 12 и приписать справа число 25. Получим $115^2 = 13\,225$.

Числовые оценки

При решении уравнений или неравенств часто приходится сравнивать числа между собой. Простейшая схема сравнения заключается в выдвижении некоторого предположения и доказательстве его истинности.

Пример 1.3. Что больше — $\frac{3+2\sqrt{2}}{6}$ или 1?

Решение. Предположим, что $\frac{3+2\sqrt{2}}{6} < 1$. Преобразуем числовое неравенство к очевидно верному или к очевидно неверному неравенству. Умножим обе части неравенства на 6 и перенесем число 3 вправо. Получим $2\sqrt{2} < 3$. Возведем обе части в квадрат $8 < 9$. Неравенство верно. Прежде чем сделать вывод об истинности нашего предположения, мы должны проверить: действительно ли из неравенства $8 < 9$ можно получить неравенство $\frac{3+2\sqrt{2}}{6} < 1$. Понятно, что если из утверждения A следует утверждение B: $A \Rightarrow B$, то это вовсе не означает, что из B следует только A: $B \Rightarrow A$. Убедившись, что обратный ход приводит к неравенству $\frac{3+2\sqrt{2}}{6} < 1$, можем сделать вывод, что наше предположение оказалось правильным.

Рациональные и иррациональные числа

Произвольные числа, с которыми имеет дело математика, в общем случае называются комплексными. Их вид $z = a + bi$, где a и b — привычные для нас положительные, отрицательные числа или нуль, а $i = \sqrt{-1}$. При $b = 0$ получаем частный случай $z = a$. Эти числа называются *действительными*. Действительные числа делятся на рациональные и иррациональные.

Определение. Все действительные числа, которые можно представить в виде $\frac{p}{q}$, где p и q — целые, а $q \neq 0$, называются *рациональными*. Все остальные действительные числа *иррациональны*.

Примеры рациональных чисел: $2; \frac{1}{3}; 0{,}5; \sqrt{4} = 2; \sin\frac{\pi}{6} = \frac{1}{2}; \ldots$.

Примеры иррациональных чисел: $\sqrt{2}; \pi; \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}; \ldots$.

Количества рациональных чисел и иррациональных чисел бесконечны. Любое действительное число может быть либо рациональным, либо иррациональным. В дальнейшем будут использованы некоторые преобразования с иррациональными числами.

Пример 1.4. Упростить $\sqrt{2\sqrt{3}+4}-\sqrt{3}$.

Решение. Преобразуем подкоренное выражение в первом слагаемом к полному квадрату и извлечем корень:

$$\sqrt{2\sqrt{3}+4}-\sqrt{3}=\sqrt{3+2\sqrt{3}+1}-\sqrt{3}=\sqrt{(\sqrt{3}+1)^2}-\sqrt{3}=$$
$$=\sqrt{3}+1-\sqrt{3}=1.$$

1.3. ЭЛЕМЕНТЫ АЛГЕБРЫ. АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ И НЕРАВЕНСТВА

Элементы алгебры

Формулы сокращенного умножения

Приведем основные формулы сокращенного умножения, которые надо очень хорошо знать:

1) $(a\pm b)^2=a^2\pm 2ab+b^2$;
2) $(a\pm b)^3=a^3\pm 3a^2b+3ab^2\pm b^3$;
3) $a^2-b^2=(a-b)(a+b)$;
4) $a^3\pm b^3=(a\pm b)(a^2\mp ab+b^2)$.

Эти четыре формулы лежат в основе всех алгебраических преобразований.

Пример 1.5. Разложить на множители выражение a^6-b^6.

Решение. Представим выражение как разность кубов и разложим по формуле 4:

$$a^6-b^6=(a^2)^3-(b^2)^3=$$
$$=(a^2-b^2)(a^4+a^2b^2+b^4)(a-b)(a+b)(a^4+a^2b^2+b^4).$$

Можно ли разложить выражение $a^4+a^2b^2+b^4$? Трудно сказать сразу.

Представим исходное выражение как разность квадратов:

$$a^6-b^6=(a^3)^2-(b^3)^2=(a^3-b^3)(a^3+b^3).$$

Продолжим разложение разности и суммы кубов:

$$(a^3-b^3)(a^3+b^3)=(a-b)(a+b)(a^2-ab+b^2)(a^2+ab+b^2).$$

Очевидно, что дальше разложить невозможно. Из сравнения разложений также становится ясно что

$$a^4+a^2b^2+b^4=(a^2-ab+b^2)(a^2+ab+b^2).$$

Итак,

$$a^6-b^6=(a-b)(a+b)(a^2-ab+b^2)(a^2+ab+b^2).$$

Деление многочлена на многочлен

Умение делить один многочлен на другой будет активно использоваться при решении уравнений высоких степеней.

Пример 1.6. Разделить многочлен $2x^4 - 3x^3 + 2x^2 - 15x + 14$ на двучлен $x - 1$.

Решение. Проведем деление уголком и объясним процесс деления:

$$
\begin{array}{ll}
\text{1-я строка} & 2x^4 - 3x^3 + 2x^2 - 15x + 14 \,\underline{|\, x - 1} \\
 & - \\
\text{2-я строка} & \underline{2x^4 - 2x^3} \\
\text{3-я строка} & 0 \quad - x^3 + 2x^2 \\
 & - \\
\text{4-я строка} & \quad \underline{- x^3 + x^2} \\
\text{5-я строка} & \qquad 0 \quad x^2 - 15x \\
 & \qquad - \\
\text{6-я строка} & \qquad\quad \underline{x^2 - x} \\
\text{7-я строка} & \qquad\quad 0 - 14x + 14 \\
 & \qquad\qquad - \\
\text{8-я строка} & \qquad\qquad \underline{- 14x + 14} \\
\text{9-я строка} & \qquad\qquad\quad 0
\end{array}
$$

Шаг 1. Подберем в частном такое выражение, которое, будучи умножено на первое слагаемое делителя, т.е. на x, даст первое слагаемое делимого, т.е. $2x^4$. Это выражение есть $2x^3$. Умножим $2x^3$ на делитель $x - 1$ и, получив $2x^4 - 2x^3$, напишем его второй строчкой в левом верхнем углу. Произведем вычитание второй строки из первой строки, у которой задействуем только два первых слагаемых. Получим $-x^3$. Снесем в эту же строку из первой строки третье слагаемое $2x^2$. Имеем $-x^3 + 2x^2$.

Шаг 2. К выражению в частном припишем с определенным знаком такое слагаемое, которое, будучи умножено на x, даст первое слагаемое третьей строки, т.е. $-x^3$. Это слагаемое есть $-x^2$. Умножим $-x^2$ на $x - 1$ и напишем четвертой строкой. Произведя вычитание четвертой строки из третьей, будем иметь x^2. Снесем в эту строку четвертое слагаемое из первой строки. Получим $x^2 - 15x$.

Шаг 3. Теперь в частном приписываем $+x$, которое умножаем на $x - 1$. Результат помещаем в шестую строку. Разность между пятой и шестой строками дает $-14x + 14$.

Шаг 4. В частном приписываем -14 и помножаем на $x - 1$, результат записывая в восьмой строке. Разность между седьмой и восьмой строками дает нуль. Деление закончено.

Модуль действительного числа

Абсолютной величиной или модулем действительного числа x называется:

- само это число, если $x > 0$;
- число, ему противоположное, если $x < 0$;
- нуль, если $x = 0$.

Общая форма записи модуля $|x| = \begin{cases} x, & x \geq 0, \\ -x, & x < 0. \end{cases}$

Третья строка отсутствует, знак равенства может быть добавлен либо в первую, либо во вторую строку, либо в обе вместе.

Из определения модуля следует, что:

1) $|x| \geq 0$ при любом действительном значении x;

2) $|x_1 - x_2| = \begin{cases} x_1 - x_2, & x_1 \geq x_2, \\ x_2 - x_1, & x_2 > x_1. \end{cases}$

Свойства модуля:

1) $|x| = |-x|$; 2) $|xy| = |x||y|$; 3) $\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$; 4) $|x|^2 = x^2$; 5) $|x + y| \leq |x| + |y|$.

Корни

Определение. *Корнем четной степени* $\sqrt[2n]{x}$ из неотрицательного числа x называется неотрицательное число y такое, что $y^{2n} = x$.

Определение. *Корнем нечетной степени* $\sqrt[2n+1]{x}$ из действительного числа x называется действительное число y такое, что $y^{2n+1} = x$.

Руководствуясь определением, можем найти:

$\sqrt{4} = 2$, а не ± 2; $\sqrt{(-2)^2} = 2$; $\sqrt[3]{8} = 2$; $\sqrt[3]{-8} = -2$.

Попробуем найти $\sqrt{x^2}$. Квадратный корень не будет в общем случае равен x, как иногда пишут. Почему? В соответствии с определением корня четной степени мы должны получить неотрицательное число. Если известно, что $x \geq 0$, тогда действительно $\sqrt{x^2} = x$. А если $x < 0$? Тогда справа получим отрицательное число, что противоречит определению четного корня. Чтобы правая часть равенства оставалась положительной при $x < 0$, надо поставить перед x знак минус. Поэтому

$$\sqrt{x^2} = \begin{cases} x \geq 0, & \text{если } x \geq 0, \\ -x > 0, & \text{если } x < 0. \end{cases}$$

Но последняя запись означает, что $\sqrt{x^2} = |x|$.

Используя последнее равенство, можем найти, что

$\sqrt{f^2(x)} = |f(x)|$, $\qquad \sqrt{(1-\sqrt{3})^2} = |1-\sqrt{3}| = \sqrt{3}-1$,

$\sqrt[4]{x^2} = \sqrt{\sqrt{x^2}} = \sqrt{|x|}$ и т.д.

Свойства арифметических корней

Будем считать $a \geq 0$, $b > 0$, n и m — натуральные числа. Тогда:

1) $\sqrt[n]{0} = 0$; 2) $\sqrt[n]{a} = a^{\frac{1}{n}}$; 3) $\sqrt[n]{a^m} = a^{\frac{m}{n}}$; 4) $\sqrt[n]{\sqrt[m]{a}} = \sqrt[mn]{a}$;

5) $\sqrt[n]{a}\sqrt[m]{a} = \sqrt[mn]{a^{m+n}}$; 6) $\sqrt[n]{ab} = \sqrt[n]{a}\sqrt[n]{b}$; 7) $\sqrt[n]{\dfrac{a}{b}} = \dfrac{\sqrt[n]{a}}{\sqrt[n]{b}}$.

З а м е ч а н и е. Если рассматривать случай $a < 0$, $b < 0$, свойства 6 и 7 следует подправить.

При $a < 0$, $b < 0$ и n — четном имеем $\sqrt[n]{ab} = \sqrt[n]{-a}\sqrt[n]{-b}$ и $\sqrt[n]{\dfrac{a}{b}} = \dfrac{\sqrt[n]{-a}}{\sqrt[n]{-b}}$.

Понятие о функции одной переменной

О п р е д е л е н и е. Если каждому значению величины x из множества X соответствует одно определенное значение другой величины y из множества Y, то величина y называется *функцией* величины x и записывается в виде $y = f(x)$ или $y = y(x)$.

1. Область определения. Множество X называется *областью определения* или областью допустимых значений аргумента (ОДЗ). В простейшем случае это отрезок $[a; b]$, т.е. множество значений x, удовлетворяющих неравенству $a \leq x \leq b$, или интервал $(a; b)$, т.е. множество всех x из неравенства $a < x < b$. Например, функция $y = \sqrt{(1-x)(x-2)}$ определена на множестве $(1-x)(x-2) \geq 0$. ОДЗ: $1 \leq x \leq 2$.

2. Область изменения. Множество Y называется *областью изменения* или *областью значений* функции. Например, функция $y = \sqrt{1-x^2}$ имеет своей областью значений отрезок $0 \leq y \leq 1$.

3. Возрастание и убывание функции. Функция называется *возрастающей* на промежутке, если большему значению аргумента на этом промежутке соответствует большее значение функции, и *убывающей*, если большему значению аргумента соответствует меньшее значение функции. Функция $y = x^2$ является убывающей на промежутке $-\infty < x < 0$ и возрастающей на промежутке $0 < x < +\infty$.

4. Корни (нули функции). Значения аргумента, при которых функция обращается в нуль, называются *корнями* или *нулями* функции. Функция $y = x^2 + 2x - 3$ имеет два корня: $x_1 = -3$; $x_2 = 1$.

5. Промежутки знакопостоянства. Промежутки на оси X, в которых функция имеет определенный знак, называются *промежутками постоянного знака*, или *промежутками знакопостоянства*. Функция $y = x^2 + 2x - 3$ положительна при $x > 1$ или $x < -3$, отрицательна при $-3 < x < 1$.

6. Четность. Функция $y = f(x)$ называется *четной*, если справедливо равенство $f(-x) = f(x)$. Функция называется *нечетной*, если верно $f(-x) = -f(x)$. Она может также не обладать свойством четности. Например, функция $y = x^2$ четная, функция $y = x^3$ нечетная, а функция $y = 2x - 3$ не обладает свойством четности.

7. Периодичность. Если существуют такие постоянные не равные нулю числа, при прибавлении которых к аргументу значение функции не изменяется, такая функция называется *периодической*. Наименьшее положительное из этих чисел называется *периодом*. Функция $y = \sin 2(x + \pi n) = \sin 2x$ периодична. Наименьшее положительное число из чисел вида πn есть число π. Оно по определению называется периодом данной функции.

8. Экстремум. Точка x_0 называется точкой *локального максимума* функции $y = f(x)$, если существует такая двусторонняя окрестность точки x_0, в которой $f(x) < f(x_0)$ при $x \neq x_0$. Точка x_0 называется точкой *локального минимума* функции $y = f(x)$, если существует такая двусторонняя окрестность точки x_0, в которой $f(x) > f(x_0)$ при $x \neq x_0$. Значение функции $f(x_0)$ называется *максимумом* или *минимумом* функции. Максимумы и минимумы функции объединяются понятием *экстремума*. Экстремум называется *строгим*, если он достигается в одной точке, и *нестрогим*, если он достигается на промежутке. Функция $y = x^2 - 2x + 3 = (x - 1)^2 + 2$ имеет строгий минимум $y_{min} = 2$ в точке $x_0 = 1$.

9. Асимптоты. Если при стремлении аргумента x к ∞ функция y стремится к постоянному числу a, то прямая $y = a$ называется *горизонтальной асимптотой*. Если при стремлении x к некоторому числу функия y неограниченно возрастает, то прямая $x = b$ называется *вертикальной асимптотой*.

Квадратный трехчлен

Напомним, что алгебраическое выражение вида $ax^2 + bx + c$ называется *квадратным трехчленом*. Функция $y = ax^2 + bx + c$ называется квадратичной. Решая в общем виде уравнение

$$ax^2 + bx + c = 0,$$

можно найти корни

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

называемые *корнями* квадратного трехчлена. Выражение $D = b^2 - 4ac$ называется *дискриминантом* квадратного трехчлена. Знак дискриминанта определяет наличие корней или их отсутствие.

Если коэффициент b четный, т.е. $b = 2k$, формулу корней можно немного упростить:

$$x_{1,2} = \frac{-k \pm \sqrt{k^2 - ac}}{a}.$$

Эту формулу можно запомнить в другом виде:

$$x_{1,2} = \frac{-\frac{b}{2} \pm \sqrt{\left(-\frac{b}{2}\right)^2 - ac}}{a} = \frac{-\frac{b}{2} \pm \sqrt{\frac{b^2}{4} - ac}}{a}.$$

Сформулируем две основные теоремы для квадратного трехчлена

Теорема 1.1. Если корни квадратного трехчлена существуют и равны x_1 и x_2, то он раскладывается на множители:

$$ax^2 + bx + c = a(x - x_1)(x - x_2).$$

Теорема 1.2 (теорема Виета). Корни квадратного трехчлена связаны с его коэффициентами следующей зависимостью:

$$\begin{cases} x_1 + x_2 = -\dfrac{b}{a}, \\ x_1 \cdot x_2 = \dfrac{c}{a}. \end{cases}$$

График квадратного трехчлена

Функция $y = ax^2 + bx + c$ описывает кривую, называемую *параболой*, общий вид которой в зависимости от знаков коэффициента a и дискриминанта $D = b^2 - 4ac$ представлен на рис. 1.1. Парабола содержит особую точку, называемую *вершиной параболы*. Координаты вершины определяются из соотношений

$$x_0 = -\frac{b}{2a}, \qquad y_0 = -\frac{b^2 - 4ac}{4a}.$$

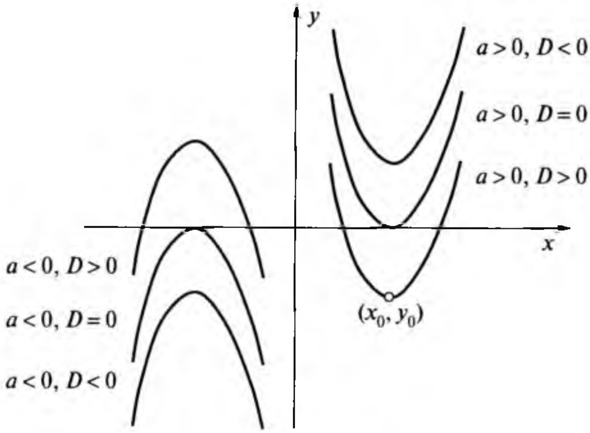

Рис. 1.1

Алгебраические уравнения

Тождество — это равенство, которое выполняется при всех допустимых значениях входящих в него букв. Например, равенство

$$\frac{a^2 - b^2}{a - b} = a + b$$

называется тождеством и справедливо при любых значениях a и b, удовлетворяющих условию $a \neq b$. Иногда, чтобы подчеркнуть тождественность левой и правой частей равенства, будем вместо знака «=» употреблять знак «≡».

Уравнение — это равенство, которое выполняется лишь при некоторых значениях входящих в него букв. Буквы уравнения по условию задачи могут быть неравноправны. Одни из них могут принимать все свои допустимые значения. Они называются *параметрами* уравнения. Значения других требуется найти. Их называют *неизвестными*. Обозначаются параметры первыми буквами латинского алфавита: a, b, c, ...; для неизвестных выбирают последние буквы латинского алфавита x, y, z, ... Примером уравнения с параметрами может служить квадратное уравнение

$$ax^2 + bx + c = 0.$$

Назовем *решениями* значения неизвестных, которые обращают уравнение в тождество. Два уравнения называются *равносильными*, если множества решений этих уравнений совпадают. Преобразование называется *равносильным*, если оно не приводит к потере корней или появлению посторонних.

Уравнения с модулем

Будем считать, что читающий это пособие умеет решать простые линейные и квадратные уравнения. Рассмотрим простейшее уравнение с модулем.

Пример 1.7. Решить уравнение $|x-1|=2$.

Решение. Чтобы решить уравнение, надо избавиться от модуля (раскрыть его). Нам неизвестен знак выражения $x-1$ под модулем, поэтому мы не можем сразу на всей числовой оси это уравнение рассмотреть. Поделим всю числовую ось OX на два промежутка. Первый промежуток получается из неравенства $x-1 \geq 0$, т.е. $x \geq 1$. В этой области значений неизвестной выражение под знаком модуля неотрицательно, модуль раскрывается со знаком «плюс». Получим

$$\begin{cases} x \geq 1, \\ x-1=2. \end{cases}$$

Корень уравнения $x=3$ удовлетворяет условию $x \geq 1$ и, следовательно, является решением задачи. Второй промежуток возникает из неравенства $x-1<0$, т.е. $x<1$. Здесь модуль следует раскрыть со знаком «минус»:

$$\begin{cases} x<1, \\ -(x-1)=2. \end{cases}$$

В качестве корня получим число $x=-1$. Оно удовлетворяет условию $x<1$ и, соответственно, является решением исходной задачи. Итак, *ответ*: $x=-1$, $x=3$.

Замечание. Мы продемонстрировали достаточно общий математический прием: если нельзя решить задачу сразу на всей числовой оси, следует разделить ось на ряд промежутков и на каждом таком промежутке сформулировать и решить конкретную задачу.

Пример 1.8. Решить уравнение $|2x+1|-|3-x|=|x-4|$.

Решение. Для решения этой задачи числовую ось придется разбить на несколько промежутков. Нарисуем одну под другой три числовые оси по числу модулей (рис. 1.2). На первой числовой оси будем заниматься первым модулем, на второй — вторым и т.д. На каждой оси поставим число, обращающее в нуль выражение под модулем, и знаки, с которыми раскрывается соответствующий модуль. На рис. 1.2 указаны четыре интервала, в каждом из которых мы теперь знаем, как раскрыть определенный модуль:

1) $x \geq 4$.

Раскрываем модули в соответствии со знаками на этом промежутке:

$2x+1+3-x=x-4$.

После приведения подобных получим $8=0$. Вывод: на данном промежутке уравнение решений не имеет;

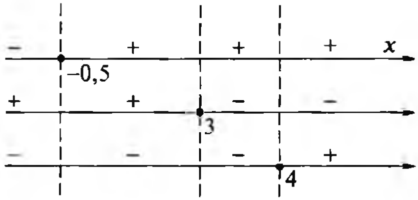

Рис. 1.2

2) $3 \le x < 4$.

Последний модуль раскрывается иначе:

$2x + 1 + 3 - x = 4 - x$

или $x = 0$. Корень уравнения $x = 0$ не попадает в рассматриваемый промежуток $3 \le x < 4$, поэтому является посторонним. Вывод: на этом промежутке также нет решений;

3) $-\frac{1}{2} \le x < 3$.

$2x + 1 - 3 + x = 4 - x$, или $x = \frac{3}{2}$. Корень уравнения удовлетворяет условию $-\frac{1}{2} \le x < 3$. Вывод: $x = \frac{3}{2}$ является решением уравнения;

4) $x < -\frac{1}{2}$.

$-2x - 1 - 3 + x = 4 - x$, или $-4 = 4$. Здесь решений нет.

Ответ: $x = \frac{3}{2}$.

Пример 1.9. Графически решить уравнение (построить на координатной плоскости множество точек (x, y), координаты которых удовлетворяют уравнению) $x = |y|$.

Решение. Раскроем модуль, рассмотрев два случая:

1) $y \ge 0$. Множество точек имеет вид $y = x$;

2) в области $y \le 0$ совокупность точек $y = -x$.

Множество точек построено на рис. 1.3.

Пример 1.10. На координатной плоскости построить множество точек $|x| + x = |y| + y$.

Решение. Начнем с раскрытия модуля:

1) $\begin{cases} x \ge 0, \\ y \ge 0. \end{cases}$ Получим равенство $y = x$;

2) $\begin{cases} x \ge 0, \\ y \le 0. \end{cases}$ Множество точек описывается прямой $x = 0$;

3) $\begin{cases} x \le 0, \\ y \ge 0. \end{cases}$ Множество точек — прямая $y = 0$;

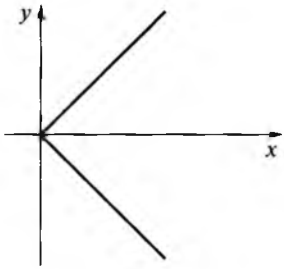

Рис. 1.3

4) $\begin{cases} x \le 0, \\ y \le 0. \end{cases}$ Раскрыв модули, получим тождество $-x + x = -y + y$, или $0 = 0$.

Множеством точек на координатной плоскости является третья четверть

$\begin{cases} x \le 0, \\ y \le 0. \end{cases}$

Решение представлено на рис. 1.4 и содержит все точки луча в первой четверти и все точки третьей четверти, включая границы.

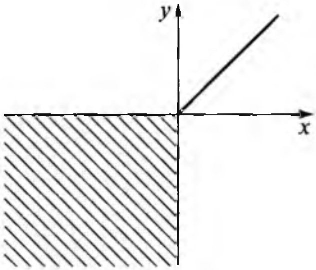

Рис. 1.4

Уравнения высоких степеней

Приведем типичные методы, с помощью которых удается решить абсолютное большинство предлагаемых уравнений.

1. Введение новой переменной с целью понижения степени в уравнении.

2. Метод деления многочлена на двучлен, заключающийся в отыскании одного из корней уравнения x_1 с последующим делением многочлена этого уравнения на двучлен $x - x_1$ или по схеме Горнера.

3. Разложение на множители.

Пример 1.11. Решить уравнение $(x^2 + x + 1)(x^2 + x + 2) - 12 = 0$.

Решение. Равенство представляет уравнение четвертой степени. В этом можно убедиться, раскрыв скобки. Но тогда будет потеряна идея решения, которая легко просматривается и которая заключается в следующем. Обозначим $x^2 + x + 1 = y$. Вторую скобку представим как функцию от y:

$$x^2 + x + 2 = x^2 + x + 1 + 1 = y + 1.$$

Нам удалось в уравнении перейти к новой переменной $y(y + 1) - 12 = 0$, понизив степень уравнения. В полученном квадратном уравнении найдем корни: $y_1 = -4$; $y_2 = 3$. Следовательно, в равенстве $x^2 + x + 1 = y$ величина y может принимать лишь значения –4 и 3. Решаем два уравнения:

$$\left[\begin{aligned} &x^2 + x + 1 = -4, \\ &x^2 + x + 1 = 3. \end{aligned}\right.$$

Первое из уравнений не имеет решений в области действительных чисел. Корни второго: $x_1 = -2$; $x_2 = 1$. Ответ: $x = -2$, $x = 1$.

Пример 1.12. Решить уравнение $x^3 + 4 = 3x^2$.

Решение. В уравнении высокой степени, где неизвестная расположена по степеням уменьшения, введение новой переменной не проходит. Попробуем отыскать один из корней, подставляя вместо неизвестной величины числа ± 1, ± 2, ... Обнаруживаем, что $x = -1$ является корнем уравнения. Применяя метод деления многочлена на двучлен, получим многочлен второй степени:

$$\begin{array}{rl}
-\,x^3 + 3x^2 + 0 \cdot x + 4 & \underline{|x + 1} \\
\underline{x^3 + \ \ x^2 \qquad\quad} & x^2 - 4x + 4 \\
0 \ -4x^2 + 0 \cdot x & \\
-\,\underline{-4x^2 - 4x \quad} & \\
0 \quad 4x + 4 & \\
-\,\underline{\quad 4x + 4} & \\
0 &
\end{array}$$

Решение уравнения $x^2 - 4x + 4 = 0$ позволяет найти еще один корень:

$$x^2 - 4x + 4 = (x - 2)^2 = 0 \quad \Rightarrow \quad x = 2.$$

Ответ: $x = -1$, $x = 2$.

Пример 1.13. Решить уравнение $x^3 + 4 = 3x^2$, раскладывая многочлен уравнения на множители.

Решение. Преобразуем многочлен $x^3 - 3x^2 + 4$:

$$x^3 - 3x^2 + 4 = x^3 + 1 - 3x^2 + 3 = (x + 1)(x^2 - x + 1) - 3(x - 1)(x + 1) =$$
$$= (x + 1)(x^2 - x + 1 - 3x + 3) = (x + 1)(x^2 - 4x + 4).$$

Уравнение примет вид $(x + 1)(x - 2)^2 = 0$. Его решение $x = -1$, $x = 2$.

Иррациональные уравнения

Уравнения, содержащие функцию $f(x)$ под знаком корня (радикала), называются *иррациональными*. При решении этих уравнений мы впервые серьезно займемся областью допустимых значений. Интересно, можно ли правильно решить задачу, не прибегая к работе с ОДЗ? Оказывается, можно. Действительно, решая задачу и совершая неравносильные преобразования, расширяющие ОДЗ, мы, получив посторонние корни, можем их отбросить, сделав проверку. Тем не менее мы будем обращать внимание на ОДЗ — это может значительно сократить решение задачи. Будем гибко подходить к этому вопросу. Выписать полностью ОДЗ иногда очень сложно, сложнее, чем решить само уравнение. Поэтому будем выписывать те неравенства, которые достаточно очевидны. Решив уравнение, найдя все возможные корни, сверив их с ОДЗ и отбросив посторонние, оставшиеся будем проверять подстановкой.

Откуда получаются посторонние корни? Рассмотрим уравнение

$\sqrt{f(x)} = \varphi(x)$.

Возведя в квадрат данное уравнение, получим

$f(x) = \varphi^2(x)$.

Это неравносильное преобразование, поскольку другое уравнение

$\sqrt{f(x)} = -\varphi(x)$

приводится к такому же виду. Отсортировать посторонние корни можно, написав ОДЗ: $\varphi(x) \geq 0$. Итак:

$$\begin{cases} \sqrt{f(x)} = \varphi(x) \\ \varphi(x) \geq 0 \end{cases} \qquad \begin{cases} \sqrt{f(x)} = -\varphi(x) \\ \varphi(x) \leq 0 \end{cases}$$

$$f(x) = \varphi^2(x)$$

Как быть с условием $f(x) \geq 0$? Должно же оно быть выполнено? Оказывается, условие $f(x) \geq 0$ записывать не нужно. Оно автоматически выполняется. Решая уравнение $f(x) = \varphi^2(x)$, вы не можете получить корни, при подстановке которых левая часть уравнения будет отрицательной, а правая — положительной.

Методы решения иррациональных уравнений сводятся к избавлению от иррациональности (от корней).

1. Возведение в соответствующую степень.
2. Введение новой переменной, иногда на определенном этапе решения.

Рассмотрим их на конкретных примерах.

Пример 1.14. Решить уравнение $\sqrt{2x+5} - \sqrt{3x-5} = 2$.

Решение. Начнем с ОДЗ и внимательно проследим за ней.

$$\begin{cases} 2x+5 \ge 0, \\ 3x-5 \ge 0, \end{cases} \quad \text{откуда} \quad x \ge \frac{5}{3}.$$

К выписанным очевидным неравенствам можно кое-что добавить: $\sqrt{2x+5} - \sqrt{3x-5} \ge 0$, или $\sqrt{2x+5} \ge \sqrt{3x-5}$. Возведя в квадрат, получим $2x+5 \ge 3x-5$, или $x \le 10$. Таким образом, решение уравнения будем искать в области $\frac{5}{3} \le x \le 10$.

Перенеся второе слагаемое направо, возведем обе части уравнения в квадрат и приведем подобные члены:

$$6 - x = 4\sqrt{3x-5}.$$

Правая часть уравнения неотрицательна. Тогда левая часть должна удовлетворять условию $6 - x \ge 0$, или $x \le 6$. Наша ОДЗ сократилась до неравенства $\frac{5}{3} \le x \le 6$.

Возведем обе части уравнения в квадрат, приведем подобные члены. Получим квадратное уравнение $x^2 - 60x + 116 = 0$, корни которого $x_1 = 2$; $x_2 = 58$. Поскольку мы постарались учесть все ограничения на неизвестную величину x, то, отбросив посторонний корень $x_2 = 58$, можем записать окончательное решение задачи без проверки. *Ответ*: $x = 2$.

З а м е ч а н и е. Ясно, что мы выбрали не самый короткий путь решения. Можно было, не обращая вовсе внимания на ОДЗ, получить корни и подставить их в исходное уравнение. Легко обнаружить, что корень $x = 58$ посторонний: $\sqrt{116+5} - \sqrt{174-5} \ne 2$. Однако, если бы корни оказались иррациональными, наш подход был бы наилучшим: подстановка иррациональных корней представляет, как правило, определенную трудность.

Рассмотрим пример, иллюстрирующий введение переменной на определенном этапе решения задачи.

Пример 1.15. Решить уравнение $\sqrt{x-1} + \sqrt{x^2-1} = \sqrt{x^3}$.

Решение. Ограничения на x: $\begin{cases} x-1 \ge 0, \\ x^2-1 \ge 0, \\ x^3 \ge 0 \end{cases}$ — дают условие $x \ge 1$. Возведем обе части уравнения в квадрат и выделим оставшийся квадратный корень в левой части уравнения:

$$2\sqrt{(x-1)(x^2-1)} = x^3 - x^2 - x + 2.$$

Дальнейшее возведение в квадрат избавит нас от корня, но получающееся при этом уравнение шестой степени дает мало шансов на успех. Раскроем под корнем скобки:

$$2\sqrt{x^3 - x^2 - x + 1} = x^3 - x^2 - x + 2.$$

Можно сообразить, что введение новой переменной $\sqrt{x^3 - x^2 - x + 1} = y$ значительно упростит уравнение. Действительно, $x^3 - x^2 - x + 1 = y^2$ и $x^3 - x^2 - x + 2 = y^2 + 1$. Уравнение примет вид $2y = y^2 + 1$. Его решение $y = 1$. Возвращаясь к старой переменной, получим $\sqrt{x^3 - x^2 - x + 1} = 1$. Возведем в квадрат, вынесем x за скобки: $x(x^2 - x - 1) = 0$. Из трех корней $x_1 = 0$, $x_{2,3} = \frac{1 \pm \sqrt{5}}{2}$ только корень $x_3 = \frac{1 + \sqrt{5}}{2}$ удовлетворяет ОДЗ. *Ответ*: $x = \frac{1 + \sqrt{5}}{2}$.

Алгебраические неравенства

Неравенствами называются выражения вида $a < b$ ($a \le b$) и $a > b$ ($a \ge b$).

Свойства неравенств.

1. Если $a > b$, то $b < a$.
2. Если $a > b$ и $b > c$, то $a > c$.
3. Пусть $a > b$:

а) $c > 0$, тогда $ac > bc$;

б) $c < 0$, тогда $ac < bc$.

4. Если $a > b$ и $c > d$, то $a + c > b + d$.
5. Пусть $a > 0$, $b > 0$, $a > b$, $n > 0$, тогда $a^n > b^n$ и $\sqrt[n]{a} > \sqrt[n]{b}$.

Замечание 1. Свойство 4 указывает, что неравенства одинакового знака можно складывать.

Замечание 2. Свойство 5 можно расширить. При $a > 0$, $b < 0$, $a > b$ и n положительном нечетном числе обе части неравенства можно возвести в n-ю степень. Этого нельзя сделать, если n — положительное четное число.

Методы решений рациональных и иррациональных неравенств значительно различаются. Иррациональные неравенства содержат неизвестную величину под знаком корня и, как правило, требуют возведения неравенства в некоторую степень. Вначале займемся более простыми рациональными неравенствами.

Рациональные неравенства

1. **Линейные неравенства.**

Все слагаемые неравенства, содержащие неизвестную величину, собираются в одной части неравенства, все свободные члены — в другой. Приводятся подобные члены. В результате получаем неравенство вида

$ax > b$,

где a и b — некоторые числа.

Для получения решения делим обе части неравенства на число $a \neq 0$ и сохраняем знак неравенства в случае $a > 0$ или меняем его при $a < 0$. Таким образом, при $a > 0$ решение $x > \frac{b}{a}$, при $a < 0$ решение $x < \frac{b}{a}$.

Пример 1.16. Решить неравенство $-x - 2 + 3x - 4 > 5x - 7$.

Решение. Перенесем все слагаемые с неизвестной величиной влево, остальные слагаемые — вправо. Получим $-x + 3x - 5x > 2 + 4 - 7$. Приведем подобные члены: $-3x > -1$. Разделим обе части неравенства на -3 и поменяем знак неравенства. Окончательно $x < \frac{1}{3}$.

2. **Квадратные и более высоких степеней неравенства**

Прежде рассмотрим несколько простых неравенств, ответы к которым желательно запомнить.

Пример 1.17. Решить неравенство $x^2 > 4$.

Решение. Извлечем квадратный корень из обеих частей: $|x| > 2$. Раскроем модуль:

1) $\begin{cases} x \geq 0, \\ x > 2 \end{cases}$ и 2) $\begin{cases} x < 0, \\ -x > 2 \end{cases}$ или $\begin{cases} x < 0, \\ x < -2. \end{cases}$ Первая система дает решение $x > 2$, вторая система — решение $x < -2$. Следует запомнить, что из неравенства $|x| > 2$ следует решение $x < -2$ или $x > 2$ и обратно.

Пример 1.18. Решить неравенство $x^2 < 9$.

Решение. При извлечении корня получим $|x| < 3$. Раскрывая модуль, рассмотрим две системы:

1) $\begin{cases} x \geq 0, \\ x < 3 \end{cases}$ и 2) $\begin{cases} x < 0, \\ -x < 3 \end{cases}$ или $\begin{cases} x < 0, \\ x > -3 \end{cases}$. Решение первой системы $0 \leq x < 3$, решение второй $-3 < x < 0$. Объединяя полученные решения, будем иметь: $-3 < x < 3$. Здесь надо запомнить, что

$|x| < 3 \Leftrightarrow -3 < x < 3$.

Пример 1.19. Решить неравенство $4 < x^2 < 9$.

Решение. Подобно предыдущим примерам, извлечем корень из всех трех частей неравенства: $2 < |x| < 3$. Выпишем две системы, получающиеся при раскрытии модуля:

1) $\begin{cases} x \geq 0, \\ 2 < x < 3 \end{cases}$ и 2) $\begin{cases} x < 0, \\ 2 < -x < 3 \end{cases}$ или $\begin{cases} x < 0, \\ -2 < x > -3. \end{cases}$ Из наших систем следует, что $2 < x < 3$ или $-3 < x < -2$. Этот вывод также следует запомнить:

$$2 < |x| < 3 \Leftrightarrow \left[\begin{array}{l} 2 < x < 3, \\ -3 < x < -2. \end{array}\right.$$

Рассмотрим простейшее неравенство $x^2 - 3x + 2 \geq 0$, на примере которого обсудим основные методы решения неравенств.

Основные методы решения рациональных неравенств

На основе простой задачи рассмотрим два основных метода решения неравенств.

Пример 1.20. Решить неравенство $x^2 - 3x + 2 \geq 0$.

1. *Метод систем.* Разложим неравенство на множители: $(x - 1) \times (x - 2) \geq 0$. Произведение двух сомножителей будет неотрицательным, если оба сомножителя либо неотрицательны, либо неположительны:

а) $\begin{cases} x - 1 \geq 0, \\ x - 2 \geq 0 \end{cases}$ или б) $\begin{cases} x - 1 \leq 0, \\ x - 2 \leq 0. \end{cases}$

В случае а) решение $x \geq 2$; в случае б) получим $x \leq 1$.

З а м е ч а н и е. Если неравенство сопровождается знаком равенства $\geq$ или $\leq$, во всех системах следует такой знак выписывать. В противном случае могут быть потеряны отдельные значения неизвестной величины. Например, если в системе а) вместо знака $\geq$ написать знак $>$, будет потеряно решение $x = 2$, если в системе б) не написать знак $\leq$, будет потеряно решение $x = 1$.

2) *Метод интервалов.* Разложим неравенство на множители и на числовой оси нанесем точки, в которых неравенство обращается в нуль (рис. 1.5). Возьмем любое значение правее точки 2, например 3. Каждый из сомножителей при $x = 3$ положителен, неравенство выполняется. На интервале $1 < x < 2$ первый сомножитель положителен, второй отрицателен, их произведение станет отрицательным. На интервале $x < 1$ оба сомножителя отрицательны, их произведение снова станет положительным. Отметим знаки, которые неравенство имеет на каждом интер-

Рис. 1.5

вале, знаками «+» или «–». Нас интересует такая область значений неизвестной величины, в которой неравенство неотрицательно. Эта область $x \le 1$ или $x \ge 2$ помечена прямоугольниками. Ниже рассмотрим метод более подробно.

Метод интервалов

Решение методом интервалов желательно проводить в следующей последовательности:

1) перенести все члены неравенства в одну часть;
2) привести к общему знаменателю и привести подобные члены;
3) разложить на множители;
4) на числовой оси расставить значения переменной, при которых числитель и знаменатель обращаются в нуль. На каждом образовавшемся промежутке определить и указать знак «+» или «–»;
5) выписать решения неравенства в ответ.

Предположим, что пункты 1) — 3) рекомендаций выполнены. Предстоит работа с числовой осью.

Пример 1.21. Решить неравенство $\dfrac{x(x-1)^2(2-x)^3}{(x-3)(x+5)^4} \ge 0.$

Решение. На числовой оси нанесем точки (значения неизвестной), в которых числитель и знаменатель обращаются в нуль. Точки числителя и знаменателя пометим по-разному. Числитель может обратиться в нуль, знаменатель — нет. Возьмем любое значение неизвестной величины $x > 3$, например число 4. В этой точке дробь отрицательна. На числовой оси ставим знак «–». При переходе через точку $x = 3$ сомножитель знаменателя $(x - 3)$ поменяет знак. Значит, знак дроби изменится также. В области $2 < x < 3$ ставим знак «+». Проходя вдоль числовой оси через точку 2 справа налево, обращаем внимание на скобку $(2 - x)^3$, которая меняет знак. Следовательно, на интервале $1 < x < 2$ надо поставить знак «–». В точке $x = 1$ знак дроби не изменится. Хотя скобка $(x - 1)$ поменяет знак, она возводится в квадрат, поэтому сомножитель $(x - 1)^2$ знака дроби не изменит. В точке $x = 0$ дробь поменяет знак за счет смены знака у сомножителя x. В точке $x = -5$ в силу четности показателя сомножитель $(x + 5)^4$ сохранит свой знак и, соответственно, знак дроби. Знаки неравенства расставлены на числовой оси рис. 1.6.

Ответ: $x < -5$, $-5 < x \le 0$, $x = 1$, $2 \le x < 3$.

Рис. 1.6

З а м е ч а н и е 1. Сомножитель, содержащий скобку в четной степени, знака дроби не меняет. Сомножитель со скобкой в нечетной степени будет менять знак дроби.

З а м е ч а н и е 2. Если знак неравенства сопровождается знаком равенства, желательно «точки числителя и знаменателя» помечать на числовой оси неодинаково, поскольку значения неизвестной, обращающие числитель в нуль, будут решениями неравенства.

Системы неравенств

Решить систему неравенств — значит найти такие значения неизвестных величин, при которых удовлетворяется каждое из неравенств, входящих в систему. Решая систему, можно придерживаться следующей последовательности действий. Каждое неравенство доводится до разложения на множители. Затем все неравенства выписываются в систему, рисуется количество числовых осей, равное числу неравенств в системе. Неравенства решаются каждое на своей числовой оси. Находится область пересечений.

Пример 1.22. Решить неравенство $\left|\dfrac{x^2-5x+4}{x^2-4}\right| \le 1.$

Решение. Раскроем модуль, используя выводы примера 1.18:

$$-1 \le \frac{x^2-5x+4}{x^2-4} \le 1.$$

Двойное неравенство запишем в виде системы:

$$\begin{cases} \dfrac{x^2-5x+4}{x^2-4} \ge -1, \\ \dfrac{x^2-5x+4}{x^2-4} \le 1. \end{cases}$$

После несложных преобразований (переноса слагаемых влево, приведения подобных членов и разложения на множители) получим систему из двух разложенных на множители неравенств:

$$\begin{cases} \dfrac{2x\left(x-\dfrac{5}{2}\right)}{(x-2)(x+2)} \ge 0, \\ \dfrac{-5\left(x-\dfrac{8}{5}\right)}{(x-2)(x+2)} \le 0. \end{cases}$$

Нарисуем две числовые оси одну под другой (рис. 1.7) и, соблюдая соответствие по вертикали, нанесем значения неизвестной. Поставим

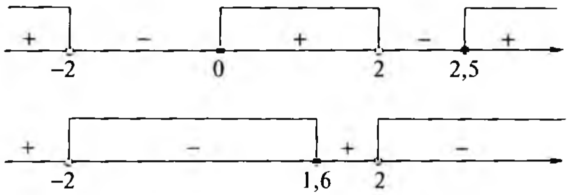

Рис. 1.7

знаки, которые имеют первое и второе неравенства на каждом промежутке. Прямоугольниками укажем решения неравенств. Решением системы будут те промежутки, которые попадают одновременно в один из верхних и один из нижних прямоугольников.

Ответ: $0 \le x \le 1{,}6$, $x \ge 2{,}5$.

Иррациональные неравенства

Путем алгебраических преобразований иррациональные неравенства можно свести к одному из двух типов:

1) $\sqrt{f(x)} < \varphi(x)$; 2) $\sqrt{f(x)} > \varphi(x)$.

Решение первого неравенства сводится к решению системы

$$\begin{cases} f(x) \ge 0, \\ \varphi(x) > 0, \\ f(x) > \varphi^2(x). \end{cases}$$

Решение второго неравенства требует решения двух систем:

А) $\begin{cases} \varphi(x) \ge 0, \\ f(x) > \varphi^2(x); \end{cases}$ Б) $\begin{cases} \varphi(x) < 0, \\ f(x) \ge 0. \end{cases}$

Действительно, если обе части неравенства неотрицательны, его можно возвести в квадрат (система А). При отрицательной правой части и неотрицательной левой неравенство верно. Следует только не забыть про область допустимых значений аргумента (система Б).

Пример 1.23. Решить неравенство $\sqrt{\frac{1}{x^2} - \frac{3}{4}} < \frac{1}{x} - \frac{1}{2}$.

Решение. Решение неравенства эквивалентно решению следующей системы:

$$\begin{cases} \frac{1}{x} - \frac{1}{2} > 0, \\ \frac{1}{x^2} - \frac{3}{4} \ge 0, \\ \frac{1}{x^2} - \frac{3}{4} < \left(\frac{1}{x} - \frac{1}{2}\right)^2. \end{cases}$$

Раскроем скобки в последнем неравенстве, преобразуя также первые два:

$$\begin{cases} \dfrac{2-x}{2x} > 0, \\ \dfrac{4-3x^2}{4x^2} \geq 0, \\ \dfrac{1}{x^2} - \dfrac{3}{4} < \dfrac{1}{x^2} - \dfrac{1}{x} + \dfrac{1}{4}. \end{cases}$$

Разложим каждое неравенство на множители:

$$\begin{cases} \dfrac{2-x}{2x} > 0, \\ \dfrac{3\left(\dfrac{2}{\sqrt{3}} - x\right)\left(\dfrac{2}{\sqrt{3}} + x\right)}{4x^2} \geq 0, \\ \dfrac{x-1}{x} > 0. \end{cases}$$

Рассмотрим решение каждого неравенства на числовой оси (рис. 1.8). Найдем область пересечений.

Ответ: $1 < x \leq 2/\sqrt{3}$.

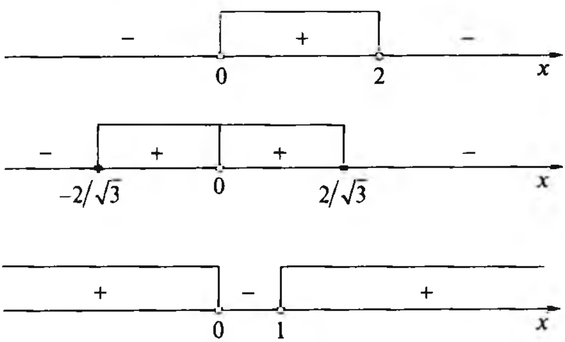

Рис. 1.8

Глава 2
ПОКАЗАТЕЛЬНАЯ И ЛОГАРИФМИЧЕСКАЯ ФУНКЦИИ

2.1. ПОКАЗАТЕЛЬНАЯ ФУНКЦИЯ

Функция вида $y = a^x$, где $a > 0$ и $a \neq 1$, называется *показательной*. Из определения показательной функции непосредственно следует, что $a^x > 0$ при $x \in R$.

Основные свойства:

1) $a^0 = 1$; 2) $a^{-x} = \dfrac{1}{a^x}$; 3) $(a^x)^y = a^{xy}$;. 4) $a^x a^y = a^{x+y}$; 5) $\dfrac{a^x}{a^y} = a^{x-y}$;
6) $(ab)^x = a^x b^x$; 7) $\left(\dfrac{a}{b}\right)^x = \dfrac{a^x}{b^x}$.

График показательной функции

Построим по точкам два графика: $y_1 = 2^x$ и $y_2 = \left(\dfrac{1}{2}\right)^x$. Взятые для построения точки представлены в табл. 2.1.

Таблица 2.1

x	-3	-2	-1	0	1	2	3
y_1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8
y_2	8	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$

Построенные по приведенным точкам кривые изображены на рис. 2.1. Показательная функция может монотонно возрастать или монотонно убывать. Это зависит от величины ее основания. При $a > 1$ функция $y = a^x$ является возрастающей, при $0 < a < 1$ — убывающей. Это свойство лежит в основе решения показательных неравенств.

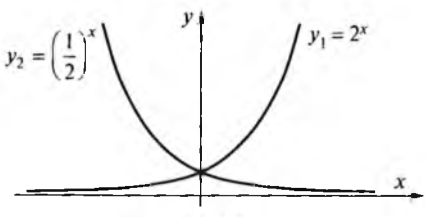

Рис. 2.1

Если $a > 1$, то из неравенства $a^{f(x)} > a^{\varphi(x)}$ следует неравенство $f(x) > \varphi(x)$.

Если $0 < a < 1$, то из неравенства $a^{f(x)} > a^{\varphi(x)}$ следует $f(x) < \varphi(x)$.

Методы преобразования показательных уравнений и неравенств

Большинство задач может быть решено одним из трех способов, достаточно четко работающих:

1) введение новой переменной, если задача содержит одну показательную функцию;

2) деление на одну из показательных функций или ее квадрат, если их две, и затем приведение к многочлену относительно новой переменной;

3) разложение на множители.

Использование методов преобразования позволяет свести показательное уравнение или неравенство к одному из основных видов.

Основные показательные уравнения

1. $a^{f(x)} = a^{\varphi(x)}$, где a — число.

Решение этого уравнения основано на следующем утверждении: если две показательные функции равны друг другу и имеют одинаковые основания, то равны и их показатели (аргументы). Действительно, пусть $a^{f(x)} = a^{\varphi(x)}$. Разделим обе части равенства на $a^{\varphi(x)} > 0$. Получим $\frac{a^{f(x)}}{a^{\varphi(x)}} = 1$ или $a^{f(x)-\varphi(x)} = 1$. При $a \neq 1$ равенство верно лишь при выполнении условия $f(x) - \varphi(x) = 0$, откуда следует $f(x) = \varphi(x)$. Итак, $a^{f(x)} = a^{\varphi(x)} \Rightarrow f(x) = \varphi(x)$. Например, из уравнения $2^x = 2^{1-x}$ следует $x = 1 - x$ или $x = \frac{1}{2}$.

2. $a^{f(x)} = b$:

а) $b \leq 0$ — нет решений;

б) $b > 0$. Представим правую часть уравнения в виде показательного выражения, для чего используем понятие логарифма. Кто с ним не знаком, не огорчайтесь. Чуть ниже мы введем понятие логарифма. А пока это пусть будет некий иероглиф.

$$a^{f(x)} = a^{\log_a b}.$$

Отсюда следует соотношение между показателями $f(x) = \log_a b$. Это уравнение решается, поскольку $\log_a b$ есть число. Например, уравнение $2^x = 3$ преобразуется к виду $2^x = 2^{\log_2 3}$, откуда следует $x = \log_2 3$.

3. $a(x)^{f(x)} = 1$. В основании стоит функция от неизвестной величины x.

Уравнение требует рассмотрения двух случаев:

$$\begin{cases} a(x) = 1, \\ f(x) - \text{определена} \end{cases} \quad \text{или} \quad \begin{cases} f(x) = 0, \\ a(x) = 0. \end{cases}$$

Например, $x^{x+1} = x$ распадается на уравнения $x + 1 = 0$ и $x = 1$.

Пример 2.1. Решить уравнение $8^{\frac{2}{x}} - 2^{\frac{3x+3}{x}} = 12 = 0$.

Решение. Перейдем в первом слагаемом к основанию 2: $(2^3)^{\frac{2}{x}} - 2^{3+\frac{3}{x}} + 12 = 0$. Выделим показательную функцию $\left(2^{\frac{3}{x}}\right)^2 - 8 \cdot 2^{\frac{3}{x}} + 1$ $+ 12 = 0$. Введем новую переменную $2^{\frac{3}{x}} = y > 0$. Получим квадратное уравнение $y^2 - 8y + 12 = 0$. Его корни: $y_1 = 2$; $y_2 = 6$. Вернемся к старой переменной $2^{\frac{3}{x}} = 2$, откуда $\frac{3}{x} = 1$, или $x = 3$. Затем $2^{\frac{3}{x}} = 6$. Представим число 6 как показательное выражение. Тогда $2^{\frac{3}{x}} = 2^{\log_2 6}$. Следовательно, $\frac{3}{x} = \log_2 6$, или $x = \frac{3}{\log_2 6}$. *Ответ*: $x_1 = 3$; $x_2 = \frac{3}{\log_2 6}$.

Пример 2.2. Решить уравнение $3 \cdot 16^x + 2 \cdot 81^x = 5 \cdot 36^x$.

Решение. Выделим в уравнении две показательные функции: $3 \cdot 4^{2x} + 2 \cdot 9^{2x} = 5 \cdot (4 \cdot 9)^x$. Особенностью уравнения является то, что каждое слагаемое представляет квадрат показательной функции $4^{2x} = (4^x)^2$ и $9^{2x} = (9^x)^2$ или произведение этих функций $(4 \cdot 9)^x = 4^x \cdot 9^x$. Такие уравнения называются *однородными*. Одним из способов решения подобных уравнений является деление на квадрат показательной функции, например на 9^{2x}. Получим $3\frac{4^{2x}}{9^{2x}} + 2 = 5\frac{4^x \cdot 9^x}{9^{2x}}$. Уравнение легко преобразуется: $3\left(\frac{4}{9}\right)^{2x} + 2 = 5\frac{4^x \cdot 9^x}{9^x \cdot 9^x}$, или $3\left(\frac{4}{9}\right)^{2x} + 2 = 5\left(\frac{4}{9}\right)^x$. Ведением переменной $\left(\frac{4}{9}\right)^x = 1$ оно сводится к квадратному $3y^2 - 5y + 2 = 0$, корни которого: $y_1 = 1$; $y_2 = \frac{2}{3}$. Для нахождения неизвестной величины x имеем два уравнения $\left(\frac{4}{9}\right)^x = 1$ и $\left(\frac{4}{9}\right)^x = \frac{2}{3}$. Решением первого является $x = 0$, второе уравнение дает решение $x = \frac{1}{2}$. *Ответ*: $x_1 = 0$; $x_2 = \frac{1}{2}$.

Пример 2.3. Решить уравнение $|x-3|^{3x^2-10x+3}=1$.

Решение. Уравнение распадается на два случая: 1) $|x-3|=1$ и 2) $\begin{cases}3x^2-10x+3=0,\\ x\neq 3.\end{cases}$. В первом случае имеем $x-3=\pm1$, или $x=2$ и $x=4$. Во втором случае из двух корней уравнения $x_1=\frac{1}{3}$ и $x_2=3$ в ответ запишем только x_1. *Ответ*: $x_1=\frac{1}{3}$; $x_2=2$; $x_3=4$.

Основные показательные неравенства

Решение показательных неравенств сводится к решению нескольких основных неравенств:

1) $a^{f(x)}>a^{\varphi(x)}$.

Воспользуемся рассмотренным выше свойством показательной функции.

Если $a>1$, то $f(x)>\varphi(x)$.

Если $0<a<1$, то $f(x)<\varphi(x)$.

Например, решая неравенство $\left(\frac{1}{2}\right)^x>\frac{1}{2}$, следует поменять знак неравенства при переходе к показателям: $x<1$;

2) $a^{f(x)}>b$:

а) $b\leq 0$. Показательная функция при всех значениях $x\in$ ОДЗ положительна. Неравенство верно. Например, неравенство $2^x>-2$ верно для $x\in R$;

б) $b>0$. Представим число как показательное выражение, используя логарифм. Получим $a^{f(x)}>a^{\log_a b}$. Далее неравенство решается в соответствии с пунктом 1). Например, неравенство $2^x>3$ можно представить как $2^x>2^{\log_2 3}$. Поэтому $x>\log_2 3$.

Пример 2.4. Решить неравенство $4^{\frac{1}{x}-1}-2^{\frac{1}{x}-2}-3\leq 0$.

Решение. Преобразуем неравенство к виду $\frac{1}{4}2^{\frac{2}{x}}-\frac{1}{4}2^{\frac{1}{x}}-3\leq 0$. Обозначим $2^{\frac{1}{x}}$ через y, которое будет положительным. Получим $\frac{1}{4}y^2-\frac{1}{4}y-3\leq 0$. Разложим неравенство на множители и учтем, что $y>0$: $\begin{cases}(y+3)(y-4)\leq 0,\\ y>0.\end{cases}$ Решение системы $0<y\leq 4$. Для неизвестной величины x это неравенство можно представить в виде $0<2^{\frac{1}{x}}\leq 4$. Отсюда следует, что $\frac{1}{x} \le 2$ или $\frac{1-2x}{x} \le 0$. Решение последнего алгебраического неравенства $0 < x < \frac{1}{2}$ дает два промежутка $(-\infty, 0)$ и $\left[\frac{1}{2}, +\infty\right)$, что и следует записать в ответ. *Ответ*: $x < 0$; $x \ge \frac{1}{2}$.

Пример 2.5. Решить неравенство $2^{x+2} - 2^{x+3} - 2^{x+4} > 5^{x+1} - 5^{x+2}$.

Решение. В неравенстве присутствуют две показательные функции 2^x и 5^x. Разделим обе части неравенства на одну из них, например на 5^x. Получим $4\frac{2^x}{5^x} - 8\frac{2^x}{5^x} - 16\frac{2^x}{5^x} > 5 - 5^2$. Введя $\frac{2^x}{5^x} = \left(\frac{2}{5}\right)^x = y$, получим линейное неравенство $4y - 8y - 16y > -20$. Тогда $y < 1$ или $\left(\frac{2}{5}\right)^x < 1$. Ответ ясен, если написать неравенство в виде $\left(\frac{2}{5}\right)^x < \left(\frac{2}{5}\right)^0$. *Ответ*: $x > 0$.

2.2. ЛОГАРИФМИЧЕСКАЯ ФУНКЦИЯ

Логарифмом числа x ($x > 0$) по основанию a ($a > 0$, $a \ne 1$) называется показатель степени y, в которую надо возвести a, чтобы получить x. Записывается логарифм так: $\log_a x$. Следовательно, из записи $\log_a x = y$ следует, что $a^y = x$. В выражение $a^y = x$ подставим $y = \log_a x$. Получим $a^{\log_a x} = x$.

Эта формула называется *основным логарифмическим тождеством*. Мы использовали эту формулу, когда число b представляли в виде показательного выражения $b = a^{\log_a b}$.

Итак, логарифм — это такой показатель степени у основания a, при возведении в которую получаем значение показательного выражения a^y, т.е. x. Вопрос стоит так: зная a и задавая x, подобрать показатель степени y. В общем случае это трудная вычислительная задача. Поэтому существуют таблицы логарифмов, в продвинутых калькуляторах имеются опции вычисления логарифмов. Для некоторых значений основания и определенных значений величины x это сделать легко. Например, пусть основание равно 2. Зададим значение $x = 8$. Ответ на вопрос, в какую степень надо возвести 2, чтобы получить 8, очевиден: в степень, равную 3. Записать это можно как $2^3 = 8$, так и в виде логарифма $\log_2 8 = 3$.

Функция вида $y = \log_a x$, где $x > 0$, $a > 0$, $a \ne 1$, называется *логарифмической*.

Логарифмы получили настолько широкое распространение в математике и ее приложениях, что для удобства работы с ними были введены частные виды логарифмов.

Логарифм по основанию десять стал называться *десятичным* логарифмом и обозначаться через сочетание букв: $\lg x$.

Логарифм по основанию 2,71828... = e (постоянная Эйлера) назван *натуральным* логарифмом и обозначается как $\ln x$.

Основные свойства логарифмов:

1) $\log_a(x) = \log_a|x| + \log_a|y|$;

2) $\log_a\left(\dfrac{x}{y}\right) = \log_a|x| - \log_a|y|$;

3) $\log_a x^k = \begin{cases} k \cdot \log_a x, & k \text{ – нечетное число;} \\ k \cdot \log_a|x|, & k \text{ – четное число;} \end{cases}$

4) $\log_a b = \dfrac{1}{\log_b a}$;

5) $\log_a x = \dfrac{\log_b x}{\log_b a}$. Свойство названо *модулем перехода*;

6) $\log_{a^k} x = \dfrac{1}{k}\log_a x$.

График логарифмической функции

Построим графики двух функций $y_1 = \log_2 x$ и $y_2 = \log_{\frac{1}{2}} x$. Возьмем некоторые значения аргумента x (табл. 2.2), вычислим соответствующие значения функций.

Таблица 2.2

x	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8
y_1	–3	–2	–1	0	0	2	3
y_2	3	2	1	0	–1	–2	–3

По найденным точкам построим кривые (рис. 2.2). Логарифмическая функция с основанием 2 возрастает, с основанием $^1/_2$ — убывает. В общем случае при $a > 1$ функция является монотонно возрастающей, при $0 < a < 1$ — монотонно убывающей. Это свойство лежит в основе решения логарифмических неравенств.

Если $a > 1$, из неравенства для функций $\log_a f(x) > \log_a \varphi(x)$ следует неравенство для аргументов $f(x) > \varphi(x)$.

Если $0 < a < 1$, из того же неравенства $\log_a f(x) > \log_a \varphi(x)$ следует $f(x) < \varphi(x)$.

Переход от логарифмических функций к их аргументам и обратно имеет названия:

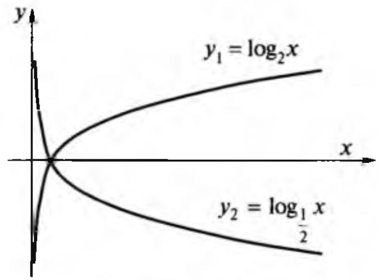

Рис. 2.2

$$\log_a f(x) \lesseqgtr \log_a \varphi(x) \underset{\text{Логарифмирование}}{\overset{\text{Потенцирование}}{\rightleftarrows}} f(x) \gtreqless \varphi(x).$$

Уравнение или неравенство можно пропотенцировать. Одним из методов решения задачи является логарифмирование.

Основные логарифмические уравнения

1. $\log_a f(x) = b$.

Надо воспользоваться определением логарифмической функции, из которого следует, что $f(x) = a^b$. Например, из уравнения $\log_2 x = 4$ сразу следует $x = 2^4 = 16$.

2. $\log_a f(x) = \log_a \varphi(x)$.

Потенцирование этого основного логарифмического уравнения основано на утверждении: если две логарифмические функции имеют одинаковые основания и равны, то равны и их выражения под знаками логарифмов, т.е. $f(x) = \varphi(x)$. В самом деле, перенесем все влево: $\log_a f(x) - \log_a \varphi(x) = 0$. Воспользуемся свойством логарифмов, приведя левую часть к одному логарифму: $\log_a \dfrac{f(x)}{\varphi(x)} = 0$. Пропотенцируем: $\dfrac{f(x)}{\varphi(x)} = 1$, или $f(x) = \varphi(x)$. Например, решая уравнение $\log_3(x+2) = \log_3(2x+1)$, получим $x + 2 = 2x + 1$, или $x = 1$.

Пример 2.6. Решить уравнение $\lg 5 + \lg(x + 10) = 1 - \lg(2x - 1) + \lg(21x - 20)$.

Решение. Напишем область допустимых значений аргумента:

$$\begin{cases} x + 10 > 0, \\ 2x - 1 > 0, \\ 21x - 20 > 0 \end{cases} \Rightarrow \quad x > \frac{20}{21}.$$

Перенесем все логарифмы влево и свернем в один логарифм:

$\lg 5 + \lg(x + 10) + \lg(2x - 1) - \lg(21x - 20) = 1$, или

$$\lg 5(x + 10) + \lg\frac{2x - 1}{21x - 20} = 1, \quad \text{или} \quad \lg\frac{5(x + 10)(2x - 1)}{21x - 20} = 1.$$

После потенцирования $\frac{5(x + 10)(21x - 20)}{2x - 1} = 10$ уравнение легко преобразуется к квадратному $2x^2 - 23x + 30 = 0$, имеющему два корня $x_1 = \frac{3}{2}$, $x_2 = 10$. Оба корня удовлетворяют ограничениям на переменную x. *Ответ*: $x_1 = \frac{3}{2}$; $x_2 = 10$.

Пример 2.7. Решить уравнение $x^{\lg x + 3} = 1000x$.

Решение. Очевидно, что ОДЗ $x > 0$. Прологарифмируем данное уравнение по основанию 10: $\lg(x^{\lg x+3}) = \lg(1000x)$. Цель этой операции — сделать сложный показатель $\lg x + 3$ сомножителем слагаемого. Вынесем показатель $\lg x + 3$ за знак логарифма: $(\lg x + 3)\lg x = \lg 1000 + \lg x$. Пусть $\lg x = y$. Имеем квадратное уравнение $y^2 + 2y - 3 = 0$ с корнями $y_1 = -3$, $y_2 = 1$. Решая два основных логарифмических уравнения: $\lg x = -3$ и $\lg x = 1$, получим решения $x_1 = 10^{-3}$, $x_2 = 10$, удовлетворяющие ОДЗ. *Ответ*: $x_1 = 10^{-3}$; $x_2 = 10$.

Пример 2.8. Решить уравнение $\log_x 9x^2 \cdot \log_3^2 x = 4$.

Решение. После выяснения ОДЗ: $x > 0$, $x \neq 1$ — воспользуемся одной из рекомендаций: перейдем в уравнении к одному основанию — лучше к основанию 3. Нам понадобится модуль перехода $\frac{\log_3 9x^2}{\log_3 x} \cdot \log_3^2 x = 4$. Логарифм числителя представим в виде суммы слагаемых $(\log_3 9 + \log_3 x^2) \times \times \log_3 x = 4$ и введем новую переменную $\log_3 x = y$. Получим $(2 + 2y)y = 4$ — квадратное уравнение, имеющее корни $y_1 = -2$, $y_2 = 1$. Отсюда $\log_3 x = -2$ и $\log_3 x = 1$. Их решения $x_1 = \frac{1}{9}$, $x_2 = 3$. Условия, наложенные ОДЗ, выполняются. *Ответ*: $x_1 = \frac{1}{9}$; $x_2 = 3$.

Основные логарифмические неравенства

1. $\log_a f(x) > \log_a \varphi(x)$.

Мы уже знаем, что логарифмическая функция является либо монотонно возрастающей, либо монотонно убывающей. Ее поведение зависит от величины основания a. Если $a > 1$, потенцирование приводит к неравенству $f(x) > \varphi(x)$. При $0 < a < 1$ знак неравенства меняется: $f(x) < \varphi(x)$.

2. $\log_a f(x) > b$.

Для понимания того, как следует потенцировать, перепишем неравенство, воспользовавшись преобразованием $b = \log_a a^b$:

$\log_a f(x) > \log_a a^b$.

Теперь можно сослаться на предыдущий пункт. В случае $a > 1$ имеем $f(x) > a^b$. Иначе, при $0 < a < 1$, получим $f(x) < a^b$.

З а м е ч а н и е. В этом основном логарифмическом неравенстве обычно не преобразуют правую часть. От логарифма избавляются, используя определение, а знак полученного неравенства сохраняют при $a > 1$ и меняют на противоположный при $0 < a < 1$.

Пример 2.9. Решить неравенство $\log^2_{0,2}(x-1)^2 > 4$.

Решение. Извлечем квадратный корень из обеих частей неравенства $|\log_{0,2}(x-1)^2| > 2$. Отсюда следует $\left[\begin{array}{l}\log_{0,2}(x-1)^2 > 2,\\ \log_{0,2}(x-1)^2 < -2.\end{array}\right.$ Избавимся от логарифмов, учитывая ОДЗ и тот факт, что основание меньше единицы: $\left[\begin{array}{l}0 < (x-1)^2 < \dfrac{1}{25},\\ (x-1)^2 > 5.\end{array}\right.$ Извлечем квадратный корень: $\left[\begin{array}{l}0 < |x-1| < \dfrac{1}{5},\\ |x-1| > 5.\end{array}\right.$ Раскрыв модули, получим ответ: $-\infty < x < -4$, $0,8 < x < 1$, $1 < x < 1,2$, $6 < x < +\infty$.

Пример 2.10. Решить неравенство $x^{0,5\log_{0,5}x-3} \geq 0,5^{3-2,5\log_{0,5}x}$.

Решение. Прологарифмируем неравенство, взяв основание 0,5. Вспомним, что логарифмическая функция с таким основанием является убывающей. Следовательно, знак неравенства надо поменять:

$$\log_{0,5}(x^{0,5\log_{0,5}x-3}) \leq \log_{0,5}(0,5^{3-2,5\log_{0,5}x}).$$

Показатели следует вынести за знаки логарифмов:

$$(0,5\log_{0,5}x-3)\log_{0,5}x \leq (3-2,5\log_{0,5}x)\log_{0,5}0,5.$$

После обозначения $\log_{0,5}x = y$ получим квадратное неравенство $y^2 - y - 6 \leq 0$. Его решение $-2 \leq y \leq 3$. Вернемся к старой переменной $-2 \leq \log_{0,5}x \leq 3$. После потенцирования $0,5^{-2} \geq x \geq 0,5^3$ получим *ответ*: $\frac{1}{8} \leq x \leq 4$.

Глава 3
ТРИГОНОМЕТРИЯ

3.1. ВВЕДЕНИЕ ТРИГОНОМЕТРИЧЕСКИХ ФУНКЦИЙ

Рассмотрим прямоугольный треугольник. Отношения сторон в прямоугольном треугольнике определены следующим образом (рис. 3.1):

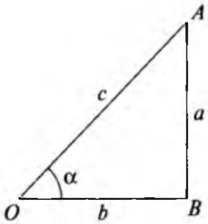

Рис. 3.1

$$\frac{a}{c} = \sin\alpha, \qquad \frac{b}{c} = \cos\alpha,$$

$$\frac{a}{b} = \operatorname{tg}\alpha, \qquad \frac{b}{a} = \operatorname{ctg}\alpha,$$

$$\frac{c}{b} = \sec\alpha, \qquad \frac{c}{a} = \operatorname{cosec}\alpha.$$

Развитие астрономических наблюдений привело к расширению этих понятий. Математики обратились к кругу единичного радиуса, в который определенным образом встраивался прямоугольный треугольник. В единичном круге расположим прямоугольный треугольник OAB так, чтобы его гипотенуза совпала с единичным радиусом (рис. 3.2). Тогда

$$\sin x = \frac{a}{c} = \frac{a}{R} = a.$$

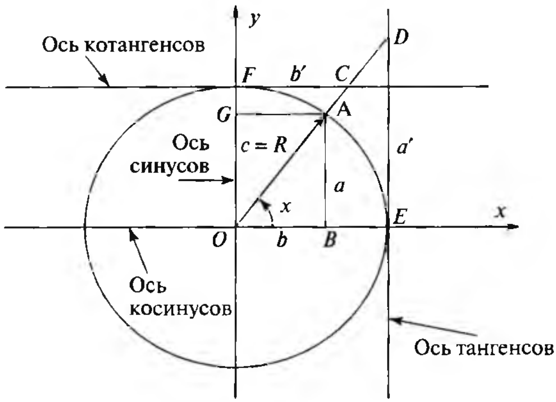

Рис. 3.2

Построив с помощью транспортира нужный угол величиной x, мы без труда определим его синус, измерив линейкой расстояние OG.

$$\cos x = \frac{b}{R} = b.$$

Значение косинуса есть длина отрезка OB.

Для нахождения значения тангенса по известному углу проведем касательную к окружности в точке E и продолжим радиус до пересечения с касательной в точке D. Поскольку треугольники OAB и ODE подобны, их стороны пропорциональны. Следовательно,

$$\operatorname{tg} x = \frac{a}{b} = \frac{a'}{R} = a'.$$

Проведем касательную к окружности в точке F. Продолжение радиуса пересечет эту касательную в точке C. Треугольники OAB и OFC подобны по двум углам. Отсюда

$$\operatorname{ctg} x = \frac{b}{a} = \frac{b'}{R} = b'.$$

Соответствующие оси получили названия: «ось синусов», «ось косинусов», «ось тангенсов» и «ось котангенсов». Принято на этих осях откладывать положительные числа вверх или вправо от точки O, отрицательные — вниз или влево. Угол x всегда отсчитывается от положительного направления оси абсцисс либо против часовой стрелки — тогда он считается положительным, либо по часовой стрелке — в этом случае его величина отрицательна. Любой точке на окружности соответствуют определенные углы. Например, точке A соответствует угол x, а также все углы, полученные полными оборотами радиуса по часовой или против часовой стрелки, т.е. углы вида $x + 2\pi n$, где n — число полных оборотов в любом из двух направлений. Обычно пишут $n \in Z$.

Измерение углов

Наиболее употребительны две единицы измерения углов.

Градус — величина угла, опирающегося на дугу, равную $^1/_{360}$ длины окружности.

Радиан — величина угла, опирающегося на дугу, длина которой равна радиусу окружности. На длине полуокружности размещается 3,14... радиуса. Эта величина обозначается как π. Следовательно, угол в $30°$ можно записать в радианах как $^{\pi}/_{6}$ радиан и т.д.

Замечание. Число π является иррациональным числом. Хотя часто пишут равенство $\pi = 3{,}14$, следует помнить, что в действительности $3{,}14 < \pi < 3{,}15$.

3.2. ГРАФИКИ И СВОЙСТВА ОСНОВНЫХ ТРИГОНОМЕТРИЧЕСКИХ ФУНКЦИЙ

I. Тригонометрическая функция $y = \sin x$.

Задавая угол x и линейкой измеряя величину синуса, построим график этой функции (рис. 3.3):

1) область определения: $x \in R$;

2) область изменения: $-1 \le y \le 1$;

3) промежутки знакопостоянства. В каждой четверти функция будет иметь определенный знак. Их удобно показать на круге (рис. 3.4);

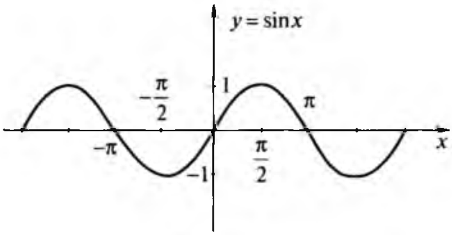

Рис. 3.3

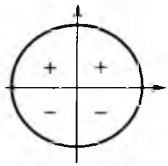

Рис. 3.4

4) четность. Из построения на единичном круге следует, что функция нечетная: $\sin(-x) = -\sin x$;

5) периодичность. Период $T = 2\pi$;

6) таблица значений. В пределах углов от нуля до 90° укажем значения синуса для некоторых углов (табл. 3.1).

Таблица 3.1

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\sin x$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

В дальнейшем мы введем обратные тригонометрические функции. Сейчас же приведем свойство, которое иногда будет использоваться:

$\arcsin(-x) = -\arcsin x$.

II. Тригонометрическая функция $y = \cos x$.

Задавая угол x и линейкой измеряя величину косинуса, построим график этой функции (рис. 3.5):

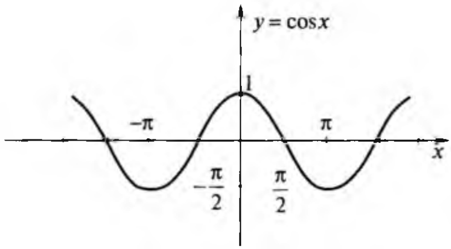

Рис. 3.5

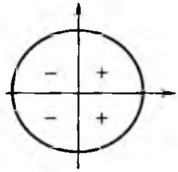

Рис. 3.6

1) область определения: $x \in R$;

2) область изменения: $-1 \le y \le 1$;

3) промежутки знакопостоянства указаны на круге (рис. 3.6);

4) четность. Из построения на единичном круге следует, что функция четная: $\cos(-x) = \cos x$;

5) периодичность. Период $T = 2\pi$.

6) таблица значений. В пределах углов от нуля до 90° укажем значения косинуса для некоторых углов (табл. 3.2).

Таблица 3.2

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\cos x$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0

Отметим свойство обратной тригонометрической функции:

$\text{atccos}(-x) = \pi - \arccos x$.

III. Тригонометрическая функция $y = \operatorname{tg} x$.

С помощью транспортира и линейки построим график этой функции (рис. 3.7):

1) область определения: $x \ne \frac{\pi}{2} + \pi n$, где $n \in Z$;

2) область изменения: $y \in R$;

3) промежутки знакопостоянства указаны на круге (рис. 3.8);

4) четность. Из построения на единичном круге следует, что функция нечетная: $\operatorname{tg}(-x) = -\operatorname{tg} x$;

5) периодичность. Период $T = \pi$;

6) таблица значений. В пределах углов от нуля до 90° укажем значения тангенса для некоторых углов (табл. 3.3).

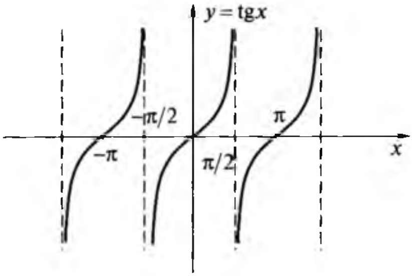

Рис. 3.7

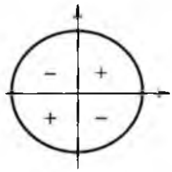

Рис. 3.8

Таблица 3.3

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\operatorname{tg}x$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	–

Свойство обратной тригонометрической функции:

$\operatorname{arctg}(-x) = -\operatorname{arctg}x$.

IV. Тригонометрическая функция $y = \operatorname{ctg}x$.

Используя транспортир и линейку, построим график котангенса (рис. 3.9):

1) область определения: $x \neq \pi n$, где $n \in Z$;

2) область изменения: $y \in R$;

3) промежутки знакопостоянства указаны на круге (рис. 3.10);

4) четность. Из построения на единичном круге следует, что функция нечетная: $\operatorname{ctg}(-x) = -\operatorname{ctg}x$;

5) периодичность. Период $T = \pi$;

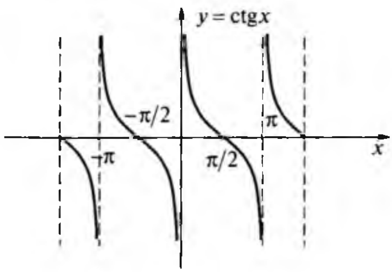

Рис. 3.9

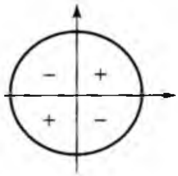

Рис. 3.10

6) таблица значений. В пределах углов от нуля до 90° укажем значения котангенса для некоторых углов (табл. 3.4).

Таблица 3.4

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\mathrm{ctg}\,x$	–	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0

Свойство обратной тригонометрической функции:

$\mathrm{arcctg}(-x) = \pi - \mathrm{arcctg}\,x$.

Что касается функций $y = \sec x$ и $y = \mathrm{cosec}\,x$, мы не будем включать их в основные тригонометрические функции. В задачах с помощью преобразований $\sec x = \dfrac{1}{\cos x}$ и $\mathrm{cosec}\,x = \dfrac{1}{\sin x}$ мы сможем избавиться от них. Графики этих функций построим, решая различные задачи на исследование функций и построение графиков.

3.3. ВЫЧИСЛЕНИЯ В ТРИГОНОМЕТРИИ

Тригонометрические формулы

Перечислим основные формулы тригонометрии, разделив их по названиям. Все другие формулы могут быть получены из основных с помощью простых тригонометрических преобразований.

Тождества:

$$\sin^2 x + \cos^2 x = 1;\ \mathrm{tg}\,x = \frac{\sin x}{\cos x};\ \mathrm{ctg}\,x = \frac{\cos x}{\sin x};\ \mathrm{tg}\,x \cdot \mathrm{ctg}\,x = 1.$$

Формулы сложения:

$\sin(x \pm y) = \sin x \cos y \pm \sin y \cos x;$

$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y;$

$$\mathrm{tg}(x \pm y) = \frac{\mathrm{tg}\,x \pm \mathrm{tg}\,y}{1 \mp \mathrm{tg}\,x \cdot \mathrm{tg}\,y}.$$

Частные случаи использования этих формул:

$\sin 2x = 2\sin x \cos x;$

$\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x;$

$$\mathrm{tg}\,2x = \frac{2\,\mathrm{tg}\,x}{1 - \mathrm{tg}^2 x};$$

$$\cos^2 x = \frac{1 + \cos 2x}{2}; \qquad \sin^2 x = \frac{1 - \cos 2x}{2}.$$

Формулы утроенного аргумента:

$$\sin 3x = 3\sin x - 4\sin^3 x;$$

$$\cos 3x = 4\cos^3 x - 3\cos x.$$

Формулы приведения

Формулы приведения позволяют избавиться от углов, кратных $\pi/2$, в аргументе вида $\frac{\pi}{2}n \pm x$ $(n \in Z)$ по определенному правилу. Если угол x отсчитывается от горизонтальной оси (углы $\pi n \pm x$), функция сохраняется, если от вертикальной $\left(\frac{\pi}{2} \pm \pi n \pm x\right)$, функция изменяется на кофункцию (например, синус на косинус). Знак правой части определяется тем знаком, который имеет исследуемая функция в рассматриваемой четверти.

Пример 3.1. Упростить $\sin\left(\frac{3\pi}{2} + x\right)$.

Решение. Угол x отсчитывается от вертикальной оси, следовательно, синус меняется на косинус. Исследуемая функция в четвертой четверти отрицательна, значит, в правой части перед косинусом следует поставить знак минус. Итак, $\sin\left(\frac{3\pi}{2} + x\right) = -\cos x$.

Формулы преобразования суммы функций в произведение и произведения в сумму:

$$\sin x + \sin y = 2\sin\frac{x + y}{2}\cos\frac{x - y}{2};$$

$$\sin x - \sin y = 2\sin\frac{x - y}{2}\cos\frac{x + y}{2};$$

$$\cos x + \cos y = 2\cos\frac{x + y}{2}\cos\frac{x - y}{2};$$

$$\cos x - \cos y = -2\sin\frac{x + y}{2}\sin\frac{x - y}{2};$$

$$\cos x \cos y = \frac{\cos(x - y) + \cos(x + y)}{2};$$

$$\sin x \sin y = \frac{\cos(x-y) - \cos(x+y)}{2};$$

$$\sin x \cos y = \frac{\sin(x-y) + \sin(x+y)}{2}.$$

3.4. ТРИГОНОМЕТРИЧЕСКИЕ УРАВНЕНИЯ И НЕРАВЕНСТВА. ОСНОВНЫЕ ТРИГОНОМЕТРИЧЕСКИЕ УРАВНЕНИЯ

Ниже рассмотрим основные уравнения (теоретические задачи), к которым сводится решение любого тригонометрического уравнения.

1. $\sin x = a$.

Чтобы уравнение имело решения, должно выполняться условие $|a| \le 1$. Решение удобно проводить на единичном круге (рис. 3.11). На оси синусов отмечаем значение a и проводим через эту точку горизонтальную линию до пересечения с окружностью. Полученные точки окружности соединяем с центром. Тот угол, который оказывается в первой или четвертой четверти, обозначается как $\arcsin a$. Угол, заканчивающийся во второй или четвертой четверти, вычисляется из геометрических соображений. Он оказывается равным $\pi - \arcsin a$. В точках A и B заканчиваются не только углы, обозначенные стрелками, но и отличающиеся от них на полное число оборотов, т.е. на 2π. Поэтому общее решение уравнения будет иметь вид

$$\left[\begin{array}{l} x = \arcsin a + 2\pi n, \\ x = \pi - \arcsin a + 2\pi k, \end{array}\right. \quad \text{где } n, k \in Z.$$

Рассмотрим некоторые частные случаи.

Пример 3.2. Решить уравнение $\sin x = -\frac{1}{2}$.

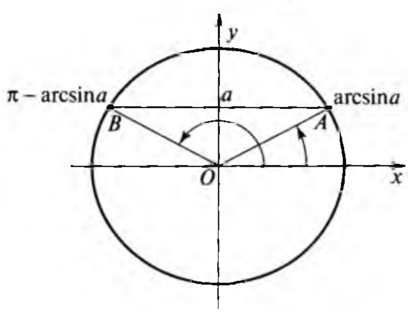

Рис. 3.11

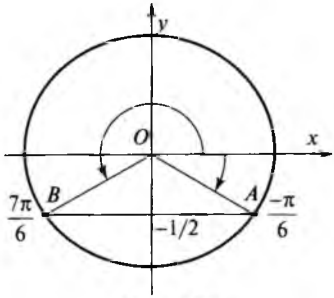

Рис. 3.12

Решение. Отметив на оси синусов точку $-\frac{1}{2}$ (рис. 3.12) и проведя горизонталь, получим на окружности точки A и B, в которых заканчиваются множества углов, описываемых уравнением. Выделим среди них углы, например $-\frac{\pi}{6}$ и $\frac{7\pi}{6}$. Общее решение будет таким: $\begin{array}{l} x = -\frac{\pi}{6} + 2\pi n, \\ x = \frac{7\pi}{6} + 2\pi n, \end{array}$ где $n \in Z$,

или в краткой форме

$$x = (-1)^n \cdot \left(-\frac{\pi}{6}\right) + \pi n = (-1)^{n+1}\frac{\pi}{6} + \pi n.$$

Точно так же рассуждая, можно для уравнения $\sin x = 0$ получить решение $x = \pi n$, для уравнения $\sin x = 1$ — решение $x = \frac{\pi}{2} + 2\pi n$ и т.д.

2. $\cos x = a$.

Решение уравнения может быть получено при выполнении условия $|a| \le 1$. На единичном круге отметим на оси косинусов точку с координатой a (рис. 3.13). Проведем вертикальную линию до пере-

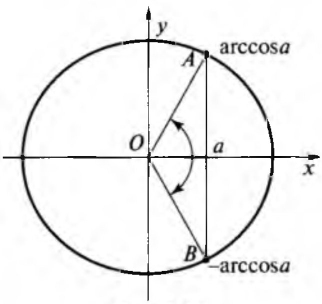

Рис. 3.13

сечения с окружностью в точках A и B. Эти точки соединим с центром. Один из образовавшихся углов, который заканчивается в первой четверти, обозначается как $\arccos x$. Другой угол отличается только знаком. Общее решение имеет вид $x = \pm\arccos x + 2\pi n$, где $n \in Z$.

Пример 3.3. Решить уравнение $\cos x = -\frac{1}{2}$.

Решение. На оси косинусов (рис. 3.14) поставим точку $-\frac{1}{2}$ и проведем вертикальную линию до пересечения с окружностью (точки A и B). Проведем радиусы в эти точки и стрелками укажем нужные нам углы. Из геометрических соображений ясно, что угол, заканчивающийся во второй четверти, равен $\pi - \frac{\pi}{3} = \frac{2\pi}{3}$. Очевидно, другой угол равен $-\frac{2\pi}{3}$. Общее решение $x = \pm\frac{2\pi}{3} + 2\pi n$, где $n \in Z$.

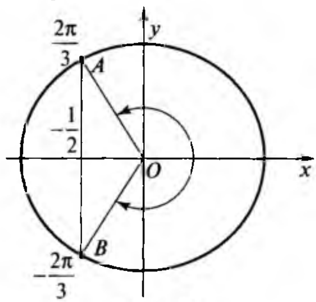

Рис. 3.14

Рассматривая уравнение $\cos x = -1$, легко получить решение $x = \pi + 2\pi n$, где $n \in Z$.

Решение уравнения $\cos x = 0$ следует запомнить: $x = \frac{\pi}{2} + \pi n$, где $n \in Z$.

3. $\operatorname{tg} x = a$.

Уравнение имеет решение при $a \in R$. Построим единичный круг и проведем ось тангенсов (рис. 3.15). Отложим на этой оси точку, соответствующую числу a, и проведем через данную точку и центр окружности прямую. Она дважды пересечет окружность (точки A и B). Стрелками укажем соответствующие углы. Угол в первой или четвертой четверти обозначается как $\operatorname{arctg} a$. Другой угол отличается ровно на π радиан (180°). Поэтому решение удобно записать в виде $x = \operatorname{arctg} x + \pi n$, где $n \in Z$.

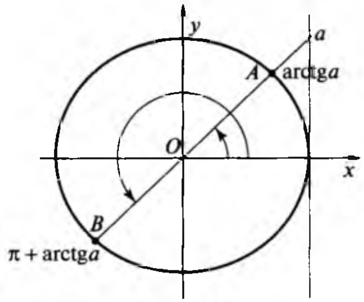

Рис. 3.15

4. $\operatorname{ctg} x = a$.

Действуем аналогично пункту 3. Но проводим вместо оси тангенсов ось котангенсов (рис. 3.16). Угол в первой или второй четверти обозначаем как $\operatorname{arcctg} a$, угол в третьей или четвертой четверти определяется как $\pi + \operatorname{arcctg} a$. Общее решение имеет вид $x = \operatorname{arcctg} x + \pi n$, где $n \in Z$.

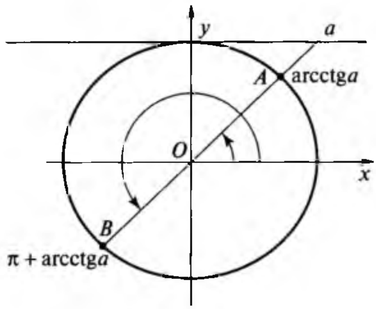

Рис. 3.16

5. $\sin x + \cos x = a$.

Укажем способ решения и проведем исследование решения в зависимости от a. Предположение, что величина a может изменяться в пределах от –2 до +2, неверно, поскольку нет такого угла, для которого синус и косинус принимали бы свои крайние значения одновременно. Умножим обе части равенства на $\frac{\sqrt{2}}{2}$, получим

$$\frac{\sqrt{2}}{2}\sin x + \frac{\sqrt{2}}{2}\cos x = \frac{\sqrt{2}}{2}a \quad \text{или} \quad \cos\frac{\pi}{4}\cdot\sin x + \sin\frac{\pi}{4}\cdot\cos x = \frac{\sqrt{2}}{2}a.$$

Свернем левую часть в синус суммы $\sin\left(x+\frac{\pi}{4}\right)=\frac{\sqrt{2}}{2}a$.

Чтобы уравнение имело решение, должно выполняться условие $\left|\frac{\sqrt{2}}{2}a\right|\le 1$ или $-\sqrt{2}\le a\le\sqrt{2}$. Решение можно записать в общем виде: $x=-\frac{\pi}{4}+(-1)^n\arcsin\frac{\sqrt{2}}{2}a+\pi n$, где $n\in Z$.

З а м е ч а н и е. Очевидно, что способ работает при любом знаке перед синусом или косинусом.

6. $\sin x+\sqrt{3}\cos x=a$.

Так же получим решение уравнения, попутно проведя исследование в зависимости от a. Умножим обе части равенства на $\frac{1}{2}$. Тогда

$$\frac{1}{2}\sin x+\frac{\sqrt{3}}{2}\cos x=\frac{a}{2}\Leftrightarrow\cos\frac{\pi}{3}\sin x+\sin\frac{\pi}{3}\cos x=\frac{a}{2}\Leftrightarrow$$

$$\Leftrightarrow\sin\left(x+\frac{\pi}{3}\right)=\frac{a}{2}.$$

Последнее уравнение имеет решение, если $\left|\frac{a}{2}\right|\le 1$, или $-2\le a\le 2$.

Оно имеет вид $x=-\frac{\pi}{3}+(-1)^n\arcsin\frac{a}{2}+\pi n$, где $n\in Z$.

Тригонометрические преобразования

Преобразования тригонометрических выражений требуют хорошего знания формул и внимательного отношения к равносильности преобразований.

Пример 3.4. Доказать тождество $\dfrac{\cos(3\pi-2\alpha)}{2\sin^2\left(\frac{5\pi}{4}+\alpha\right)}=\operatorname{tg}\left(\alpha-\frac{5\pi}{4}\right)$.

Решение. Преобразуем вначале левую часть равенства:

$$\frac{\cos(3\pi-2\alpha)}{2\sin^2\left(\frac{5\pi}{4}+\alpha\right)}=\frac{-\cos 2\alpha}{1-\cos\left(\frac{5\pi}{2}+\alpha\right)}=\frac{-\cos 2\alpha}{1+\sin 2\alpha}=$$

$$=-\frac{\cos^2\alpha-\sin^2\alpha}{\sin^2\alpha+\cos^2\alpha+2\sin\alpha\cos\alpha}=-\frac{(\cos\alpha-\sin\alpha)(\cos\alpha+\sin\alpha)}{(\cos\alpha+\sin\alpha)^2}=$$

$$=\frac{\sin\alpha-\cos\alpha}{\sin\alpha+\cos\alpha}.$$

Теперь преобразуем правую часть:

$$\operatorname{tg}\left(\alpha-\frac{5\pi}{4}\right)=\frac{\sin\left(\alpha-\frac{5\pi}{4}\right)}{\cos\left(\alpha-\frac{5\pi}{4}\right)}=\frac{\sin\alpha\cos\frac{5\pi}{4}-\cos\alpha\sin\frac{5\pi}{4}}{\cos\alpha\cos\frac{5\pi}{4}+\sin\alpha\sin\frac{5\pi}{4}}=$$

$$=\frac{\sin\alpha\left(-\frac{\sqrt{2}}{2}\right)-\cos\alpha\left(-\frac{\sqrt{2}}{2}\right)}{\cos\alpha\left(-\frac{\sqrt{2}}{2}\right)+\sin\alpha\left(-\frac{\sqrt{2}}{2}\right)}=\frac{\sin\alpha-\cos\alpha}{\sin\alpha+\cos\alpha}.$$

Левая часть равенства оказалась равной его правой части. Тождество доказано.

Тригонометрические вычисления

Пример 3.5. Зная, что $\sin x=\frac{1}{3}$ и $\frac{\pi}{2}<x<\pi$, вычислить $\cos x$, $\sin 2x$, $\cos 2x$, $\operatorname{ctg} 3x$.

Решение. Из тригонометрического тождества имеем $|\cos x|=$ $=\sqrt{1-\sin^2 x}=\sqrt{1-\frac{1}{9}}=\frac{2\sqrt{2}}{3}$. Поскольку угол x лежит во второй четверти, где $\cos x<0$, то $-\cos x=\frac{2\sqrt{2}}{3}$, или $\cos x=-\frac{2\sqrt{2}}{3}$.

Разложим функции $\sin 2x$ и $\cos 2x$ по формулам двойного аргумента:

$$\sin 2x=2\sin x\cos x=2\cdot\frac{1}{3}\cdot\left(-\frac{2\sqrt{2}}{3}\right)=-\frac{4\sqrt{2}}{9};$$

$$\cos 2x=1-2\sin^2 x=1-2\cdot\frac{1}{9}=\frac{7}{9}.$$

Формулу, выражающую $\operatorname{ctg} 3x$ через $\operatorname{tg} x$ или $\operatorname{ctg} x$, получим разложением этих функций по формулам тройного угла:

$$\operatorname{ctg} 3x=\frac{\cos 3x}{\sin 3x}=\frac{4\cos^3 x-3\cos x}{3\sin x-4\sin^3 x}=\frac{4\left(-\frac{2\sqrt{2}}{3}\right)^3+3\frac{2\sqrt{2}}{3}}{3\frac{1}{3}-4\left(\frac{1}{3}\right)^3}=-\frac{10\sqrt{2}}{23}.$$

Решение тригонометрических уравнений

Решение тригонометрического уравнения представляет собой проведение тригонометрических преобразований с целью приведения уравнения к одному из перечисленных основных уравнений. Поскольку в тригонометрии достаточно формул, многие из них могут быть использованы в данной задаче. Однако произвольное их использование чаще всего приводит задачу в тупик, и только одна или две дают верное направление преобразованию. Нахождению верного пути могут помочь некоторые рекомендации по проведению преобразований. Перечислив рекомендации, сопроводим их иллюстрирующими примерами.

1. Разложение уравнения на множители.
2. Приведение уравнения к одной функции и введение новой переменной.
3. Преобразование суммы функций в произведение или произведения в сумму.
4. Свертывание выражений после умножения на $\frac{\sqrt{2}}{2}, \frac{1}{2}, \frac{1}{\sqrt{a^2+b^2}}$.
5. Сведение к однородному уравнению.

Пример 3.6. Решить уравнение $2\cos 2x = \sqrt{6}(\cos x - \sin x)$.

Решение. Воспользуемся разложением на множители:

$$2\cos 2x - \sqrt{6}(\cos x - \sin x) = 0 \Leftrightarrow$$
$$\Leftrightarrow 2(\cos^2 x - \sin^2 x) - \sqrt{6}(\cos x - \sin x) = 0 \Leftrightarrow$$
$$\Leftrightarrow (\cos x - \sin x)\left(2(\cos x + \sin x) - \sqrt{6}\right) = 0:$$

а) $\cos x - \sin x = 0$. Это частный случай основного тригонометрического уравнения 5. Разделим обе части на $\cos x$:

$$\operatorname{tg} x = 1 \Leftrightarrow x = \frac{\pi}{4} + \pi n, \ n \in Z;$$

б) $2(\cos x + \sin x) - \sqrt{6} = 0 \Leftrightarrow \cos x + \sin x = \frac{\sqrt{6}}{2}$. Уравнение также приведено к основному. Условие существования решения $\left|\frac{\sqrt{6}}{2}\right| \le \sqrt{2}$ выполнено. Умножим обе части равенства на $\frac{\sqrt{2}}{2}$:

$$\frac{\sqrt{2}}{2}\cos x + \frac{\sqrt{2}}{2}\sin x = \frac{\sqrt{2}}{2}\frac{\sqrt{6}}{2} \Leftrightarrow \sin\left(x + \frac{\pi}{4}\right) = \frac{\sqrt{3}}{2} \Leftrightarrow$$

$$\Leftrightarrow \left[\begin{aligned} & x + \frac{\pi}{4} = \frac{\pi}{3} + 2\pi n; \\ & x + \frac{\pi}{4} = \pi - \frac{\pi}{3} + 2\pi k, \end{aligned}\right. \quad n, k \in Z \Leftrightarrow \left[\begin{aligned} & x = \frac{\pi}{12} + 2\pi n; \\ & x = \frac{5\pi}{12} + 2\pi k. \end{aligned}\right.$$

Пример 3.7. Решить уравнение $8\cos^4 x - \cos 4x = 1$.

Решение. Уменьшим показатель в первом слагаемом и коэффициент перед x во втором в два раза:

$$8(\cos^2 x)^2 - \cos 2(2x) = 1 \Leftrightarrow 8\left(\frac{1+\cos 2x}{2}\right)^2 - (2\cos^2 2x - 1) = 1.$$

Введем новую переменную $\cos 2x = y$. Получим

$$8\left(\frac{1+y}{2}\right)^2 - (2y^2 - 1) = 1 \Leftrightarrow 2 + 4y + 2y^2 - 2y^2 + 1 = 1 \Leftrightarrow y = -\frac{1}{2}.$$

Перейдем к старой переменной:

$$\cos 2x = -\frac{1}{2} \Leftrightarrow 2x = \pm\frac{2\pi}{3} + 2\pi n, \quad n \in Z \Leftrightarrow x = \pm\frac{\pi}{3} + \pi n.$$

Пример 3.8. Решить уравнение $\cos 3x \cdot \cos 4x + \sin 2x \cdot \sin 5x = \frac{\cos 2x + \cos 4x}{2}$.

Решение. Преимущественным преобразованием при наличии произведения функций с разными аргументами является преобразование в сумму:

$$\frac{\cos(-x) + \cos 7x}{2} + \frac{\cos(-3x) - \cos 7x}{2} = \frac{\cos 2x + \cos 4x}{2} \Leftrightarrow$$

$$\Leftrightarrow \cos x + \cos 3x = \cos 2x + \cos 4x.$$

Для суммы функций чаще других срабатывает преобразование в произведение:

$$2\cos\frac{x+3x}{2}\cos\frac{x-3x}{2} = 2\cos\frac{2x+4x}{2}\cos\frac{2x-4x}{2} \Leftrightarrow$$

$$\Leftrightarrow 2\cos x(\cos 2x - \cos 3x) = 0 \Leftrightarrow -4\cos x \sin\left(-\frac{x}{2}\right)\sin\frac{5x}{2} = 0:$$

а) $\cos x = 0 \Leftrightarrow x = \frac{\pi}{2} + 2\pi n, \quad n \in Z;$

б) $\sin\frac{x}{2} = 0 \Leftrightarrow x = 2\pi k, \quad k \in Z;$

в) $\sin\frac{5x}{2} = 0 \Leftrightarrow x = \frac{2\pi}{5}l, \quad l \in Z.$

Множество $x = 2\pi k$ включено во множество $x = \frac{2\pi}{5}l$, поэтому отдельно может не выписываться.

Ответ: $x = \frac{\pi}{2} + 2\pi n,\ x = \frac{2\pi}{5}l, n, l \in Z.$

Пример 3.9. Решить уравнение $\sin 8x - \cos 6x = \sqrt{3}(\sin 6x + \cos 8x)$.

Решение. Умножение уравнения на $\frac{1}{2}$ и перекомпоновка слагаемых позволит свернуть левую и правую части уравнения:

$$\frac{1}{2}\sin 8x - \frac{1}{2}\cos 6x = \frac{\sqrt{3}}{2}\sin 6x + \frac{\sqrt{3}}{2}\cos 8x \Leftrightarrow$$

$$\Leftrightarrow \frac{1}{2}\sin 8x - \frac{\sqrt{3}}{2}\cos 8x = \frac{\sqrt{3}}{2}\sin 6x + \frac{1}{2}\cos 6x \Leftrightarrow$$

$$\Leftrightarrow \sin\left(8x - \frac{\pi}{3}\right) = \sin\left(6x + \frac{\pi}{6}\right).$$

Если $\sin f(x) = \sin\varphi(x)$, легко получить, что $f(x) = \varphi(x) + 2\pi n$ и $f(x) = \pi - \varphi(x) + 2\pi n$, $n \in Z$. Воспользуемся этими результатами:

$$\left[\begin{array}{ll} 8x - \frac{\pi}{3} = 6x + \frac{\pi}{6} + 2\pi n, & n \in Z; \\ 8x - \frac{\pi}{3} = \pi - \left(6x + \frac{\pi}{6}\right) + 2\pi k, & k \in Z \end{array}\right. \Leftrightarrow \left[\begin{array}{l} x = \frac{\pi}{4} + \pi n; \\ x = \frac{\pi}{12} + \frac{\pi}{7}k. \end{array}\right.$$

Пример 3.10. Решить уравнение $12\sin^2 x + 3\sin 2x - 2\cos^2 x = 2$.

Решение. Легко убедиться в однородности этого уравнения:

$$12\sin^2 x + 6\sin x\cos x - 2\cos^2 x = 2(\sin^2 x + \cos^2 x) \Leftrightarrow$$

$$\Leftrightarrow 5\sin^2 x + 3\sin x\cos x - 2\cos^2 x = 0.$$

Уравнение можно разделить без потери корней на $\cos^2 x$. В самом деле, если бы $\cos^2 x = 0$, то наше уравнение стало бы иметь вид $5\sin^2 x = 0$. Но нет такого угла, для которого одновременно выполняются условия $\cos x = 0$ и $\sin x = 0$.

Итак, перейдем к тангенсу:

$$5\,\mathrm{tg}^2 x + 3\,\mathrm{tg}\,x - 2 = 0.$$

Корни уравнения: $\mathrm{tg}\,x_1 = -1$; $\mathrm{tg}\,x_2 = \frac{2}{5}$.

Окончательные решения: $\left[\begin{array}{l} x = -\frac{\pi}{4} + \pi n; \\ x = \mathrm{arctg}\frac{2}{5} + \pi k, \end{array}\right. \quad n, k \in Z.$

Решение тригонометрических неравенств

Решение тригонометрического неравенства делится на два этапа: нахождение корней и определение промежутков между корнями. Надо помнить следующее:

1) решение тригонометрического неравенства представляет собой двойное неравенство $a < x < b$;

2) написав неравенство $a < x < b$, следует убедиться, что $a < b$.

Пример 3.11. Решить неравенство $\sin x > \frac{1}{2}$.

Решение. На единичном круге (рис. 3.17) укажем некоторые решения уравнения $\sin x = \frac{1}{2}$: $x = \frac{\pi}{6}$ и $x = \frac{5\pi}{6}$. Поскольку $\sin x > \frac{1}{2}$, на оси синусов следует взять точки с координатами от $\frac{1}{2}$ до 1. Им соответствуют углы, заканчивающиеся на меньшей дуге AB. Следовательно, решение неравенства имеет вид

$$\frac{\pi}{6} + 2\pi n < x < \frac{5\pi}{6} + 2\pi n,\ n \in Z.$$

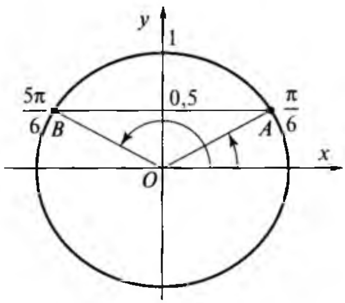

Рис. 3.17

3.5. ОБРАТНЫЕ ТРИГОНОМЕТРИЧЕСКИЕ ФУНКЦИИ

Рассмотрим функцию $y = 2x + 1$. Найдем переменную x: $x = \frac{y-1}{2}$. В этой зависимости $x = x(y)$ мы задаем значения переменной y и вычисляем значения переменной x. Если построить зависимости $x = \frac{y-1}{2}$ и $y = 2x + 1$, то, очевидно, получим одну и ту же совокупность точек, т.е. графики будут совпадать. Было бы привычней независимую переменную обозначать через x, зависимую — через y. Поэтому сделаем замену переменных и получим $y = \frac{x-1}{2}$. Назовем эту функцию обратной по отношению к функции $y = 2x + 1$.

В общем виде найдем аргумент x у функции $y = f(x)$, используя для этой цели специальное обозначение arc: $x = \text{arc}f(y)$. Сделав замену, будем иметь $y = \text{arc}f(x)$. Полученная функция $y = \text{arc}f(x)$ называется обратной к функции $y = f(x)$.

Перейдем к тригонометрии. Для функции $y = \sin x$ найдем обратную. Выразим аргумент x: $x = \arcsin y$. После замены переменных получим $y = \arcsin x$ (табл. 3.5). Построим график этой зависимости. Мы пока ничего о ней не знаем, кроме способа ее получения. Поэтому воспользуемся обратным переходом и получим $x = \sin y$. Эту зависимость мы и построим по точкам. Ее график приведен на рис. 3.18. Отличительная особенность зависимости — каждому значению переменной x соответствует бесконечное множество значений y, т.е. определение функции не выполнено. Чтобы можно было использовать эту зависимость как функцию, ограничим область переменной y. Наиболее удобное ограничение: $|y| \le \frac{\pi}{2}$. Тогда при изменении $-1 \le x \le 1$ каждому значению аргумента x ставится в соответствие одно значение переменной y из области $|y| \le \frac{\pi}{2}$ и величина y есть функция величины x.

Таблица 3.5

x	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\arcsin x$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$

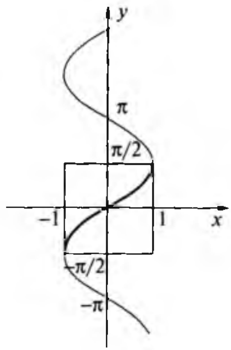

Рис. 3.18

Укажем важное свойство функции: для всех $x \in [-1; 1]$ функция $y = \arcsin x$ является возрастающей.

Составим таблицу некоторых значений $y = \arcsin x$ (см. табл. 3.5).

На рис. 3.18 изображен жирной линией в окне график обратной функции $y = \arcsin x$.

Аналогично определяется обратная функция для $y = \cos x$. Находим: $x = \arccos y$; меняем переменные: $y = \arccos x$; строим зависимость: $x = \cos y$. Задаем ограничения на переменную y: $0 \le y \le \pi$.

Важным свойством определенной таким образом функции $y = \arccos x$ является ее убывание для всех $x \in [-1; 1]$.

Составим также таблицу некоторых значений $y = \arccos x$ (табл. 3.6).

График обратной функции $y = \arccos x$ при условии $|x| \le 1$ и $0 \le y \le \pi$ изображен на рис. 3.19.

Таблица 3.6

x	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\arccos x$	$\frac{\pi}{2}$	$\frac{\pi}{3}$	$\frac{\pi}{4}$	$\frac{\pi}{6}$	0

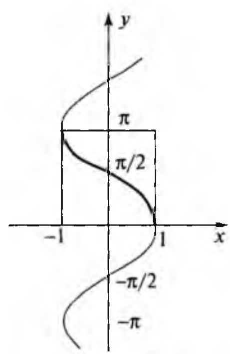

Рис. 3.19

Определим функцию, обратную тангенсу. Из равенства $y = \operatorname{tg} x$ найдем: $x = \operatorname{arctg} y$; заменим переменные: $y = \operatorname{arctg} x$; построим: $x = \operatorname{tg} y$. Из множества кривых выберем ту, для которой $|y| < \frac{\pi}{2}$. Приведем таблицу некоторых значений $y = \operatorname{arctg} x$ (табл. 3.7). График обратной функции $y = \operatorname{arctg} x$ при условии $x \in R$ и $|y| < \frac{\pi}{2}$ изображен на рис. 3.20 жирной линией. Функция является возрастающей.

Таблица 3.7

x	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$
$\operatorname{arctg} x$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$

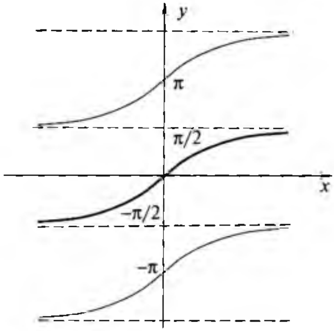

Рис. 3.20

Наконец, определим функцию $y = \operatorname{arcctg} x$, которая является обратной котангенсу.

Приведем таблицу некоторых значений $y = \operatorname{arcctg} x$ (табл. 3.8).

Таблица 3.8

x	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$
$\operatorname{arcctg} x$	$\frac{\pi}{2}$	$\frac{\pi}{3}$	$\frac{\pi}{4}$	$\frac{\pi}{6}$

График обратной функции $y = \operatorname{arcctg} x$ при условии $x \in R$ и $0 < y < \pi$ изображен на рис. 3.21 жирной линией. Функция убывает при всех $x \in (0; \pi)$.

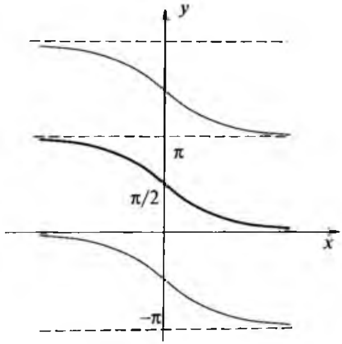

Рис. 3.21

Основные формулы для обратных функций

1. $\arcsin(-x) = -\arcsin x$; $\arccos(-x) = \pi - \arccos x$; $\operatorname{arctg}(-x) = -\operatorname{arctg} x$; $\operatorname{arcctg}(-x) = \pi - \operatorname{arcctg} x$	2. $\sin(\arcsin x) = x$; $\cos(\arccos x) = x$; $\operatorname{tg}(\operatorname{arctg} x) = x$; $\operatorname{ctg}(\operatorname{arcctg} x) = x$
3. $\sin(\arccos x) = \sqrt{1 - x^2}$; $\cos(\arcsin x) = \sqrt{1 - x^2}$; $\operatorname{tg}(\operatorname{arcctg} x) = \dfrac{1}{x}$; $\operatorname{ctg}(\operatorname{arctg} x) = \dfrac{1}{x}$	4. $\arcsin x + \arccos x = \dfrac{\pi}{2}$; $\operatorname{arctg} x + \operatorname{arcctg} x = \dfrac{\pi}{2}$

Пример 3.12. Вычислить $\operatorname{tg}^2\left(\frac{1}{2}\arccos\frac{2}{3}\right)$.

Решение. Вспомним, что $\operatorname{tg}^2\frac{x}{2} = \dfrac{\sin^2\frac{x}{2}}{\cos^2\frac{x}{2}} = \dfrac{1-\cos x}{1+\cos x}$.

Поэтому $\operatorname{tg}^2\left(\frac{1}{2}\arccos\frac{2}{3}\right) = \dfrac{1-\cos\left(\arccos\frac{2}{3}\right)}{1+\cos\left(\arccos\frac{2}{3}\right)} = \dfrac{\frac{1}{3}}{\frac{5}{3}} = \dfrac{1}{5}$.

Глава 4
АРИФМЕТИЧЕСКАЯ И ГЕОМЕТРИЧЕСКАЯ ПРОГРЕССИИ

4.1. АРИФМЕТИЧЕСКАЯ ПРОГРЕССИЯ

Арифметической прогрессией называется такая последовательность чисел, в которой каждое число, начиная со второго, равно предыдущему, сложенному с одним и тем же числом, постоянным для этой последовательности. Это постоянное число называется *разностью* прогрессии, а числа, составляющие прогрессию, — ее членами.

Пусть члены прогрессии $a_1, a_2, \ldots, a_n$. Тогда формула n-го члена арифметической прогрессии имеет вид

$$a_n = a_1 + (n-1)d,$$

где d — разность прогрессии.

Сумма n первых членов арифметической прогрессии

$$S_n = \frac{a_1 + a_n}{2} n = \frac{2a_1 + (n-1)d}{2} n.$$

Свойство арифметической прогрессии:

$$\frac{a_1 + a_3}{2} = a_2, \qquad \frac{a_{n-1} + a_{n+1}}{2} = a_n, \qquad \frac{a_{n-k} + a_{n+k}}{2} = a_n,$$

где $k < n$.

4.2. ГЕОМЕТРИЧЕСКАЯ ПРОГРЕССИЯ

Геометрической прогрессией называется такая последовательность чисел, в которой каждое число, начиная со второго, равно предыдущему, умноженному на одно и то же число, постоянное для этой последовательности и не равное нулю. Это постоянное число называется *знаменателем* прогрессии, а числа, составляющие прогрессию, — ее членами.

Пусть члены прогрессии $b_1, b_2, \ldots, b_n$. Тогда формула n-го члена геометрической прогрессии имеет вид

$$b_n = b_1 q^{n-1},$$

где q — знаменатель прогрессии.

Сумма n первых членов геометрической прогрессии

$$S_n = \frac{b_n q - b_1}{q - 1} = \frac{b_1(q^n - 1)}{q - 1},$$

где $q \neq 0$, $q \neq 1$.

Свойство геометрической прогрессии:

$$b_1 b_3 = b_2^2, \quad b_{n-1}b_{n+1} = b_n^2, \quad b_{n-k}b_{n+k} = b_n^2,$$

где $k < n$.

Геометрическая прогрессия называется *бесконечно убывающей*, если число членов ее бесконечно, а $|q| < 1$.

Сумма бесконечного числа членов такой прогрессии равна

$$S = \frac{b_1}{1-q}.$$

Пример 4.1. Найти сумму всех трехзначных чисел, делящихся на 7.

Решение. Минимальное трехзначное число 100 делится на 7 с остатком 2. Следовательно, на 7 без остатка делится число 98, следующее за ним — 105. Последнее трехзначное число 999 делится на 7 с остатком 5. Значит, число 994 делится на 7 без остатка.

Составим ряд трехзначных чисел, делящихся на 7: 105, 112, ..., 994. Он представляет собой арифметическую прогрессию, у которой $a_1 = 105$, $d = 7$, $a_n = 994$. Найдем число членов n из уравнения

$$a_n = 105 + 7(n-1) = 994.$$

Отсюда $n = 128$. Сумма всех трехзначных чисел, делящихся на 7,

$$S_{128} = \frac{105+994}{2}128 = 70\,336.$$

Пример 4.2. Найти все бесконечно убывающие геометрические прогрессии, у которых каждый член в 4 раза больше суммы всех последующих членов.

Решение. Рассмотрим задачу в общем виде. Пусть b_n в 4 раза больше суммы последующих членов. Тогда имеет место равенство

$$b_n = 4(b_{n+1} + b_{n+2} + \ldots).$$

Представим $b_{n+1}, b_{n+2}, \ldots, b_{n+1}, b_{n+2}, \ldots$ через b_n:

$$b_n = 4(b_n q + b_n q^2 + \ldots) = 4b_n(q + q^2 + \ldots).$$

Разделим обе части равенства на $b_n \neq 0$ и найдем в правой части равенства сумму бесконечно убывающей прогрессии с первым членом q:

$$1 = 4\frac{q}{1-q}.$$

Решая уравнение, получим $q = \frac{1}{5}$. Общий вид бесконечно убывающей геометрической прогрессии, у которой каждый член в 4 раза больше суммы всех последующих членов, таков:

$$b_1, b_1 \cdot \frac{1}{5}, b_1 \cdot \frac{1}{5^2}, b_1 \cdot \frac{1}{5^3}, \ldots,$$

где $b_1 \in R$, $b_1 \neq 0$.

Глава 5
ЭЛЕМЕНТАРНЫЕ ФУНКЦИИ И ГРАФИКИ

Основными функциями называются следующие аналитически заданные функции.

Степенная функция $y = x^{\alpha}$, где α — любое действительное число a.

Показательная функция $y = a^x$, где $a > 0$, $a \neq 1$.

Логарифмическая функция $y = \log_a x$, где $a > 0$, $a \neq 1$.

Тригонометрические функции: $y = \sin x$, $y = \cos x$, $y = \mathrm{tg}\, x$, $y = \mathrm{ctg}\, x$. $y = \sec x$, $y = \mathrm{cosec}\, x$.

Обратные тригонометрические функции: $y = \arcsin x$, $y = \arccos x$, $y = \mathrm{arctg}\, x$, $y = \mathrm{arcctg}\, x$.

В математическом анализе для иллюстрации некоторых идей используется функция $y = \mathrm{sgn}(x)$, которая определяется следующим образом: $y = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$

График этой функции представлен на рис. 5.0.

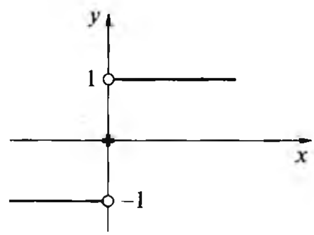

Рис. 5.0

Элементарной называется функция, которая может быть задана одной формулой вида $y = f(x)$, где стоящее справа выражение составлено из основных функций и постоянных при помощи конечного числа операций сложения, вычитания, умножения, деления и взятия функции от функции.

5.1. ТЕОРИЯ ЛИНЕЙНОЙ ФУНКЦИИ

Функция вида $y = kx + b$ называется *линейной* и описывает прямую линию. Легко найти *расстояние между точками* на прямой:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Углом наклона прямой называется угол между положительным направлением оси x и прямой, отсчитанный против часовой стрелки. Пусть прямая $y = kx + b$ проходит через точки с координатами (x_1, y_1) и (x_2, y_2) (рис. 5.1). Найдем тангенс угла наклона:

$$\operatorname{tg}\alpha = \frac{y_2 - y_1}{x_2 - x_1} = \frac{kx_2 + b - (kx_1 + b)}{x_2 - x_1} = \frac{k(x_2 - x_1)}{x_2 - x_1} = k.$$

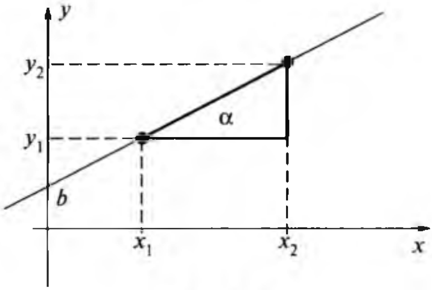

Рис. 5.1

Итак, коэффициент k перед аргументом в уравнении прямой линии $y = kx + b$ имеет геометрический смысл тангенса угла наклона. Из рис. 5.1 видно, что геометрический смысл числа b — ордината пересечения прямой и оси y.

5.2. ЗАДАЧИ НА ЭЛЕМЕНТЫ ИССЛЕДОВАНИЯ ФУНКЦИЙ

Пример 5.1. Найти область определения функции $y = \sqrt{\arcsin(\log_2 x)}$.

Решение. Область определения или область допустимых значений аргумента (ОДЗ) находится из решения системы

$$\begin{cases} \arcsin(\log_2 x) \geq 0, \\ -1 \leq \log_2 x \leq 1, \\ x > 0, \end{cases}$$

откуда следует $\begin{cases} \log_2 x \geq 0, \\ 0 \leq \log_2 x \leq 1, \\ x > 0, \end{cases}$ или $1 \leq x \leq 2$.

Пример 5.2. Найти область изменения или область допустимых значений функции $y = \sqrt{-x^2 + x + 2}$.

Решение. Заметим, что $y \geq 0$. Далее воспользуемся следующим приемом. Посредством преобразований получим величину x как функцию величины y и найдем ОДЗ на y. Для этого возведем равенство в квадрат и решим квадратное уравнение относительно x:

$$y^2 = -x^2 + x + 2 \Leftrightarrow -x^2 + x + 2 - y^2 = 0 \Leftrightarrow x_{1,2} = \frac{1 \pm \sqrt{9 - 4y^2}}{2}.$$

Очевидны ограничения на переменную y: $9 - 4y^2 \geq 0$, или $|y| \leq \frac{3}{2}$.

С учетом неравенства $y \geq 0$ окончательно получим $0 \leq y \leq \frac{3}{2}$.

Пример 5.3. Исследовать на четность функцию $y = \lg\frac{1+x}{1-x}$.

Решение. Проверим для этой функции справедливость равенства $f(-x) = f(x)$ или $f(-x) = -f(x)$, что устанавливает четность или нечетность функции:

$$y(-x) = \lg\frac{1-x}{1+x} = \lg\left(\frac{1+x}{1-x}\right)^{-1} = -\lg\frac{1+x}{1-x} = -y(x).$$

Ответ: функция $y = \lg\frac{1+x}{1-x}$ является нечетной.

Пример 5.4. Найти асимптоты графика функции $y = \frac{2-x}{x-1}$.

Решение. Преобразуем числитель, выделив в нем ту часть, которая нацело делится на $x + 1$:

$$y = \frac{2-x}{x-1} = \frac{1-(x-1)}{x-1} = \frac{1}{x-1} - 1.$$

При стремлении аргумента x к бесконечности функция y стремится к -1. Прямая $y = -1$ является горизонтальной асимптотой. Когда $x \to 1$, функция y неограниченно возрастает по модулю. Следовательно, прямая $x = 1$ есть вертикальная асимптота.

Ответ: горизонтальная асимптота $y = -1$; вертикальная асимптота $x = 1$.

5.3. ПОСТРОЕНИЕ ГРАФИКОВ ЭЛЕМЕНТАРНЫХ ФУНКЦИЙ

Рассмотрим в этом параграфе некоторые общие подходы к построению графиков функций. Построение графика исключительно по точкам используется только для простейших функций либо таких, общих ход которых мы знаем, требуется лишь уточнить локальное местоположение. В общем случае необходимо провести предварительное исследование, может быть, провести некоторые преобразования с функцией, чтобы выявить характерные особенности графика. Иначе график может быть построен неверно.

Графики на основе линейной функции

Пример 5.5. Построить график функции $y = \frac{x^2 - 3x + 2}{|x-2|}$.

Решение. Разложим числитель на множители и раскроем модуль:

$$y = \frac{(x-2)(x-1)}{|x-2|} = \left[\begin{array}{ll} x-1, & x > 2, \\ 1-x, & x < 2. \end{array}\right.$$

В области $x > 2$ построим график функции $y = x - 1$, в области $x < 2$ — график $y = 1 - x$ (рис. 5.2).

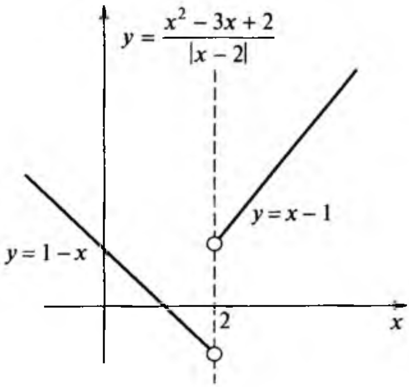

Рис. 5.2

Графики на основе квадратного трехчлена

Пример 5.6. Построить график функции $|y| = |x^2 - 3|x| - 2|$.

Решение. Функция $x^2 - 3|x|$ является четной, следовательно, симметричной относительно оси Oy. Функция $|y|$ четна относительно переменной y, а значит, симметрична относительно оси Ox. Таким образом, можно рассмотреть поведение функции в первой четверти, а затем построенный график отразить справа налево, потом результат сверху вниз (рис. 5.3).

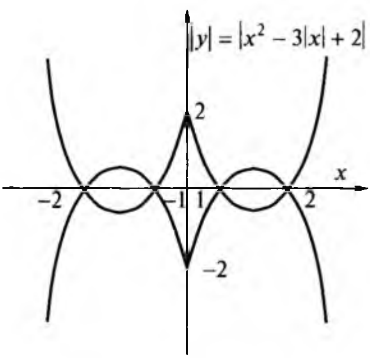

Рис. 5.3

В области $x \geq 0$, $y \geq 0$ функция имеет вид $y = |x^2 - 3x + 2| = |(x - 2) \times \times (x - 1)|$:

1) $0 \leq x \leq 1$ или $x \geq 2 \Rightarrow y = x^2 - 3x + 2 = \left(x - \frac{3}{2}\right)^2 - \frac{1}{4}$;

2) $1 \leq x \leq 2 \Rightarrow y = -(x^2 - 3x + 2) = -\left(x - \frac{3}{2}\right)^2 + \frac{1}{4}$.

График окружности

Уравнение окружности радиуса R с центром в точке с координатами (x_0, y_0) имеет вид $(x - x_0)^2 + (y - y_0)^2 = R^2$. Однако уравнение окружности может быть записано иначе.

Пример 5.7. Построить график функции $y^2 - 4y + x^2 - 2x + 1 = 0$.

Решение. Выделим в равенстве полные квадраты:

$y^2 - 4y + 4 + x^2 - 2x + 1 = 4 \Leftrightarrow (y - 2)^2 = (x - 1)^2 = 2^2$.

Построим окружность радиуса 2 с центром в точке (1, 2) (рис. 5.4).

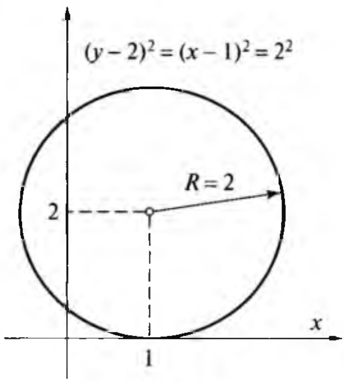

Рис. 5.4

Графики на основе гиперболы

Общий вид уравнения гиперболы $y = \dfrac{ax + b}{cx + d}$. Для построения графика гиперболы рекомендуется найти асимптоты и затем, ориентируясь на них, строить ветви гиперболы по нескольким точкам.

Пример 5.8. Построить график функции $|y| = \dfrac{1}{\frac{x - 1}{x}}$.

Решение. Функция симметрична относительно оси Ox и при условии $x \neq 0$, $y \geq 0$ может быть записана в виде $y = \dfrac{x}{x - 1}$. Выделим асимптоты:

$y = 1 + \dfrac{1}{x-1}$. Горизонтальная — $y = 1$, вертикальная — $x = 1$. Подставив несколько точек, нарисуем график выше оси абсцисс, который отразим вниз (рис. 5.5).

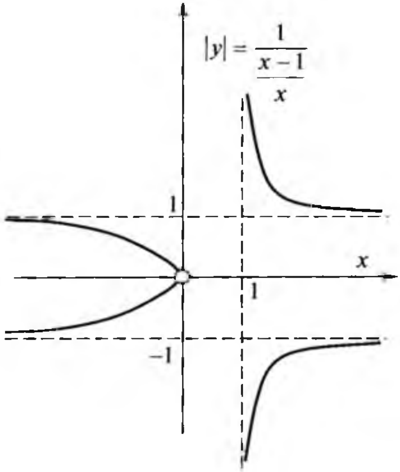

Рис. 5.5

Графики на основе показательной и логарифмической функций

Пример 5.9. Построить график функции $y = 2^x + \left(\dfrac{1}{2}\right)^x$.

Решение. График функции $y = 2^x + \left(\dfrac{1}{2}\right)^x$ получим, построив пунктиром два вспомогательных графика $y_1 = 2^x$ и $y_2 = \left(\dfrac{1}{2}\right)^x$, и затем сложим их ординаты при одинаковых значениях x (рис. 5.6).

Пример 5.10. Построить график функции $y = \log_x 2$.

Решение. График этой функции будем строить по такой схеме. Вначале приведем график функции $y = \log_x 2$ к виду $y = \log_2 x$, а уже на его основе построим $y = \dfrac{1}{\log_2 x}$ по частям (рис. 5.7).

I. Взяв несколько больших значений x, построим часть I графика.

II. При приближении аргумента x к единице справа функция $y = \log_2 x$ приближается к нулю, оставаясь положительной. Дробь $\dfrac{1}{\log_2 x}$ неограниченно возрастает (часть II).

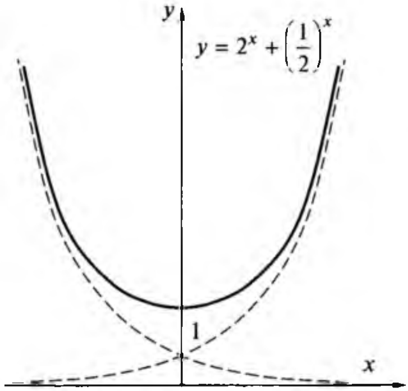

Рис. 5.6

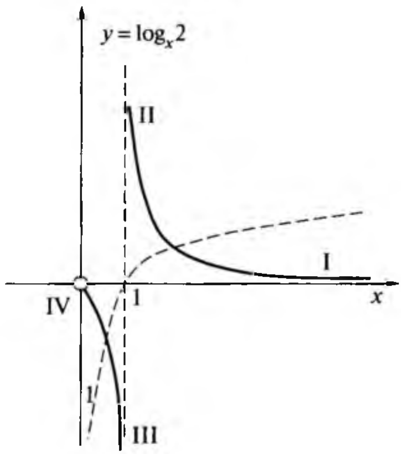

Рис. 5.7

III. При стремлении аргумента x к единице слева функция $y = \log_2 x$ приближается к нулю, оставаясь отрицательной. Дробь $\frac{1}{\log_2 x}$, будучи отрицательной, неограниченно убывает (часть III). Таким образом, прямая $x = 1$ является асимптотой графика.

IV. Когда $x \to 0$, функция $y = \log_2 x$ стремится к $-\infty$. Но тогда функция $y = \frac{1}{\log_2 x}$ будет стремиться к нулю, оставаясь отрицательной (часть IV).

Графики на основе тригонометрических функций

Пример 5.11. Построить график функции $y = 4\left(\sin^4\frac{x}{2} + \cos^4\frac{x}{2}\right)$.

Решение. Упростим правую часть:

$$4\left(\sin^4\frac{x}{2} + \cos^4\frac{x}{2}\right) = 4\left(\frac{1-\cos x}{2}\right)^2 + 4\left(\frac{1+\cos x}{2}\right)^2 =$$

$$= 2 + 2\cos^2 x = 2 + 2\frac{1+\cos 2x}{2} = 3 + \cos 2x.$$

График функции $y = 3 + \cos 2x$ получим из графика функции $y = \cos 2x$, сдвинутого вдоль оси ординат на три единицы вверх (рис. 5.8).

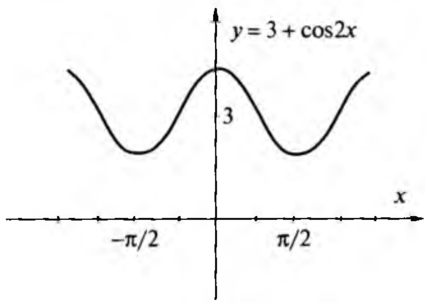

Рис. 5.8

Пример 5.12. Построить график функции $y = f(x)$, заданной неявно: $y^4 - (y\log_2 x)^2 + (2^x\log_2 x)^2 - 4^x y^2 = 0$.

Решение. Перед нами наглядный пример того, как необычайно трудно было бы строить график исключительно по точкам. Сгруппируем первое слагаемое со вторым, третье — с четвертым и вынесем общие множители за скобки:

$$y^2(y^2 - \log_2^2 x) - 4^x(y^2 - \log_2^2 x) = 0 \Leftrightarrow (y^2 - \log_2^2 x)(y^2 - 4^x) = 0 \Leftrightarrow$$

$$\Leftrightarrow \left[\begin{array}{l} y^2 - \log_2^2 x = 0, \\ y^2 - 4^x = 0, \end{array}\right. \Leftrightarrow \left[\begin{array}{l} |y| = |\log_2 x| = 0, \\ |y| = 2^x. \end{array}\right.$$

В области $x > 0$ построим графики функций $y = \pm\log_2 x$, $y = \pm 2^x$.

5.4. ПОСТРОЕНИЕ МНОЖЕСТВА ТОЧЕК

Как правило, построение множества точек основано на построении графиков функций, которые рассматриваются как границы некоторой области или нескольких областей. Равенство, в котором заключены две или более функций, может также рассматриваться как множество точек (см. пример 5.6).

Пример 5.13. На плоскости xOy изобразить множество точек, координаты x и y каждой из которых удовлетворяют условию

$$\sin(\pi|x| + \pi|y|) \cdot \sqrt{16 - x^2 - y^2} = 0.$$

Решение. Указав на ОДЗ $x^2 + y^2 \le 16$, приравняем нулю каждый из двух сомножителей:

$$\left[\begin{array}{l} \sin(\pi|x| + \pi|y|) = 0, \\ \sqrt{16 - x^2 - y^2} = 0 \end{array}\right. \Leftrightarrow \left[\begin{array}{l} \pi|x| + \pi|y| = \pi n, \\ 16 - x^2 - y^2 = 0, \end{array}\right. n \in Z \Leftrightarrow \left[\begin{array}{l} |x| + |y| = n, \\ x^2 + y^2 = 16. \end{array}\right.$$

Условие $x^2 + y^2 \le 16$ означает, что ОДЗ представляет собой все точки круга радиусом 4. В этой области жирной линией прорисуем окружность $x^2 + y^2 = 16$. Равенство $|x| + |y| = n$ при $n = 0$ дает точку в начале координат, при $n = 1, 2, 3, 4$ — квадраты, при $n = 5$ — хорды окружности радиуса 4, которые изображены на рис. 5.9.

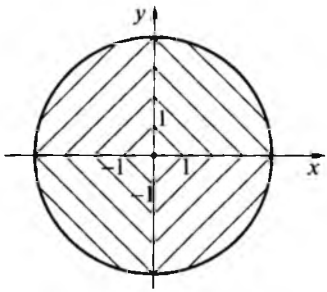

Рис. 5.9

Пример 5.14. Построить множество точек, удовлетворяющих системе неравенств $\begin{cases} x - \sqrt{1 - y^2} \ge 0, \\ x + |y| - 4 \le 0. \end{cases}$

Решение. Перепишем неравенство в виде $x \ge \sqrt{1 - y^2}$. Его решение эквивалентно решению системы

$$\begin{cases} x \ge 0, \\ |y| \le 1, \\ x^2 + y^2 \ge 1. \end{cases}$$

Множество точек представляет собой полосу: $|y| \le 1$ в правой полуплоскости; $x \ge 0$ вне круга; $x^2 + y^2 \ge 1$ (рис. 5.10).

Второе неравенство исходной системы $x + |y| - 4 \le 0$ приводится к виду

$$\begin{cases} x - 4 \le y \le 4 - x, \\ x \le 4. \end{cases}$$

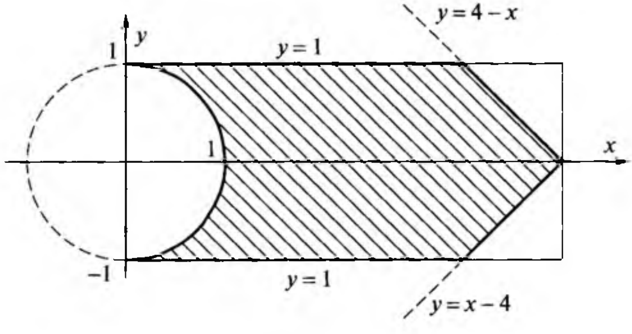

Рис. 5.10

5.5. ЗАДАЧИ ЭКОНОМИКИ И УПРАВЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

Задачи экономики и управления с использованием элементарной математики — один из наиболее распространенных типов задач в производственной деятельности. Они часто не требуют знаний, выходящих за пределы средней школы. Умение логически мыслить, из условий выводить следствия, не ошибаться в арифметических и алгебраических преобразованиях — условия правильного решения. Часто менеджеру приходится использовать проценты. Как известно, процент — это сотая часть числа. Если число a составляет b процентов от числа A, то справедливо равенство:

$$\frac{a}{A} = \frac{b\%}{100\%}, \text{ или } a = \frac{b\%}{100\%}A, \text{ или } b\% = \frac{a}{A}100\%, \text{ или } A = a\frac{100\%}{b\%}.$$

Пример 5.15. 90% стоимости пакета акций инвестора ценой в 22 млн руб. составляли бумаги компании А, после падения котировок ее акций — только 12%. Сколько теперь стоит пакет акций?

Решение. Найдем количество акций других компаний. 22 млн руб. × × 10% = 2,2 млн руб. Если после падения котировок пакет содержит 12% акций компании А, значит, он содержит 88% акций других компаний. Составим пропорцию:

2,2 млн руб. — 88%,

X млн руб. — 100%.

Из нее следует: $X = \dfrac{2{,}2 \cdot 100\%}{88\%} = 2{,}5$ млн руб.

Ответ: новая стоимость пакета акций 2,5 млн руб.

В текстовых экономических задачах используются сложные проценты. Пусть вклад в a ден. ед. положен в банк под p процентов в год. Сумма вклада через год станет равной

$$A_1 = a + a\frac{p\%}{100\%} = a\left(1 + \frac{p}{100}\right).$$

Через два года его величина составит

$$A_2 = a\left(1 + \frac{p}{100}\right) + a\left(1 + \frac{p}{100}\right)\frac{p\%}{100\%} = a\left(1 + \frac{p}{100}\right)^2.$$

Продолжая вычисления, получим через n лет величину вклада равной

$$A_n = a\left(1 + \frac{p}{100}\right)^n.$$

Мы получили формулу сложных процентов, по которой начисляются, например, проценты в коммерческом банке.

Пример 5.16. Сбербанк начисляет ежегодно 10% от суммы вклада. Через сколько лет внесенная сумма удвоится?

Решение. Пусть в банк положены x ден. ед. под 10% годовых. Через n лет вклад удвоится и станет равным $2a$. Следовательно, по формуле сложных процентов имеем

$$2a = a\left(1 + \frac{10\%}{100\%}\right)^n.$$

Разделим равенство на a и прологарифмируем по основанию 10:

$$n\lg\left(1 + \frac{10}{100}\right) = \lg 2 \text{ или } n = \frac{\lg 2}{\lg 1{,}1} \approx 7{,}3 \text{ года.}$$

ЛИТЕРАТУРА

1. *Бачурин В.А.* Задачи по элементарной математике и началам математического анализа. — М.: Физматлит, 2005.
2. *Выгодский М.Я.* Справочник по элементарной математике. — М.: Астрель: АСТ, 2001.
3. *Ермолицкий А.А.* Большой справочник по элементарной математике: Универсальное пособие. — Минск: Харвест, 2003.
4. *Куланин Е., Норин В., Федин С., Шевченко Ю.* 3000 конкурсных задач по математике. — М.: Айрис-Пресс, 2007.
5. Полный сборник решений задач для поступающих в вузы. Группа А / Под ред. *М.И. Сканави.* — М.: Мир и образование; Минск: Харвест, 2003. В 2 кн.
6. Полный сборник решений задач для поступающих в вузы. Группа Б / Под ред. *М.И. Сканави.* — М.: Мир и образование; Минск: Харвест, 2003. В 2 кн.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Раздел II
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Глава 6
ОСНОВНЫЕ ПОНЯТИЯ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ

6.1. ИДЕИ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Как уже указывалось в историческом очерке, понятия *предела, производной, дифференциала и интеграла*, появившиеся в XVII в., составили фундамент новой математической науки — математического анализа. Период от зарождения математики до введения Р. Декартом переменных величин можно назвать статическим периодом в развитии математики. Исследователи оперировали только числами, получали только численные решения. Буквы, которыми систематически начали пользоваться математики со времен арабского ученого Хорезми, не рассматривались в своем изменении. Р. Декарт придал буквам возможность непрерывно изменять свое значение в математическом выражении, положив начало динамическому периоду развития.

Обозначим буквой некоторую величину, которой придаются разные числовые значения. Обозначим другой буквой вычисляемое выражение. Тогда ее значение может быть вычислено для любого значения введенной в математическом выражении буквы. Со временем буква, которой придавалось определенное числовое значение, стала называться *аргументом*, а другая буква, названная *функцией*, вычислялась из математического выражения. Получающуюся зависимость одной переменной величины от другой Декарт стал изображать на так называемой координатной плоскости, о чем ниже.

Введение переменной величины, а также появление функции в добавление к имевшемуся понятию числа привели к возникновению интереса к исследованию функций. Был развит математический инструмент, позволяющий изучать функции, находить особенности их поведения, характерные точки. Золотые страницы этого раздела математики — теория бесконечно малых, пределов, производной, дифференциала и интеграла — написаны Исааком Ньютоном и Готфридом Лейбницем.

Первые шаги в этом направлении сделал создатель математического естествознания Исаак Ньютон. Изучая закономерности механического движения, он оперировал понятиями переменной величины (координаты движущейся точки) и скорости ее изменения. Пусть координата точки при ее движении со временем t описывается математическим выражением $y = y(t)$. За отрезок времени от момента t_1 до момента t_2, т.е. за время $t_2 - t_1$, точка пройдет путь $y(t_2) - y(t_1)$. Скорость перемещения на этом отрезке составит величину $v = \frac{y(t_2) - y(t_1)}{t_2 - t_1}$. Однако это средняя скорость на временном отрезке $t_2 - t_1$.

Будем приближать величину t_2 к t_1. Тогда при плавном движении точки значение функции $y(t_2)$ станет приближаться к значению функции $y(t_1)$. Средняя скорость движения точки будет определяться на все более короткой длине пройденного пути. Возникает вопрос: насколько близко могут быть взяты значения t_2 и t_1, чтобы можно было определить величину скорости? Этот вопрос оставался открытым, пока не был сделан предельный переход.

Пусть функция $y = x^2$ описывает некоторый процесс. Изменим аргумент на величину $x_2 - x_1$, которую обозначим как Δx. Тогда $x_2 = x_1 + \Delta x$. Значение функции изменится на величину

$$y_2 - y_1 = x_2^2 - x_1^2 = (x_1 + \Delta x)^2 - x_1^2.$$

Скорость изменения функции будет равна

$$v = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(x_1 + \Delta x)^2 - x_1^2}{\Delta x}. \quad (6.1)$$

Приближение значения x_2 к x_1 означает, что величина Δx стремится к нулю. Подстановка нуля вместо Δx в формулу (6.1) приводит к непонятному выражению $\frac{0}{0}$, из которого ничего определить нельзя. Вместо подстановки попробуем раскрыть скобки:

$$v = \frac{(x_1 + \Delta x)^2 - x_1^2}{\Delta x} = \frac{x_1^2 + 2x_1\Delta x + (\Delta x)^2 - x_1^2}{\Delta x} = \frac{2x_1\Delta x + (\Delta x)^2}{\Delta x}.$$

Упростим выражение, вынеся за скобки Δx, затем сократим этот сомножитель со знаменателем:

$$v = \frac{2x_1\Delta x + (\Delta x)^2}{\Delta x} = \frac{\Delta x(2x_1 + \Delta x)}{\Delta x} = 2x_1 + \Delta x. \quad (6.2)$$

Теперь можно приближать величину Δx к нулю и даже положить ее равной нулю, не получая проблем. Смысл полученного выражения $v = 2x_1$ в том, что скорость изменения функции $y = x^2$ определена в точке x_1. Впоследствии операция нахождения скорости изменения функции при стремлении Δx к нулю была оформлена символически следующим образом:

$$v = \lim_{\Delta x \to 0} \frac{(x_1 + \Delta x)^2 - x_1^2}{\Delta x} = 2x_1. \qquad (6.3)$$

Величина скорости v изменения функции y была названа *производной функцией* от исходной функции или просто *производной*. В настоящее время наиболее употребляемое в математическом анализе обозначение для скорости изменения функции есть y'.

Возвратимся к задаче о движении. Скорость перемещения точки может быть определена в любой момент времени t_1. Достаточно найти предел при стремлении t_2 к t_1 или $t_2 - t_1 = \Delta t \to 0$. Это уже не средняя скорость на отрезке, а мгновенная скорость, с которой точка проходит через координату $y(t_1)$ в момент времени t_1.

Выражение (6.3) перепишем в общем виде для функции $y = y(x)$:

$$y' = \lim_{\Delta x \to 0} \frac{y(x + \Delta x) - y(x)}{\Delta x}. \qquad (6.4)$$

Главные понятия математического анализа содержатся в этой формуле. Устремляя Δx к нулю, мы вступаем в область *бесконечно малых* величин. Понятие *предела* и умение совершать предельные переходы на основе знания теории бесконечно малых — важнейшие шаги освоения математического анализа. Умение находить производную через нахождение пределов, а затем с использованием таблицы производных — дальнейшие важные шаги. Эти основные понятия, на основе которых построена теория их применения и различного рода приложения, составляют первый раздел математического анализа — дифференциальное исчисление.

Немецкий ученый Г. Лейбниц начал свои исследования бесконечно малых несколькими годами позже. Подход Лейбница был иным. Основным понятием явились так называемые дифференциалы — бесконечно малые приращения переменных величин. Он нашел, что связь между бесконечно малым приращением аргумента dx и соответствующим ему бесконечно малым приращением функции $df(x)$ имеет вид произведения производной на приращение аргумента:

$$df(x) = f'(x) \cdot dx.$$

Эта символика сохранилась до настоящего времени. Он развил теорию дифференциалов, вследствие чего само новое исчисление получило название *дифференциального*. Найти дифференциал функции все равно, что найти производную, ибо когда производная найдена, то дифференциал функции получается через простое умножение на приращение dx независимой переменной. Используя дифференциалы, Лейбниц пришел к тем же результатам, что и Ньютон. Тем самым И. Ньютон и Г. Лейбниц, независимо проводя исследования, впервые в общем виде рассмотрели основные положения нового исчисления.

Спор из-за чести открытия дифференциального исчисления, возникший между английскими и континентальными учеными, будоражил умы математиков того времени и рассматривался на заседаниях английского королевского общества. Высоко уважая друг друга, эти великие ученые обменивались почтительными письмами, разрешая спор решением сложных задач по дифференциальному исчислению.

При изучении дифференциального исчисления студенты познакомятся с обоими подходами в математическом анализе. Будет найдена связь между дифференциалами функций и их производными. При изучении дифференциального исчисления функций одной переменной обычно следуют рассуждениям И.Ньютона, в изучении функций многих переменных упор делается на понятии дифференциала.

Прямой задаче нахождения производной была противопоставлена обратная задача нахождения переменных по заданным соотношениям между производными. Другими словами, по уравнению, содержащему производные, решалась задача восстановления функции, названной в таких задачах *первообразной*.

Пусть дана производная функция $y'(x) = 2x$. Требуется по производной восстановить исходную функцию $y(x)$ (найти первообразную). Умея отыскивать производные, можно такую функцию подобрать (см. выше). Она выражается формулой $y(x) = x^2$. Найденная функция оказывается не единственной. Например, функция вида $y(x) = x^2 + 1$ имеет такую же производную. Следовательно, обратная задача: по производной найти исходную функцию — имеет не единственное решение.

Была развита теория нахождения первообразных, изучены различные приложения. Самое популярное из них — нахождение площади фигур, ограниченных произвольными линиями. Усилия Ньютона и Лейбница привели к появлению интегрального исчисления с его основной формулой нахождения первообразной функции по ее производной — формулой Ньютона—Лейбница.

6.2. ПРЕДЕЛ И НЕПРЕРЫВНОСТЬ ФУНКЦИИ

Понятие предела функции является одним из основных в математическом анализе. Определения производной, интеграла, непрерывности и т.д. основаны на использовании предела.

Пусть функция $f(x)$ определена в некоторой окрестности X точки x_0, кроме, быть может, самой точки x_0. На рис. 6.1 изображен график функции $y = f(x)$. Предположим, что функция приближается при $x \to x_0$ к числу b. Возьмем произвольное сколь угодно малое число $\varepsilon > 0$. Окружим число b ε-окрестностью $(b - \varepsilon;\ b + \varepsilon)$. Найдем на оси OX такую окрестность точки x_0: $(x_0 - \delta;\ x_0 + \delta)$, при попадании в которую значений аргумента x соответствующие значения функции попадут в ε-окрестность числа b. При уменьшении числа ε интервал $(b - \varepsilon;\ b + \varepsilon)$ будет стягиваться к числу b. Соответствующий ему интервал $(x_0 - \delta;\ x_0 + \delta)$ будет стягиваться к числу x_0. Эту математическую процедуру запишем коротко словами следующим образом: пределом функции $f(x)$ при приближении аргумента x к точке x_0 является число b. Запись с помощью математических значков выглядит так:

$$\lim_{x \to x_0} f(x) = b.$$

Поскольку в математике приняты самые короткие и отточенные формулировки, исчерпывающе выражающие определенную идею,

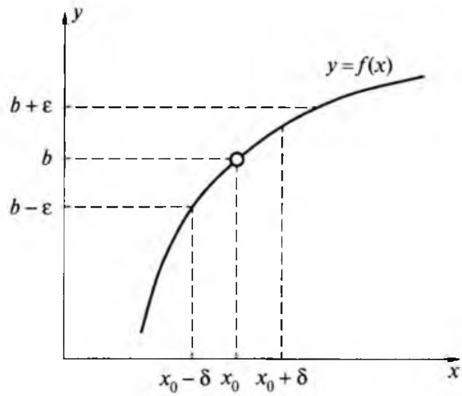

Рис. 6.1

где каждое слово стоит на своем месте, где убрать хотя бы одно слово означает — исказить мысль, то точное математическое определение предела звучит следующим образом.

Определение. Число b называется ***пределом функции*** $f(x)$ при $x \to x_0$, если для любой сходящейся к x_0 последовательности $x_1, x_2, ..., x_n, ...$ $(x_i \neq x_0)$ соответствующая последовательность $f(x_1), f(x_2), ..., f(x_n)$ значений функции $f(x)$ сходится к b.

Переменная x может стремиться к числу x_0 не только по произвольному закону, но и, например, только справа: $x \to x_0 + 0$, или только слева: $x \to x_0 - 0$. Переменная x может стремиться к бесконечности: $x \to \infty$, $x \to +\infty$, $x \to -\infty$. Соответствующие значения функции могут приближаться к числу b по произвольному закону ($y \to b$), сверху ($y \to b + 0$), снизу ($y \to b - 0$). Функция может неограниченно возрастать ($y \to +\infty$), убывать ($y \to -\infty$), неограниченно возрастать по модулю ($|y| \to +\infty$).

Это приводит к существованию определений 36 пределов функции в точке.

$$\lim_{x \to \begin{bmatrix} x_0 \\ x_0+0 \\ x_0-0 \\ \infty \\ +\infty \\ -\infty \end{bmatrix}} f(x) = \begin{bmatrix} b \\ b+0 \\ b-0 \\ \infty \\ +\infty \\ -\infty \end{bmatrix}.$$

Бесконечный предел. Функция имеет бесконечный предел при $x \to x_0$, если $\lim\limits_{x \to x_0} f(x) = \infty$. Такая функция называется бесконечно большой в окрестности точки x_0.

Односторонний предел. Функция имеет односторонний предел справа при $x \to x_0 + 0$, если $\lim\limits_{x \to x_0+0} f(x) = b$.

Предел функции на бесконечности. Функция имеет предел при $x \to -\infty$, если $\lim\limits_{x \to -\infty} f(x) = b$.

Функция $y = \alpha(x)$ называется *бесконечно малой функцией* при $x \to x_0$, если $\lim\limits_{x \to x_0} \alpha(x) = 0$. Например, функция $y = x^2$ является бесконечно малой при $x \to 0$, а $y = \dfrac{1}{x}$ является бесконечно малой при $x \to \infty$.

Свойства бесконечно малых функций

1. Если $\alpha(x)$ и $\beta(x)$ — бесконечно малые функции при $x \to x_0$, то функции $\alpha(x) \pm \beta(x)$ есть также бесконечно малые при $x \to x_0$.

2. Если функция $\alpha(x)$ является бесконечно малой функцией при $x \to x_0$, а функция $f(x)$ ограничена в окрестности точки x_0, то произведение $\alpha(x) \cdot f(x)$ есть бесконечно малая при $x \to x_0$.

3. Пусть функция $f(x)$ определена в некоторой окрестности точки x_0, кроме, быть может, самой точки x_0, и $\lim\limits_{x \to x_0} f(x) = b$. Тогда $f(x) = b + \alpha(x)$, где $\alpha(x)$ — бесконечно малая функция при $x \to x_0$.

«Связь» между существованием функции в точке x_0 и существованием предела при $x \to x_0$

В определении предела функции в точке x_0 значение функции в точке x_0 не влияет на предел функции в этой точке. Функция может быть не определена в этой точке. Проиллюстрируем это с помощью графиков, рассмотрев несколько случаев. Пусть $x \geq 0$ и $x \to 2$.

1. Пусть дана функция $y = x$. График изображен на рис. 6.2.

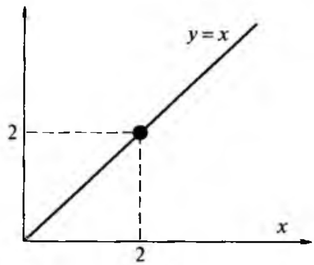

Рис. 6.2

При $x = 2$ функция существует и равна 2. При $x \to 2$ предел существует и $\lim\limits_{x \to 2} x = 2$. Значение функции и предел в точке $x = 2$ равны.

Таким образом, $\begin{cases} f(2) = 2, \\ \lim\limits_{x \to 2} f(x) = 2. \end{cases}$

2. Пусть дана функция $y = \begin{cases} x & \text{при } x \neq 2, \\ 1 & \text{при } x = 2 \end{cases}$. График изображен на рис. 6.3. При $x = 2$ функция существует, ее значение равно единице. При $x \to 2$ предел существует и $\lim\limits_{x \to 2} x = 2$. Значение функции и предел в точке $x = 2$ существуют, но не равны.

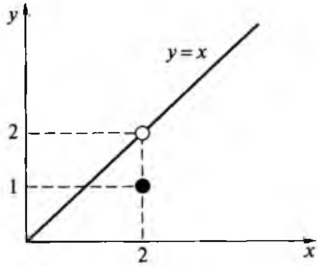

Рис. 6.3

Таким образом, $\begin{cases} f(2) = 1, \\ \lim\limits_{x \to 2} f(x) = 2. \end{cases}$

3. Пусть дана функция $y = \begin{cases} x & \text{при } x \neq 2, \\ \text{не существует} & \text{при } x = 2. \end{cases}$ График изображен на рис. 6.4. При $x = 2$ функция не существует. При $x \to 2$ предел существует и $\lim\limits_{x \to 2} x = 2$. Итак, $\begin{cases} f(2) = \text{не сущ.}, \\ \lim\limits_{x \to 2} f(x) = 2. \end{cases}$

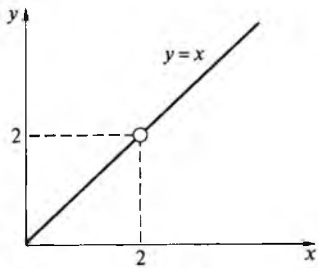

Рис. 6.4

4. Пусть дана функция $y = \dfrac{x(x-2)}{|x-2|}$. График изображен на рис. 6.5. При $x = 2$ функция не существует. При $x \to 2$ предел не существует, но существует предел справа $\lim\limits_{x \to 2+0} x = 2$ и предел слева $\lim\limits_{x \to 2-0} x = -2$.

Таким образом, $\begin{cases} f(2) = \text{не сущ.}, \\ \lim\limits_{x \to 2+0} f(x) = 2, \quad \lim\limits_{x \to 2-0} f(x) = -2. \end{cases}$

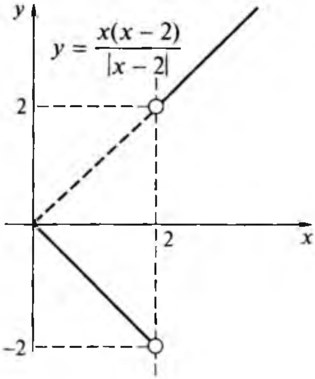

Рис. 6.5

5. Пусть дана функция $y = \begin{cases} \dfrac{x(x-2)}{|x-2|} & \text{при } x \neq 2, \\ 2 & \text{при } x = 2. \end{cases}$ График изображен на рис. 6.6. При $x = 2$ функция существует и равна 2. При $x \to 2$ предел не существует, но существует предел справа $\lim\limits_{x \to 2+0} x = 2$ и предел слева $\lim\limits_{x \to 2-0} x = -2$.

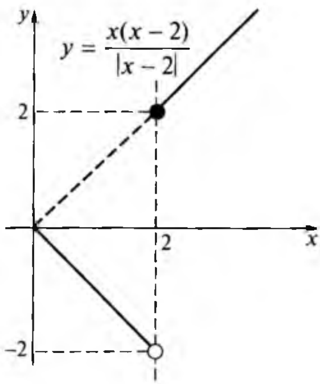

Рис. 6.6

Таким образом, $\begin{cases} f(2) = 2 \\ \lim\limits_{x \to 2+0} f(x) = 2, \ \lim\limits_{x \to 2-0} f(x) = -2. \end{cases}$

Вывод, который можно сделать из рассмотрения этих случаев: никакой связи между существованием функции в точке x_0 и существованием предела при $x \to x_0$ нет.

Свойства пределов функций

Приведем свойства пределов без доказательства.

1. Если функция $f(x)$ имеет предел в точке x_0, то этот предел единственный.

2. Если $f(x) \le \varphi(x)$ для всех x из некоторой окрестности точки x_0, кроме, быть может, самой точки x_0, и каждая из функций $f(x)$ и $\varphi(x)$ в точке x_0 имеет предел, то $\lim\limits_{x\to x_0} f(x) \le \lim\limits_{x\to x_0} \varphi(x)$.

3. Если $f(x) \le \varphi(x) \le a(x)$ для всех x в некоторой окрестности точки x_0, кроме, быть может, самой точки x_0, и функции $f(x)$ и $a(x)$ в точке x_0 имеют один и тот же предел b, то и функция $\varphi(x)$ в точке x_0 имеет предел, равный этому же числу b.

Пусть функции $f(x)$ и $\varphi(x)$ определены в некоторой окрестности точки x_0, кроме, быть может, самой точки x_0, и пусть существуют $\lim\limits_{x\to x_0} f(x) = a$ и $\lim\limits_{x\to x_0} \varphi(x) = b$. Тогда:

4. $\lim\limits_{x\to x_0} (f(x) \pm \varphi(x)) = \lim\limits_{x\to x_0} f(x) \pm \lim\limits_{x\to x_0} \varphi(x) = a \pm b.$

5. $\lim\limits_{x\to x_0} (f(x) \cdot \varphi(x)) = \lim\limits_{x\to x_0} f(x) \cdot \lim\limits_{x\to x_0} \varphi(x) = a \cdot b.$

6. $\lim\limits_{x\to x_0} \dfrac{f(x)}{\varphi(x)} = \dfrac{\lim\limits_{x\to x_0} f(x)}{\lim\limits_{x\to x_0} \varphi(x)} = \dfrac{a}{b}$, если $b \ne 0$.

7. $\lim\limits_{x\to x_0} kf(x) = ka.$

При выполнении некоторых дополнительных условий (непрерывности функций, о чем ниже) имеют место свойства:

8. $\lim\limits_{x\to x_0} f(\varphi(x)) = f\left(\lim\limits_{x\to x_0} \varphi(x)\right) = f(b).$

9. $\lim\limits_{x\to x_0} (f(x))^{\varphi(x)} = \lim\limits_{x\to x_0} (f(x))^{\lim\limits_{x\to x_0} \varphi(x)} = a^b.$

Если предельные значения оказываются равными 0 или ∞, то могут возникнуть неопределенности различных видов. При этом надо четко отделять возникающие неопределенности от внешне похожих вполне определенных получаемых величин — определенностей. При вычислении пределов могут появиться неопределенности вида:

$$\frac{0}{0},\ \frac{\infty}{\infty},\ \infty-\infty,\ 0^0,\ (+\infty)^0,\ 1^\infty$$

и др. Соотношения вида:

$$\frac{0}{\infty},\ \frac{\infty}{0},\ \infty+\infty,\ 0^{\pm\infty},\ (+\infty)^{\pm\infty}$$

являются примерами определенностей. Действительно,

$$\left[\frac{0}{\infty}\right]\Rightarrow 0,\quad \left[\frac{\infty}{0}\right]\Rightarrow\infty,\quad [0^{+\infty}]\Rightarrow 0,\quad [+\infty]^{-\infty}\Rightarrow 0$$

и т.д.

Для преодоления этих трудностей (раскрытия неопределенностей) существуют различные методы. Некоторые сводятся к определенным очевидным преобразованиям. Другие требуют теоретического обоснования.

Первый замечательный предел

Пусть x — угол, выраженный в радианах. Тогда $\lim\limits_{x\to 0}\frac{\sin x}{x}=1$.

Для доказательства рассмотрим круг радиусом R с центром в точке O. Пусть OB — подвижный радиус, образующий угол x $\left(0<x<\frac{\pi}{2}\right)$ с осью OX (рис. 6.7). Из геометрических соображений следует, что площадь треугольника AOB меньше площади сектора

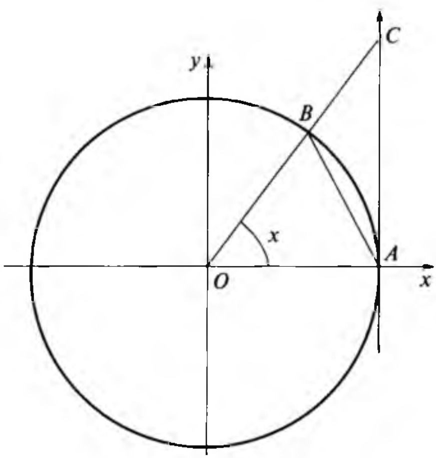

Рис. 6.7

AOB, которая, в свою очередь, меньше площади прямоугольного треугольника *AOC*, т.е.

$$\frac{1}{2}R^2\sin x < \frac{1}{2}R^2 x < \frac{1}{2}R^2\operatorname{tg} x.$$

Разделим части двойного неравенства на $\frac{1}{2}R^2\sin x$, получим $1 < \frac{x}{\sin x} < \frac{1}{\cos x}$, откуда $1 > \frac{\sin x}{x} > \cos x$. Функция $y = \cos x$ имеет при $x \to 0$ $\lim\limits_{x\to 0}\cos x = 1$. Таким образом, две функции $f(x) = 1$ и $\varphi(x) = \cos x$ имеют при $x \to 0$ предел, равный единице. По третьему свойству пределов получаем $\lim\limits_{x\to 0}\frac{\sin x}{x} = 1$.

Второй замечательный предел

Бином Ньютона. Введем определение *факториала* как произведения n первых членов натурального ряда: $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n = n!$ Напишем выражение для суммы двух слагаемых в квадрате, затем в кубе в таком виде:

$$(a+b)^2 = a^2 + 2ab + b^2 = a^2 + \frac{2}{1!}ab + \frac{2\cdot 1}{2!}b^2,$$

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 = a^3 + \frac{3}{1!}a^2b + \frac{3\cdot 2}{2!}ab^2 + \frac{3\cdot 2\cdot 1}{3!}b^3.$$

При возведении в степень n суммы двух слагаемых получим формулу

$$(a+b)^n = a^n + \frac{n}{1!}a^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \ldots +$$
$$+ \frac{n(n-1)\ldots(n-k+1)}{k!}a^{n-k}b^k + \ldots + \frac{n(n-1)\ldots 2\cdot 1}{n!}b^n,$$

где коэффициент перед b^n, очевидно, равен единице. Эта формула называется *биномом Ньютона*.

Рассмотрим числовую последовательность $a_n = \left(1+\frac{1}{n}\right)^n$. Вычислим значения нескольких членов последовательности: $a_1 = 2$, $a_2 = 2{,}25$, $a_3 = 2{,}37$, $a_4 = 2{,}44$, $a_5 = 2{,}49$. Можно предположить, что последовательность является возрастающей. Для доказательства воспользуемся формулой бинома Ньютона:

$$a_n = \left(1+\frac{1}{n}\right)^n = 1 + \frac{n}{1!}\frac{1}{n} + \frac{n(n-1)}{2!}\frac{1}{n^2} + \ldots +$$
$$+ \frac{n(n-1)(n-2)\ldots(n-(n-1))}{n!}\frac{1}{n^n}.$$

Преобразуем полученное выражение:

$$a_n = 2+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\ldots+\frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{n-1}{n}\right).$$

При увеличении n увеличивается как число положительных слагаемых, так и величина каждого слагаемого; следовательно, последовательность $\{a_n\}$ является монотонно возрастающей. Эта последовательность ограничена сверху:

$$a_n < 2+\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{1}{n!} < 2+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{n-1}}.$$

Стоящая в правой части сумма, начиная со второго слагаемого, представляет собой сумму $(n-1)$ члена геометрической прогрессии:

$$S_{n-1} = \frac{b_1(q^{n-1}-1)}{q-1} = \frac{0{,}5(0{,}5^{n-1}-1)}{0{,}5-1} = 1-\frac{1}{2^{n-1}} < 1.$$

Поэтому

$$a_n = \left(1+\frac{1}{n}\right)^n < 2+S_{n-1} < 2+1 = 3.$$

Последовательность $\{a_n\}$ монотонно возрастает и ограничена сверху. Следовательно, она имеет предел. Теория числовых рядов позволяет вычислить его величину. Это иррациональное число, приблизительно равное 2,71828... . Обозначим его через e.

Можно доказать, что функция

$$y = \left(1+\frac{1}{x}\right)^x$$

не только при натуральном n, но при любом действительном числе $x \to +\infty$ или $x \to -\infty$ имеет предел, равный e: $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e$.

Число e (число Эйлера) было названо именем швейцарского математика и физика Эйлера (1707–1783), который свыше 30 лет прожил в Петербурге и нашел там благоприятные условия для своей научной деятельности. Число e используется при выводе производных от некоторых функций и играет важную роль в математическом анализе. К использованию числа e приводит анализ таких процессов, как рост народонаселения, распад радия, размножение бактерий и т.д.

Асимптотические равенства

Пусть переменная $x \to 0$. Некоторые элементарные функции ведут себя в окрестности нуля достаточно просто:

$\sin x \approx x.$

$\operatorname{tg} x \approx x.$

$\cos x \approx 1 - \dfrac{x^2}{2}.$

$\arcsin x \approx x.$

$\operatorname{arctg} x \approx x.$

$\ln(1 + x) \approx x.$

$e^x \approx 1 + x.$

$(1 + x)^p \approx 1 + px.$

Эти соотношения легко доказываются, но мы ограничимся одной иллюстрацией. Например, из первого замечательного предела $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1$ следует, что в малой окрестности нуля отношение $\dfrac{\sin x}{x}$ уже не будет равно единице, но, тем не менее, достаточно близко к этому числу. Иначе говоря, $\dfrac{\sin x}{x} \approx 1$. Отсюда $\sin x \approx x$. Другие соотношения можно рассмотреть как задачи.

Использование полученных асимптотических формул дает мощный метод раскрытия неопределенностей при вычислении пределов.

Пример 6.1. Найти $\lim\limits_{x\to 0}\dfrac{\operatorname{tg}10x - 5x}{2x}$.

Решение. При $x \to 0$ имеем неопределенность $\left[\dfrac{0}{0}\right]$. Воспользуемся заменой функции $\operatorname{tg}10x$ на асимптотически равную ей функцию $10x$. Получим

$$\lim_{x\to 0}\frac{\operatorname{tg}10x - 5x}{2x} = \lim_{x\to 0}\frac{10x - 5x}{2x} = \lim_{x\to 0}\frac{5x}{2x} = \frac{5}{2}.$$

Пример 6.2. Найти $\lim\limits_{x\to 0}\dfrac{a^x - 1}{x}$.

Решение. Здесь также неопределенность типа $\left[\dfrac{0}{0}\right]$. Представим $a^x = e^{\ln a^x} = e^{x\ln a}$ и заменим асимптотической формулой $1 + x\ln a$:

$$\lim_{x\to 0}\frac{a^x - 1}{x} = \lim_{x\to 0}\frac{e^{x\ln a} - 1}{x} = \lim_{x\to 0}\frac{1 + x\ln a - 1}{x} = \ln a.$$

Пример 6.3. Найти $\lim\limits_{x\to+\infty}\dfrac{\sqrt[3]{8x^3+5}}{x}$.

Решение. Для раскрытия неопределенности вида $\left[\dfrac{\infty}{\infty}\right]$ воспользуемся последней формулой в списке асимптотических равенств:

$$\lim_{x\to+\infty}\frac{\sqrt[3]{8x^3+5}}{x}=\lim_{x\to+\infty}\frac{\sqrt[3]{8x^3\left(1+\dfrac{5}{8x^3}\right)}}{x}=\lim_{x\to+\infty}\frac{2x\left(1+\dfrac{5}{8x^3}\right)^{\frac{1}{3}}}{x}=$$

$$=2\lim_{x\to+\infty}\left(1+\frac{1}{3}\cdot\frac{5}{8x^3}\right)=2.$$

Непрерывность функции

Понятие непрерывности функции, так же как и понятие предела, является одним из основных понятий математического анализа. Когда устанавливалось понятие предела функции $f(x)$ при $x\to x_0$, то считалось, что $x\neq x_0$. А если функция существовала в этой точке, то это не учитывалось. Рассмотрим теперь случай, когда $f(x_0)$ существует, причем $\lim\limits_{x\to x_0}f(x)=f(x_0)$. Итак, пусть функция $f(x)$ определена в самой точке x_0 и ее некоторой окрестности.

Функция $f(x)$ называется ***непрерывной в точке*** x_0, если:

а) $f(x)$ существует в точке x_0;

б) $f(x)$ имеет предел в точке x_0;

в) $\lim\limits_{x\to x_0}f(x)=f(x_0)$.

Пример 6.4. Исследовать непрерывность в точке $x = 0$ следующих функций:

1) $y=\dfrac{x}{x}$; 2) $y=\operatorname{sgn}(x)$; 3) $y=|\operatorname{sgn}(x)|$, 4) $y=\begin{cases}\dfrac{x}{x}, & x\neq 0,\\ 1, & x=0.\end{cases}$

Решение.

1. В точке $x=0$ функция $y=\dfrac{x}{x}$ не является непрерывной, так как нарушено первое условие непрерывности — существование $f(0)$ (рис. 6.8).

2. В точке $x = 0$ функция $y=\operatorname{sgn}(x)$ (рис. 6.9) не является непрерывной — первое условие непрерывности выполнено, $f(0)$ существует и равно нулю, второе условие нарушено — не существует $\lim\limits_{x\to 0}\operatorname{sgn}(x)$, хотя существуют конечные односторонние пределы $\lim\limits_{x\to+0}\operatorname{sgn}(x)=1$ и $\lim\limits_{x\to-0}\operatorname{sgn}(x)=-1$.

3. В точке $x = 0$ функция $y=|\operatorname{sgn}(x)|$ не является непрерывной (рис. 6.10). Первые два условия непрерывности выполнены, поскольку

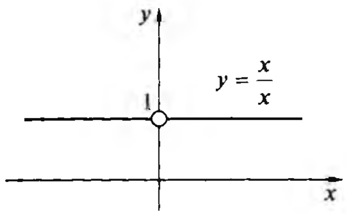

Рис. 6.8

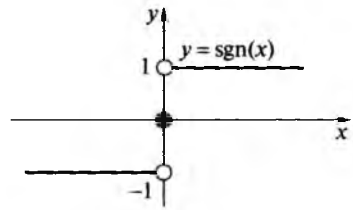

Рис. 6.9

$f(0) = |\mathrm{sgn}(0)| = 0,$

$\lim\limits_{x\to 0} f(x) = \lim\limits_{x\to 0} |\mathrm{sgn}(x)| = 1,$

но нарушено третье условие:

$\lim\limits_{x\to 0} |\mathrm{sgn}(x)| \neq f(0).$

4. В точке $x = 0$ функция $y = \begin{cases} \dfrac{x}{x}, & x \neq 0, \\ 1, & x = 0 \end{cases}$ (рис.6.11) непрерывна, так как выполнены все три условия непрерывности: $f(0) = 1$, $\lim\limits_{x\to 0} f(x) = 1$ и $\lim\limits_{x\to 0} f(x) = f(0)$.

График непрерывной функции в точке $x = 0$ рисуется при прохождении этой точки без отрыва ручки от листа бумаги.

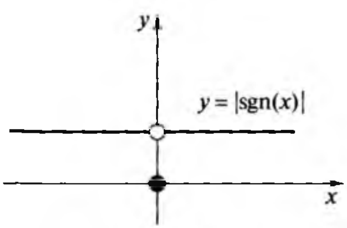

Рис. 6.10

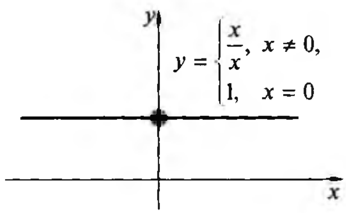

Рис. 6.11

Свойства непрерывных функций

Пусть функции $f(x)$ и $g(x)$ определены в точке x_0 и ее некоторой окрестности и непрерывны в точке x_0.

Если, кроме того, $f(x)$ непрерывна в окрестности точки x_0 и $f(x_0) \neq 0$, то существует окрестность точки x_0, в которой функция не обращается в нуль и сохраняет свой знак (знак числа $f(x_0)$).

Функции $f(x) \pm g(x)$, $f(x) \cdot g(x)$, $\dfrac{f(x)}{g(x)}$ (при дополнительном условии $g(x_0) \neq 0$) непрерывны в точке x_0.

Сложная функция $f(g(x))$ непрерывна в точке x_0, т.е.

$$\lim_{x\to x_0} f(g(x)) = f\left(\lim_{x\to x_0} g(x)\right) = f(g(x_0)).$$

Можно доказать, что все основные элементарные функции непрерывны в каждой точке своей области определения. Функции, которые получаются из основных с помощью конечного числа арифметических операций, а также операций взятия функции от функции, примененных конечное число раз, также являются непрерывными функциями.

Точки разрыва функции. Их классификация

Пусть функция $f(x)$ определена в некоторой окрестности точки x_0. Согласно определению непрерывность функции $f(x)$ в точке x_0 выражается соотношением

$$\lim_{x\to x_0} f(x) = f(x_0).$$

Пользуясь односторонними пределами функции, это равенство можно заменить равносильным ему равенством

$$\lim_{x\to x_0-0} f(x) = f(x_0-0) = \lim_{x\to x_0+0} f(x) = f(x_0+0) = f(x).$$

Таким образом, функция $f(x)$ непрерывна в точке x_0 тогда и только тогда, когда в этой точке существуют пределы слева и справа, они равны между собой и равны значению функции $f(x)$ в точке x_0.

Если в точке x_0 функция $f(x)$ не является непрерывной, то говорят, что $f(x)$ *разрывна* в этой точке. Точку x_0 называют *точкой разрыва* функции $f(x)$, причем функция $f(x)$ может быть не определена в точке x_0.

Точки разрыва функции классифицируются в зависимости от того, какое условие непрерывности нарушено.

Если в точке x_0 функция $f(x)$ имеет предел слева и предел справа и они равны между собой, но не равны значению функции в точке x_0, т.е.

$$\lim_{x\to x_0-0} f(x) = \lim_{x\to x_0+0} f(x) \neq f(x_0),$$

то точка x_0 называется точкой *устранимого разрыва* функции $f(x)$. Подобный разрыв можно устранить, если дополнить разрывную функцию до непрерывности так:

$$F(x) = \begin{cases} f(x), & x \neq x_0, \\ \lim\limits_{x\to x_0} f(x), & x = x_0. \end{cases}$$

Пример 6.5. Пусть $y = \dfrac{x^2 - 1}{x - 1}$.

Имеем

$$\lim_{x\to 1-0} \frac{x^2-1}{x-1} = \lim_{x\to 1-0} \frac{(x-1)(x+1)}{x-1} = \lim_{x\to 1-0} (x+1) = 2.$$

Точно так же

$$\lim_{x\to 1+0} \frac{x^2-1}{x-1} = \lim_{x\to 1+0} \frac{(x-1)(x+1)}{x-1} = \lim_{x\to 1+0} (x+1) = 2.$$

В точке $x = 1$ функция не существует, но правый и левый пределы существуют и равны между собой (рис. 6.12). Если изменить значение данной функции в точке $x = 1$, положив $f(1) = 2$, то получим непрерывную всюду функцию $y = \begin{cases} \dfrac{x^2-1}{x-1}, & x \neq 1, \\ 2, & x = 1. \end{cases}$

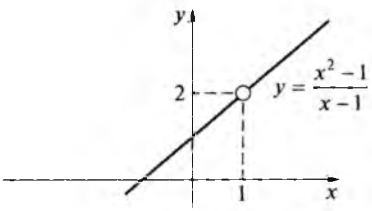

Рис. 6.12

Пример 6.6. Функция $y = \dfrac{\sin x}{x}$ имеет в точке $x = 0$ устранимый разрыв. Дополнив ее значением $y = 1$ в точке $x = 0$, получим непрерывную функцию

$$y = \begin{cases} \dfrac{\sin x}{x}, & x \neq 0, \\ 1, & x = 0. \end{cases}$$

Мы устранили разрыв, изменив значение функции в одной точке $x = 0$.

Определение. Если в точке x_0 функция $f(x)$ имеет конечные пределы слева и справа, но они не равны между собой:

$$\lim_{x\to x_0-0} f(x) \neq \lim_{x\to x_0+0} f(x),$$

то точка x_0 называется точкой неустранимого *разрыва первого рода*, или разрыва с конечным скачком функции. (При этом безразлично, совпадает или нет $f(x_0)$ с одним из односторонних пределов.)

Пример 6.7. Пусть $y = \operatorname{sgn} x$. В точке $x = 0$ функция имеет различные правый и левый пределы:

$$\lim_{x\to -0} \operatorname{sgn} x = -1, \quad \lim_{x\to +0} \operatorname{sgn} x = +1.$$

Значит, в точке $x = 0$ существует неустранимый разрыв первого рода.

Определение. Если в точке x_0 функция $f(x)$ имеет бесконечный предел слева или справа или один из этих пределов не существует, то точка x_0 называется точкой *разрыва второго рода*.

Пример 6.8. Пусть $y = e^{\frac{1}{x}}$.

Для данной функции точка $x = 0$ есть точка разрыва второго рода. Это обнаруживается при нахождении правостороннего предела:

$$\lim_{x\to +0} e^{\frac{1}{x}} = [e^{+\infty}] = +\infty.$$

Левосторонний предел оказывается конечной величиной:

$$\lim_{x\to -0} e^{\frac{1}{x}} = [e^{-\infty}] = 0.$$

Свойства функций, непрерывных на отрезке

В первой половине XIX в. происходило становление математического анализа из разрозненных знаний как систематизированной, логически стройной науки. Значительную роль в этом сыграли чешский теолог и математик Бернард Больцано (1781–1848), французский математик, член Парижской АН Огюстен Луи Коши (1789–1857) и немецкий математик, профессор Берлинского университета Карл Вейерштрасс (1815–1897).Приведем свойства непрерывных функций без доказательства.

Теорема 6.1 (первая теорема Больцано—Коши). Если функция $f(x)$ непрерывна на отрезке $[a; b]$ и на концах его имеет значения, противоположные по знаку, то $f(x)$ обращается в нуль по крайней мере в одной точке интервала $(a; b)$.

Геометрически результат теоремы очевиден. Если $f(a)f(b) < 0$, то точки $A(a, f(a))$ и $B(b, f(b))$ лежат в разных полуплоскостях, на которые ось Ox делит плоскость xOy. График непрерывной функции $y = f(x)$, соединяющий эти точки, обязательно пересечет ось Ox по крайней мере в одной точке (рис. 6.13).

Требование непрерывности функции $f(x)$ на $[a; b]$ является необходимым: функция, имеющая разрыв хотя бы в одной точке, может перейти от отрицательного значения к положительному, и не обращаясь в нуль. Так будет, например, с функцией $y = \begin{cases} 1, & x \ge 0, \\ -1, & x < 0. \end{cases}$

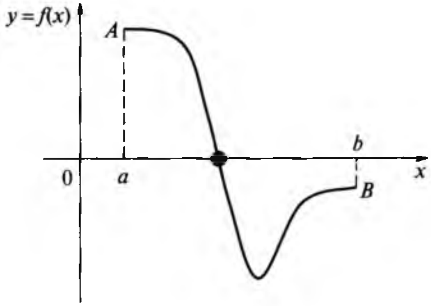

Рис. 6.13

Теорема 6.2 (вторая теорема Больцано—Коши). Пусть функция $f(x)$ непрерывна на отрезке $[a;\ b]$, причем $f(a) = A$, $f(b) = B$. Тогда, каким бы ни было число C, заключенное между числами A и B, на отрезке $[a;\ b]$ найдется по крайней мере одна точка c такая, что $f(c) = C$.

Эти теоремы устанавливают, что, переходя от одного своего значения к другому, функция хоть раз принимает каждое свое промежуточное значение между ее значениями на концах отрезка.

Теорема 6.3 (первая теорема Вейерштрасса). Если функция $y = f(x)$ непрерывна на отрезке $[a;\ b]$, то она ограничена на нем сверху и снизу, т.е. существуют такие числа m и M, что для всех $x \in [a;\ b]$ верно неравенство $m \leq f(x) \leq M$.

Замечание 1. Если функция $f(x)$ непрерывна на интервале $(a;\ b)$, то она необязательно ограничена на нем. Например, функция $f(x) = \dfrac{1}{x}$ непрерывна на полуинтервале $(0;\ 1]$, но не ограничена на нем.

Замечание 2. Если функция $f(x)$ не является непрерывной на отрезке $[a;\ b]$, то она может не достигать своих наименьшего и наибольшего значений. Например, функция $y = \begin{cases} x, & 0 \leq x < 1, \\ 0, & x = 1 \end{cases}$ не достигает своего наибольшего значения на отрезке $[0;\ 1]$ (рис. 6.14).

Как уже указывалось, функция называется *ограниченной сверху*, если существует $M \in R : f(x) \leq M$. Наименьшее из всех возможных значений M называется точной верхней гранью и обозначается $\sup f(x) = \min M$ (*supremum* — наивысший).

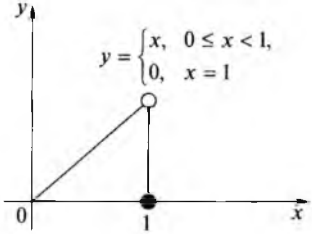

Рис. 6.14

Функция называется *ограниченной снизу*, если существует$M \in R$: $: f(x) \ge m$. Наибольшее из всех возможных значений m называется точной нижней гранью и обозначается $\inf f(x) = \max m$ (*infimum* — наинизший).

Теорема 6.4. (вторая теорема Вейерштрасса). Если функция $f(x)$ непрерывна на отрезке $[a;\ b]$, то она достигает на этом отрезке своих точных нижней и верхней граней, т.е. на отрезке $[a;\ b]$ найдутся такие точки x_1 и x_2, что $f(x_1) = m = \inf\limits_{x \in [a;b]} f(x)$ и $f(x_2) = M = \sup\limits_{x \in [a;b]} f(x)$. Иными словами, найдутся такие точки x_1 и x_2, что значения в этих точках будут наименьшими и наибольшими из всех возможных значений функции $f(x)$ на отрезке.

Замечание 1. Условие $x \in [a;\ b]$ существенно: функция $f(x) = x$ непрерывна на интервале $(0;\ 1)$ и ограничена на нем, но ее точная верхняя грань $\sup f(x) = 1$ не достигается, т.е. нет такого $x_2 \in (0;\ 1)$, что $f(x_2) = 1$.

Замечание 2. Условие непрерывности на отрезке также является важным. Рассмотренная ранее функция $y = \begin{cases} x, & 0 \le x < 1, \\ 0, & x = 1 \end{cases}$ не достигает своего наибольшего значения на отрезке $[0;\ 1]$, хотя $\sup f(x) = 1$.

6.3. ПРОИЗВОДНАЯ ФУНКЦИИ И ЕЕ ГЕОМЕТРИЧЕСКИЕ ПРИЛОЖЕНИЯ

Производная функции одной переменной

Основной задачей дифференциального исчисления является нахождение по заданной функции $y = f(x)$ ее производной $y' = f'(x)$. Пусть задана функция $y = f(x)$. Возьмем какое-нибудь значение x_0 из области определения функции. Соответствующее значение функции в этой точке $y=f(x_0)$. Зададим аргументу x приращение Δx. Получим значение функции в новой точке $y=f(x_0+\Delta x)$. *Приращением функции* $y=f(x)$ в точке x_0, соответствующим приращению аргумента $\Delta x = x - x_0$, называется величина $\Delta y = f(x_0+\Delta x) - f(x_0)$.

Производной $y'(x)$ от функции $y=f(x)$ называется предел отношения $\frac{\Delta y}{\Delta x}$ при $\Delta x \to 0$, если он существует. Таким образом, по определению

$$y'(x_0) = f'(x_0) = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} =$$

$$= \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0},$$

где $x = x_0 + \Delta x$. Эти обозначения производной принадлежат Лагранжу.

Функция $y=f(x)$ имеет производную на интервале $(a; b)$, если производная $f'(x_0)$ существует в каждой точке x_0 этого интервала. Учитывая это, будем опускать индекс у величины x_0 и записывать производную так: $f'(x)$.

Пример 6.9. Найдем производную функции $y=x^2$ в любой точке x области определения:

$$y'(x) = \lim_{\Delta x\to 0}\frac{(x+\Delta x)^2-x^2}{\Delta x} = \lim_{\Delta x\to 0}\frac{x^2+2x\cdot\Delta x+(\Delta x)^2-x^2}{\Delta x} =$$

$$= \lim_{\Delta x\to 0}(2x+\Delta x) = 2x.$$

Следовательно, функция $y=x^2$ имеет во всякой точке x производную $y'(x)=2x$.

Введем понятия левой, правой и бесконечной производной.

Правой производной y'_+ функции $y=f(x)$ в данной точке x называется величина

$$y'_+ = f'_+(x) = f'(x+0) = \lim_{\Delta x\to +0}\frac{\Delta y}{\Delta x},$$

левой производной — величина

$$y'_- = f'_-(x) = f'(x-0) = \lim_{\Delta x\to -0}\frac{\Delta y}{\Delta x},$$

если эти пределы существуют.

Пользуясь понятием односторонних пределов функции, получаем: для того чтобы в точке x существовала производная $f'(x)$, необходимо и достаточно, чтобы в точке x функция $y=f(x)$ имела правую и левую производные и эти производные были равны между собой:

$y'_-(x) = y'_+(x) = y'(x).$

Функция $y=f(x)$ имеет в точке x *бесконечную производную*, если в этой точке

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \infty.$$

Символ ∞ может быть уточнен: $f'(x) = +\infty$ или $f'(x) = -\infty$.

Пример 6.10. Исследуем существование производной у функции $f(x) = \sqrt{x}$ в точке $x_0 = 0$. К этой точке можно подойти только справа. Правая производная

$$f_+(0) = \lim_{\Delta x \to +0} \frac{\sqrt{x_0 + \Delta x} - \sqrt{x_0}}{\Delta x} = \lim_{\Delta x \to +0} \frac{\sqrt{\Delta x}}{\Delta x} = \lim_{\Delta x \to +0} \frac{1}{\sqrt{\Delta x}} = +\infty.$$

Левый предел, соответственно левая производная, не существует. Вывод: функция $f(x) = \sqrt{x}$ в точке $x_0 = 0$ не имеет производной, но имеет правую бесконечную производную, равную $+\infty$.

Операцию нахождения производной функции называют также *дифференцированием* этой функции. Функция обладает тем свойством, что если она имеет конечную производную в точке x_0, то непрерывна в этой точке.

Дифференциал функции

Как указывалось, функция $f(x)$, имеющая предел b, может быть представлена в виде $f(x) = b + \alpha(x)$, где $\alpha(x)$ — бесконечно малая функция при $x \to x_0$. Тогда предел $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x_0)$ запишем в виде $\Delta y = f'(x_0)\Delta x + \alpha \cdot \Delta x$.

Дифференциалом функции называется линейная относительно Δx часть приращения функции. Она обозначается как dy или $df(x)$. Таким образом,

$$dy = f'(x) \cdot \Delta x.$$

Замечание 1. Дифференциал функции составляет основную часть ее приращения. Например, приращение функции $y = x^2$ в точке $x_0 = 1$, вызванное приращением аргумента $x = 0,1$, есть величина $\Delta y = (x_0 + \Delta x)^2 - x_0^2 = (1 + 0,1)^2 - 1^2 = 0,21$. Дифференциал функции в этой точке равен $dy = f'(x_0) \cdot \Delta x = 2 \cdot x_0 \cdot \Delta x = 2 \cdot 1 \cdot 0,1 = 0,2$. Таким образом, на нелинейную часть приращения приходится величина 0,01 из полной величины приращения 0,21. Поэтому дифференциал называют *главной частью* приращения функции, *линейной* относительно Δx.

Замечание 2. Наряду с понятием дифференциала функции, вводится понятие дифференциала аргумента. По определению *дифференциал аргумента* есть приращение аргумента:

$$dx = \Delta x.$$

З а м е ч а н и е 3. Формулу для дифференциала функции можно записать в виде

$$dy = f'(x) \cdot dx.$$

Отсюда получим $\frac{dy}{dx} = f'(x)$,

т.е. производная может быть представлена как обыкновенная дробь — отношение дифференциалов функции и аргумента. Подобное обозначение для производной ввел Г.В. Лейбниц (1646–1716) — немецкий математик и философ.

Правила вычисления производных

Пусть функции $u(x)$ и $v(x)$ имеют производные в точке x. Тогда:

1. $(c \cdot u(x))' = c \cdot u'(x)$, если c — постоянная величина.
2. $(u(x) \pm v(x))' = u'(x) \pm v'(x)$.
3. $(u(x) \cdot v(x))' = u'(x)v(x) + u(x)v'(x)$.
4. $\left(\frac{u(x)}{v(x)}\right) = \frac{u'(x)v(x) - u(x)v'(x)}{v^2(x)}, \quad v(x) \neq 0.$
5. Пусть функции $y = y(u)$ и $u = u(x)$ имеют производные соответственно в точках $u_0 = u(x_0)$ и x_0.

Тогда $(y(u(x)))'_x\big|_{x=x_0} = y'(u)\big|_{u=u_0} \cdot u'(x)\big|_{x=x_0}$.

Более короткая запись выглядит так: $y'_x = y'_u \cdot u'_x$ или в дифференциалах: $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$.

6. Пусть для функции $y = y(x)$ имеется взаимно однозначное соответствие между значениями переменных y и x в области задания этих переменных, причем $y_0 = y(x_0)$.

Тогда $x'_y\big|_{y=y_0} = \frac{1}{y'_x\big|_{x=x_0}}$, или $\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$.

Докажем некоторые из этих соотношений.

1. Пусть задана функция $y = c \cdot u(x)$. Ее производная находится по определению

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{c \cdot u(x + \Delta x) - c \cdot u(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{c \cdot \Delta u}{\Delta x} =$$

$$= c \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = c \cdot u'.$$

2. Для функции $y = u(x) + v(x)$ производную можно получить так:

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{[u(x + \Delta x) + v(x + \Delta x)] - [u(x) + v(x)]}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{[u(x + \Delta x) - u(x)] + [v(x + \Delta x) - v(x)]}{\Delta x}.$$

Далее, перейдя к сумме пределов, получим ответ

$$\lim_{\Delta x \to 0} \frac{\Delta u + \Delta v}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u'(x) + v'(x).$$

3. Если функция имеет вид $y = u(x) \cdot v(x)$, то ее производная вычисляется следующим образом:

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{u(x + \Delta x) \cdot v(x + \Delta x) - u(x) \cdot v(x)}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{(u(x) + \Delta u) \cdot (v(x) + \Delta v) - u(x) \cdot v(x)}{\Delta x}.$$

Раскрыв скобки и приведя подобные, получим

$$\lim_{\Delta x \to 0} \frac{u(x) \cdot \Delta v + v(x) \cdot \Delta u}{\Delta x} = u(x) \cdot \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} + v(x) \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} =$$

$$= u(x) \cdot v'(x) + v(x) \cdot u'(x).$$

Мы учли, что $u(x)$ и $v(x)$ можно вынести за знак предела как постоянные величины, не зависящие от Δx.

4. Производную функции $y = \dfrac{u(x)}{v(x)}$ вычисляем по такой схеме:

$$y' = \lim_{\Delta x \to 0} \left[\frac{u(x + \Delta x)}{v(x + \Delta x)} - \frac{u(x)}{v(x)} \right] \cdot \frac{1}{\Delta x} = \lim_{\Delta x \to 0} \left[\frac{u(x) + \Delta u}{v(x) + \Delta v} - \frac{u(x)}{v(x)} \right] \cdot \frac{1}{\Delta x}.$$

Приведем выражение в квадратных скобках к общему знаменателю, опустим для упрощения записи аргумент x. Получим

$$\lim_{\Delta x \to 0} \frac{\Delta u \cdot v - u \cdot \Delta v}{v \cdot (v + \Delta v) \cdot \Delta x} = \lim_{\Delta x \to 0} \frac{v \cdot \dfrac{\Delta u}{\Delta x} - u \cdot \dfrac{\Delta v}{\Delta x}}{v^2 + v \cdot \Delta v} = \frac{\lim\limits_{\Delta x \to 0} \left[v \cdot \dfrac{\Delta u}{\Delta x} - u \cdot \dfrac{\Delta v}{\Delta x} \right]}{\lim\limits_{\Delta x \to 0} [v^2 + v \cdot \Delta v]} =$$

$$= \frac{v \cdot \lim\limits_{\Delta x \to 0} \dfrac{\Delta u}{\Delta x} - u \cdot \lim\limits_{\Delta x \to 0} \dfrac{\Delta v}{\Delta x}}{v^2} = \frac{vu' - uv'}{v^2}.$$

5. Правило нахождения производной сложной функции можно проиллюстрировать, используя дифференциалы. Пусть заданы функции $y = y(u)$ и $u = u(x)$. Для функции $y = y(u)$ имеем $y'_u = \dfrac{dy}{du}$.

Для функции $u = u(x)$, аналогично, $u'_x = \dfrac{du}{dx}$.

Тогда $y'_x = \dfrac{dy}{dx} = \dfrac{dy}{du} \cdot \dfrac{du}{dx} = y'_u \cdot u'_x$.

6. Для нахождения x'_y в зависимости $y = y(x)$ также воспользуемся дифференциалами:

$$x'_y = \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{y'_x}.$$

Замечание. Пункты 5 и 6 нельзя рассматривать как доказательства правил нахождения производной сложной и обратной функций. Мы использовали здесь факт инвариантности первого дифференциала (о чем ниже), который устанавливается именно при помощи теоремы о дифференцируемости сложной функции.

Правила вычисления дифференциалов

1. $d(c \cdot u(x)) = c \cdot du(x)$, где c — постоянная величина.
2. $d(u(x) \pm v(x)) = du(x) \pm dv(x)$.
3. $d(u(x) \cdot v(x)) = u(x) \cdot dv(x) + v(x) \cdot du(x)$.
4. $d\left(\frac{u(x)}{v(x)}\right) = \frac{v \cdot du - u \cdot dv}{v^2}$, где $v(x) \neq 0$.

Производные некоторых элементарных функций (таблица производных)

Используя правила вычисления производных, можно составить таблицу производных основных функций.

1. $(c)' = 0$, где c — постоянная.
2. $(x^\alpha)' = \alpha \cdot x^{\alpha-1}$.
3. $(a^x)' = a^x \cdot \ln a$.
4. $(e^x)' = e^x$.
5. $(\log_a x)' = \frac{1}{x \cdot \ln a}$.
6. $(\ln x)' = \frac{1}{x}$.
7. $(\sin x)' = \cos x$.
8. $(\cos x)' = -\sin x$.
9. $(\mathrm{tg}\, x)' = \frac{1}{\cos^2 x}$.
10. $(\mathrm{ctg}\, x)' = -\frac{1}{\sin^2 x}$.
11. $(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}$.
12. $(\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}$.

13. $(\operatorname{arctg} x)' = \dfrac{1}{1+x^2}$.

14. $(\operatorname{arcctg} x)' = -\dfrac{1}{1+x^2}$.

Для примера докажем справедливость седьмой формулы. Ищем производную, пользуясь формулой разности синусов и первым замечательным пределом:

$$y' = \lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin x}{\Delta x} = \lim_{\Delta x\to 0}\frac{2\sin\frac{\Delta x}{2}\cos\frac{2x+\Delta x}{2}}{\Delta x} =$$

$$= \lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\cdot\lim_{\Delta x\to 0}\cos\frac{2x+\Delta x}{2} = \cos x.$$

Геометрические приложения производной

Уравнение касательной к кривой

Для вывода *уравнения касательной* будем кривую описывать уравнением $z = z(x)$, а прямую — уравнением $y = kx + b$. Назовем касательной прямую линию $y = kx + b$, наилучшим образом описывающую исходную функцию $z = z(x)$ в окрестности точки x_0. «Наилучшим образом» означает, что $z(x) - (kx + b) = \alpha$ в окрестности точки x_0, где $\alpha = \alpha(x - x_0)$ есть бесконечно малая функция при $x \to x_0$.

В соответствии с определением производной

$$z'(x_0) = \lim_{x\to x_0}\frac{z(x)-z(x_0)}{x-x_0} = \lim_{x\to x_0}\frac{z-z_0}{x-x_0}.$$

Здесь введено обозначение $z(x_0) = z_0$.

По определению предела

$$\frac{z-z_0}{x-x_0} - z_0' = \beta,$$

где $\beta = \beta(x - x_0)$ — бесконечно малая функция при $x \to x_0$.

Отсюда

$$z = z_0'\cdot x + z_0 - z_0'\cdot x_0 + \beta(x - x_0).$$

Здесь $\alpha = \beta(x - x_0)$ — бесконечно малая функция при $x \to x_0$.

Обозначим $k = z_0'$, $b = z_0 - z_0'\cdot x_0$. Тогда в окрестности точки x_0 уравнение кривой $z = z(x)$ можно представить как $z = kx + b$.

Следовательно, $y = kx + b$ есть по определению уравнение касательной к кривой $z = z(x)$ в точке x_0.

Геометрический смысл производной (производная как тангенс угла наклона)

Вернемся к прежнему обозначению произвольной функции $y = f(x)$. ***Углом наклона*** между кривой и осью x в точке x_0 называется угол между касательной к этой кривой в точке x_0 и положительным направлением оси x. Рассмотрим график касательной $y = kx + b$ к кривой $y = f(x)$ на рис. 6.15. Дадим аргументу x значения x_1 и x_2. Получим соответствующие значения функции y_1 и y_2. Построим прямоугольный треугольник ABC с острым углом φ. Вычислим $\operatorname{tg}\varphi$:

$$\operatorname{tg}\varphi = \frac{y_2 - y_1}{x_2 - x_1} = \frac{kx_2 + b - (kx_1 + b)}{x_2 - x_1} = k.$$

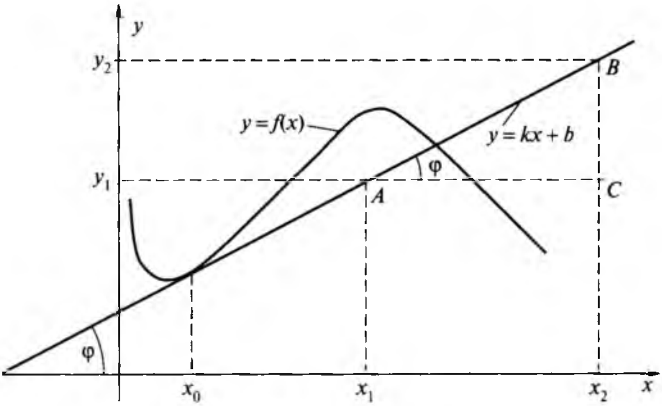

Рис. 6.15

Таким образом, величина $k = y'_0 = y'(x_0) = f'(x_0)$ есть *тангенс угла наклона* кривой $y = f(x)$ в точке x_0 к оси x.

Если в точке x_0 функция $y = f(x)$ непрерывна и имеет правую и левую производные f'_+ и f'_-, причем $f'_+ \neq f'_-$, то в точке x_0 (рис. 6.16) график функции $y = f(x)$ касательной не имеет. Но существуют две односторонние полукасательные, или, что то же самое, правая и левая касательные. Точку на графике функции, в которой происходит излом графика, называют в этом случае *угловой точкой* кривой $y = f(x)$.

Если функция $y = f(x)$ непрерывна в точке x_0, а ее правая и левая производные в этой точке бесконечны, то возможны четыре различных случая:

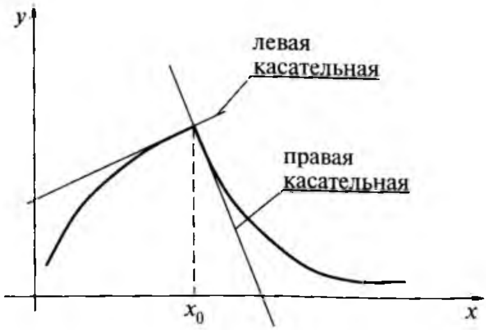

Рис. 6.16

1) $f'_+(x_0) = f'_-(x_0) = +\infty$;
2) $f'_+(x_0) = f'_-(x_0) = -\infty$;
3) $f'_+(x_0) = +\infty, f'_-(x_0) = -\infty$;
4) $f'_+(x_0) = -\infty, f'_-(x_0) = +\infty$.

На рис. 6.17–6.20 представлены графики кривых $y = f(x)$, которые проходят через точку M под углом 90° и отвечают случаям 1–4. Очевидно, во всех случаях касательная перпендикулярна оси x, в последних двух случаях образуются нижний и верхний «клювики».

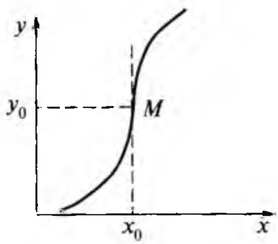

Рис. 6.17

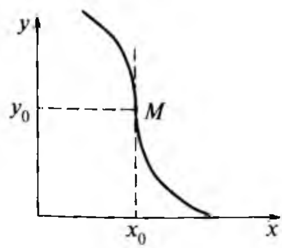

Рис. 6.18

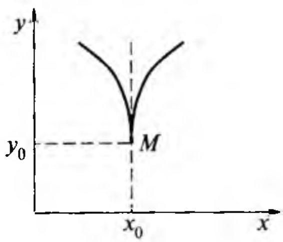

Рис. 6.19

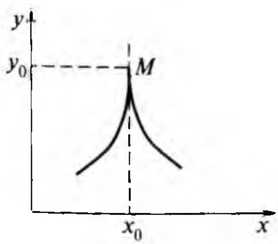

Рис. 6.20

Угол между кривыми

Углом между кривыми на плоскости в их общей точке $M(x_0, y_0)$ называется наименьший из двух возможных угол между касательными к этим кривым в данной точке (рис. 6.21). Пусть к кривым, описываемым уравнениями $C = f_1(x)$ и $y = f_2(x)$, проведены касательные в их общей точке $M(x_0, y_0)$, уравнения которых $y = k_1x + b_1$ и $y = k_2x + b_2$. Угол между касательными будет равен $\varphi = \alpha - \beta$. Чтобы можно было использовать производные в данной точке для вычисления угла, найдем тангенс угла φ:

$$\operatorname{tg}\varphi = \operatorname{tg}(\alpha - \beta) = \frac{\operatorname{tg}\alpha - \operatorname{tg}\beta}{1 + \operatorname{tg}\alpha\operatorname{tg}\beta} = \frac{f_1'(x_0) - f_2'(x_0)}{1 + f_1'(x_0)f_2'(x_0)} = \frac{k_1 - k_2}{1 + k_1k_2}.$$

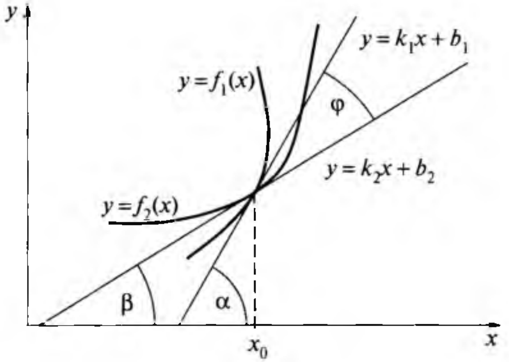

Рис. 6.21

Тангенс острого угла — величина положительная. Если вычисленный $\operatorname{tg}\varphi$ оказался меньше нуля, значит, найден тупой угол $\psi = \pi - \varphi$.

$\operatorname{tg}\varphi = \operatorname{tg}(\pi - \psi) = -\operatorname{tg}\psi$.

Таким образом, угол между кривыми в точке их пересечения может быть найден по формуле

$$\varphi = \operatorname{arctg}\left|\frac{f_1'(x_0) - f_2'(x_0)}{1 + f_1'(x_0)f_2'(x_0)}\right|.$$

Условие параллельности двух прямых

$\varphi = 0 \to \operatorname{tg}\varphi = 0 \to k_1 = k_2$, т.е. коэффициенты при переменной x должны быть равны. Поэтому уравнение прямой, параллельной данной прямой $\frac{y-y_1}{x-x_1} = k$, имеет вид $\frac{y-y_2}{x-x_2} = k$.

Здесь первая прямая проходит через точку с координатами (x_1, y_1), вторая — через точку с координатами (x_2, y_2).

Условие перпендикулярности двух прямых

$$\varphi = \frac{\pi}{2} \to \operatorname{tg}\varphi = \infty \to 1 + k_1 k_2 = 0, \text{ т.е. } k_2 = -\frac{1}{k_1}.$$

Отсюда уравнение прямой, перпендикулярной данной прямой

$$\frac{y-y_1}{x-x_1} = k$$

и проходящей через ту же точку (x_1, y_1), имеет вид

$$\frac{y-y_1}{x-x_1} = -\frac{1}{k}.$$

Основные теоремы дифференциального исчисления

В подготовке современных методов дифференциального исчисления значительную роль сыграли французские ученые: юрист и математик, автор выдающихся работ Пьер Ферма (1601–1665), математики, члены Парижской АН Мишель Ролль (1652–1719) и Жозеф Луи Лагранж (1736–1813), автор метода решения задач на условный экстремум, Огюстен Луи Коши (1789–1857).

Теоремы Ферма, Ролля, Лагранжа, Коши

Рассмотрим основные теоремы дифференциального исчисления, лежащие в основе приложений производной.

Теорема 6.5 (теорема Ферма). Пусть функция $y = f(x)$: 1) дифференцируема на интервале (a, b); 2) достигает наибольшего или наименьшего значения в точке $x_0 \in (a, b)$. Тогда производная функции в этой точке равна нулю, т.е. $f'(x_0) = 0$.

Пусть функция $y = f(x)$ дифференцируема на интервале $(a; b)$ и в точке x_0 (рис. 6.22) принимает наибольшее значение при $x_0 \in (a; b)$. По определению производной

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0},$$

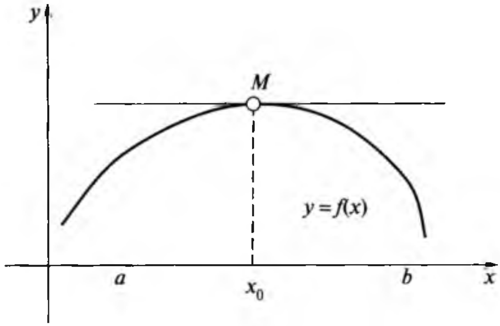

Рис. 6.22

причем предел не зависит от того, будет ли $x \to x_0$ справа или слева. Но при $x > x_0$

$$\frac{f(x) - f(x_0)}{x - x_0} \le 0,$$

откуда следует, что $f'(x_0) \le 0$. При $x < x_0$ имеем

$$\frac{f(x) - f(x_0)}{x - x_0} \ge 0,$$

следовательно, $f'(x_0) \ge 0$.

По условию функция $y = f(x)$ дифференцируема в точке x_0, следовательно, ее предел при $x \to x_0$ не должен зависеть от выбора направления приближения аргумента x к точке x_0, т.е.

$$\lim_{x \to x_0 + 0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0 - 0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0).$$

Получаем систему $\begin{cases} f'(x_0) \le 0 \\ f'(x_0) \ge 0, \end{cases}$ из которой следует $f'(x_0) = 0$. Аналогично рассматривается случай, когда функция принимает в точке x_0 наименьшее значение.

Геометрический смысл теоремы Ферма очевиден: в точке наибольшего или наименьшего значения, достигаемого внутри промежутка, касательная к графику функции параллельна оси абсцисс.

Теорема 6.6 (теорема Ролля). Пусть функция $y = f(x)$: 1)непрерывна на отрезке $[a; b]$; 2) дифференцируема на интервале $(a; b)$; 3) на концах отрезка $[a; b]$ принимает равные значения: $f(a) = f(b)$.

Тогда на интервале $(a; b)$ найдется по крайней мере одна точка x_0, в которой $f'(x_0) = 0$.

Теорема 6.7 (теорема Лагранжа). Пусть функция $y = f(x)$: 1) непрерывна на отрезке $[a; b]$; 2) дифференцируема на интервале $(a; b)$. Тогда на интервале $(a; b)$ найдется по крайней мере одна точка x_0 такая, что

$$\frac{f(b) - f(a)}{b - a} = f'(x_0).$$

Формула называется формулой Лагранжа или формулой конечных приращений. Она может быть переписана в виде

$$f(b) - f(a) = f'(x_0) \cdot (b - \alpha).$$

Раскрытие неопределенностей (правило Лопиталя)

Сформулированные и доказанные теоремы легли в основу доказательства мощного метода раскрытия неопределенностей, который нашел швейцарский ученый И. Бернулли, но опубликовал французский математик Мишель Лопиталь (1661–1704).

Теорема 6.8 (правило Лопиталя). Пусть функции $y = f(x)$ и $y = g(x)$: 1) дифференцируемы в окрестности точки a, кроме, быть может, самой точки a; 2) $g(x) \neq 0$ и $g'(x) \neq 0$ в этой окрестности; 3) $\lim\limits_{x \to a} f(x) = 0$, $\lim\limits_{x \to a} g(x) = 0$; 4) $\lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)}$ существует конечный или бесконечный. Тогда существует $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)}$, причем $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = \lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)}$.

Замечание 1. Правило Лопиталя распространяется на случай неопределенности типа $\left[\dfrac{\infty}{\infty}\right]$ при $x \to a$, поскольку можно доказать теорему Лопиталя при условии, что

$$\lim_{x \to a} f(x) = \infty \text{ и } \lim_{x \to a} g(x) = \infty.$$

Пример 6.11. Найти $\lim\limits_{x \to 0} \dfrac{\ln x}{\operatorname{ctg} x}$.

Решение.

$$\lim_{x \to 0} \frac{\ln x}{\operatorname{ctg} x} = \left[\frac{\infty}{\infty}\right] = -\lim_{x \to 0} \frac{\sin^2 x}{x} = -\lim_{x \to 0} \frac{\sin x}{x} \cdot \lim_{x \to 0} \sin x = 1 \cdot 0 = 0.$$

Замечание 2. Правило Лопиталя распространяется на случай $x \to \infty$. Чтобы убедиться в этом, достаточно сделать замену $x = \dfrac{1}{t}$ и воспользоваться результатом теоремы.

Пример 6.12. Найти $\lim\limits_{x\to\infty}\dfrac{\ln x}{\sqrt[3]{x}}$.

Решение. $\lim\limits_{x\to\infty}\dfrac{\ln x}{\sqrt[3]{x}}=\left[\dfrac{\infty}{\infty}\right]=\lim\limits_{x\to\infty}\dfrac{\frac{1}{x}}{\frac{1}{3}x^{-\frac{2}{3}}}=3\lim\limits_{x\to\infty}\dfrac{1}{\sqrt[3]{x}}=0.$

З а м е ч а н и е 3. Иногда приходится применять правило Лопиталя последовательно несколько раз (делать несколько шагов), если от неопределенности не удается избавиться на первом шаге. Однако условия теоремы на каждом шаге должны оставаться справедливыми.

Пример 6.13. Найти $\lim\limits_{x\to\infty}\dfrac{e^x}{x^5}$.

Решение.

$$\lim_{x\to\infty}\frac{e^x}{x^5}=\left[\frac{\infty}{\infty}\right]=\lim_{x\to\infty}\frac{e^x}{5x^4}=\left[\frac{\infty}{\infty}\right]=\lim_{x\to\infty}\frac{e^x}{20x^3}=\left[\frac{\infty}{\infty}\right]=\lim_{x\to\infty}\frac{e^x}{60x^2}=$$

$$=\left[\frac{\infty}{\infty}\right]=\lim_{x\to\infty}\frac{e^x}{120x}=\left[\frac{\infty}{\infty}\right]=\lim_{x\to\infty}\frac{e^x}{120}=+\infty.$$

З а м е ч а н и е 4. Хотя правило Лопиталя работает только с неопределенностями $\left[\frac{0}{0}\right]$ и $\left[\frac{\infty}{\infty}\right]$, неопределенности других типов могут быть раскрыты с помощью этого правила, если путем преобразований удастся привести изучаемую неопределенность к типу $\left[\frac{0}{0}\right]$ или $\left[\frac{\infty}{\infty}\right]$.

Пример 6.14. Найти $\lim\limits_{x\to\infty}\left(x\cdot\sin\dfrac{a}{x}\right)$.

Решение.

$$\lim_{x\to\infty}\left(x\cdot\sin\frac{a}{x}\right)=[\infty\cdot 0]=\lim_{x\to\infty}\frac{\sin\frac{a}{x}}{\frac{1}{x}}=\left[\frac{0}{0}\right]=\lim_{x\to\infty}\frac{\cos\frac{a}{x}\cdot\frac{-a}{x^2}}{\frac{-1}{x^2}}=$$

$$=a\lim_{x\to\infty}\cos\frac{a}{x}=a.$$

Данную задачу можно решить, используя первый замечательный предел. Действительно,

$$\lim_{x\to\infty}\left(x\cdot\sin\frac{a}{x}\right)=\lim_{x\to\infty}\frac{\sin\frac{a}{x}}{\frac{1}{x}}=a\cdot\lim_{x\to\infty}\frac{\sin\frac{a}{x}}{\frac{a}{x}}=a\cdot 1=a.$$

Сравнение функций по скорости роста (теоретические задачи)

Рассмотрим некоторые функции, возрастающие при $x \to +\infty$. Составим из них ряд:

$$y = \log_a x,\ a > 1; \quad y = x^k,\ k > 0; \quad y = a^x,\ a > 1; \quad y = x!; \quad y = x^x.$$

Используя рассмотренные методы нахождения пределов, можно доказать, что чем правее в ряду находится функция, тем быстрее она растет. Например,

$$\lim_{x\to+\infty} \frac{\log_a x}{x^k} = \begin{bmatrix} \text{правило} \\ \text{Лопиталя} \end{bmatrix} = \lim_{x\to+\infty} \frac{\frac{1}{x\ln a}}{kx^{k-1}} = \lim_{x\to+\infty} \frac{1}{kx^k \ln a} = 0,$$

следовательно, функция $y = x^k$, $k > 0$, растет быстрее при $x \to +\infty$, чем $y = \log_a x$, $a > 1$.

Формулы Маклорена и Тейлора

В 1715 г. Брук Тейлор (1685–1731) опубликовал формулу для разложения функции в степенной ряд, которая явилась мощным инструментом для исследования функций и приближенных вычислений. Частный случай этой формулы, наиболее часто используемый в практике математического анализа, исторически неправильно приписывается шотландскому математику Колину Маклорену (1698–1746). Формулы Тейлора и Маклорена являются одними из основных формул математического анализа и имеют многочисленные приложения.

Формула разложения функции $y = f(x)$ в ряд по степеням x в окрестности 0 имеет вид

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + R(x).$$

Эту формулу называют формулой Маклорена разложения функции $y = f(x)$ по степеням x. $R(x)$ называется остаточным членом. Его величина мала по сравнению с x^n, что можно записать так $o(x^n)$. Для остаточного члена получены выражения, позволяющие дать оценку его величине.

При разложении функции $y = f(x)$ в окрестности точки $x = a$ имеем формулу Тейлора:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \ldots +$$
$$+ \frac{f^{(n)}(a)}{n!}(x-a)^n + o((x-a)^n).$$

Отсюда вывод: поведение любой n раз дифференцируемой функции в окрестности точки $x = a$ (в частности, $x = 0$) можно описать многочленом достаточно точно, а при $n \to +\infty$ — со сколь угодно высокой степенью точности.

Разложение по формуле Маклорена элементарных функций

Разложим в ряд Маклорена следующие элементарные функции: e^x; $\sin x$; $\cos x$; $\ln(1+x), x > -1$; $(1+x)^{\pm}, x > -1$.

С этой целью составим таблицу производных этих функций и значений производных в точке $x = 0$.

$f(x)$	$f(0)$	$f'(x)$	$f'(0)$	$f''(x)$	$f''(0)$	$f'''(x)$	$f'''(0)$
e^x	1	e^x	1	e^x	1	e^x	1
$\sin x$	0	$\cos x$	1	$-\sin x$	0	$-\cos x$	-1
$\cos x$	1	$-\sin x$	0	$-\cos x$	-1	$\sin x$	0
$\ln(1+x)$	0	$\dfrac{1}{1+x}$	1	$-\dfrac{1}{(1+x)^2}$	-1	$\dfrac{1\cdot 2}{(1+x)^3}$	$2!$
$(1+x)^{\alpha}$	1	$\alpha(1+x)^{\alpha-1}$	α	$\alpha(\alpha-1)(1+x)^{\alpha-2}$	$\alpha(\alpha-1)$		

Подставляя в формулу Маклорена значения производных, взятые из четных столбцов таблицы, получим разложения в ряд для каждой функции:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + o(x^3);$$

$$\sin x = 0 + \frac{1}{1!}x + \frac{0}{2!}x^2 + \frac{-1}{3!}x^3 + \frac{0}{4!}x^4 + o((x^4)) = x - \frac{x^3}{3!} + o(x^4);$$

$$\cos x = 1 + \frac{-1}{2!}x^2 + \frac{0}{3!}x^3 + \frac{1}{4!}x^4 + \frac{0}{5!}x^5 + o((x^5)) =$$

$$= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + o(x^5);$$

$$\ln(1+x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} + o(x^3);$$

$$(1+x)^{\alpha} = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + o(x^3).$$

З а м е ч а н и е 1. Анализируя ряд разложения функции, легко заметить закономерности образования ряда и выписать следующие члены разложения.

Пример 6.15. Разложить по формуле Маклорена функцию $\sin x$.

Решение. $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ... + (-1)^{k-1}\frac{x^{2k-1}}{(2k-1)!} + ...$

На рис. 6.23 изображен жирной линией график функции $Y_1 = \sin x$, тонкими линиями его приближение одним членом ряда Маклорена $Y_2 = x$, приближение четырьмя отличными от нуля членами ряда

$$Y_3 = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}$$

и, наконец, приближение семью не равными нулю членами ряда

$$Y_4 = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \frac{x^{13}}{13!}.$$

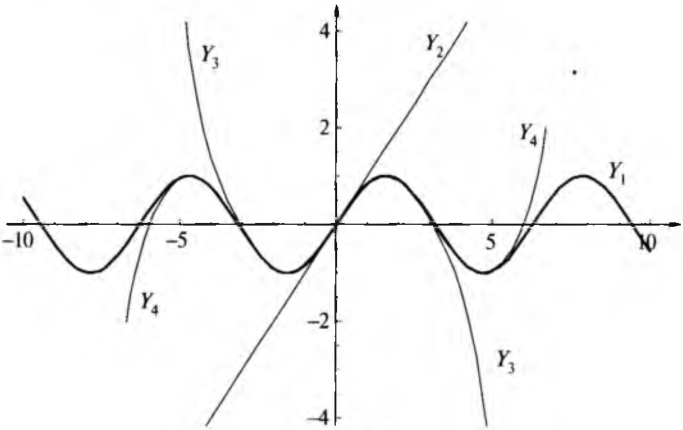

Рис. 6.23

Производные высших порядков

Если функция $f(x)$ имеет производную в каждой точке x области определения, то $f'(x)$ есть функция от x. Функция $y = f'(x)$, в свою очередь, может иметь производную, которую называют ***производной второго порядка*** функции $f(x)$ (или второй производной) и обозначают символом $f''(x)$.

Таким образом, $f''(x_0) = \lim\limits_{x \to x_0} \frac{f'(x) - f'(x_0)}{x - x_0}$.

Функция $y = f'(x)$ имеет производную на интервале $(a; b)$, если ее производная $f''(x_0)$ существует в каждой точке x_0 этого интервала. Учитывая это, будем опускать индекс у величины x_0 и записывать производную так: $f''(x)$. Таким образом,

$f''(x) = (f'(x))'.$

Производные более высоких порядков определяются аналогично, а именно производная n-го порядка функции $f(x)$ есть производная от производной $(n-1)$-го порядка этой функции:

$f^{(n)}(x) = (f^{(n-1)}(x))'.$

Число n, указывающее порядок производной, заключают в скобки.

Дифференциалы высших порядков

Пусть функция $y = f(x)$ зависит от переменной x и дифференцируема в точке x. Может оказаться, что в точке x дифференциал $dy = f'(x)dx$, рассматриваемый как функция от x, есть также дифференцируемая функция. Тогда существует дифференциал от дифференциала $d(dy)$ данной функции, который называется ***дифференциалом второго порядка*** функции $y = f(x)$ и обозначается $d(dy) \equiv d^2y$.

Аналогично определяются дифференциалы более высоких порядков. Дифференциалом n-го порядка d^ny функции $y = f(x)$ называется дифференциал от дифференциала $(n-1)$-го порядка этой функции:

$d^ny = d(d^{n-1}y).$

Производные функций, заданных неявно

Пусть функция $y = y(x)$ задана в неявном виде, т.е. уравнением, не разрешенным относительно y:

$F(x, y) = 0.$

Подставим функцию $y = y(x)$ в уравнение $F(x, y) = 0$, получив тем самым тождество $F(x, y(x)) \equiv 0$. Возьмем дифференциал от обеих частей тождества. Получим

$$dF(x, y) = F'_x(x, y) \cdot dx + F'_y(x, y) \cdot dy = 0. \qquad (1)$$

Полагая, что $F'_y(x, y) \neq 0$, и деля обе части уравнения на $F'_y(x, y) \times dx$, будем иметь

$$\frac{dy}{dx} = y'_x(x) = -\frac{F'_x(x, y)}{F'_y(x, y)}.$$

Другой подход к нахождению производной от функции, заданной неявно, заключается в следующем. Дифференциал функции $y = y(x)$ есть величина $dy = y'_x(x) \cdot dx$. Подставим величину dy в левую часть уравнения (1) и вынесем dx за скобки:

$F'_x(x, y) \cdot dx + F'_y(x, y) \cdot dy = F'_x(x, y) \cdot dx + F'_y(x, y) \cdot y'_x(x) \cdot dx =$

$= (F'_x(x, y) + F'_y(x, y) \cdot y'_x(x)) \cdot dx.$

Тогда наше уравнение будет иметь вид

$(F'_x(x, y) + F'_y(x, y) \cdot y'_x(x)) \cdot dx = 0.$

Поскольку величина dx является постоянной и в общем случае не равной нулю, последнее уравнение принимает вид $F'_x(x, y) + F'_y(x, y) \cdot y'_x(x) = 0.$

Отсюда $y'_x(x) = -\dfrac{F'_x(x, y)}{F'_y(x, y)}.$

Подобный подход дает алгоритм нахождения производной функции $y(x)$, заданной в неявном виде. Необходимо продифференцировать обе части уравнения, рассматривая y как функцию от x, а затем из полученного уравнения найти производную y'.

Пример 6.16. Найти производную функции $y = y(x)$, заданную уравнением $\sin x - x \cdot y(x) + \ln y(x) - 2 = 0.$

Решение. Дифференцируя обе части равенства по аргументу x, получим:

$$\cos x - y(x) - xy'(x) + \frac{y'(x)}{y(x)} = 0.$$

Отсюда $y'(x) = \dfrac{\cos x - y}{x - y^{-1}}.$

Для нахождения второй производной $y''(x)$ мы обычным образом находим производную от первой производной, заменяя получаемые в правой части равенства функции $y'(x)$ их значениями. Для сокращения записи зависимость y от x часто не указывается в явном виде, но подразумевается.

Пример 6.17. Функция $y = y(x)$ задана уравнением $y^2 = 2px$. Найти $y''(x)$.

Решение. Находим первую производную $2y \cdot y' = 2p$, откуда

$$y' = \frac{p}{y} = p \cdot y^{-1}.$$

Находим вторую производную $y'' = p \cdot (-1) \cdot y^{-2} \cdot y'$.

Заменяем в полученном выражении величину y' на $p \cdot y^{-1}$.

Получим $y'' = p \cdot (-1) \cdot y^{-2} \cdot p \cdot y^{-1} = -\dfrac{p^2}{y^3}.$

6.4. ИССЛЕДОВАНИЕ ФУНКЦИЙ С ПОМОЩЬЮ ПРОИЗВОДНЫХ

Условия возрастания и убывания функции

Изучим условия возрастания (неубывания) и убывания (невозрастания) функций. Напомним, что функция $y = f(x)$ называется *воз-*

растающей на промежутке, если большему значению аргумента из этого промежутка соответствует большее значение функции, т.е. из неравенства $x_2 > x_1$ следует неравенство $f(x_2) > f(x_1)$. Функция называется *убывающей* на промежутке, если из $x_2 > x_1$ следует $f(x_2) < f(x_1)$.

Функция $y = f(x)$ называется ***неубывающей на промежутке***, если из неравенства $x_2 > x_1$ следует неравенство $f(x_2) \geq f(x_1)$, и ***невозрастающей***, если из условия $x_2 > x_1$ следует $f(x_2) \leq f(x_1)$.

Теорема 6.9. Если $f'(x) > 0$ на некотором промежутке X, то функция $y = f(x)$ возрастает на этом промежутке; если $f'(x) < 0$ на промежутке X, то функция $y = f(x)$ убывает на этом промежутке.

Пример 6.18. Функция $y = x^3$ возрастает на всей числовой оси, соответственно $f'(x) > 0$, но в точке $x = 0$ производная $f'(0) = 0$.

Пример 6.19. Функция $y = \begin{cases} 2x, & x \geq 0, \\ x, & x < 0 \end{cases}$ не имеет производной в точке $x = 0$ (левая и правая производные различны), однако она возрастает при всех значениях x, в том числе и в точке $x = 0$.

Понятие экстремума

Точка x_0 называется точкой *локального максимума* функции $f(x)$, если существует такая окрестность точки x_0, что для всех x из этой окрестности $f(x) \leq f(x_0)$.

Точка x_0 называется точкой *локального минимума* функции $f(x)$, если существует такая окрестность точки x_0, что для всех x из этой окрестности $f(x) \geq f(x_0)$. Значение функции в точке максимума называется локальным максимумом, значение функции в точке минимума — локальным минимумом данной функции. Максимум и минимум функции называются ее локальными экстремумами (*extremum* — крайний).

Термин «локальный (относительный, местный) экстремум» обусловлен тем, что введенное понятие экстремума связано с окрестностью данной точки в области определения функции, а не со всей этой областью. Функция может иметь несколько экстремумов, причем может случиться, что минимум в одной точке больше максимума в другой. Мы будем рассматривать лишь точки строгого максимума и минимума.

Точка x_0 называется точкой строгого *локального максимума* (*минимума*) функции $y = f(x)$, если для всех x из окрестности точки x_0 верно строгое неравенство $f(x) < f(x_0)$ (соответственно $f(x) > f(x_0)$).

Пример 6.20. Функция $y = \begin{cases} x^2, & x \neq 0, \\ 1, & x = 0 \end{cases}$ разрывна в точке $x = 0$, но имеет в этой точке локальный максимум, поскольку существует окрестность точки $x = 0$, в которой $f(x) < f(x_0)$.

Наибольшее либо наименьшее значение функции на промежутке называется *глобальным экстремумом*. Глобальный экстремум может достигаться либо в точках локального экстремума, либо на концах отрезка.

Необходимое условие экстремума

Теорема 6.10. Если функция $y = f(x)$ имеет экстремум в точке x_0, то ее производная $f'(x_0)$ либо равна нулю, либо не существует.

Если в точке x_0 функция имеет экстремум и дифференцируема, то в некоторой окрестности этой точки выполнены условия теоремы Ферма, следовательно, производная функции в этой точке равна нулю.

Но функция $y = f(x)$ может иметь экстремум и не быть дифференцируемой в этой точке. Достаточно указать пример. Примером может служить функция $y = |x|$, которая имеет минимум в точке $x = 0$, однако недифференцируема в этой точке.

Замечание 1. Геометрическую иллюстрацию теоремы дает рис. 6.24. Функция $y = f(x)$, график которой представлен на этом рисунке, имеет экстремумы в точках x_1, x_3, x_4, при этом в точке x_1 производная не существует, в точке x_3 она равна нулю, в точке x_4 производная по модулю обращается в бесконечность. В точках x_2, x_5 функция экстремума не имеет, причем в точке x_2 производная обращается в бесконечность, в точке x_5 производная равна нулю.

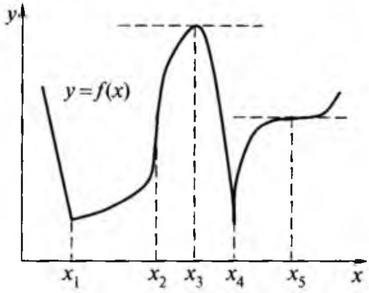

Рис. 6.24

Замечание 2. Точки, в которых выполняется необходимое условие экстремума для непрерывной функции, называются *крити-*

ческими точками этой функции. Они определяются как решения уравнения $f'(x) = 0$ (стационарные точки) или $f'(x) = \infty$.

Замечание 3. Не в каждой своей критической точке функция обязательно имеет максимум или минимум.

Пример 6.21. Рассмотрим функцию $y = x^3$. Критической для этой функции является точка $x = 0$, что следует из уравнения $f'(x) = 3x^2 = 0$. Однако эта функция при всех x является возрастающей и экстремума не имеет.

Теорема 6.11. Первое достаточное условие экстремума.

Пусть для функции $y = f(x)$ выполнены следующие условия: 1) $y = f(x)$ непрерывна в окрестности точки x_0; 2) $f'(x) = 0$ или $|f'(x)| = \infty$ в точке x_0; 3) $f'(x)$ при переходе через точку x_0 меняет свой знак. Тогда в точке $x = x_0$ функция $y = f(x)$ имеет экстремум, причем это: минимум, если при переходе через точку x_0 слева направо производная меняет свой знак с минуса на плюс; максимум, если при переходе через точку x_0 производная меняет свой знак с плюса на минус. Если производная $f'(x)$ при переходе через точку x_0 не меняет своего знака, экстремума в точке $x = x_0$ нет.

Условия теоремы можно свести в следующую таблицу:

Знак производной до и после перехода через точку x_0		*Экстремум*
–	+	Минимум
+	–	Максимум
–	–	Нет
+	+	Нет

Так как по условию $f'(x) < 0$ при $x < x_0$, то на левом относительно точки x_0 интервале функция $f(x)$ убывает. Так как $f'(x) > 0$ при $x > x_0$, то на правом относительно точки x_0 интервале функция $f(x)$ возрастает. Следовательно, $f(x_0)$ есть наименьшее значение функции $f(x)$ в окрестности точки x_0, а это означает, что $f(x_0)$ есть локальный минимум функции $f(x)$. Если при переходе с левого интервала на правый функция продолжает убывать, то в точке x_0 не будет достигаться минимальное значение функции (экстремума нет).

Аналогично доказывается существование максимума.

На рис. 6.25, *а–з* представлены возможные случаи наличия или отсутствия экстремума непрерывной функции, производная которой в критической точке равна нулю или обращается в бесконечность.

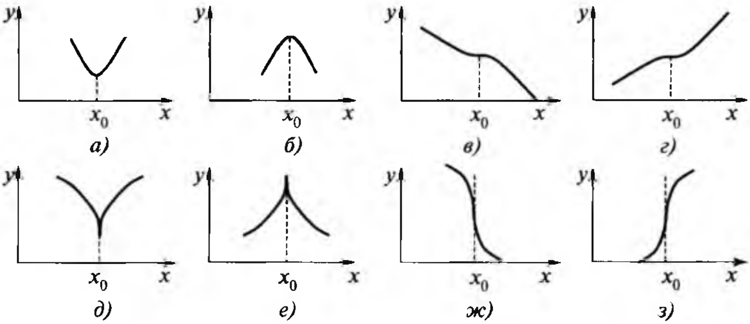

Рис. 6.25

Теорема 6.12. Второе достаточное условие экстремума. Пусть для функции $y = f(x)$ выполнены следующие условия: 1) $y = f(x)$ непрерывна в окрестности точки x_0; 2) $f'(x) = 0$ в точке x_0; 3) $f''(x) \neq 0$ в точке x_0. Тогда в точке x_0 достигается экстремум, причем если $f''(x) > 0$, то в точке $x = x_0$ функция $y = f(x)$ имеет минимум, если $f''(x) < 0$, то в точке $x = x_0$ функция $y = f(x)$ имеет максимум.

По определению второй производной

$$f''(x_0) = \lim_{x \to x_0} \frac{f'(x) - f'(x_0)}{x - x_0}.$$

Но величина $f'(x) = 0$ по условию теоремы.

Поэтому $f''(x_0) = \lim\limits_{x \to x_0} \dfrac{f'(x)}{x - x_0}$.

Если $f''(x) > 0$, то дробь $\dfrac{f'(x)}{x - x_0} > 0$ в некоторой окрестности точки $x = x_0$. При $x < x_0$ дробь положительна, если $f'(x) < 0$. При $x > x_0$ дробь положительна, если $f'(x) > 0$. Следовательно, производная $f'(x)$ при переходе через точку $x = x_0$ меняет знак, поэтому есть экстремум. Если знак производной меняется с минуса на плюс, значит, это минимум. Аналогично доказывается случай $f''(x) < 0$.

Пример 6.22. Исследовать на экстремум функцию $y = x^2 + 2x + 3$.

1. Находим производную $y' = 2x + 2$.

2. Находим критические точки, для чего приравниваем к нулю производную:

$y' = 2x + 2 = 0, \to x_0 = -1$.

3. Изучаем знак производной слева и справа от этой точки (рис. 6.26). Поскольку знак производной меняется с минуса на плюс, в точке $x_0 = -1$ достигается минимум.

4. Находим величину минимума: $y_{\min}(-1) = 2$.

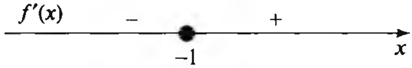

Рис. 6.26

Наибольшее и наименьшее значения функции, непрерывной на отрезке

Если функция $f(x)$ определена и непрерывна на отрезке $[a; b]$, то согласно второй теореме Вейерштрасса она на этом отрезке достигает своих наибольшего и наименьшего значений.

Если свое наибольшее значение M функция $f(x)$ принимает во внутренней точке x_0 отрезка $[a; b]$, то $M = f(x_0)$ будет локальным максимумом функции $f(x)$, так как в этом случае существует окрестность точки x_0 такая, что значения $f(x)$ для всех точек x из этой окрестности будут не больше $f(x_0)$.

Однако свое наибольшее значение M функция $f(x)$ может принимать и на концах отрезка $[a; b]$. Поэтому, чтобы найти наибольшее значение M непрерывной на отрезке $[a; b]$ функции $f(x)$, надо найти все максимумы функции в интервале $(a; b)$ и значения $f(x)$ на концах отрезка $[a; b]$ и выбрать среди них наибольшее число. Вместо исследования на максимум можно ограничиться нахождением значений функции в критических точках. Наименьшее значение m непрерывной на отрезке $[a; b]$ функции $f(x)$ разыскивается среди всех минимумов функции $f(x)$ в интервале $(a; b)$ и значений $f(a)$ и $f(b)$.

Выпуклость функции. Точки перегиба

График функции $y = f(x)$, дифференцируемой на интервале $(a; b)$, имеет на этом интервале *выпуклость, направленную вверх* (*вниз*), если график этой функции в пределах интервала $(a; b)$ лежит не выше (не ниже) любой своей касательной (рис. 6.27).

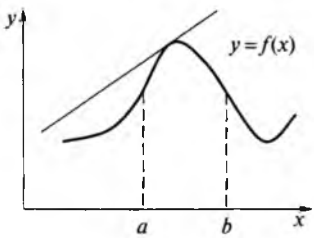

Рис. 6.27

Теорема 6.13. Пусть функция $y = f(x)$ определена на интервале $(a; b)$ и имеет непрерывную, не равную нулю в точке $x_0 \in (a; b)$ вторую производную. Тогда если $f''(x) > 0$ всюду на интервале $(a; b)$, то функция имеет выпуклость вниз на этом интервале, если $f''(x) < 0$, то функция имеет выпуклость вверх.

Точкой перегиба графика функции $y = f(x)$ называется точка $M(x_1, f(x_1))$, разделяющая промежутки выпуклости вверх и вниз. Иными словами, точка $M(x_1, f(x_1))$ — точка перегиба кривой, если в этой точке кривая переходит с одной стороны касательной на другую, меняя направление выпуклости (рис. 6.28).

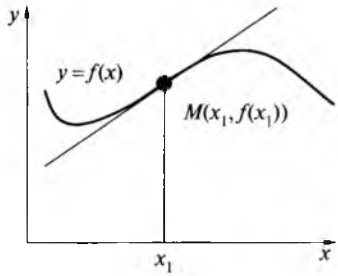

Рис. 6.28

Теорема 6.14. Если: 1) первая производная $f'(x)$ непрерывна в окрестности точки x_1; 2) вторая производная $f''(x) = 0$ или не существует в точке x_1; 3) $f''(x)$ при переходе через точку x_1 меняет свой знак, тогда в точке $M(x_1, f(x_1))$ функция $y = f(x)$ имеет перегиб.

Пример 6.23. Найти интервалы выпуклости и точки перегиба функции $y = x^3 - 3x^2 + x - 1$.

Находим вторую производную функции:

$y' = 3x^2 - 6x + 1, \quad y'' = 6x - 6.$

Находим точку, в которой вторая производная равна нулю: $y'' = 0$ при $x = 1$. Исследуем знак второй производной слева и справа от найденной точки. Для этого рисуем числовую ось и указываем на ней знаки второй производной (рис. 6.29). Делаем заключение об интервале выпуклости вверх слева от точки $x=1$ и интервале выпуклости вниз справа от этой точки.

Делаем вывод о наличии перегиба в точке $(1, -2)$.

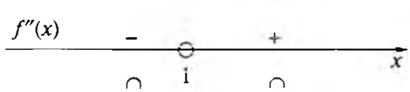

Рис. 6.29

Асимптоты графика функции

Напомним определения вертикальной и горизонтальной асимптот, а также введем понятие наклонной асимптоты.

Прямая $x = x_0$ называется *вертикальной асимптотой* графика функции $y = f(x)$, если хотя бы одно из предельных значений $\lim\limits_{x \to x_0 - 0} f(x)$ или $\lim\limits_{x \to x_0 + 0} f(x)$ равно $+\infty$ или $-\infty$, т.е.

$$\lim_{\left[\begin{array}{l} x \to x_0 - 0 \\ x \to x_0 + 0 \end{array}\right)} f(x) = \left[\begin{array}{l} +\infty \\ -\infty \end{array}\right. .$$

Прямая $y = y_0$ называется *горизонтальной асимптотой* графика функции $y = f(x)$, если хотя бы одно из предельных значений $\lim\limits_{x \to +\infty} f(x)$ или $\lim\limits_{x \to -\infty} f(x)$ равно b. График функции может иметь только правую горизонтальную асимптоту или только левую.

Прямая $y = kx + b$ называется *наклонной асимптотой* графика функции $y = f(x)$, если

$$\lim_{x \to \infty} (f(x) - kx - b) = 0,$$

т.е. когда функция при $x \to \infty$ представима в виде

$$f(x) = kx + b + \alpha(x), \text{ где } \lim_{x \to +\infty} \alpha(x) = 0.$$

Существование асимптоты $y = kx + b$ у кривой $y = f(x)$ при $x \to \infty$ означает, что при $x \to \infty$ функция ведет себя «почти как линейная», т.е. отличается от линейной функции $y = kx + b$ бесконечно мало (рис. 6.30). Наклонная асимптота может быть как правой так и левой.

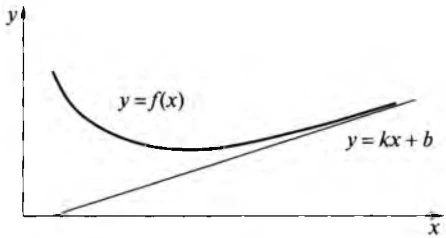

Рис. 6.30

Исследование функций и построение их графиков

При построении графика функции необходимо провести ее предварительное исследование. Построение сразу по точкам, за исключением элементарных случаев, может привести к потере на графике важных свойств функции. Примерная схема исследования функции с целью построения ее графика представлена ниже.

1. Область определения $D(y)$ и область допустимых значений $E(y)$ функции.
2. Симметрия и периодичность.
3. Точки разрыва и промежутки непрерывности функции.
4. Нули функции и промежутки постоянного знака.
5. Экстремумы и промежутки монотонности.
6. Точки перегиба и промежутки выпуклости.
7. Асимптоты.

Замечание 1. Схема представлена как примерная. Пункты исследования можно опускать, если они дают банальную информацию, или переставлять, если обнаруживаются интересные особенности поведения графика. Однако без нахождения разрывов, экстремумов, асимптот и исследования на выпуклость часто невозможно получить график, правильно отражающий поведение функции.

Замечание 2. Для уточнения графика можно найти некоторые дополнительные точки, но иногда удается обойтись и без них.

Замечание 3. Рекомендуется строить график одновременно с исследованием функции, нанося на координатную плоскость информацию по завершении каждого пункта исследования.

Пример 6.24. Провести полное исследование функции $y = \dfrac{1}{1+x^2}$ и построить график.

1. Областью определения является вся числовая ось.

2. Функция четная: $f(-x) = f(x)$, так что ее график симметричен относительно оси ординат. Из четности функции следует, что достаточно построить ее график в правой полуплоскости, а затем отразить его в левую полуплоскость.

3. Точек разрыва нет, функция непрерывная на всей числовой оси.

4. При $x = 0$ имеем $y = 1$. Функция положительна при всех x, так что график функции лежит в верхней полуплоскости.

5. $y' = -\dfrac{2x}{(1+x^2)^2}$. Функция возрастает при $x < 0$ и убывает при $x > 0$. Точка $x = 0$ — критическая. При переходе x через точку $x = 0$ производная $y'(x)$ меняет знак с плюса на минус (рис. 6.31). Следовательно, точка $x = 0$ — точка максимума, $y(0) = 1$.

6. $y'' = -2\dfrac{1-3x^2}{(1+x^2)^3}$. Вторая производная обращается в нуль в точках $x = \pm\dfrac{1}{\sqrt{3}}$. Исследуем точку $x = \dfrac{1}{\sqrt{3}}$. При $x > \dfrac{1}{\sqrt{3}}$ имеем $y'' > 0$, т.е. кривая

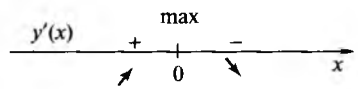

Рис. 6.31

выпукла вниз; при $x < \frac{1}{\sqrt{3}}$ получаем $y'' < 0$ (кривая выпукла вверх) (рис. 6.32). Следовательно, точка $x = \frac{1}{\sqrt{3}}$ — точка перегиба графика функции.

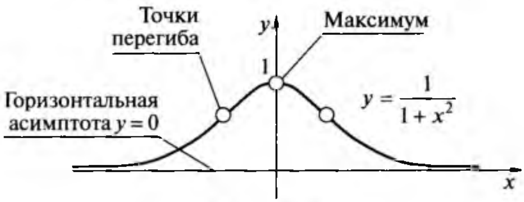

Рис. 6.32

7. $\lim\limits_{x \to +\infty} \frac{1}{1 + x^2} = 0$. График имеет горизонтальную асимптоту $y = 0$, наклонных асимптот нет.

Строим график в правой полуплоскости и симметрично отражаем его в левую полуплоскость. График функции изображен на рис. 6.32.

6.5. ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Понятие функции как отображения

Расширение понятия функции одной переменной на случай функции нескольких переменных совершенствует математический аппарат, позволяя изучать зависимость исследуемого объекта одновременно от многих причин.

Определение. *Отображением* множества E в множество F, или *функцией*, определенной на E со значениями в F, называется правило, по которому каждому элементу $x \in E$ ставится в соответствие определенный элемент $y \in F$.

Элемент $x \in E$ называют *независимым переменным*, или *аргументом* функции, элемент $y \in F$ называют *значением* функции, или *образом*; при этом элемент $x \in E$ называется *прообразом* элемента $y \in F$.

Отображение (функцию) обычно обозначают буквой f или символом $f\colon E \to F$, указывая тем самым, что f отображает множество E в

F. Иногда функцию удобно задавать посредством равенства $y = f(x)$, в котором содержится закон соответствия. Например, можно задать функцию равенством $y = x^2 + 1$.

Как образ, так и прообраз могут быть либо скалярной величиной, либо вектором (рис. 6.33).

Рис. 6.33

По этому признаку функции можно классифицировать так.

1. Скалярная функция скалярного аргумента. Эти функции изучались нами как функции одной переменной.
2. Скалярная функция векторного аргумента. Обычно говорят: функция нескольких или многих переменных. Эти функции рассматриваются в настоящей главе.
3. Векторная функция скалярного аргумента.
4. Векторная функция векторного аргумента.

Понятие функции нескольких переменных

Переменная величина z называется *функцией двух переменных* x, y, если каждой совокупности их значений из данной области D соответствует единственное определенное значение z. Соответствующая зависимость записывается в виде $z = f(x, y)$ или $z = z(x, y)$. Если имеется n переменных величин $x_1, x_2, \ldots, x_n$, то функциональная зависимость имеет вид $z = f(x_1, x_2, \ldots, x_n)$. Введем обозначение $x = (x_1, x_2, \ldots, x_n)$. Назовем совокупность n переменных величин *вектором*, а сами величины — *координатами* вектора. Тогда функция нескольких переменных может быть названа *функцией векторного аргумента*. Множество D называется *областью определения* функции.

Пусть функция $z = f(x, y)$ определена в некоторой области D на плоскости xOy. Каждой точке (x, y) на плоскости будет соответствовать точка $M(x, y, z)$ трехмерного пространства. Множество таких точек $M(x, y, z)$ в трехмерной декартовой системе координат представляет собой некоторую поверхность и называется графиком функции $z = f(x, y)$. Для построения графика функции $z = f(x, y)$ следует рассматривать функции одной переменной $z = f(x_0, y)$ и $z = f(x, y_0)$, представляющие сечения графика плоскостями, параллельными координатным плоскостям xOz и yOz. Поверхности в трехмерном пространстве за малым исключением сложно нарисовать без компьютерной поддержки. Поэтому многие рассуждения, нуждающиеся в геометрическом представлении, будут построены на простых объемных геометрических фигурах. К их числу относятся шар, конус, цилиндр, параболоид вращения.

Среди функций нескольких переменных перечислим следующие.

1. Линейная функция $z = \sum_{i=1}^{n} a_i x_i + b$.

2. Квадратичная функция $z = \sum_{i,j=1}^{n} a_{ij} x_i x_j$. При выполнении условия $a_{ij} = a_{ji}$ квадратичная функция называется квадратичной формой переменных $x_1, x_2, \ldots, x_n$. Она играет важную роль в линейной алгебре.

3. Функция Кобба—Дугласа $z = \prod_{i=1}^{n} a x_i^{\alpha_i}$. Для двух переменных она принимает вид $z = a x_1^{\alpha_1} x_2^{\alpha_2}$. С помощью функции Кобба—Дугласа строятся производственные функции, выражающие результат производственной деятельности в зависимости от различных факторов $x_1, x_2, \ldots, x_n$.

Линии уровня

Определение. *Линией уровня* функции $z = f(x, y)$ называется такая линия $f(x, y) = C$ на плоскости xOy, в точках которой функция принимает постоянное значение $z = C$. Например, функция $z = x^2 + y^2$, описывая поверхность, называемую параболоидом вращения, имеет линии уровня вида $x^2 + y^2 = C$. Задавая параметру C различные значения из области $C \geq 0$, получим несколько линий уровня в виде совокупности концентрических окружностей с центром в начале координат (рис. 6.34). Эта совокупность называется фрагментом карты линий уровня.

Линия уровня может быть получена при пересечении графика функции $z = f(x, y)$ плоскостью $z = C$, параллельной плоскости xOy. Затем эту линию следует спроектировать на плоскость xOy.

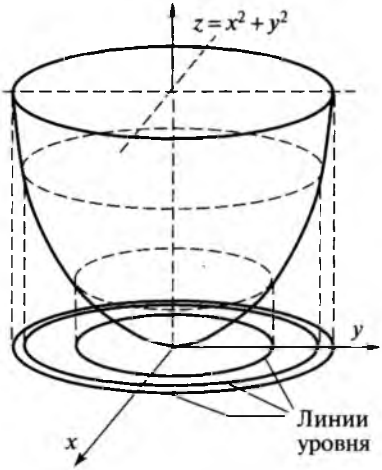

Рис. 6.34

Пример 6.25. Построить фрагмент карты линий уровня функции $z = \arcsin xy$ в области $|x| \le 1$, $|y| \le 1$.

Решение. Пусть $z = C$. Тогда $xy = \sin C$. При $C = 0$ линиями уровня будут прямые $x = 0$ и $y = 0$. При $x \ne 0$ имеем $y = \dfrac{\sin C}{x}$. Задавая параметру C различные значения от $-\dfrac{\pi}{2}$ до $\dfrac{\pi}{2}$, получим фрагмент карты линий уровня в области $|x| \le 1$, $|y| \le 1$ (рис. 6.35). На рисунке оттенками серого цвета закрашены области между линиями разного уровня.

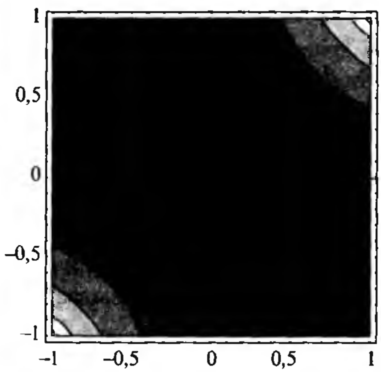

Рис. 6.35

Пример 6.26. Построить фрагмент карты линий уровня функции Кобба—Дугласа $z = x_1^{1/4}x_2^{3/4}$.

Решение. Функция Кобба—Дугласа играет важную роль в экономико-математических задачах, поэтому рассмотрим ее подробнее. Возьмем $z = C$. Соответствующие линии уровня имеют вид $x_1 = \frac{C^4}{x_2^3}$. При $C = 1$ и $C = 2$ получим $x_1 = \frac{1}{x_2^3}$ и $x_1 = \frac{16}{x_2^3}$. На рис. 6.36 представлена функция Кобба—Дугласа $z = x_1^{1/4}x_2^{3/4}$, которую пересекают плоскости $z = 1$ и $z = 2$. Линии уровня для $z = 1$ и $z = 2$ представлены красным и оранжевым цветами. Вид сверху на трехмерный график спроектирован на рис. 6.37, где на плоскости x_1Ox_2 построена совокупность этих линий. Так получается карта линий уровня функции. На рис. 6.37 выделены линии уровня, соответствующие $z = 1$ и $z = 2$. Оттенками серого цвета закрашены области между линиями разного уровня.

Предел функции нескольких переменных

Определение. Число a называется пределом функции $z = f(x, y)$ при $x \to x_0$, $y \to y_0$ (или в точке (x_0, y_0)), если для любой последовательности точек (x_n, y_n), где $n = 1, 2, \ldots$, сходящейся к точке (x_0, y_0), соответствующая последовательность значений функции $f(x_n, y_n)$ сходится к a. Математическое обозначение:

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = a.$$

Замечание. Вычисление пределов функции двух переменных является более сложной задачей по сравнению с вычислением пределов функции одной переменной. Это связано с тем, что точка (x_n, y_n) может стремиться к точке (x_0, y_0) по любому направлению на плоскости в отличие от функции одной переменной, где переменная x может стремиться к числу x_0 на числовой прямой только справа или слева. Получающиеся при этом многочисленные значения пределов функции двух переменных должны совпадать друг с другом. Легче доказать отсутствие предела функции $z = f(x, y)$ при $x \to x_0$, $y \to y_0$. Достаточно выбрать два таких направления, движение по которым приводит к различным значениям пределов.

Пример 6.27. Найти предел функции $f(x, y) = \frac{\sin xy}{y}$ при $x \to 0$, $y \to 0$.

Решение. Функция $f(x, y)$ определена всюду, кроме линии $y = 0$. Функция в точке (0, 0) не определена. При нахождении предела следует умножить числитель и знаменатель на x, сделать замену $xy = \rho$, а затем воспользоваться первым замечательным пределом:

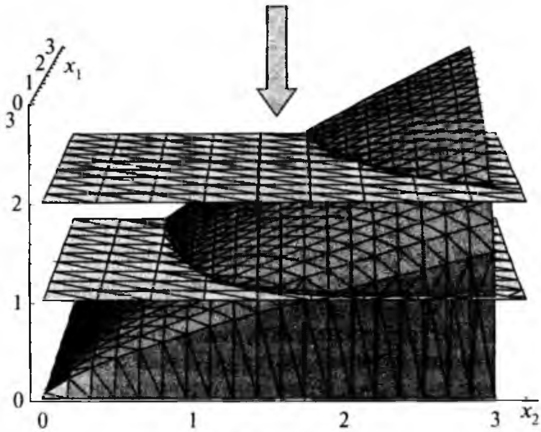

Рис. 6.36

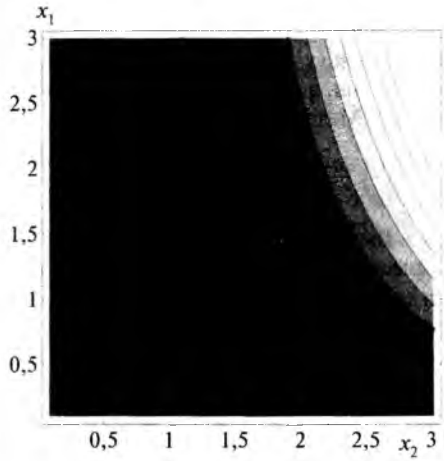

Рис. 6.37

$$\lim_{\substack{x \to 0 \\ y \to 0}} \frac{\sin xy}{y} = \lim_{\substack{x \to 0 \\ y \to 0}} \frac{x \sin xy}{xy} = \lim_{\substack{x \to 0 \\ \rho \to 0}} \left(x \cdot \frac{\sin \rho}{\rho} \right) = 0.$$

Непрерывность

Непрерывность функции нескольких переменных

Определение. Функция $f(x, y)$ называется непрерывной в точке (x_0, y_0), если она:

1) определена в точке (x_0, y_0);
2) имеет конечный предел при $x \to x_0$, $y \to y_0$;
3) $\lim\limits_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = f(x_0, y_0)$.

Функция называется непрерывной в области определения, если она непрерывна в каждой точке области. Точки, в которых функция не является непрерывной, называются *точками разрыва*. Эти точки могут быть как изолированными, так и составлять *линии разрыва*.

Свойства непрерывных функций нескольких переменных

Пусть функции $f(x, y)$ и $g(x, y)$ непрерывны в точке $M(x_0, y_0)$. Тогда:

1) функции $f(x, y) \pm g(x, y)$ непрерывны в точке $M(x_0, y_0)$;
2) функция $f(x, y) \cdot g(x, y)$ непрерывна в точке $M(x_0, y_0)$;
3) функция $\dfrac{f(x, y)}{g(x, y)}$ при условии $g(x_0, y_0) \neq 0$ непрерывна в точке $M(x_0, y_0)$.

Частные производные

Дадим аргументам функции $z = f(x, y)$ приращения Δx и Δy, оставаясь в области определения. Функция в точке $(x + \Delta x, y + \Delta y)$ будет равна $f(x + \Delta x, y + \Delta y)$. Разность между значениями функции в точках $(x + \Delta x, y + \Delta y)$ и (x, y) называется *полным приращением* функции Δf:

$$\Delta f = f(x + \Delta x, y + \Delta y) - f(x, y).$$

Частным приращением функции *по аргументу* x называется величина

$$\Delta_x f = f(x + \Delta x, y) - f(x, y).$$

Частным приращением функции *по аргументу* y называется величина

$$\Delta_y f = f(x, y + \Delta y) - f(x, y).$$

Пример 6.28. Найти полное приращение и частные приращения функции $z = xy$.

Решение. Полное приращение функции

$$\Delta z = (x + \Delta x)(y + \Delta y) - xy = x\Delta y + y\Delta x + \Delta x\Delta y.$$

Частные приращения функции

$$\Delta_x z = (x + \Delta x)y - xy = y\Delta x, \quad \Delta_y z = x(y + \Delta y) - xy = x\Delta y.$$

Пример показывает, что $\Delta z \neq \Delta_x z + \Delta_y z$.

Составим отношения $\dfrac{\Delta_x f}{\Delta x}$ и $\dfrac{\Delta_y f}{\Delta y}$. Они являются функциями соответственно Δx и Δy. Аргументы x и y в данной точке (x, y) являются постоянными величинами. Устремим Δx и Δy к нулю.

Определение. Частной производной функции $f(x, y)$ в точке (x, y) по переменной x называется предел отношения $\dfrac{\Delta_x f}{\Delta x}$ при $\Delta x \to 0$ и обозначается символами f'_x или $\dfrac{\partial f}{\partial x}$:

$$f'_x \equiv \frac{\partial f}{\partial x} = \lim_{\Delta x \to 0} \frac{\Delta_x f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}.$$

Определение. Частной производной функции $f(x, y)$ в точке (x, y) по переменной y называется предел отношения $\dfrac{\Delta_y f}{\Delta y}$ при $\Delta y \to 0$ и обозначается символами f'_y или $\dfrac{\partial f}{\partial y}$:

$$f'_y \equiv \frac{\partial f}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta_y f}{\Delta y} = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}.$$

Замечание 1. При нахождении частной производной по одной переменной другая переменная остается постоянной. Поэтому правила вычисления частных производных совпадают с правилами вычисления производной функции одной переменной с параметрами.

Замечание 2. В приблизительных расчетах частные производные можно заменять отношениями приращений $\dfrac{\partial f}{\partial x} \approx \dfrac{\Delta_x f}{\Delta x}$, $\dfrac{\partial f}{\partial y} \approx \dfrac{\Delta_y f}{\Delta y}$. Результат вычисления будет тем более точным, чем меньшими по абсолютной величине будут приращения аргументов.

Пример 6.29. Найти частные производные функции $z = \sin(x^2 y^3)$.

Решение. Считаем аргумент y постоянной величиной, берем производную по x:

$z'_x = 2xy^3\cos(x^2 y^3)$.

Теперь x — константа, изменяется аргумент y:

$z'_y = 3x^2 y^2\cos(x^2 y^3)$.

Понятие дифференцируемости

Понятие дифференцируемости формализуется в следующем определении.

Определение. Функция $f(x, y)$ называется дифференцируемой в точке (x, y) области ее определения, если полное приращение функции можно представить в виде

$$\Delta f = f(x+\Delta x, y+\Delta y) - f(x, y) = A\cdot\Delta x + B\cdot\Delta y + \alpha\cdot\Delta x + \beta\cdot\Delta y,$$

где A и B — некоторые постоянные для точки (x, y) числа; α и β — бесконечно малые функции, зависящие от Δx и Δy и стремящиеся к нулю при $\Delta x \to 0$ и $\Delta y \to 0$.

Дифференцируемая функция, согласно определению, есть сумма линейных по Δx и Δy слагаемых и нелинейной бесконечно малой по сравнению с Δx и Δy добавки.

Выясним смысл постоянных A и B.

Теорема 6.15 (о существовании частных производных у дифференцируемой функции).

Если функция $f(x, y)$ дифференцируема в точке (x, y) области ее определения, то она имеет в этой точке частные производные $\frac{\partial f}{\partial x}$ и $\frac{\partial f}{\partial y}$, причем $\frac{\partial f}{\partial x} = A$ и $\frac{\partial f}{\partial y} = B$.

Следствие (о приращении дифференцируемой функции)

Если функция $f(x, y)$ дифференцируема в точке (x, y) области ее определения, то полное приращение функции может быть представлено в виде

$$\Delta f = \frac{\partial f}{\partial x}\cdot\Delta x + \frac{\partial f}{\partial y}\cdot\Delta y + \alpha\cdot\Delta x + \beta\cdot\Delta y.$$

Дифференциал

Определение. Пусть функция $z = f(x, y)$ дифференцируема в точке (x, y). *Полным дифференциалом dz* функции называется **линейная** относительно Δx и Δy часть приращения $\frac{\partial f}{\partial x}\cdot\Delta x + \frac{\partial f}{\partial y}\cdot\Delta y$ этой функции $dz = \frac{\partial f}{\partial x}\cdot\Delta x + \frac{\partial f}{\partial y}\cdot\Delta y$.

Следовательно, полный дифференциал функции отличается от ее приращения на сумму последних слагаемых $\alpha\cdot\Delta x + \beta\cdot\Delta y$, которые являются бесконечно малыми по сравнению с первыми членами.

З а м е ч а н и е 1. Последняя фраза означает, что при задании достаточно малых приращений Δx и Δy соответствующее приращение функции может быть заменено полным дифференциалом, который легко рассчитывается.

З а м е ч а н и е 2. Изучая поведение функции нескольких переменных, мы будем опираться на дифференциал функции. В этом существенное отличие от функции одной переменной, где главным инструментом в изучении поведения функции являлась производная.

Полный дифференциал называется **главной** частью приращения функции при условии $dz \neq 0$.

Назовем приращения аргументов Δx и Δy дифференциалами этих аргументов: $\Delta x \equiv dx$, $\Delta y \equiv dy$ — и перепишем формулу полного дифференциала в виде

$$dz = \frac{\partial f}{\partial x} \cdot dx + \frac{\partial f}{\partial y} \cdot dy. \tag{6.5}$$

Пример 6.30. Найти приращение функции $z = x^2 + y^2$ в точке (1, 1).

Решение. Заменим приращение функции дифференциалом. Получим

$$\Delta z \approx dz = \frac{\partial f}{\partial x} \cdot \Delta x + \frac{\partial f}{\partial y} \cdot \Delta y = 2x\Big|_{\substack{x=1\\y=1}} \Delta x + 2y\Big|_{\substack{x=1\\y=1}} \Delta y = 2\Delta x + 2\Delta y.$$

Возьмем, например, $\Delta x = -0{,}1$ и $\Delta y = -0{,}1$, т.е. выйдем из точки (1, 1) по биссектрисе первой четверти по направлению к началу координат. Функция изменится на величину $\approx -0{,}4$.

Пример 6.31. Для функции Кобба—Дугласа $z = kx^\alpha y^\beta$ найти связь между дифференциалами аргументов при условии, что дифференциал функции останется равным нулю.

Решение. Найдем полный дифференциал функции и приравняем его нулю:

$$dz = z'_x dx + z'_y dy = 0.$$

Отсюда

$$\frac{dx}{dy} = -\frac{z'_y}{z'_x}.$$

Подставим частные производные функции Кобба—Дугласа:

$$\frac{dx}{dy} = -\frac{k\beta x^\alpha y^{\beta-1}}{k\alpha x^{\alpha-1} y^\beta} = -\frac{\beta}{\alpha}\frac{x}{y}.$$

Полученная формула содержит важное экономическое понятие — предельную норму замены трудовых ресурсов y капиталом x. На языке экономического анализа пример 31 формулируется так: на какую величину Δx следует изменить объем вложенного капитала x, чтобы при изменении трудовых ресурсов y на величину Δy выпуск продукции z оставался постоянным: $\Delta z = 0$?

Сложные функции. Их производные

Дифференцируемость сложной функции нескольких переменных

Теорема 6.16. (о дифференцируемости сложной функции). Пусть функции $x = x(t)$ и $y = y(t)$ дифференцируемы в некоторой точке t, а функция $z = z(x, y)$ дифференцируема в соответствующей точке (x, y). Тогда сложная функция $z = z(x(t), y(t))$ дифференцируема в точке t, причем

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Производная функции $z = z(x, y)$ при $x = x(t)$ и $y = y(t)$

1. Пусть $z = z(x, y)$, где $x = x(t)$ и $y = y(t)$. Тогда дифференциал функции

$$dz = z'_x dx + z'_y dy. \tag{6.6}$$

Для функций $x = x(t)$ и $y = y(t)$ дифференциалы равны $dx = x'dt$ и $dy = y'dt$. Подставим их в (6.6):

$$dz = z'_x dx + z'_y dy = z'_x x'dt + z'_y y'dt = (z'_x x' + z'_y y')dt.$$

Отсюда

$$\frac{dz}{dt} = z'_t = z'_x x' + z'_y y'.$$

Производная функции $z = z(u, v)$ при $u = u(x, y)$ и $v = v(x, y)$

2. Пусть $z = z(u, v)$, где $u = u(x, y)$ и $v = v(x, y)$, и условия непрерывности всех частных производных выполнены.

С одной стороны, $dz = z'_u du + z'_v dv$, причем $du = u'_x dx + u'_y dy$ и $dv = v'_x dx + v'_y dy$.

Подставим du, dv в dz:

$$dz = z'_u(u'_x dx + u'_y dy) + z'_v(v'_x dx + v'_y dy) = (z'_u u'_x + z'_v v'_x)dx + (z'_u u'_y + z'_v v'_y)dy.$$

С другой стороны,

$$z = z(u(x, y), v(x, y)) = z(x, y)$$

есть функция двух конечных аргументов x и y. Поэтому

$$dz = z'_x dx + z'_y dy.$$

Сравним дифференциалы dz, полученные различными путями. В силу единственности дифференциала, а также учитывая, что равенства должны выполняться при произвольных задаваемых нами малых приращениях аргументов dx, dy, будем иметь

$$z'_x = z'_u u'_x + z'_v v'_x,$$
$$z'_y = z'_u u'_y + z'_v v'_y.$$

Производная функции $z = z(u, v)$ при произвольном задании аргументов

3. Пусть $z = z(u, v, t)$, причем $u = u(x, t)$ и $v = v(y, t)$, $x = x(t)$, $y = y(t)$. Найдем все дифференциалы:

$$dz = z'_u du + z'_v dv + z'_t dt;$$
$$du = u'_x dx + u'_t dt;$$
$$dv = v'_y dy + v'_t dt;$$
$$dx = x' dt;$$
$$dy = y' dt.$$

Подставим последние дифференциалы в первое равенство:

$$dz = (z'_u(u'_x x' + u'_t) + z'_v(v'_y y' + v'_t) + z'_t)dt$$

или $\dfrac{dz}{dt} = z'_u(u'_x x' + u'_t) + z'_v(v'_y y' + v'_t) + z'_t$,

где $\dfrac{dz}{dt}$ — полная производная функции z по независимой переменной t, а $z'_t = \dfrac{\partial z}{\partial t}$ — частная производная по t, при вычислении которой зависимые аргументы u, v, x, y принимаются за постоянные.

Производная по направлению

Мы познакомились с частными производными $\dfrac{\partial z}{\partial x}$ и $\dfrac{\partial z}{\partial y}$. Каждая из них представляет приращение функции вдоль соответствующей оси (x или y) при неизменной второй переменной:

$$\frac{\partial z}{\partial x} \approx \left.\frac{\Delta z}{\Delta x}\right|_{y=\text{const}} \quad \text{и} \quad \frac{\partial z}{\partial y} \approx \left.\frac{\Delta z}{\Delta y}\right|_{x=\text{const}}.$$

Во многих приложениях функций векторного аргумента, включая экономические, требуется определять изменение функции не только вдоль оси x или y, но вдоль любого направления на координатной плоскости xOy.

Пусть функция $z = z(x, y)$ задана в некоторой окрестности точки $M_0(x_0, y_0)$ и описывает поверхность S. Изобразим на рисунке ось l, в направлении которой нужно найти производную. При перемещении в направлении l от точки $M_0(x_0, y_0)$ в точку $M(x, y)$ функция получит приращение

$$\Delta_l z = z(x, y) - z(x_0, y_0),$$

соответствующее приращению Δl.

Поскольку $x = x_0 + \Delta x$, $y = y_0 + \Delta y$, приращение в направлении l составит $\Delta l = \sqrt{(\Delta x)^2 + (\Delta y)^2}$.

Определение. Предел отношения $\frac{\Delta_l z}{\Delta l}$ при $\Delta l \to 0$ называется *производной по направлению* l от функции $z = z(x, y)$ в точке $M_0(x_0, y_0)$:

$$\frac{\partial z}{\partial l} = \lim_{\Delta l \to 0} \frac{\Delta_l z}{\Delta l}.$$

Возьмем произвольную перемещающуюся вдоль оси l точку $M(x, y)$, отстоящую от стационарной точки $M_0(x_0, y_0)$ на расстояние l (рис. 6.38). Координаты точки $M(x, y)$ связаны с координатами точки $M_0(x_0, y_0)$ соотношениями $x = x_0 + l\cos\alpha$, $y = y_0 + l\cos\beta$. Рассмотрим функцию $z = z(x, y)$, как сложную функцию переменной l и найдем ее производную по l:

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}\frac{dx}{dl} + \frac{\partial z}{\partial y}\frac{dy}{dl} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta.$$

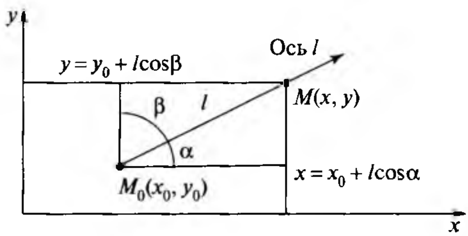

Рис. 6.38

Из рис. 6.38 видно, что $\beta = \frac{\pi}{2} - \alpha$. Производную по направлению в двухмерном случае можно записать иначе:

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\sin\alpha.$$

Для трехмерного случая производная по направлению для функции $u = u(x, y, z)$ имеет вид

$$\frac{\partial u}{\partial l} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma,$$

где α, β, γ — углы между направлением l и координатными осями соответственно Ox, Oy, Oz.

Пример 6.32. Найти производную функции $z = x^2 + y^2$ в точке $M(1, 2)$ в направлении радиуса-вектора этой точки.

Решение. Найдем углы, задающие направление радиус-вектора:

$$\cos\alpha = \frac{x_0}{\sqrt{x_0^2 + y_0^2}} = \frac{1}{\sqrt{5}}, \quad \sin\alpha = \frac{y_0}{\sqrt{x_0^2 + y_0^2}} = \frac{2}{\sqrt{5}}.$$

Производная по направлению равна

$$\frac{\partial z}{\partial l} = \left.\frac{\partial z}{\partial x}\right|_{\substack{x_0=1\\ y_0=2}} \cos\alpha + \left.\frac{\partial z}{\partial y}\right|_{\substack{x_0=1\\ y_0=2}} \sin\alpha = 2\cdot\frac{1}{\sqrt{5}} + 4\cdot\frac{2}{\sqrt{5}} = 2\sqrt{5}.$$

Градиент

Вспомним из школьной программы, что *скалярным произведением* двух векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ называется число, которое определяется по правилу $(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}||\boldsymbol{b}|\cos\varphi$, где $|\boldsymbol{a}|$, $|\boldsymbol{b}|$ — длины векторов; φ — угол между векторами. *Координатами* вектора называются координаты его конечной точки (x, y), если начальная точка вектора совпадает с началом координат.

Пусть $\boldsymbol{a} = (x_1, y_1)$, $\boldsymbol{b} = (x_2, y_2)$. Тогда

1) $|\boldsymbol{a}| = \sqrt{x_1^2 + y_1^2}$, $\quad |\boldsymbol{b}| = \sqrt{x_2^2 + y_2^2}$;
2) $\boldsymbol{a} + \boldsymbol{b} = (x_1 + x_2, y_1 + y_2)$;
3) $(\boldsymbol{a}, \boldsymbol{b}) = x_1x_2 + y_1y_2$.

Определение. *Градиентом функции* $z = z(x, y)$ в точке $M(x, y)$ называется вектор с координатами $\frac{\partial z}{\partial x}$ и $\frac{\partial z}{\partial y}$. Для градиента введено обозначение

$$\operatorname{grad} z = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right).$$

Построим в направлении l единичный вектор $\boldsymbol{a} = (\cos\alpha, \cos\beta)$ (рис. 6.39) и найдем скалярное произведение векторов $\operatorname{grad} z$ и $\boldsymbol{a}$:

$$(\operatorname{grad} z, \boldsymbol{a}) = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta.$$

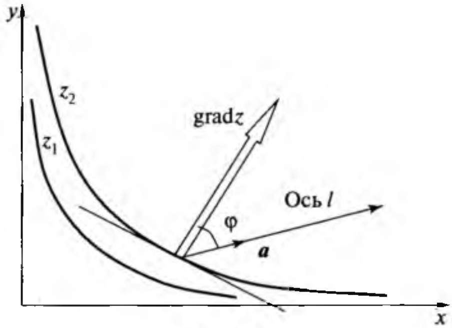

Рис. 6.39

Следовательно, скалярное произведение градиента grad$z$ и единичного вектора $\boldsymbol{a}$, задающего направление l, есть производная по этому направлению:

$$\frac{\partial z}{\partial l} = (\operatorname{grad} z, \boldsymbol{a}) = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta.$$

Введем угол φ между градиентом grad$z$ и вектором $\boldsymbol{a}$ и перепишем скалярное произведение в другом виде:

$$\frac{\partial z}{\partial l} = (\operatorname{grad} z, \boldsymbol{a}) = |\operatorname{grad} z| \cdot |\boldsymbol{a}| \cdot \cos\varphi = |\operatorname{grad} z| \cdot \cos\varphi.$$

Исследование зависимости $\frac{\partial z}{\partial l}$ от величины угла φ показывает, что при $\varphi = 0$ производная по направлению достигает своего максимального значения $\left.\frac{\partial z}{\partial l}\right|_{\max} = |\operatorname{grad} z|$ (см. рис. 6.39).

Таким образом, градиент характеризует направление и величину максимальной скорости изменения функции в точке. Направление вектора grad$z$ указывает направление возрастания функции. Для функции $z = z(x, y)$ на рис. 6.39 изображены две линии уровня $z(x, y)$ и $z(x, y) = z_2$, причем $z_2 > z_1$.

Если $\varphi = \pm\frac{\pi}{2}$, то $\frac{\partial z}{\partial l} = 0$, т.е. в этом направлении функция не изменяется. Это направление линии уровня функции.

Например, у функции $z = x^2 + y^2$ линии уровня представляют концентрические окружности $x^2 + y^2 = C$ радиусами $\sqrt{C}$. В каждой точке окружности градиент направлен перпендикулярно касательной к окружности. Поле градиентов функции $z = x^2 + y^2$ изображено на рис. 6.40.

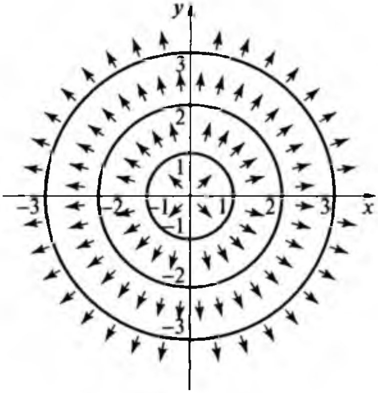

Рис. 6.40

Свойства градиента

Сведем в таблицу основные свойства градиента.

1. Градиент перпендикулярен к линии уровня.
2. Градиент направлен в сторону возрастания функции.
3. Длина градиента равна максимальной величине производной по направлению в данной точке. Другими словами, производная по направлению принимает максимальное значение в том направлении, куда «смотрит» градиент, причем

$$\left.\frac{\partial z}{\partial l}\right|_{\max} = |\operatorname{grad} z| = \sqrt{z_x'^2 + z_y'^2}.$$

Производные и дифференциалы высших порядков

Производные высших порядков

Пусть частная производная $\dfrac{\partial z(x, y)}{\partial x}$ функции $z = z(x, y)$ существует в некоторой области. Тогда в этой области можно определить частную производную от функции $\dfrac{\partial z(x, y)}{\partial x}$ по аргументам x или y:

$$\frac{\partial\left(\dfrac{\partial z}{\partial x}\right)}{\partial x} = \lim_{\Delta x \to 0} \frac{\dfrac{\partial z(x + \Delta x, y)}{\partial x} - \dfrac{\partial z(x, y)}{\partial x}}{\Delta x};$$

$$\frac{\partial\left(\dfrac{\partial z}{\partial x}\right)}{\partial y} = \lim_{\Delta y \to 0} \frac{\dfrac{\partial z(x, y + \Delta y)}{\partial x} - \dfrac{\partial z(x, y)}{\partial x}}{\Delta y}.$$

Эти производные называются ***вторыми частными производными*** или ***частными производными второго порядка*** функции $z = z(x, y)$ и обозначаются

$$\frac{\partial\left(\frac{\partial z}{\partial x}\right)}{\partial x} \equiv \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) \equiv \frac{\partial^2 z}{\partial x^2} \equiv z''_{xx} \quad \text{и} \quad \frac{\partial\left(\frac{\partial z}{\partial x}\right)}{\partial y} \equiv \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) \equiv \frac{\partial^2 z}{\partial y \partial x} \equiv z''_{xy}.$$

Так же вводятся вторые частные производные при дифференцировании функции $\frac{\partial z(x, y)}{\partial y}$:

$$\frac{\partial\left(\frac{\partial z}{\partial y}\right)}{\partial y} \equiv \frac{\partial^2 z}{\partial y^2} \equiv z''_{yy} \quad \text{и} \quad \frac{\partial\left(\frac{\partial z}{\partial y}\right)}{\partial x} \equiv \frac{\partial^2 z}{\partial x \partial y} \equiv z''_{yx}.$$

Среди четырех полученных частных производных z''_{xx}, z''_{xy}, z''_{yy}, z''_{yx} две из них: z''_{yx} и z''_{xy} — называются смешанными, поскольку получаются последовательным дифференцированием по обеим переменным.

Подобным образом определяются частные производные более высоких порядков.

Теорема 6.17 (о порядке дифференцирования непрерывных смешанных производных). Если функция $z = z(x, y)$ имеет непрерывные смешанные производные n-го порядка, значение любой смешанной производной не зависит от того порядка, в котором производятся последовательные дифференцирования.

Дифференциалы высших порядков

Пусть функция $z = z(x, y)$ имеет в области определения непрерывные частные производные до второго порядка включительно. Определим d^2z как дифференциал от дифференциала $d(dz)$. В этом случае получится тот же результат. Действительно,

$$d^2z = d(dz) = d\left(\frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy\right) = d\left(\frac{\partial z}{\partial x}\right)dx + d\left(\frac{\partial z}{\partial y}\right)dy =$$

$$= \frac{\partial^2 z}{\partial x^2}dx^2 + \frac{\partial^2 z}{\partial x \partial y}dxdy + \frac{\partial^2 z}{\partial y \partial x}dxdy + \frac{\partial^2 z}{\partial y^2}dy^2,$$

откуда и получается формула.

Введем формальный символ (оператор) $\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy$, условное умножение которого на функцию z приводит к выражению

$\frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$, а квадрат символа (и не только квадрат) раскрывается по формулам сокращенного умножения:

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 = \frac{\partial^2}{\partial x^2}dx^2 + 2\frac{\partial^2}{\partial x \partial y}dxdy + \frac{\partial^2}{\partial y^2}dy^2.$$

Тогда

$$dz(x, y) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)z;$$

$$d^2z(x, y) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 z;$$

$$d^2u(x, y, z) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)^2 u.$$

В общем случае для функции $u = u(x_1, x_2, ..., x_k)$, имеющей непрерывные частные производные n-го порядка:

$$d^n u(x_1, x_2, ..., x_k) = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + ... + \frac{\partial}{\partial x_k}dx_k\right)^n u.$$

Пример 6.33. Найти третий дифференциал от функции двух переменных.

Решение. Пусть функция $z = z(x, y)$ имеет непрерывные третьи частные производные. Для раскрытия скобок в выражении для третьего дифференциала

$$d^3z(x, y) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^3 z$$

воспользуемся алгебраической формулой сокращенного умножения:

$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.

Получим

$$d^3z = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^3 z =$$

$$= \frac{\partial^3 z}{\partial x^3}dx^3 + 3\frac{\partial^3 z}{\partial x^2 \partial y}dx^2dy + 3\frac{\partial^3 z}{\partial x \partial y^2}dxdy^2 + \frac{\partial^3 z}{\partial y^3}dy^3.$$

Глава 7
ОСНОВНЫЕ ПОНЯТИЯ ИНТЕГРАЛЬНОГО ИСЧИСЛЕНИЯ

7.1. НЕОПРЕДЕЛЕННЫЕ ИНТЕГРАЛЫ

Понятие первообразной

В интегральном исчислении основной задачей является нахождение функции $y=f(x)$ по ее известной производной $y'=f'(x)$.

Функция $F(x)$ называется *первообразной* для функции $y=f(x)$ на промежутке X, конечном или бесконечном, если функция $F(x)$ дифференцируема в каждой точке этого промежутка и ее производная $F'(x)=f(x)$.

Равенство $F'(x)=f(x)$ запишем через дифференциалы:

$$\frac{dF}{dx} = f(x) \text{ или } dF=f(x)\cdot dx.$$

Пример 7.1. Функция $F(x) = \dfrac{x^2}{2}$ является первообразной для функции $f(x)=x$, так как $F'(x) = \left(\dfrac{x^2}{2}\right)' = x.$

Замечание. Первообразная $F(x)$ имеет конечную производную. Следовательно, она является непрерывной функцией. Это следует учитывать при нахождении неопределенных интегралов в случаях, когда область интегрирования переменной x разделяется на несколько промежутков.

Теорема 7.1. Если $F(x)$ является первообразной для функции $f(x)$ на промежутке X, то и функция $\Phi(x) = F(x) + C$, где C — произвольная постоянная, будет первообразной для $f(x)$ на этом промежутке.

В самом деле, $\Phi(x) = (F(x) + C)' = F'(x) = f(x)$.

Таким образом, если функция $y=f(x)$ имеет первообразную, то она имеет бесконечное множество первообразных. Между двумя различными первообразными для одной и той же функции существует тесная связь, которая устанавливается следующей теоремой.

Теорема 7.2. Если $F(x)$ и $\Phi(x)$ — две любые первообразные для функции $y=f(x)$, то их разность равна некоторой постоянной:

$\Phi(x) - F(x) = C, \qquad C = \text{const}.$

Эту теорему можно было сформулировать так: каждая функция, первообразная для $f(x)$, может быть представлена в виде $F(x) + C$.

Совокупность всех первообразных функции $f(x)$, определенных на промежутке, называется *неопределенным интегралом* от функции $f(x)$ и обозначается символом

$$\int f(x)dx.$$

Таким образом, $\int f(x)dx = F(x) + C$.

Здесь знак $\int$ называется знаком интеграла, выражение $f(x)dx$ — подынтегральным выражением, сама функция $f(x)$ — подынтегральной функцией, а x называется переменной интегрирования.

Операцию нахождения первообразной или неопределенного интеграла от функции $f(x)$ называют интегрированием функции $f(x)$. Интегрирование представляет собой операцию, обратную дифференцированию.

Свойства неопределенного интеграла

1. Дифференциал от неопределенного интеграла равен подынтегральному выражению:

$$d\left(\int f(x)dx\right) = d(F(x) + C) = dF(x) = F'(x)dx = f(x)dx.$$

2. Производная от неопределенного интеграла равна подынтегральной функции:

$$\left(\int f(x)dx\right)' = (F(x) + C)' = F'(x) = f(x).$$

3. Неопределенный интеграл от дифференциала некоторой функции равен этой функции плюс произвольная постоянная:

$$\int dF(x) = \int F'(x)dx = \int f(x)dx = F(x) + C.$$

4. Постоянный множитель можно выносить за знак неопределенного интеграла или вносить под знак интеграла:

$$\int \alpha \cdot f(x)dx = \alpha \int f(x)dx.$$

По второму свойству левая часть равенства $\left(\int \alpha \cdot f(x)dx\right)' = \alpha$ $= \alpha \cdot f(x)$. По свойству производной правая часть равенства

$$\left(\alpha \int f(x)dx\right)' = \alpha\left(\int f(x)dx\right)' = \alpha \cdot f(x).$$

Таким образом, левая и правая части равенства в четвертом свойстве равны.

5. Неопределенный интеграл от алгебраической суммы двух функций равен алгебраической сумме неопределенных интегралов от этих функций:

$$\int(f(x) \pm g(x))dx = \int f(x)dx \pm \int g(x)dx.$$

По свойству 2

$$\left(\int(f(x) \pm g(x))dx\right)' = f(x) \pm g(x).$$

По свойству производной

$$\left(\int f(x)dx \pm \int g(x)dx\right)' = \left(\int f(x)dx\right)' \pm \left(\int g(x)dx\right)' = f(x) \pm g(x).$$

Таким образом, $\int(f(x) \pm g(x))dx$ и $\int f(x)dx \pm \int g(x)dx$ являются первообразными для одних и тех же функций $f(x) \pm g(x)$. Следовательно, они отличаются друг от друга не более чем на некоторую постоянную величину.

Замечание. Равенства, описывающие свойства 4 и 5, справедливы и в том случае, если к любой части уравнения добавить константу C. Это означает, что знак равенства при нахождении интегралов носит условный характер: левая часть равенства равна правой части с точностью до произвольной постоянной величины.

Табличные интегралы

Рассмотрим некоторые элементарные функции $F(x)$ и их производные $f(x)$. Формулу, их связывающую: $F'(x) = f(x)$ — запишем в виде $dF(x) = f(x)dx$. По известному дифференциалу $dF(x)$ функции $F(x)$ восстановим саму функцию, пользуясь свойством 3:

$$\int dF(x) = F(x) + C = \int f(x)dx,$$

где C — const. Таким путем получаются основные интегралы, представленные ниже в виде таблицы.

1. $\int 0 \cdot dx = C.$
2. $\int 1 \cdot dx = x + C.$
3. $\int x^\alpha dx = \dfrac{x^{\alpha+1}}{\alpha+1} + C$, где $\alpha \neq -1,\ x > 0.$
4. $\int \dfrac{dx}{x} = \ln|x| + C,\ x \neq 0.$

5. $\int a^x dx = \frac{a^x}{\ln a} + C.$

6. $\int e^x dx = e^x + C.$

7. $\int \sin x dx = -\cos x + C.$

8. $\int \cos x dx = \sin x + C.$

9. $\int \frac{dx}{\sin^2 x} = -\mathrm{ctg}\, x + C.$

10. $\int \frac{dx}{\cos^2 x} = \mathrm{tg}\, x + C.$

11. $\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin\frac{x}{a} + C, \quad |x| < |a|.$

12. $\int \frac{dx}{a^2 + x^2} = \frac{1}{a}\mathrm{arctg}\frac{x}{a} + C.$

13. $\int \frac{dx}{a^2 - x^2} = \frac{1}{2a}\ln\left|\frac{a + x}{a - x}\right| + C, \quad |x| \neq a$ «высокий» логарифм

14. $\int \frac{dx}{\sqrt{x^2 + a}} = \ln\left|x + \sqrt{x^2 + a}\right| + C$ «длинный» логарифм

Замечание 1. Не все интегралы можно свести к табличным. Если операция дифференцирования элементарных функций всегда приводит к элементарным функциям, то операция интегрирования непрерывных функций может привести к таким функциям, которые не выражаются через элементарные. Например, интеграл Пуассона $\int e^{-x^2} dx$, играющий большую роль в теории вероятностей, существует в силу непрерывности подынтегральной функции, однако он не выражается через известные элементарные функции.

Замечание 2. Операция сведения интегралов к табличным может быть достаточно сложной как с идейной, так и с технической точки зрения. Существует огромное количество различных методов, позволяющих это сделать. Ниже мы рассмотрим некоторые из них.

Методы нахождения неопределенных интегралов

1. С помощью тождественных преобразований подынтегральной функции интеграл сводится к интегралу, к которому применимы основные правила интегрирования и возможно использование таблицы основных интегралов.

Пример 7.2. Найти $\int(ax^2+bx+c)dx$.

Решение.

$$\int(ax^2+bx+c)dx=\int ax^2dx+\int bxdx+\int cdx=\frac{ax^3}{3}+\frac{bx^2}{2}+cx+C.$$

Пример 7.3. Найти $\int 2^{3x-1}dx$.

Решение. $\int 2^{3x-1}dx=\int\frac{8^x}{2}dx=\frac{2^{3x}}{6\ln 2}+C.$

2. Подведение под знак дифференциала.

В формуле неопределенного интеграла величина dx означает, что берется дифференциал от переменной x. Можно использовать некоторые свойства дифференциала, чтобы, усложнив выражение под знаком дифференциала, тем самым упростить нахождение самого интеграла. Для этого воспользуемся формулой $y'(x)dx=dy(x)$. Если нужная функция $y(x)$ отсутствует, иногда ее можно образовать путем алгебраических преобразований.

Пример 7.4. Доказать, что $2dx=d(2x+3)$.

Решение. Если нужно образовать под знаком дифференциала $2x+3$, то делим и умножаем dx на 2:

$$dx=\frac{1}{2}\cdot 2\cdot dx=\frac{1}{2}\cdot d(2x)=\frac{1}{2}\cdot d(2x+3).$$

Пример 7.5. Внести переменную x под знак дифференциала.

Решение. Если перед дифференциалом dx стоит переменная x, то легко получить под знаком дифференциала функцию x^2:

$$x\cdot dx=d\left(\frac{x^2}{2}\right)=\frac{1}{2}\cdot d(x^2).$$

Пример 7.6. Внести функцию $\cos 2x$ под знак дифференциала.

Решение. $\cos 2x\cdot dx=\cos 2x\cdot\frac{1}{2}\cdot d(2x)=\frac{1}{2}\cdot d(\sin 2x).$

Пример 7.7. Найти $\int\frac{dx}{3-5x}$.

Решение. $\int\frac{dx}{3-5x}=\int\frac{\frac{1}{-5}d(-5x)}{3-5x}=-\frac{1}{5}\int\frac{d(3-5x)}{3-5x}=-\frac{1}{5}\ln|3-5x|+C.$

Пример 7.8. Найти $\int xe^{x^2}dx$.

Решение. $\int xe^{x^2}dx=\int e^{x^2}d\left(\frac{x^2}{2}\right)=\frac{1}{2}\int e^{x^2}d(x^2)=\frac{1}{2}e^{x^2}+C.$

3. Интегрирование заменой переменной, или подстановка.

Пусть $x = \varphi(t)$, где функция $\varphi(t)$ имеет непрерывную производную $\varphi'(t)$, а между переменными x и t существует взаимно-однозначное соответствие. Тогда справедливо равенство

$$\int f(x)dx = \int f(\varphi(t))\varphi'(t)dt.$$

Пример 7.9. Найти $\int \frac{dx}{1-2x}$.

Решение. Положим $t = 1 - 2x$, откуда $x = 0,5 - 0,5t$, т.е. вводится функция $x = \varphi(t)$, имеющая непрерывную производную и однозначно связывающая переменные x и t. Находим $dx = -0,5dt$.

Подставляем это в интеграл:

$$\int \frac{dx}{1-2x} = \int \frac{-0,5dt}{t} = -0,5\int \frac{dt}{t} = -0,5\ln|1-2x| + C.$$

4. Интегрирование по частям

Интегрированием по частям называется интегрирование по формуле

$$\int udv = uv - \int vdu.$$

Пусть функции $u(x)$ и $v(x)$ имеют непрерывные производные $u'(x)$ и $v'(x)$. Тогда по правилу дифференцирования произведения будем иметь

$$(u(x)v(x))' = u(x)v'(x) + u'(x)v(x).$$

Отсюда, используя $v'(x)dx = dv$ и $u'(x)dx = du$, получим

$$\int (u(x)v(x))'dx = uv;$$

$$\int (u(x)v'(x) + u'(x)v(x))dx = \int u(x)v'(x)dx + \int u'(x)v(x)dx =$$
$$= \int udv + \int vdu.$$

Таким образом, $\int udv + \int vdu = uv$.

Окончательно $\int udv = uv - \int vdu$.

З а м е ч а н и е 1. Использование формулы интегрирования по частям целесообразно в тех случаях, когда дифференцирование упрощает один из сомножителей, в то время как интегрирование не усложняет другой.

З а м е ч а н и е 2. При интегрировании по частям удобно использовать определенную форму записи, позволяющую избежать ошибок подстановки типа «берем не то и подставляем не туда».

Пример 7.10. Найти интеграл $\int x \cdot \cos x dx$.

Решение. Здесь следует ввести $u(x) = x$ и $dv = \cos x dx = d(\sin x)$. Тогда

$$\int x \cdot \cos x dx = \begin{Bmatrix} u = x, & du = dx, \\ dv = \cos x dx, & v = \sin x \end{Bmatrix} = x\sin x - \int \sin x dx =$$

$$= x\sin x + \cos x + C.$$

Выше в фигурных скобках стоят 4 равенства. Далее рассматриваем правые части этих равенств (члены равенств). Берем произведение левого верхнего члена равенства на правый нижний ($uv = x\sin x$) минус интеграл от произведения правого нижнего члена равенства на правый верхний ($\int vdu = \int \sin x dx$).

Замечание 3. К нахождению интеграла $\int vdu$ в правой части формулы можно снова применять интегрирование по частям.

Пример 7.11. Найти $\int x^2 e^x dx$.

Решение. $\int x^2 e^x dx = \begin{Bmatrix} u = x^2, & du = 2xdx, \\ dv = e^x dx, & v = e^x \end{Bmatrix} = x^2 e^x - 2\int xe^x dx.$

Последний интеграл возьмем тем же методом:

$$\int xe^x dx = \begin{Bmatrix} u = x, & du = dx, \\ dv = e^x dx, & v = e^x \end{Bmatrix} = xe^x - \int e^x dx = xe^x - e^x.$$

Окончательно получим

$$\int x^2 e^x dx = x^2 e^x - 2\int xe^x dx = x^2 e^x - 2xe^x + 2e^x + C.$$

5. Интегрирование рациональных дробей.

Рациональная дробь $\dfrac{P_m(x)}{Q_n(x)}$ называется *правильной*, если степень многочлена числителя меньше степени многочлена знаменателя, т.е. $m < n$. Если $m \geq n$, дробь называется неправильной. Если дробь неправильная, то, разделив числитель на знаменатель по правилу деления многочленов, можно получить многочлен плюс правильную дробь.

Пример 7.12. Найти $\int \dfrac{x^2 - 2x + 2}{x - 1} dx$.

Решение.

$$\int \frac{x^2 - 2x + 2}{x - 1} dx = \int \left(x - 1 + \frac{1}{x - 1} \right) dx = 0{,}5x^2 - x + \ln|x - 1| + C.$$

Если дробь правильная, рассмотрим следующие случаи:

а) в знаменателе подынтегральной функции стоит линейный двучлен $f(x) = \frac{1}{ax+b}$:

$$\int \frac{dx}{ax+b} = \frac{1}{a}\int \frac{d\left(x+\frac{b}{a}\right)}{x+\frac{b}{a}} = \frac{1}{a}\ln\left|x+\frac{b}{a}\right| + C;$$

б) в знаменателе интегрируемой функции — квадратный трехчлен $f(x) = \frac{1}{ax^2+bx+c}$.

Преобразуем его, выделив полный квадрат:

$$\int \frac{dx}{ax^2+bx+c} = \frac{1}{a}\int \frac{dx}{x^2+2\cdot\frac{b}{2a}\cdot x+\frac{b^2}{4a^2}-\frac{b^2}{4a^2}+\frac{c}{a}} =$$

$$= \frac{1}{a}\int \frac{d\left(x+\frac{b}{2a}\right)}{\left(x+\frac{b}{2a}\right)^2+\frac{4ac-b^2}{4a^2}}.$$

Мы получили табличный интеграл, величина которого равна арктангенсу, если $\frac{4ac-b^2}{4a^2} > 0$, и «высокому» логарифму, если $\frac{4ac-b^2}{4a^2} < 0$.

Другой подход к решению заключается во введении новой переменной, за которую принимается производная квадратного трехчлена $t = 2ax + b$. Это приведет к исчезновению в знаменателе линейного слагаемого, в результате чего получится табличный интеграл. Найдем $x = \frac{t-b}{2a}$ и $dx = \frac{dt}{2a}$. Сделаем подстановку в исходный интеграл:

$$\int \frac{dx}{ax^2+bx+c} = \int \frac{\frac{dt}{2a}}{a\left(\frac{t-b}{2a}\right)^2+b\left(\frac{t-b}{2a}\right)+c}.$$

После раскрытия скобок и приведения подобных получим интеграл

$$\int \frac{2dt}{t^2+(4ac-b^2)},$$

который в зависимости от знака числа $4ac - b^2$ даст либо арктангенс, либо «высокий» логарифм;

в) подынтегральная функция имеет вид $f(x) = \dfrac{mx+n}{ax^2+bx+c}$. Аналогично вводится подстановка $t = 2ax + b$, в результате интеграл приводится к сумме двух интегралов:

$$\int \frac{mx+n}{ax^2+bx+c}dx = \alpha\int\frac{tdt}{t^2+\gamma} + \beta\int\frac{dt}{t^2+\gamma},$$

где α, β, γ — некие коэффициенты.

Первый из этих интегралов приводит к логарифму:

$$\alpha\int\frac{tdt}{t^2+\gamma} = \frac{\alpha}{2}\int\frac{d(t^2+\gamma)}{t^2+\gamma} = \frac{\alpha}{2}\ln\left|t^2+\gamma\right|,$$

а значение второго интеграла есть либо арктангенс, либо «высокий» логарифм.

7.2. ОПРЕДЕЛЕННЫЕ ИНТЕГРАЛЫ

Площадь криволинейной трапеции

К понятию определенного интеграла приводит задача отыскания площади криволинейной трапеции. Фигуру, ограниченную графиком положительно определенной функции $y = f(x)$, вертикальными прямыми $x = a$, $x = b$ и осью OX, назовем криволинейной трапецией (рис. 7.1).

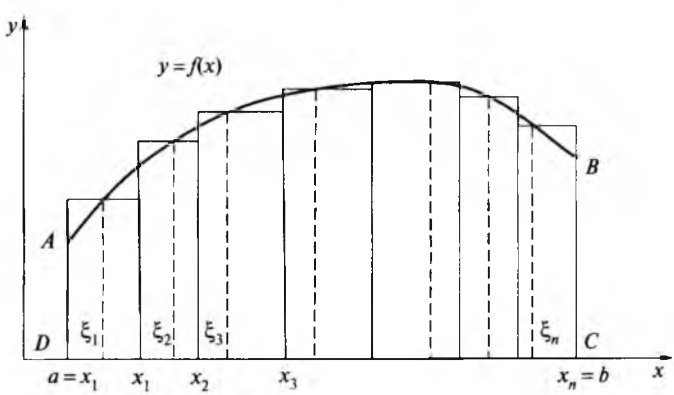

Рис. 7.1

Для нахождения площади плоской фигуры $ABCD$ разобьем отрезок $[a; b]$ на n частей точками:

$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b.$

На каждом частичном отрезке $[x_{k-1}; x_k]$ возьмем по одной произвольной точке ξ_k и построим прямоугольник с основанием $\Delta x_k = x_k - x_{k-1}$ и высотой, равной $f(\xi_k)$. Площадь этого прямоугольника будет равна $\Delta S = f(\xi_k)\Delta x_k$. Таких прямоугольников, покрывающих площадь криволинейной трапеции, будет n штук. В результате такого построения получим «ступенчатую» фигуру, площадь которой будет равна

$$S_n = f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \ldots + f(\xi_n)\Delta x_n = \sum_{k=1}^{n} f(\xi_k)\Delta x_k.$$

Величина $S_n = \sum\limits_{k=1}^{n} f(\xi_k)\Delta x_k$ называется *интегральной суммой.*

Будем теперь увеличивать число n делений отрезка $[a; b]$. Тогда «ступенчатая» фигура будет все меньше отклоняться от криволинейной трапеции $ABCD$. Введем $\max\limits_{1\le k\le n} \Delta x_k$ — длину наибольшего из рассматриваемых частичных отрезков. При $\max\limits_{1\le k\le n} \Delta x_k \to 0$ число частичных отрезков будет неограниченно увеличиваться, а их длины будут стремиться к нулю. Пусть предел интегральной суммы при $\max\limits_{1\le k\le n} \Delta x_k \to 0$ существует, конечен и не зависит ни от способа разбиения отрезка $[a; b]$, ни от выбора точек ξ_k, тогда он принимается за площадь криволинейной трапеции $ABCD$ и называется *определенным интегралом*, т.е.

$$S = \lim_{\substack{\max \Delta x_k \to 0 \\ 1\le k\le n}} S_n = \lim_{\substack{\max \Delta x_k \to 0 \\ 1\le k\le n}} \sum_{k=1}^{n} f(\xi_k)\Delta x_k = \int_a^b f(x)dx.$$

Числа a и b называются соответственно ***нижним*** и ***верхним пределами интеграла***; x называется переменной интегрирования, $f(x)$ — *подынтегральной* функцией, $f(x)dx$ — *подынтегральным* выражением.

Замечание 1. Возникает закономерный вопрос. Почему предел интегральной суммы записывается как $\lim\limits_{\substack{\max \Delta x_k \to 0 \\ 1\le k\le n}} S_n$, а не в привычном нам виде $\lim\limits_{n\to+\infty} S_n$? Условие $\max\limits_{1\le k\le n} \Delta x_k \to 0$ означает, что **ни один из промежутков** не исключается из рассмотрения. Запись $\lim\limits_{n\to+\infty} S_n$ допускает изъятие одного или нескольких промежутков при неограниченном делении остальных.

З а м е ч а н и е 2. Введенное понятие определенного интеграла дает возможность в некоторых случаях узнать, чему он равен, хотя мы пока не знаем, как его вычислить в общем случае.

Пример 7.13. $\int\limits_0^1 1 \cdot dx = 1$, так как площадь, ограниченная прямой $y = 1$, вертикальными прямыми $x = 0$ и $x = 1$ и осью OX (площадь прямоугольника), равна 1 (рис. 7.2).

Пример 7.14. $\int\limits_0^1 x \cdot dx = \frac{1}{2}$. Интеграл есть площадь под прямой $y = x$ (рис. 7.3).

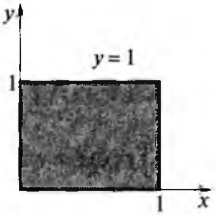

Рис. 7.2

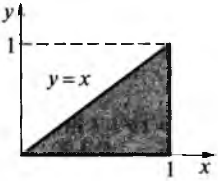

Рис. 7.3

З а м е ч а н и е 3. Определенный и неопределенный интегралы — это разные понятия. Неопределенный интеграл есть семейство функций, а определенный интеграл есть число.

З а м е ч а н и е 4. Непрерывность функции $f(x)$ на отрезке $[a; b]$ является достаточным условием ее интегрирования. Однако требования к функции можно ослабить. Если функция ограничена на $[a; b]$ и имеет конечное число точек разрыва, то она интегрируема. В дальнейшем будем считать подынтегральную функцию непрерывной.

Свойства определенного интеграла

1. Если поменять местами пределы интегрирования, интеграл поменяет знак.

Действительно, если построить интегральную сумму так, что $b < a$, т.е. все $\Delta x_k = x_k - x_{k-1} < 0$, тогда $\int\limits_a^b f(x)dx = -\int\limits_b^a f(x)dx$.

2. Интеграл, пределы интегрирования которого равны, равен нулю.

Если $a = b$, то $\int\limits_a^a f(x)dx = -\int\limits_a^a f(x)dx$, откуда следует $\int\limits_a^a f(x)dx = 0$.

3. Определенный интеграл зависит только от величины нижнего и верхнего пределов интегрирования и от вида подынтегральной функции, он не зависит от переменной интегрирования. Поэтому величина определенного интеграла не изменится, если переменную x заменить любой другой переменной:

$$\int\limits_a^b f(x)dx = \int\limits_a^b f(y)dy.$$

4. Постоянный множитель можно выносить за знак определенного интеграла:

$$\int\limits_a^b \alpha \cdot f(x)dx = \lim_{\substack{\max \Delta x_k \to 0 \\ 1 \le k \le n}} \sum_{k=1}^{n} \alpha \cdot f(\xi_k)\Delta x_k =$$

$$= \alpha \cdot \lim_{\substack{\max \Delta x_k \to 0 \\ 1 \le k \le n}} \sum_{k=1}^{n} f(\xi_k)\Delta x_k = \alpha \int\limits_a^b f(x)dx.$$

5. Определенный интеграл от алгебраической суммы двух функций равен алгебраической сумме интегралов от этих функций:

$$\int\limits_a^b (f(x) \pm g(x))dx = \int\limits_a^b f(x)dx \pm \int\limits_a^b g(x)dx.$$

Для доказательства следует обратиться к пределу интегральной суммы двух функций и воспользоваться тем, что предел суммы (разности) функций равен сумме (разности) пределов этих функций.

6. Для любых чисел a, b, c имеет место равенство

$$\int\limits_a^b f(x)dx = \int\limits_a^c f(x)dx + \int\limits_c^b f(x)dx.$$

Рассмотрим два случая.

а) Пусть $a < c < b$. Отрезок $[a; b]$ разобьем на два отрезка $[a; c]$ и $[c; b]$ — и составим интегральные суммы на каждом отрезке. Пределы этих интегральных сумм и будут определенными интегралами на каждом таком отрезке, а их сумма есть определенный интеграл на отрезке $[a; b]$;

б) $a < b < c$. Из пункта а) следует

$$\int\limits_a^c f(x)dx = \int\limits_a^b f(x)dx + \int\limits_b^c f(x)dx,$$

откуда находим, что

$$\int_a^b f(x)dx = \int_a^c f(x)dx - \int_b^c f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

7. Если $f(x) \le g(x)$, то $\int_a^b f(x)dx \le \int_a^b g(x)dx$,

т.е. неравенство можно интегрировать.

Это следует из того, что справедливо неравенство

$$\lim_{\substack{\max\limits_{1\le k\le n} \Delta x_k \to 0}} \sum_{k=1}^n f(\xi_k)\Delta x_k \le \lim_{\substack{\max\limits_{1\le k\le n} \Delta x_k \to 0}} \sum_{k=1}^n g(\xi_k)\Delta x_k.$$

8. Если функция $y = f(x)$ непрерывна и ограничена на отрезке $[a; b]$, т.е. $m \le f(x) \le M$, то

$$m \le \frac{1}{b-a}\int_a^b f(x)dx \le M.$$

Поскольку $m \le f(x) \le M$, то по свойству 7 получим

$$\int_a^b m\,dx \le \int_a^b f(x)dx \le \int_a^b M\,dx.$$

Левый и правый интегралы легко берутся (см. пример 7.13).

$$m(b-a) \le \int_a^b f(x)dx \le M(b-a),$$

откуда $m \le \dfrac{1}{b-a}\displaystyle\int_a^b f(x)dx \le M$.

В силу непрерывности функция $y = f(x)$ принимает все промежуточные значения, заключенные между m и M. Поэтому по второй теореме Больцано—Коши найдется такое число x_0 ($a \le x_0 \le b$), что

$$f(x_0) = \frac{1}{b-a}\int_a^b f(x)dx.$$

Это значение функции называется *средним значением* на отрезке $[a; b]$. Последнее выражение можно переписать в виде

$$f(x_0)\cdot(b-a) = \int_a^b f(x)dx.$$

Формула Ньютона—Лейбница

На этих страницах не впервые звучат имена Лейбница и Ньютона. Следует указать на выдающуюся роль этих математиков в развитии дифференциального и интегрального исчислений. Готфрид Вильгельм Лейбниц (1646—1716) — немецкий философ и математик, физик и юрист, историк и языковед. Он ввел определения дифференциала и интеграла, ему мы обязаны использованием знаков дифференциала d и интеграла $\int$, использованием терминов «функция», «переменная», «координаты», «абсцисса» и многим другим. Английский физик и математик Исаак Ньютон (1643—1727) — родом из деревни, в юности бедный студент, чудом спасшийся во время эпидемии чумы. С. Вавилов писал, что «на всей физике лежал индивидуальный отпечаток его мысли; без Ньютона наука развивалась бы иначе». Закон тяготения Ньютона, бином Ньютона, формула Ньютона—Лейбница — перечень научных открытий, которыми мы обязаны ему, огромен. Если составить список выдающихся умов, внесших наибольший вклад в науки, в частности в математический анализ, Ньютон и Лейбниц, видимо, разделили бы первое-второе места.

Теорема 6.3. Пусть функция $f(x)$ непрерывна на отрезке $[a; b]$, а функция $F(x)$ есть одна из первообразных на этом отрезке, тогда

$$\int_a^b f(x)dx = F(b) - F(a).$$

Эта формула Ньютона—Лейбница называется также *основной формулой интегрального исчисления*.

Замечание 1. Формула Ньютона—Лейбница сводит вычисление определенного интеграла от функции $f(x)$ к нахождению ее первообразной $F(x)$.

Замечание 2. Первым шагом при вычислении определенного интеграла является нахождение первообразной, вторым — вычисление значения первообразной функции в точках b и a. Поэтому удобно формулу Ньютона—Лейбница записать в таком виде:

$$\int_a^b f(x)dx = F(x)\Big|_a^b = F(b) - F(a).$$

Пример 7.15. Найти $\int\limits_0^1 x\,dx$.

Решение. $\int\limits_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$.

Замечание. Этот интеграл был приведен в качестве примера в начале раздела для иллюстрации того факта, что, не зная способа нахождения интеграла, мы, тем не менее, смогли узнать, чему равна его величина. Теперь у нас есть инструмент для его вычисления. Это формула Ньютона—Лейбница.

Пример 7.16. Найти $\int\limits_0^\pi \sin x\,dx$.

Решение. $\int\limits_0^\pi \sin x\,dx = -\cos x|_0^\pi = -\cos\pi - (-\cos 0) = 2$.

Формула замены переменной в определенном интеграле

Пусть дан интеграл $\int\limits_a^b f(x)dx$, где функция $f(x)$ непрерывна на отрезке $[a; b]$. Введем новую переменную равенством $x = \varphi(t)$. Между переменными x и t существует взаимно-однозначное соответствие; $x = \varphi(t)$ непрерывна на отрезке $[\alpha; \beta]$; $\varphi(\alpha) = a$, $\varphi(\beta) = b$; $\varphi'(t)$ непрерывна на $[\alpha; \beta]$. Тогда $\int\limits_a^b f(x)dx = \int\limits_\alpha^\beta f(\varphi(t))\varphi'(t)dt$.

Пример 7.17. Найти $\int\limits_0^1 \sqrt{1-x^2}dx$.

Решение. Сделаем замену $x = \sin t$. Функция $\sin t$ является непрерывной вместе со своей производной. При $0 \le x \le 1$ справедливо неравенство $0 \le \sin t \le 1$, решение которого $2\pi n \le t \le \pi + 2\pi n$, где $n \in Z$. Из этого бесконечного множества промежутков выберем промежуток $0 \le t \le \frac{\pi}{2}$, в пределах которого каждому значению переменной t соответствует единственное значение переменной x и обратно, причем при изменении переменной x от 0 до 1 переменная t изменяется от 0 до $\frac{\pi}{2}$. Получим

$$\int\limits_0^1 \sqrt{1-x^2}dx = \int\limits_0^{\frac{\pi}{2}} \sqrt{1-\sin^2 t}\cos t\,dt = \int\limits_0^{\frac{\pi}{2}} |\cos t|\cos t\,dt.$$

На промежутке $0 \le t \le \frac{\pi}{2}$ функция $\cos t$ неотрицательна, поэтому модуль раскрывается со знаком плюс. Далее используем формулу $\cos^2 t = \frac{1 + \cos 2t}{2}$.

Тогда $\int\limits_0^{\frac{\pi}{2}} |\cos t| \cos t\, dt = \int\limits_0^{\frac{\pi}{2}} \frac{1 + \cos 2t}{2} dt = \frac{t|_0^{\pi/2}}{2} + \frac{\sin 2t|_0^{\pi/2}}{4} = \frac{\pi}{4}$.

Формула интегрирования по частям

Пусть функции $u = u(x)$ и $v = v(x)$ имеют на отрезке $[a; b]$ непрерывные производные. Тогда имеет место равенство

$$\int\limits_a^b u\,dv = uv|_a^b - \int\limits_a^b v\,du.$$

Пример 7.18. Вычислить $\int\limits_0^{\pi} x \cos x\,dx$.

Решение. $\int\limits_0^{\pi} x \cos x\,dx = \int\limits_0^{\pi} x\,d(\sin x) = x \sin x|_0^{\pi} - \int\limits_0^{\pi} \sin x\,dx = \cos x|_0^{\pi} = -2$.

Приближенное вычисление определенных интегралов

Точное вычисление определенного интеграла может представлять собой трудоемкую задачу, а иногда и невозможно. Поэтому развиты приближенные методы вычисления, позволяющие с заранее заданной точностью найти значение интеграла. Рассмотрим простейший из них — метод трапеций.

Пусть требуется вычислить интеграл $\int\limits_a^b f(x)dx$, где функция $f(x)$ непрерывна на отрезке $[a; b]$. Для упрощения рассуждений будем считать, что $f(x) \ge 0$.

Разобьем отрезок $[a; b]$ на n равных частей точками:

$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$.

Через эти точки проведем вертикальные прямые до пересечения с графиком функции $f(x)$ и соседние точки пересечения соединим между собой (рис. 7.4). Получим n прямолинейных и прямоугольных трапеций. Пусть $f(a) = y_0, f(x_1) = y_1, \ldots, f(x_n) = y_n$ — основания трапеций, их высоты $\frac{b-a}{n}$. Сумма площадей этих трапеций приближенно равна площади криволинейной трапеции $ABCD$.

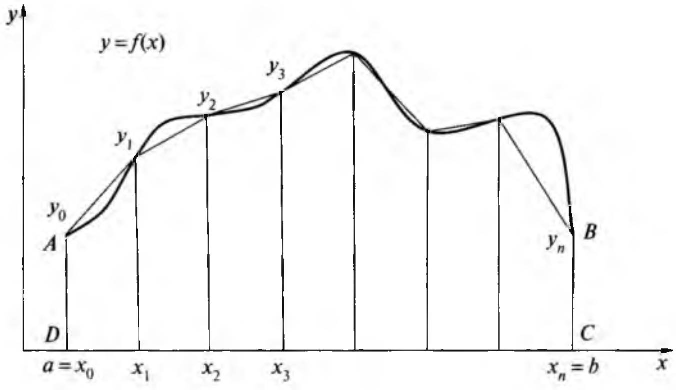

Рис. 7.4

Получим

$$\int_a^b f(x)dx \approx \frac{b-a}{2n}(y_0 + y_1) + \frac{b-a}{2n}(y_1 + y_2) + \ldots + \frac{b-a}{2n}(y_{n-1} + y_n) =$$

$$= \frac{b-a}{n}\left(\frac{y_0 + y_n}{2} + y_1 + y_2 + \ldots + y_{n-1}\right),$$

или $\int_a^b f(x)dx \approx \frac{b-a}{2n}\left(y_0 + y_n + 2\sum_{k=1}^{n-1} y_k\right).$

Замечание. Данная формула получила название формулы трапеций. Ее точность зависит от n, и при возрастании n погрешность формулы убывает.

Пример 7.19. Вычислить приближенно интеграл $\int_0^5 e^{-x^2}dx.$

Решение. Разобьем отрезок [0; 5] на пять равных частей.

Получим $\int_0^5 e^{-x^2}dx \approx \frac{5-0}{5}\left(\frac{e^0 + e^{-25}}{2} + e^{-1} + e^{-4} + e^{-9} + e^{-16}\right) = 0{,}9.$

Погрешность при $n = 5$ довольно велика (~100%).

Оценка определенных интегралов

В некоторых случаях можно отказаться от приближенного вычисления, достаточно сделать оценку интеграла. Она основана на восьмом свойстве определенных интегралов. Если функция $y = f(x)$ ограничена на отрезке $[a; b]$, т.е. $m \le f(x) \le M$, то

$$m(b-a) \le \int_a^b f(x)dx \le M(b-a).$$

Пример 7.20. Оценить интеграл $\int_0^{\pi/2} \sqrt{2-\sin^2 x}dx$.

Решение. Оценим подынтегральное выражение. Поскольку $x \in \left[0; \frac{\pi}{2}\right]$, то

$0 \le \sin^2 x \le 1$,

$1 \le 2 - \sin^2 x \le 2$,

$1 \le \sqrt{2-\sin^2 x} \le \sqrt{2}$.

Теперь мы сможем оценить сам интеграл:

$$1 \cdot \left(\frac{\pi}{2} - 0\right) \le \int_0^{\frac{\pi}{2}} \sqrt{2-\sin^2 x}dx \le \sqrt{2}\left(\frac{\pi}{2} - 0\right).$$

Окончательно $1{,}57 \approx \frac{\pi}{2} \le \int_0^{\frac{\pi}{2}} \sqrt{2-\sin^2 x}dx \le \frac{\pi}{\sqrt{2}} \approx 2{,}23$.

Вычисление площадей плоских фигур

1. Пусть функция $y = f(x)$ непрерывна и неотрицательна на отрезке $[a; b]$. Тогда по геометрическому смыслу определенного интеграла площадь S под кривой (площадь криволинейной трапеции) численно равна определенному интегралу

$$\int_a^b f(x)dx.$$

Пример 7.21. Найти площадь фигуры, ограниченной параболой $y = x^2$, прямыми $x = 0$, $x = 1$ и осью OX.

Решение. $S = \int_0^1 x^2 dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}$.

2. Пусть функция $y = f(x)$ непрерывна и неположительна на отрезке $[a; b]$ (рис. 7.5). Отразим функцию $y = f(x)$ относительно оси OX, получим функцию $y = -f(x)$, расположенную над осью OX. Площадь под кривой $y = -f(x)$ будет равна

$$S = \int_a^b (-f(x))dx = -\int_a^b f(x)dx.$$

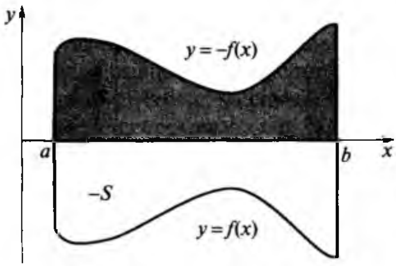

Рис. 7.5

Следовательно, величина найденного интеграла $\int_a^b f(x)dx$ будет отличаться знаком от величины площади S.

Пример 7.22. Найти площадь фигуры, ограниченной параболой $y = -x^2$, прямыми $x = 0$, $x = 1$ и осью Ox.

Решение. $S = -\int_0^1 (-x^2)dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}.$

3. Пусть функции $y = f_1(x)$ и $y = f_2(x)$ непрерывны на отрезке $[a; b]$, причем $f_2(x) \geq f_1(x)$. Тогда площадь фигуры, ограниченная кривыми $y = f_1(x)$ и $y = f_2(x)$, будет равна разности площадей криволинейных трапеций, ограниченных этими линиями:

а) рис. 7.6, *а*: $S = \int_a^b (f_2(x) - f_1(x))dx;$

б) рис. 7.6, *б*: $S = \int_a^b f_2(x)dx - \int_a^b f_1(x)dx = \int_a^b (f_2(x) - f_1(x))dx;$

в) рис. 7.6, *в*: $S = \int_a^b (-f_1(x))dx - \int_a^b (-f_2(x))dx = \int_a^b (f_2(x) - f_1(x))dx;$

г) рис. 7.6, *г*: $S = \int_a^b f_2(x)dx + \int_a^b (-f_1(x))dx = \int_a^b (f_2(x) - f_1(x))dx.$

Площадь на последнем рисунке может быть вычислена как сумма площадей криволинейных трапеций видов, изображенных на рис. 7.6, *а–г*, и, следовательно, по той же формуле.

Пример 7.23. Найти площадь фигуры, ограниченной линиями

$$\begin{cases} y = 2 - x^2, \\ y = -x. \end{cases}$$

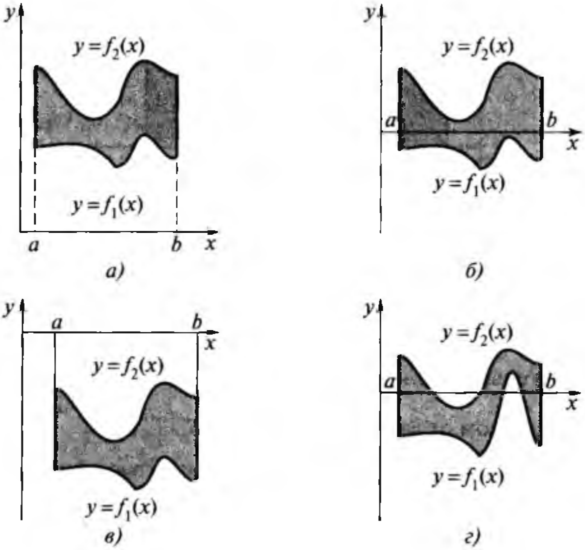

Рис. 7.6

Решение. Найдем точки пересечения графиков заданных функций (рис. 7.7): $2 - x^2 = -x$, откуда $x_1 = -1$, $x_2 = 2$.

Искомая площадь

$$S = \int_{-1}^{2} (2 - x^2 - (-x))dx = \left(2x - \frac{x^3}{3} + \frac{x^2}{2} \right)\Bigg|_{-1}^{2} = 4{,}5.$$

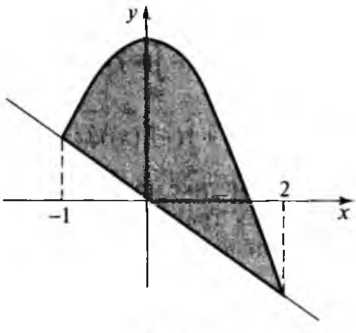

Рис. 7.7

7.3. НЕСОБСТВЕННЫЕ ИНТЕГРАЛЫ

Мы изучали понятие определенного интеграла для случая конечного промежутка и непрерывной ограниченной функции. Обобщим понятие определенного интеграла на случаи бесконечного промежутка и неограниченной на промежутке функции.

Несобственные интегралы 1-го рода

Пусть функция $f(x)$ определена для всех $x \geq a$ и интегрируема на каждом конечном отрезке $[a; t]$.

Несобственным интегралом 1-го рода $\int\limits_a^{+\infty} f(x)dx$ называется $\lim\limits_{t\to+\infty}\int\limits_a^t f(x)dx$. Если этот предел существует и конечен, несобственный интеграл называется *сходящимся*, в противном случае — *расходящимся*. Таким же образом вводятся понятия несобственного интеграла 1-го рода на неограниченных промежутках $[-\infty; a]$ и $[-\infty; +\infty]$.

$$\int\limits_{-\infty}^{a} f(x)dx = \lim_{t\to-\infty}\int\limits_t^a f(x)dx;$$

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \lim_{\substack{t\to-\infty\\ z\to+\infty}}\int\limits_t^z f(x)dx = \lim_{t\to-\infty}\int\limits_t^a f(x)dx + \lim_{z\to+\infty}\int\limits_a^z f(x)dx.$$

Последний интеграл называется сходящимся, если сходятся оба интеграла в правой части равенства независимо от выбора числа a.

Пример 7.24. Вычислить интеграл или установить его расходимость $\int\limits_0^{+\infty}\frac{dx}{1+x^2}$.

Решение.

$$\int\limits_0^{+\infty}\frac{dx}{1+x^2} = \lim_{t\to+\infty}\int\limits_0^t\frac{dx}{1+x^2} = \lim_{t\to+\infty}(\operatorname{arctg} x|_0^t) =$$

$$= \lim_{t\to+\infty}(\operatorname{arctg} t - \operatorname{arctg} 0) = \frac{\pi}{2}.$$

Данный интеграл сходится, его величина равна $\frac{\pi}{2}$.

Эталонный интеграл 1-го рода

Рассмотрим интеграл $\int\limits_1^{+\infty}\frac{dx}{x}$.

По определению $\int\limits_1^{+\infty} \frac{dx}{x} = \lim\limits_{t\to+\infty} \int\limits_1^t \frac{dx}{x} = \lim\limits_{t\to+\infty} (\ln x|_1^t) = \lim\limits_{t\to+\infty} (\ln t) = +\infty$, т.е. интеграл расходится.

Рассмотрим интеграл $\int\limits_1^{+\infty} \frac{dx}{x^p}$, где $p \neq 1$. Выясним условия сходимости этого интеграла.

$$\int\limits_1^{+\infty} \frac{dx}{x^p} = \lim_{t\to+\infty} \int\limits_1^t \frac{dx}{x^p} = \lim_{t\to+\infty} \left(\frac{x^{1-p}}{1-p} \Bigg|_1^t \right) = \begin{cases} +\infty, & p < 1, \\ \frac{1}{p-1}, & p > 1. \end{cases}$$

Обобщая результаты исследования сходимости двух последних интегралов, имеем: несобственный интеграл 1-го рода $\int\limits_1^{+\infty} \frac{dx}{x^p}$ сходится при $p > 1$ и расходится при $p \leq 1$. Он называется *эталонным интегралом 1-го рода*).

Замечание. Нижний предел интегрирования был взят из соображений простоты вычислений. Рассуждения останутся справедливыми, если вместо числа $x = 1$ взять любое число a, удовлетворяющее условию $a > 0$.

Полученные результаты имеют простой геометрический смысл. Рассмотрим область S, ограниченную сверху кривой $y = \frac{1}{x^p}$, снизу — осью OX, слева прямой $x = 1$ (рис. 7.8). Ее площадь оказывается ко-

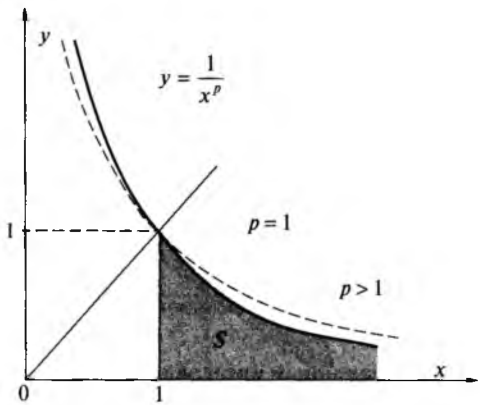

Рис. 7.8

нечной величиной, если $p > 1$, и бесконечно большой величиной, если $p \le 1$.

Несобственные интегралы 2-го рода

Пусть функция $y = f(x)$ неограниченна на конечном промежутке $[a; b)$, причем $\lim\limits_{x \to b-0} f(x) = \infty$.

Несобственным интегралом 2-го рода $\int\limits_a^b f(x)dx$ на промежутке $[a; b)$ называется

$$\lim_{\delta \to +0} \int\limits_a^{b-\delta} f(x)dx.$$

Если предел существует и конечен, несобственный интеграл называется *сходящимся*, в противном случае — *расходящимся*.

Аналогично, если функция $y = f(x)$ неограниченна на конечном промежутке $(a; b]$, причем $\lim\limits_{x \to a+0} f(x) = \infty$, или если функция $y = f(x)$ неограниченна на конечном промежутке $[a; b]$, причем во внутренней точке этого промежутка обращается в бесконечность $\lim\limits_{x \to c} f(x) = \infty$ $a < c < b$, то несобственный интеграл 2-го рода определяется так:

$$\int\limits_a^b f(x)dx = \lim_{\delta \to +0} \int\limits_{a+\delta}^{b} f(x)dx,$$

или $$\int\limits_a^b f(x)dx = \lim_{\delta \to +0} \int\limits_a^{c-\delta} f(x)dx + \lim_{\gamma \to +0} \int\limits_{c+\gamma}^{b} f(x)dx.$$

Последний интеграл называется сходящимся, если сходятся оба интеграла в правой части равенства.

Пример 7.25. Вычислить интеграл или установить его расходимость $\int\limits_0^1 \frac{dx}{\sqrt{1-x^2}}$.

Решение. Подынтегральная функция $f(x) = \frac{1}{\sqrt{1-x^2}}$ интегрируется на конечном промежутке. При $x \to 1-0$ функция $f(x) \to +\infty$.

Следовательно,

$$\int\limits_0^1 \frac{dx}{\sqrt{1-x^2}} = \lim_{\delta \to +0} \int\limits_0^{1-\delta} \frac{dx}{\sqrt{1-x^2}} = \lim_{\delta \to +0} (\arcsin x|_0^{1-\delta}) =$$

$$= \lim_{\delta \to +0} \arcsin(1-\delta) = \frac{\pi}{2}.$$

Эталонный интеграл 2-го рода

Рассмотрим два интеграла: $\int\limits_0^1 \frac{dx}{x}$ и $\int\limits_0^1 \frac{dx}{x^p}$, где $p \neq 1$.

Их величины $\int\limits_0^1 \frac{dx}{x} = \lim\limits_{\delta \to +0} \int\limits_\delta^1 \frac{dx}{x} = \lim\limits_{\delta \to +0} \ln x\Big|_\delta^1 = \lim\limits_{\delta \to +0} (\ln 1 - \ln \delta) = +\infty$,

$$\int\limits_0^1 \frac{dx}{x^p} = \lim_{\delta \to +0} \int\limits_\delta^1 \frac{dx}{x^p} = \lim_{\delta \to +0} \left(\frac{x^{1-p}}{1-p}\Bigg|_\delta^1 \right) = \lim_{\delta \to +0} \left(\frac{1}{1-p}(1 - \delta^{1-p}) \right) =$$

$$= \left[\begin{array}{ll} \dfrac{1}{1-p}, & p < 1, \\ +\infty, & p > 1. \end{array} \right.$$

Суммируя результаты, можем сказать, что несобственный интеграл 2-го рода (он называется *эталонным интегралом 2-го рода*) $\int\limits_0^1 \frac{dx}{x^p}$ сходится при $p < 1$ и расходится при $p \geq 1$.

Замечание. Верхний предел интегрирования был взят из соображений простоты вычислений, как и в случае эталонного интеграла 1-го рода. Рассуждения останутся справедливыми, если вместо числа $x = 1$ взять любое число a, удовлетворяющее условию $a > 0$.

Геометрическая интерпретация интеграла заключается в следующем. Рассмотрим область S, ограниченную сверху кривой $y = \frac{1}{x^p}$, снизу — осью OX, слева и справа — прямыми $x = 0$ и $x = 1$ (рис. 7.9). Ее площадь оказывается конечной величиной, если $p < 1$, и бесконечно большой величиной, если $p \geq 1$ (сравните с рис. 7.9).

Использование интегралов в экономике

Предположим, что проведен статистический опрос населения с целью определения величины доходов на душу населения. По результатам опроса, используя методы математической статистики, выведем формулу, связывающую долю доходов y и долю имеющего эти доходы населения x. При равном распределении доходов зависимость доли населения от доли имеющего их населения должна выражаться прямой $y = x$. Наши исследования дают формулу $y = 2^x - 1$. Оценим степень неравенства в распределении доходов. Она характеризуется коэффициентом Джини, равным отношению площади фигуры OAB к площади треугольника OAC (рис. 7.10).

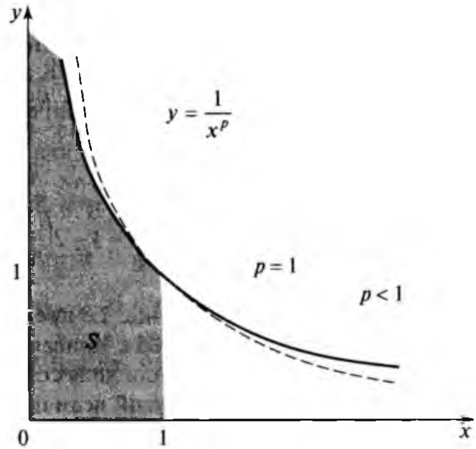

Рис. 7.9

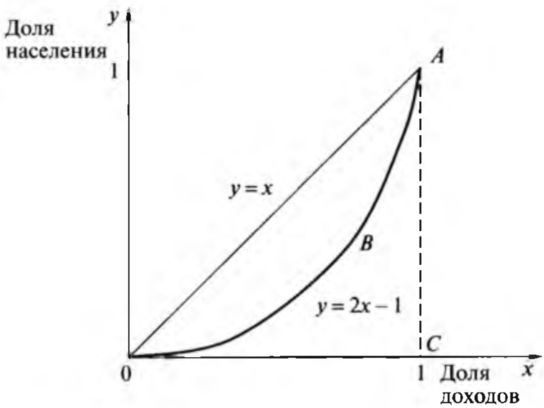

Рис. 7.10

$$k_{\text{Джини}} = \frac{S_{OAB}}{S_{OAC}}.$$

Коэффициент Джини удовлетворяет условию $0 < k_{\text{Джини}} < 1$. Чем он выше, тем больше неравномерность распределения доходов среди населения и, следовательно, тем сильнее концентрация денежных средств в руках немногих.

Вычислим этот коэффициент:

$$k_{\text{Джини}} = \frac{S_{OAB}}{S_{OAC}} = \frac{S_{OAC} - S_{OBAC}}{S_{OAC}} = 1 - \frac{S_{OBAC}}{\frac{1}{2}}.$$

Площадь фигуры S_{OBAC} удобно найти, взяв определенный интеграл:

$$k_{\text{Джини}} = 1 - \frac{\int_0^1 (2^x - 1)dx}{\frac{1}{2}} = 1 - 2\left(\frac{2^x}{\ln 2} - x\right)\Big|_0^1 = 1 - 2\left(\frac{1}{\ln 2} - 1\right) \approx 0{,}11.$$

Линия *OBA* называется кривой Лоренца. Ее предложил нидерландский физик и математик Лоренц в 1905 г. Кривая Лоренца строится как нарастающая доля в процентах или долях одной величины в зависимости от нарастающей доли другой величины, начиная с наименьшей. Например, кривая Лоренца может отражать относительное неравенство размера фирм на рынке.

ЛИТЕРАТУРА

1. *Абчук В.А.* Математика для менеджеров и экономистов. — М.: Изд-во Михайлова, 2002.
2. *Воронцов М.В., Мещеряков Г.П.* Высшая математика для экономистов и менеджеров. — М.: Феникс, 2004.
3. *Ильин В.А., Ким Г.Д.* Линейная алгебра и аналитическая геометрия. — М.: Изд-во Моск. ун-та, 1998.
4. *Краснов М.Л. и др.* Вся высшая математика. Т.1. — М.: Эдиториал: УРСС, 2003.
5. *Краснов М.Л. и др.* Вся высшая математика. Т. 2. — М.: Изд-во ЛКИ, 2007.
6. *Красс М.С., Чупрынов Б.П.* Математика для экономистов. СПб.: Питер, 2004.
7. *Кремер Н.Ш.* Высшая математика для экономистов. — М.: ЮНИТИ-ДАНА, 2006.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Раздел III
ЭЛЕМЕНТЫ ЛИНЕЙНОЙ АЛГЕБРЫ

Глава 8
МАТРИЧНАЯ АЛГЕБРА

8.1. ИДЕИ ЛИНЕЙНОЙ АЛГЕБРЫ

Матричная алгебра. Линейная алгебра формировалась позже математического анализа. Связано это было с задачами, которые решались исследователями. Исторически первым вопросом линейной алгебры был вопрос о решении систем линейных уравнений. Системы из двух-трех уравнений решались с помощью подстановки, но решение системы из десяти—двадцати уравнений требовало новых идей. Ответом на запросы практики явилось развитие теории матриц.

Попытки математически описать окружающий наш мир предпринимались еще Евклидом. Но только с изобретением Декартом метода координат появилась возможность описывать точки на плоскости, а затем и в пространстве совокупностями чисел, названных координатами точек. В соответствии с методом координат на плоскости строятся две взаимно перпендикулярные прямые (оси координат), точка пересечения которых принимается за начало отсчета. Оси масштабируются. Каждой точке на оси ставится в соответствие определенное число. Прямая превращается в числовую ось. Теперь можно определить местоположение любой точки на плоскости. Достаточно из этой точки опустить перпендикуляры на координатные оси и найти числа, соответствующие точкам пересечения перпендикуляров с числовыми осями. На декартовой координатной плоскости какая-нибудь точка, поименованная как точка M, записывается с указанием своих координат, например так: $M(2, 3)$. Возникли новые вопросы. Как определить расстояние между точками на плоскости? Под какими углами те или иные точки видны из начала координат?

Векторная алгебра. На эти и другие вопросы легко получить ответ, если каждой точке поставить в соответствие направленный отрезок, начало которого совпадает с началом координат, а конец совпадает с рассматриваемой точкой. Английский ученый Гамильтон ввел для таких отрезков термин «вектор». Для векторов были разработаны правила сложения и вычитания, умножения вектора на число и на вектор. Развилась новая дисциплина — векторная алгебра. Поскольку вектор иначе можно описать матрицей-строкой или матрицей-столбцом, матричная алгебра и векторная алгебра объединились при описании многих явлений.

Векторные пространства. Если рассматриваемые векторы имеют по две координаты, они описывают множество точек плоскости, если по три — множество точек трехмерного пространства. Приписывая вектору четыре, пять и более координат, исследователи выходили за рамки трехмерных пространств. Так вводились четырех-, пятимерные пространства и даже n-мерные, где n — сколь угодно большое натуральное число. Появилась теория векторных пространств. Из одного векторного пространства векторы научились переводить (отображать) в другое векторное пространство посредством изобретенного математического оператора. Среди векторных пространств были выделены и математически описаны пространства, обладающие свойствами, присущими нашему физическому трехмерному пространству. Они были названы евклидовыми пространствами. Кроме них, были придуманы и другие пространства с иными свойствами. Физически существование их не доказано, хотя ученые подозревают, например, о существовании евклидовых пространств с размерностью, большей трех. Развитие теории векторных пространств представляет не только абстрактный интерес. Теория находит многочисленные приложения в математике, естественных науках, экономике.

Линейный мир. По мере развития наук с используемым математическим аппаратом исследователи приходили к пониманию того, что одной из наиболее общих естественно-научных идей является *идея линейности*. Аппарат линейной алгебры, развитый для описания векторных пространств, оказался востребованным почти во всех науках. Любой непрерывный процесс при малых изменениях линеен. Эта идея, возможно интуитивно, легла в основу математического анализа. Вспомним линейные приращения или касательные к кривым, описывающие эти кривые в окрестности точки касания. Идея линейности положена физиками в основу формулировки фундаментальных законов природы. Вершина экономико-математической мысли — теория оптимизации в экономике базируется на идее линейности.

Для объективности укажем на следующее. Исследования XX столетия показали (теория относительности Эйнштейна), что физическое пространство, на которое исследователи опирались в разработке аппарата линейной алгебры, является линейным лишь в малой окрестности наблюдателя. Не исключено, что понимание линейности окружающего нас мира, линейности процессов в нем и стремление адекватно описать линейный мир были бы затруднены, если бы малая окрестность наблюдателя не была достаточно велика.

8.2. МАТРИЦЫ И ОПЕРАЦИИ НАД НИМИ

Матричная алгебра играет большую роль в экономических расчетах. Многие экономико-математические модели рассматриваются и решаются в матричной форме.

Матрицей с размерами $m \times n$ называется прямоугольная таблица чисел, содержащая m строк и n столбцов. Числа, составляющие матрицу, называются элементами матрицы. Матрицы обычно обозначаются заглавными буквами латинского алфавита, например A, B, C, ..., а для обозначения элементов матрицы используются строчные буквы с двойной индексацией: a_{ij}, где i — номер строки; j — номер столбца. Числа i и j определяют расположение элемента a_{ij} в матрице A и играют роль координат этого элемента в прямоугольной таблице чисел.

Например, матрица

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

имеет m строк и n столбцов.

Набор

$$(a_{i1}, a_{i2}, \dots, a_{in})$$

называется i-й строкой матрицы A, а набор

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix}$$

называется j-м столбцом матрицы A. Любые строки и столбцы матрицы A, в свою очередь, являются матрицами.

Две матрицы A и B одинаковых размеров называются *равными*, если они совпадают поэлементно. Равенство записывается как $A = B$.

Виды матриц

Матрица произвольных размеров, все элементы которой равны нулю, называется *нулевой* и обозначается O.

Матрица, состоящая из одной строки $A = (a_{11}, a_{12}, ..., A_{1n})$, называется матрицей-строкой или вектором.

Матрица, состоящая из одного столбца $A = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{n1} \end{pmatrix}$, называется матрицей-столбцом или также вектором.

Матрица называется *квадратной* n-го порядка, если число ее строк равно числу столбцов и равно n.

Элементы квадратной матрицы a_{ij}, у которых номер строки совпадает с номером столбца, называются *диагональными* и образуют *главную диагональ*.

Если все недиагональные элементы квадратной матрицы равны нулю, то матрица называется *диагональной*. Например,

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

является диагональной матрицей 3-го порядка.

Квадратная матрица, на главной диагонали которой стоят единицы, а все остальные элементы равны нулю, называется *единичной* и обозначается E. Например, матрица

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

является единичной матрицей 4-го порядка.

Квадратная матрица, у которой все элементы выше или ниже главной диагонали равны нулю, называется *треугольной*.

Произвольная матрица вида $C = (A|B)$, составленная из двух матриц, разделенных вертикальной чертой, называется *расширенной*. Например, матрица

$$C = \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right)$$

является расширенной. Она составлена из квадратной матрицы 3-го порядка и единичной матрицы 3-го порядка.

Матрица может содержать своими элементами другие матрицы. Например, матрица

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

может быть записана в виде $A = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \dots \\ \boldsymbol{a}_n \end{pmatrix}$, где $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ — матрицы-строки исходной матрицы.

Операции над матрицами

Над матрицами возможно проведение некоторых арифметических операций.

1. Умножение числа на матрицу производится по следующему правилу: число умножается на каждый элемент матрицы.

Замечание. Общий множитель всех элементов матрицы можно выносить за знак матрицы.

2. Сложение матриц одинаковых размеров — соответствующие элементы матриц складываются.

3. Вычитание матриц одинаковых размеров — соответствующие элементы матриц вычитаются.

4. Умножение матрицы на матрицу — элемент новой матрицы, стоящий на пересечении i-й строки и j-го столбца, равен сумме произведений элементов i-й строки первой матрицы на соответствующие элементы j-го столбца второй матрицы. Операция определена при условии, что число столбцов первой матрицы равно числу строк второй).

Пример 8.1.

$$\begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -2 & 2 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 \cdot 0 + 0 \cdot 1 + (-1) \cdot (-2) & 1 \cdot (-1) + 0 \cdot 0 + (-1) \cdot 2 \\ 2 \cdot 0 + 1 \cdot 1 + 0 \cdot (-2) & 2 \cdot (-1) + 1 \cdot 0 + 0 \cdot 2 \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

З а м е ч а н и е 1. Используя знак сокращенного суммирования, формулу умножения можно записать в виде

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n.$$

З а м е ч а н и е 2. Введем обозначение матрицы в виде $\underset{m \times n}{A}$, означающее, что матрица содержит m строк и n столбцов. Тогда произведение матриц можно записать следующим образом:

$$\underset{m \times k}{A} \cdot \underset{k \times n}{B} = \underset{m \times n}{C}.$$

З а м е ч а н и е 3. Порядок матриц-сомножителей существен. Поэтому говорят об умножении матрицы A на матрицу B справа или слева.

Если произведение матриц $A \cdot B$ существует, то произведение матриц $B \cdot A$ может не существовать.

Если существуют произведения матриц $A \cdot B$ и $B \cdot A$, они могут быть матрицами разных размеров.

Если матрицы A и B квадратные, то их произведения $A \cdot B$ и $B \cdot A$ существуют и имеют одинаковый порядок, но в общем случае $A \cdot B \neq$ $\neq B \cdot A$.

З а м е ч а н и е 4. Умножение единичной матрицы E на квадратную матрицу A не изменяет последней: $E \cdot A = A \cdot E = A$.

З а м е ч а н и е 5. Произведение двух ненулевых матриц может дать нулевую матрицу O, например:

$$A \cdot B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = O.$$

5. Возведение матрицы в целую положительную степень k сводится к произведению k одинаковых матриц:

$$A^k = \underbrace{A \cdot A \cdots A}_{k}.$$

Дополнительно определим $A^0 = E$, $A^1 = A$.

З а м е ч а н и е 1. Возведение в степень матрицы может привести к нулевой матрице. Например:

$$\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}^2 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

З а м е ч а н и е 2. Операция возведения в степень определена только для квадратных матриц.

6. Транспонирование матрицы — переход к матрице, у которой строки и столбцы меняются местами.

Матрица

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

называется *транспонированной* по отношению к матрице

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

и обозначается A^T.

З а м е ч а н и е. Из определения следует, что если матрица A имеет размеры $m \times n$, то транспонированная матрица A^T имеет размеры $n \times m$.

Операции транспонирования, а также операции сложения и умножения матриц обладают легко проверяемыми свойствами.

Свойства транспонирования

1. $(A^T)^T = A$.
2. $(a \cdot A)^T = a \cdot A^T$, где a — число.
3. $(A + B)^T = A^T + B^T$.
4. $(A \cdot B)^T = B^T A^T$.

Свойства операций сложения и умножения

1. $A + B = B + A$.
2. $(A + B) + C = A + (B + C)$.
3. $\alpha(A + B) = \alpha A + \alpha B$.
4. $A(B + C) = AB + AC$.
5. $(A + B)C = AC + BC$.
6. $C(AB) = (CA)B$.
7. $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

8.3. ОПРЕДЕЛИТЕЛИ И ИХ СВОЙСТВА

Свяжем с каждой квадратной матрицей A некоторое число, вводимое по определенному правилу. Назовем это число *определителем матрицы* и обозначим $|A|$.

Определителем матрицы 1-го порядка $A = (a_{11})$ назовем число

$$|A| = a_{11}.$$

Определителем матрицы 2-го порядка $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ назовем число, равное

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}M_{11} - a_{12}M_{12},$$

где M_{1j} (индекс j равен 1 или 2) — определитель матрицы 1-го порядка, полученный вычеркиванием из матрицы A первой строки и j-го столбца.

Например, определитель M_{11} получен из матрицы A вычеркиванием первой строки и первого столбца. Следовательно, величина определителя M_{11} равна a_{22}:

$$M_{11} = a_{22}.$$

Тогда

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}M_{11} - a_{12}M_{12} = a_{11}a_{22} - a_{12}a_{21}.$$

Определителем матрицы 3-го порядка $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ назовем число, равное

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13},$$

где M_{1j} (индекс j равен 1, 2 или 3) — определитель матрицы 2-го порядка, полученный вычеркиванием из матрицы A первой строки и j-го столбца. Например, определитель M_{11} получен из матрицы A вычеркиванием первой строки и первого столбца:

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32},$$

$$M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = a_{21}a_{33} - a_{23}a_{31},$$

$$M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = a_{21}a_{32} - a_{22}a_{31}.$$

Подставим полученные соотношения в выражение для определителя:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) +$$

$$+ a_{13}(a_{21}a_{32} - a_{22}a_{31}) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} -$$

$$- a_{13}a_{22}a_{31} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}.$$

Из структуры формулы видно, что в каждое слагаемое в правой части равенства входит по одному элементу из каждой строки и каждого столбца матрицы. Формулу несложно запомнить, если воспользоваться правилом треугольников (рис. 8.1). Берутся произведения элементов, соединенных линиями. На верхнем рисунке линиями указаны произведения элементов, которые следует взять со знаком «+», на нижнем — со знаком «–».

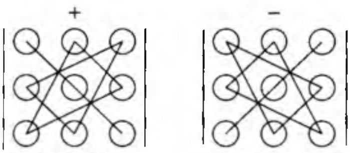

Рис. 8.1

Например, величина определителя матрицы

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ -2 & 1 & 3 \end{pmatrix}$$

равна

$$|A| = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ -2 & 1 & 3 \end{vmatrix} = 1 \cdot 1 \cdot 3 + 0 \cdot 0 \cdot (-2) + 2 \cdot 1 \cdot (-1) - (-2) \cdot 1 \cdot (-1) -$$

$$- 1 \cdot 1 \cdot 0 - 0 \cdot 2 \cdot 3 = -1.$$

Предположим, что определители матриц, порядок которых меньше n, введены. *Определителем* квадратной матрицы n-го порядка

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

назовем число

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{11}M_{11} - a_{12}M_{12} + \dots + (-1)^{n+1}a_{1n}M_{1n},$$

где M_{1j} — определитель матрицы $(n-1)$-го порядка, полученной из матрицы A вычеркиванием первой строки и j-го столбца.

Введем понятия минора и алгебраического дополнения.

Минором M_{ij} элемента a_{ij} матрицы A n-го порядка называется определитель $(n-1)$-го порядка, полученный из матрицы A вычеркиванием i-й строки и j-го столбца. Например, минор M_{23} элемента a_{23} матрицы 3-го порядка получается вычеркиванием из матрицы второй строки и третьего столбца:

$$M_{23} = \begin{vmatrix} a_{11} & a_{12} & \cdots \\ \cdots & \cdots & \cdots \\ a_{31} & a_{32} & \cdots \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = a_{11}a_{32} - a_{12}a_{31}.$$

Алгебраическим дополнением A_{ij} элемента a_{ij} матрицы A n-го порядка называется минор M_{ij}, взятый со знаком $(-1)^{i+j}$:

$$A_{ij} = (-1)^{i+j}M_{ij}.$$

Используя понятие алгебраического дополнения, формулу можно записать в виде

$$|A| = a_{11}M_{11} - a_{12}M_{12} + \dots + (-1)^{n+1}a_{1n}M_{1n} = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n}.$$

Замечание 1. Рассмотренные нами выше определители M_{11}, M_{12}, ... есть миноры соответствующих элементов матрицы.

Замечание 2. Формула допускает сокращенную запись:

$$|A| = \sum_{j=1}^{n}(-1)^{j+1}a_{1j}M_{1j} = \sum_{j=1}^{n}a_{1j}A_{1j}.$$

Иными словами, определитель квадратной матрицы равен сумме произведений элементов первой строки на их алгебраические дополнения.

Замечание 3. Формула называется разложением определителя по первой строке.

Замечание 4. Величина алгебраического дополнения A_{ij} элемента a_{ij} зависит только от положения этого элемента (i, j) в матрице A. При замене элемента a_{ij} матрицы на другое число величина алгебраического дополнения A_{ij} не изменяется.

Теорема 8.1. Определитель квадратной матрицы равен сумме произведений элементов любой строки или любого столбца на их алгебраические дополнения:

$$|A| = \sum_{j=1}^{n}(-1)^{i+j} a_{ij} M_{ij} = \sum_{j=1}^{n} a_{ij} A_{ij}, \text{ где } i = 1, 2, ..., n$$

и

$$|A| = \sum_{i=1}^{n}(-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}, \text{ где } j = 1, 2, ..., n.$$

При доказательстве ограничимся для простоты рассмотрением матрицы 3-го порядка. Мы получили формулу разложения определителя по первой строке. Разложим теперь определитель, например, по второму столбцу:

$$|A|_{2\text{-й ст.}} = -a_{12}M_{12} + a_{22}M_{22} - a_{32}M_{32}.$$

Каждый минор $M_{i2}(i = 1, 2, 3)$ является определителем 2-го порядка:

$$M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = a_{12}a_{33} - a_{31}a_{23},$$

$$M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = a_{11}a_{33} - a_{31}a_{13},$$

$$M_{32} = \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = a_{11}a_{23} - a_{21}a_{13}.$$

Подставим эти выражения в формулу, раскроем скобки и соберем положительные слагаемые, затем отрицательные. Получим

$$|A|_{2\text{-й ст.}} = -a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{22}(a_{11}a_{33} - a_{31}a_{13}) -$$
$$- a_{32}(a_{11}a_{23} - a_{21}a_{13}) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} -$$
$$- a_{13}a_{22}a_{31} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}.$$

Сравнивая правые части , убеждаемся в том, что $|A| = |A|_{2\text{-й ст}}$.

Подобным же образом проверяются и другие равенства, получаемые разложением определителя по определенной строке или столбцу.

Пример 8.2. Вычислить определитель треугольной матрицы n-го порядка

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn} \end{vmatrix}.$$

Решение. Имеем

$$|A| = a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11}a_{22} \begin{vmatrix} a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots \\ 0 & \dots & a_{nn} \end{vmatrix} = \dots = a_{11}a_{22}\dots a_{nn}.$$

Мы убедились, что определитель треугольной матрицы равен произведению ее диагональных элементов.

Свойства определителей

1. Определитель с нулевой строкой или нулевым столбцом равен нулю.

Для доказательства этого свойства достаточно разложить определитель по нулевой строке или нулевому столбцу.

2. Умножение определителя на число равносильно умножению какой-либо строки или столбца определителя на это число.

Умножим любую строку или столбец исходного определителя на число, разложим определитель по этой строке или столбцу, вынесем это число за скобки и свернем оставшееся в скобках выражение в исходный определитель.

3. При транспонировании матрицы величина ее определителя не изменяется: $|A| = |A^T|$.

Разложим определитель $|A|$ по первой строке, транспонируем его. Разложим полученный определитель $|A^T|$ по первому столбцу. Из доказанной выше теоремы следует, что результат будет одинаков.

4. При перестановке двух строк или столбцов определитель меняет знак.

В определителе

$$|A|_1 = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

переставим, например, первую и вторую строки. Получим

$$|A|_2 = \begin{vmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Разложим определитель $|A|_1$ по второй строке, а определитель $|A|_2$ — по первой. Получим

$$|A|_1 = -a_{21}M_{21} + a_{22}M_{22} + \dots + (-1)^{n+2}a_{2n}M_{2n},$$

$$|A|_2 = a_{21}M_{21} - a_{22}M_{22} + \dots + (-1)^{n+1}a_{2n}M_{2n},$$

откуда следует $|A|_1 = -|A|_2$.

Теперь переставим i-ю строку с $(i + k)$-й. Для этого сместим i-ю строку на k строк вниз. Определитель изменит знак k раз. Строка с номером $i + k$ окажется при этом на $(i + k - 1)$-м месте. Переставим эту строку на место i-й строки, для чего поднимем ее на $k - 1$ строк вверх. Определитель изменит знак $k - 1$ раз. В результате процедуры определитель изменит знак нечетное число раз: $k + k - 1 = 2k - 1$, т.е. знак определителя при любой перестановке строк изменится.

5. Определитель матрицы с двумя одинаковыми строками (столбцами) равен нулю.

При перестановке двух строк определитель изменит знак. Переставим местами одинаковые строки. Определитель останется таким же. Значит, $-|A| = |A|$. Отсюда следует, что $|A| = 0$.

6. Определитель, содержащий две пропорциональные строки (столбца), равен нулю.

Вынесем коэффициент пропорциональности за знак определителя. В нем образуются две одинаковые строки. Поэтому такой определитель равен нулю.

7. Определитель можно разложить на сумму определителей.

Представим элементы i-й строки определителя в виде суммы двух слагаемых. Получим

$$|A| = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ \alpha b_1 + \beta c_1 & \dots & \alpha b_n + \beta c_n \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix},$$

где α, β — некоторые коэффициенты, равные в частном случае единице. Разложим определитель $|A|$ по i-й строке, используя алгебраические дополнения, и преобразуем полученную сумму. Тогда

$$|A| = \sum_{j=1}^{n} (\alpha b_j + \beta c_j) A_{ij} = \sum_{j=1}^{n} \alpha b_j A_{ij} + \sum_{j=1}^{n} \beta c_j A_{ij} =$$

$$= \alpha \sum_{j=1}^{n} b_j A_{ij} + \beta \sum_{j=1}^{n} c_j A_{ij} = \alpha \cdot |A|_1 + \beta \cdot |A|_2,$$

где $|A|_1 = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ b_1 & \dots & b_n \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$, $\quad |A|_2 = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ c_1 & \dots & c_n \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$.

8. Определитель матрицы не изменится, если к элементам одной строки прибавить элементы другой строки, умноженные на одно и то же число.

Полученный определитель можно разложить на сумму двух определителей. Один из них является исходным. Другой содержит две пропорциональные строки и, следовательно, равен нулю.

9. Сумма произведений произвольных чисел $b_1, b_2, \dots, b_n$ на алгебраические дополнения элементов любой строки (столбца) равна определителю матрицы, полученной из данной заменой элементов этой строки (столбца) на числа $b_1, b_2, \dots, b_n$.

Свойство вытекает из замечания 4 к алгебраическим дополнениям и доказанной теоремы.

10. Сумма произведений элементов одной строки матрицы на алгебраические дополнения к элементам другой строки этой матрицы равна нулю.

Умножим элементы i-й строки исходной матрицы на алгебраические дополнения к j-й строке и составим сумму:

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn} = \sum_{k=1}^{n} a_{ik} A_{jk}.$$

Подобная сумма получается из матрицы, у которой на месте j-й строки стоит i-я строка:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow i\text{-я строка} \\ \\ \leftarrow j\text{-я строка} \\ \\ \\ \end{array}$$

Эта матрица имеет две одинаковые строки, поэтому величина ее определителя равна нулю.

11. Определитель произведения квадратных матриц равен произведению определителей этих матриц, т.е.

$$|A \cdot B| = |A| \cdot |B|.$$

Вычисление определителя

Существует несколько способов вычисления величины определителя. Выбор способа диктуется видом и порядком определителя. Удачно выбранный способ позволяет существенно сократить вычисления. Рассмотрим их на примере определителя матрицы 3-го порядка.

Пример 8.3. Вычислить определитель матрицы

$$A = \begin{pmatrix} 2 & 2 & -1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{pmatrix}.$$

Первый способ. Использование теоремы о разложении определителя по любой строке или столбцу.

Разложим определитель, например, по третьему столбцу:

$$\begin{vmatrix} 2 & 2 & -1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{vmatrix} = -1(-1)^{1+3}\begin{vmatrix} 2 & -4 \\ -2 & -2 \end{vmatrix} + 1(-1)^{2+3}\begin{vmatrix} 2 & 2 \\ -2 & -2 \end{vmatrix} + 2(-1)^{3+3}\begin{vmatrix} 2 & 2 \\ 2 & -4 \end{vmatrix} =$$

$$= -(-4-8) - (-4+4) + 2(-8-4) = -12.$$

Второй способ. Использование правила треугольников.

$$\begin{vmatrix} 2 & 2 & -1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{vmatrix} = 2 \cdot (-4) \cdot 2 + 2 \cdot 1 \cdot (-2) + 2 \cdot (-2) \cdot (-1) - (-1) \cdot (-4) \cdot (-2) -$$

$$- 2 \cdot 2 \cdot 2 - (-2) \cdot 1 \cdot 2 = -16 - 4 + 4 + 8 - 8 + 4 = -12.$$

Третий способ. Использование свойств определителя для преобразования его к виду, когда определитель содержит строку или столбец с максимальным количеством нулей. Затем разложение определителя по этой строке (столбцу).

$$\begin{vmatrix} 2 & 2 & -1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{vmatrix} = (\text{прибавим к первой строке третью}) = \begin{vmatrix} 0 & 0 & 1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{vmatrix} =$$

$= (\text{разложим определитель по первой строке}) =$

$$= 1 \cdot (-1)^{1+3} \cdot \begin{vmatrix} 2 & -4 \\ -2 & -2 \end{vmatrix} = -4 - 8 = -12.$$

Четвертый способ. Использование свойств определителя для преобразования его к треугольному виду. Тогда величина определителя вычисляется как произведение элементов, стоящих на главной диагонали.

$$\begin{vmatrix} 2 & 2 & -1 \\ 2 & -4 & 1 \\ -2 & -2 & 2 \end{vmatrix} = (\text{первую строку умножим на -1 и сложим со второй строкой, разместив результат на месте второй строки}) = \begin{vmatrix} 2 & 2 & -1 \\ 0 & -6 & 2 \\ -2 & -2 & 2 \end{vmatrix} =$$

$= (\text{первую строку сложим с третьей и разместим результат на месте третьей строки}) = \begin{vmatrix} 2 & 2 & -1 \\ 0 & -6 & 2 \\ 0 & 0 & 1 \end{vmatrix} = (\text{перемножим элементы главной диагонали}) = -12.$

8.4. ОБРАТНАЯ МАТРИЦА

Матрица A называется ***невырожденной***, если ее определитель отличен от нуля, т.е. $|A| \neq 0$. В противном случае она называется ***вырожденной***.

Матрица A^{-1} называется *обратной* по отношению к квадратной матрице A, если выполняется равенство

$$A^{-1} \cdot A = A \cdot A^{-1} = E.$$

Следующая теорема устанавливает условия существования обратной матрицы.

Теорема 8.2. Обратная матрица A^{-1} существует и единственна тогда и только тогда, когда исходная матрица невырожденна.

Необходимость. Пусть матрица A имеет обратную матрицу A^{-1}. Тогда $A^{-1} \cdot A = A \cdot A^{-1} = E$. Используя свойство 11 определителя, получаем $|A| \cdot |A^{-1}| = |E| = 1$, откуда вытекает $|A^{-1}| = \frac{1}{|A|}$. Следовательно, $|A| \neq 0$. Матрица A является невырожденной.

Достаточность. Пусть матрица A является невырожденной: $|A| \neq 0$. Матрицу A транспонируем и на основе транспонированной матрицы A^T построим новую матрицу A^P, элементами которой являются алгебраические дополнения элементов матрицы A^T. Назовем эту матрицу A^P *присоединенной*. Итак,

$$A^P = \begin{pmatrix} A_{11}^T & A_{12}^T & \ldots & A_{1n}^T \\ A_{21}^T & A_{22}^T & \ldots & A_{2n}^T \\ \ldots & \ldots & \ldots & \ldots \\ A_{n1}^T & A_{n2}^T & \ldots & A_{nn}^T \end{pmatrix}.$$

Найдем новую матрицу B как произведение матриц A^P и A: $B = A^P \cdot A$. Она имеет вид

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} A_{11}^T & A_{12}^T & \ldots & A_{1n}^T \\ A_{21}^T & A_{22}^T & \ldots & A_{2n}^T \\ \ldots & \ldots & \ldots & \ldots \\ A_{n1}^T & A_{n2}^T & \ldots & A_{nn}^T \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Элементы матрицы B вычислим по отдельности и воспользуемся равенством $A_{ij}^T = A_{ji}$, которое легко проверяется:

$$b_{11} = A_{11}^T \cdot a_{11} + A_{12}^T \cdot a_{21} + \ldots + A_{1n}^T \cdot a_{n1} = \sum_{j=1}^{n} A_{1j}^T \cdot a_{j1} = \sum_{j=1}^{n} A_{j1} \cdot a_{j1} =$$

$$= \sum_{j=1}^{n} a_{j1} \cdot A_{j1} = |A|;$$

$$b_{21} = A_{21}^T \cdot a_{11} + A_{22}^T \cdot a_{21} + \ldots + A_{2n}^T \cdot a_{n1} = \sum_{j=1}^{n} A_{2j}^T \cdot a_{j1} = \sum_{j=1}^{n} A_{j2} \cdot a_{j1} =$$

$$= \sum_{j=1}^{n} a_{j1} \cdot A_{j2} = 0.$$

Продолжая вычисления дальше, обратим внимание на то, что отличными от нуля окажутся только диагональные элементы матрицы B:

$$b_{ij} = A_{i1}^T \cdot a_{1j} + A_{i2}^T \cdot a_{2j} + \ldots + A_{in}^T \cdot a_{nj} = \sum_{k=1}^{n} A_{ik}^T \cdot a_{kj} = \sum_{k=1}^{n} A_{ki} \cdot a_{kj} =$$

$$= \begin{cases} |A|, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

Поэтому матрица B имеет вид

$$B = \begin{pmatrix} |A| & 0 & \ldots & 0 \\ 0 & |A| & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & |A| \end{pmatrix} = |A| \cdot \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{pmatrix} = |A| \cdot E.$$

Следовательно, $A^P \cdot A = |A| \cdot E$.

Аналогично можно доказать, что $A \cdot A^P = |A| \cdot E$.

Рассмотрим соотношение $A^P \cdot A = A \cdot A^P = |A| \cdot E$.

Разделив его на $|A| \neq 0$, получим

$$\frac{A^P}{|A|} \cdot A = A \cdot \frac{A^P}{|A|} = E.$$

Поскольку для матрицы $\frac{A^P}{|A|}$ выполнено определение, эта матрица является обратной $\frac{A^P}{|A|} \equiv A^{-1}$.

Единственность обратной матрицы. Пусть кроме обратной матрицы A^{-1} к матрице A существует еще одна обратная матрица $X \neq A^{-1}$. Тогда выполняется равенство $X \cdot A = E$. Умножим это равенство справа на A^{-1}. Получим $X \cdot A \cdot A^{-1} = E \cdot A^{-1}$, откуда $X \cdot E = E \cdot A^{-1}$ или $X = A^{-1}$. Таким образом, не существует обратной матрицы X, отличной от A^{-1}. Аналогично доказывается, что равенство $A \cdot Y = E$ выполняется в том единственном случае, когда $Y = A^{-1}$.

Свойства обратных матриц

1. $(A^{-1})^{-1} = A$.
2. $(A^{-1})^T = (A^T)^{-1}$.
3. $(A^{-1})^m = (A^m)^{-1}$.
4. $|A^{-1}| = \frac{1}{|A|}$.
5. $(AB)^{-1} = B^{-1}A^{-1}$.

Доказанная теорема дает способ вычисления обратной матрицы.

Пример 8.4. Найти матрицу, обратную данной:

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 1 \\ -2 & -2 & 1 \end{pmatrix}.$$

Решение. Обратную матрицу будем искать, делая последовательно следующие шаги.

1. Находим определитель матрицы A. Его величина $|A| = -1$. Следовательно, обратная матрица существует.
2. Находим транспонированную к A матрицу A^T:

$$A^T = \begin{pmatrix} 1 & 2 & -2 \\ 2 & -1 & -2 \\ -1 & 1 & 1 \end{pmatrix}.$$

3. Находим алгебраические дополнения к элементам матрицы A^T:

$$A_{11}^T = (-1)^{1+1}\begin{vmatrix} -1 & -2 \\ 1 & 1 \end{vmatrix} = 1, \quad A_{12}^T = (-1)^{1+2}\begin{vmatrix} 2 & -2 \\ -1 & 1 \end{vmatrix} = 0, \ldots,$$

$$A_{33}^T = (-1)^{3+3}\begin{vmatrix} 1 & 2 \\ 2 & -1 \end{vmatrix} = -5.$$

Выписываем присоединенную матрицу:

$$A^P = \begin{pmatrix} 1 & 0 & 1 \\ -4 & -1 & -3 \\ -6 & -2 & -5 \end{pmatrix}.$$

4. Вычисляем обратную матрицу:

$$A^{-1} = \frac{1}{|A|}A^P = \frac{1}{-1}\begin{pmatrix} 1 & 0 & 1 \\ -4 & -1 & -3 \\ -6 & -2 & -5 \end{pmatrix} = \begin{pmatrix} -1 & 0 & -1 \\ 4 & 1 & 3 \\ 6 & 2 & 5 \end{pmatrix}.$$

8.5. РАНГ МАТРИЦЫ

Понятие ранга матрицы — одно из фундаментальных в линейной алгебре. В матрице A размерами $m \times n$ вычеркиванием каких-либо строк или столбцов можно образовать квадратную матрицу k-го порядка ($k \times k$). Определитель M_k такой матрицы называется *минором k-го порядка*. У матрицы размерами $m \times n$ есть миноры 1-го порядка, 2-го порядка и т.д. до k-го порядка, где $k = \min(m, n)$. Например, у матрицы $\underset{5\times 3}{A}$ имеются миноры 1-го, 2-го и 3-го порядков.

Рангом матрицы $\underset{m\times n}{A}$ называется наивысший порядок отличных от нуля миноров этой матрицы. Обозначение: *rang A* или *r*(*A*).

Пример 8.5. Вычислить ранг матрицы $\underset{3\times 4}{A} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$.

Решение. Для матрицы $\underset{3\times 4}{A}$ ранг $r(A) \le \min(3, 4)$. Чтобы проверить, может ли ранг быть равным 3, вычислим все миноры 3-го порядка, которые можно образовать из матрицы вычеркиванием одного столбца:

$$M_3^{(1)} = \begin{vmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} = 0, \quad M_3^{(2)} = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = 0, \quad M_3^{(3)} = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0,$$

$$M_3^{(4)} = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 0.$$

Следовательно, ранг не может быть более 2. Легко найти минор 2-го порядка, отличный от нуля. Например, $M_2 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$. Но тогда $r(A) = 2$.

Поиск ранга матрицы большого порядка перебором миноров является трудоемкой задачей. Развиты эффективные методы определения ранга матрицы.

Назовем *элементарными* следующие преобразования с матрицами.

1. Изменение порядка строк или столбцов.
2. Умножение элементов одной строки или столбца на любое не равное нулю число.
3. Сложение строк с предварительным умножением любой из них на произвольное не равное нулю число.
4. Транспонирование.
5. Отбрасывание нулевой строки или столбца.

Теорема 8.3. Элементарные преобразования не изменяют ранга матрицы.

Теорема дает возможность посредством элементарных преобразований привести матрицу к определенному виду, когда ее ранг вычисляется без труда. Рассмотрим задачу эффективного вычисления ранга подробнее.

Матрица *A* называется матрицей *ступенчатого вида* или *ступенчатой* матрицей, если она имеет вид

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2r} & \dots & a_{2n} \\ 0 & 0 & \dots & a_{3r} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr} & \dots & a_{rn} \end{pmatrix}$$

или

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2r} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr} & \dots & a_{rn} \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix},$$

где $a_{ii} \neq 0$, $i = 1, 2, \dots, r$; $r \leq k = \min(m, n)$. Ранг ступенчатой матрицы равен r, так как существует минор порядка r, отличный от нуля:

$$M_r = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ 0 & a_{22} & \dots & a_{2r} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr} \end{vmatrix} = a_{11} \cdot a_{22} \cdots a_{rr} \neq 0.$$

Таким образом, произвольную матрицу $\underset{m \times n}{A}$ следует привести к ступенчатому виду. Число ненулевых строк матрицы будет равно ее рангу. Если квадратная матрица $\underset{n \times n}{A}$ примет треугольный вид, ее ранг будет равен n. При проведении элементарных преобразований с матрицей знак равенства ставиться не может (матрицы не равны), ставится обычно знак тильды «~» .

Пример 8.6. Найти ранг матрицы

$$A = \begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 4 & -2 & 5 & 1 & 7 \\ 2 & -1 & 1 & 8 & 2 \end{pmatrix}.$$

Решение. К второй строке прибавим первую, предварительно умноженную на –2, к третьей строке прибавим первую, предварительно умноженную на –1. Получим

$$\begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 4 & -2 & 5 & 1 & 7 \\ 2 & -1 & 1 & 8 & 2 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 0 & 0 & -1 & 5 & -1 \\ 0 & 0 & -2 & 10 & -2 \end{pmatrix}.$$

К третьей строке прибавим вторую, предварительно умноженную на –2:

$$\begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 0 & 0 & -1 & 5 & -1 \\ 0 & 0 & -2 & 10 & -2 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 0 & 0 & -1 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 & -2 & 4 \\ 0 & 0 & -1 & 5 & -1 \end{pmatrix}.$$

Число ненулевых строк равно 2. Тогда $r(A) = 2$. При ином обосновании выделим из матрицы минор максимального порядка, не равный нулю. Это, например, $M_2 = \begin{vmatrix} -1 & 3 \\ 0 & -1 \end{vmatrix} = 1$. Тогда $r(A) = 2$.

Линейные комбинации строк или столбцов

Познакомимся с понятием линейной зависимости строк или столбцов. В матрице

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

введем обозначения строк:

$$e_1 = (a_{11} \quad a_{12} \quad \dots \quad a_{1n})$$

$$e_2 = (a_{21} \quad a_{22} \quad \dots \quad a_{2n})$$

....

$$e_m = (a_{m1} \quad a_{m2} \quad \dots \quad a_{mn})$$

Эти строки $e_1, e_2, \dots, e_m$ являются n-мерными и представляют собой матрицы размерами $1 \times n$. В новых обозначениях исходная матрица записывается в виде

$$A = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_m \end{pmatrix}.$$

Строка $e = (b_1 \quad b_2 \quad \dots \quad b_n)$, определяемая равенством

$$e = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_m e_m,$$

называется *линейной комбинацией* строк $e_1, e_2, \dots, e_m$, где $\lambda_1, \lambda_2, \dots, \lambda_m$ — любые действительные числа.

В развернутом матричном виде последнее равенство выглядит так:

$$b_1 \quad b_2 \quad \dots \quad b_n = \lambda_1 \cdot (a_{11} \quad a_{12} \quad \dots \quad a_{1n}) + \lambda_2 \cdot (a_{21} \quad a_{22} \quad \dots \quad a_{2n}) +$$
$$+ \dots + 1_m \cdot (a_{m1} \quad a_{m2} \quad \dots \quad a_{mn}).$$

Для элементов строки $\boldsymbol{e}$ имеем систему уравнений

$$\begin{cases} b_1 = \lambda_1 a_{11} + \lambda_2 a_{21} + \ldots + \lambda_m a_{m1}, \\ b_2 = \lambda_1 a_{12} + \lambda_2 a_{22} + \ldots + \lambda_m a_{m2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ b_n = \lambda_1 a_{1n} + \lambda_2 a_{2n} + \ldots + \lambda_m a_{mn}. \end{cases}$$

Строки $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ называются *линейно зависимыми*, если существуют такие $\lambda_1, \lambda_2, \ldots, \lambda_m$, не равные нулю одновременно, что линейная комбинация этих строк равна нулевой строке:

$$\lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2 + \ldots + \lambda_m \boldsymbol{e}_m = \boldsymbol{0}.$$

Строки $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ называются *линейно независимыми*, если линейная комбинация этих строк равна нулевой строке только при $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$.

Теорема 8.4. Если строки матрицы линейно зависимы, то одна из них является линейной комбинацией остальных.

Пусть строки $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ линейно зависимы. Тогда найдутся числа $\lambda_1, \lambda_2, \ldots, \lambda_m$, не все равные нулю одновременно и такие, что

$$\lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2 + \ldots + \lambda_m \boldsymbol{e}_m = \boldsymbol{0}.$$

Пусть, например, $\lambda_m \neq 0$. Перенесем первые $m - 1$ слагаемых направо и разделим равенство на $\lambda_m \neq 0$. Получим

$$\boldsymbol{e}_m = -\frac{\lambda_1}{\lambda_m}\boldsymbol{e}_1 - \frac{\lambda_2}{\lambda_m}\boldsymbol{e}_2 - \ldots - \frac{\lambda_{m-1}}{\lambda_m}\boldsymbol{e}_{m-1}$$

или

$$\boldsymbol{e}_m = \lambda_1' \boldsymbol{e}_1 + \lambda_2' \boldsymbol{e}_2 + \ldots + \lambda_{m-1}' \boldsymbol{e}_{m-1},$$

где $\lambda_i' = -\dfrac{\lambda_i}{\lambda_m}$, $i = 1, 2, \ldots, m - 1$.

З а м е ч а н и е 1. Верно и обратное утверждение: если одна из строк является линейной комбинацией остальных, то эти строки линейно зависимы.

З а м е ч а н и е 2. Аналогичными свойствами обладает множество m-мерных столбцов.

Теорема 8.5. Ранг матрицы равен числу ее независимых строк (столбцов).

Пусть матрица $\underset{m\times n}{A}$ имеет ранг r. По определению ранга матрицы существует минор порядка r, отличный от нуля. Пусть для определенности это минор

$$M_r = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \dots & \dots & \dots & \dots \\ a_{r1} & a_{r2} & \dots & a_{rr} \end{vmatrix} \neq 0.$$

Тогда строки $e_1, e_2, \dots, e_r$ линейно независимы. Предположим противное. Например, строка с номером r есть линейная комбинация остальных строк. В этом случае

$$e_r = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_{r-1} e_{r-1}.$$

Проведем элементарные преобразования, не изменяющие величину определителя. Прибавим к этой строке первую строку, предварительно умноженную на $-\lambda_1$, вторую строку, умноженную на $-\lambda_2$ и т.д., наконец, $(r-1)$-ю строку, умноженную на $-\lambda_{r-1}$. Получим на месте строки с номером r последовательно строку

$$e_r - \lambda_1 e_1,$$

$$e_r - \lambda_1 e_1 - \lambda_2 e_2,$$

....

$$e_r - \lambda_1 e_1 - \lambda_2 e_2 - \dots - \lambda_{r-1} e_{r-1}.$$

Последняя строка теперь будет состоять из одних нулей. Но тогда $M_r = 0$, что невозможно. Наше предположение о том, что строки $e_1, e_2, \dots, e_r$ линейно зависимы, неверно.

Глава 9
СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

9.1. ОБЩИЙ ПОДХОД К РЕШЕНИЮ СИСТЕМ УРАВНЕНИЙ

Решение многих экономических задач сводится к исследованию и решению систем линейных уравнений.

Система m линейных уравнений с n переменными имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases}$$

или в сокращенной записи

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, 2, \ldots, m.$$

Здесь роль переменных, подлежащих определению, играют величины x_j, называемые *неизвестными*. Параметрами являются переменные a_{ij} и b_i, которые могут принимать любые действительные значения. Все a_{ij} называются *коэффициентами* при переменных, все b_i — *свободными членами* уравнений.

Решением системы называется упорядоченная совокупность чисел $\alpha_1, \alpha_2, \ldots, \alpha_n$, которая при подстановке в систему обращает каждое уравнение в тождество. Линейная система уравнений называется *совместной*, если она имеет хотя бы одно решение, и *несовместной*, если она не имеет решений. Совместная система уравнений называется *определенной*, если она имеет единственное решение, и *неопределенной*, если она имеет более одного решения. Две системы уравнений называются *равносильными*, если они имеют одно и то же множество решений.

Системы уравнений удобно решать в матричной форме. Запишем систему в виде матричного уравнения. Введем обозначения:

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}; \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}; \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix},$$

где A — матрица коэффициентов при переменных, или матрица системы; X — матрица-столбец неизвестных; B — матрица-столбец свободных членов. Тогда система уравнений может быть представлена в матричном виде

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$$

или в компактной матричной форме

$$A \cdot X = B.$$

Нахождение единственного решения системы

Рассмотрим вначале частные случаи решения системы линейных уравнений. Пусть $m = n$ и $|A| \neq 0$.

Метод обратной матрицы

Решим матричное уравнение в случае $m = n$ и $|A| \neq 0$. Для матрицы A существует обратная матрица A^{-1}. Умножим слева обе части матричного уравнения на A^{-1}:

$$A^{-1} \cdot A \cdot X = A^{-1} \cdot B$$

или

$$X = A^{-1} \cdot B.$$

Уравнение в матричном виде решено. Для нахождения элементов матрицы неизвестных X следует найти обратную матрицу коэффициентов и умножить ее на столбец свободных членов B.

Пример 9.1. Решить систему уравнений

$$\begin{cases} x_1 - x_2 + 2x_3 = 1, \\ -2x_1 + x_2 + x_3 = 2, \\ x_1 - x_2 + x_3 = -1. \end{cases}$$

Решение. Напишем систему в матричном виде:

$$\begin{pmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

Матрица коэффициентов

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$$

не вырождена (ее определитель $|A| = 1$). Поэтому существует обратная матрица, которая легко может быть найдена одним из рассмотренных в главе 1 способов:

$$A^{-1} = \begin{pmatrix} 2 & -1 & -3 \\ 3 & -1 & -5 \\ 1 & 0 & -1 \end{pmatrix}.$$

Находим далее произведение

$$A^{-1} \cdot B = \begin{pmatrix} 2 & -1 & -3 \\ 3 & -1 & -5 \\ 1 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 2 \end{pmatrix}.$$

Матрица неизвестных равна

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A^{-1} \cdot B = \begin{pmatrix} 3 \\ 6 \\ 2 \end{pmatrix}.$$

Ответ можно записать также в виде

$$\begin{pmatrix} x_1 = 3, \\ x_2 = 6, \\ x_3 = 2. \end{pmatrix}.$$

Метод с использованием расширенной матрицы

Более эффективный способ решения системы из n уравнений с n неизвестными можно осуществить с помощью расширенной матрицы. Составим расширенную матрицу

$(A|B)$.

Элементарными преобразованиями строк расширенной матрицы приведем матрицу A к единичной. Тогда матрица B обратится в матрицу

$A^{-1} \cdot B$.

Расширенная матрица примет вид

$(E|A^{-1} \cdot B)$.

Извлекая из расширенной матрицы матричное произведение $A^{-1} \cdot B$ и приравнивая его матрице неизвестных, получаем для неизвестных равенство

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A^{-1} \cdot B.$$

Пример 9.2. Решить систему уравнений предыдущего примера с помощью расширенной матрицы.

Решение. Составим расширенную матрицу:

$$(A \mid B) = \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ -2 & 1 & 1 & 2 \\ 1 & -1 & 1 & -1 \end{array}\right).$$

Приведем вначале матрицу A к треугольному виду:

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & -1 & 5 & 4 \\ 0 & 0 & -1 & -2 \end{array}\right).$$

Далее образуем треугольник нулей выше главной диагонали:

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & -1 & 5 & 4 \\ 0 & 0 & -1 & -2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -1 & 0 & -3 \\ 0 & -1 & 0 & -6 \\ 0 & 0 & -1 & -2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & -1 & 0 & -6 \\ 0 & 0 & -1 & -2 \end{array}\right).$$

Наконец, получим единицы на главной диагонали:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 2 \end{array}\right).$$

Следовательно,

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 2 \end{pmatrix}.$$

Метод с использованием формул Крамера

Третий способ решения системы из n уравнений с n неизвестными в случае $|A| \neq 0$ дает теорема швейцарского математика Габриэля Крамера (1704–1752).

Теорема 9.1. Пусть у квадратной матрицы коэффициентов при неизвестных в системе из n линейных уравнений с n неизвестными определитель $|A| \neq 0$. Пусть $|A_j|$ — определитель матрицы, получаемой из матрицы A заменой j-го столбца столбцом свободных членов. Тогда система имеет единственное решение, имеющее вид

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}, \quad \ldots, \quad x_n = \frac{|A_n|}{|A|}.$$

Развернем матричное уравнение $X = A^{-1} \cdot B$ и запишем обратную матрицу через алгебраические дополнения:

$$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \frac{1}{|A|} \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

Перемножив матрицы, получим

$$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \frac{1}{|A|} \begin{pmatrix} A_{11}b_1 + A_{21}b_2 + \dots + A_{n1}b_n \\ A_{12}b_1 + A_{22}b_2 + \dots + A_{n2}b_n \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ A_{1n}b_1 + A_{2n}b_2 + \dots + A_{nn}b_n \end{pmatrix}.$$

Сумма $A_{11}b_1 + A_{21}b_2 + \dots + A_{n1}b_n$ представляет собой произведение чисел $b_1, b_2, \dots, b_n$ на алгебраические дополнения элементов первого столбца. Она равна определителю матрицы, полученной из данной заменой элементов этого столбца на числа $b_1, b_2, \dots, b_n$ (свойство 9 определителей). Следовательно,

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & \dots & a_{1n} \\ b_2 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ b_n & a_{n2} & \dots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}} = \frac{|A_1|}{|A|}.$$

Аналогично сумма $A_{12}b_1 + A_{22}b_2 + \dots + A_{n2}b_n$ есть произведение чисел $b_1, b_2, \dots, b_n$ на алгебраические дополнения элементов второго столбца. Тогда

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & \dots & a_{1n} \\ a_{21} & b_2 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & b_n & \dots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}} = \frac{|A_2|}{|A|}.$$

Продолжив вычисления, окончательно получим

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}, \quad \dots, \quad x_n = \frac{|A_n|}{|A|}.$$

Способ решения системы линейных уравнений, основанный на формулах Крамера, получил название метода или правила Крамера.

Пример 9.3. Решить систему уравнений предыдущей задачи методом Крамера.

Решение. Условия, при которых правило Крамера работает ($m = n$, $|A| \neq 0$), выполнены. Воспользуемся формулами Крамера:

$$x_1 = \frac{|A_1|}{|A|} = \frac{\begin{vmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ -1 & -1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & -1 & 1 \end{vmatrix}} = \frac{3}{1} = 3; \quad x_2 = \frac{|A_2|}{|A|} = \frac{\begin{vmatrix} 1 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & -1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & -1 & 1 \end{vmatrix}} = \frac{6}{1} = 6;$$

$$x_3 = \frac{|A_3|}{|A|} = \frac{\begin{vmatrix} 1 & -1 & 1 \\ -2 & 1 & 2 \\ 1 & -1 & -1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & -1 & 1 \end{vmatrix}} = \frac{2}{1} = 2.$$

Замечание. Перечисленные методы решения систем линейных уравнений становятся трудоемкими при ручном счете уже при $n \geq 4$. Однако они удобны при решении задач на компьютере.

Рассмотренные методы являются решением систем частного вида, для которых выполняются условия $m = n$, $|A| \neq 0$. Перейдем к рассмотрению решения линейных систем общего вида. В дальнейшем будем оперировать понятиями матричной алгебры.

Общий подход к решению систем уравнений

Теорема 9.2. При элементарных преобразованиях строк первых четырех типов линейные системы остаются равносильными.

Рассмотрим элементарные преобразования каждого типа по отдельности.

1. Если система линейных уравнений получена из исходной системы элементарными преобразованиями первого типа, т.е. изменен порядок уравнений в системе, то решения системы не изменятся.

2. Если система линейных уравнений получена из исходной системы элементарными преобразованиями второго типа, т.е. одно из уравнений умножено на число $a \neq 0$, это не приведет к изменению решений системы.

3. Если система линейных уравнений получена из исходной системы элементарными преобразованиями третьего типа, т.е. одно из уравнений представляет собой сумму двух уравнений, одно из которых предварительно умножено на число λ, это также не приведет к изменению решений системы.

Действительно, пусть $x_1^0, x_2^0, ..., x_n^0$ — решение системы уравнений. Тогда уравнения с произвольными номерами i и j

$$a_{i1}x_1 + ... + a_{in}x_n = b_i$$
$$a_{j1}x_1 + ... + a_{jn}x_n = b_j$$

при подстановке чисел $x_1^0, x_2^0, ..., x_n^0$ обратятся в тождества. Сложим оба уравнения, предварительно умножив первое из них на число λ:

$$\lambda(a_{i1}x_1 + ... + a_{in}x_n - b_i) + (a_{j1}x_1 + ... + a_{jn}x_n - b_j) = 0.$$

Подставив сюда числа $x_1^0, x_2^0, ..., x_n^0$, получим $\lambda \cdot 0 + 0 = 0$.

4. Если система линейных уравнений получена из исходной системы элементарными преобразованиями четвертого типа, т.е. одно из уравнений, содержащее нулевые коэффициенты и нулевой свободный член, вычеркнуто, это, очевидно, не изменит решений системы.

Применение элементарных преобразований при решении систем линейных уравнений приводит к мощному методу решения произвольных линейных систем — методу немецкого математика и физика, профессора Геттингенского университета Карла Фридриха Гаусса (1777–1855).

Метод Гаусса

Метод Гаусса заключается в последовательном исключении неизвестных с помощью элементарных преобразований.

Для системы уравнений образуем расширенную матрицу:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right).$$

Посредством элементарных преобразований приведем расширенную матрицу к ступенчатому виду:

$$\begin{pmatrix} \dot a_{11} & \dot a_{12} & \dots & \dot a_{1r} & \dots & \dot a_{1n} & \vert & \dot b_1 \\ 0 & \dot a_{22} & \dots & \dot a_{2r} & \dots & \dot a_{2n} & \vert & \dot b_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \vert & \dots \\ 0 & 0 & \dots & \dot a_{rr} & \dots & \dot a_{rn} & \vert & \dot b_r \\ 0 & 0 & \dots & 0 & \dots & 0 & \vert & \dot b_{r+1} \\ 0 & 0 & \dots & 0 & \dots & 0 & \vert & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \vert & \dots \\ 0 & 0 & \dots & 0 & \dots & 0 & \vert & 0 \end{pmatrix}.$$

Рассмотрим различные случаи.

1. $\dot b_{r+1} \neq 0$. Расширенная матрица имеет вид флажка (рис. 9.1, *а*). Тогда система уравнений несовместна. Действительно, уравнение с номером $r + 1$ содержит нулевые коэффициенты перед неизвестными, тогда как свободный член отличен от нуля. Пусть далее $\dot b_{r+1} = 0$.

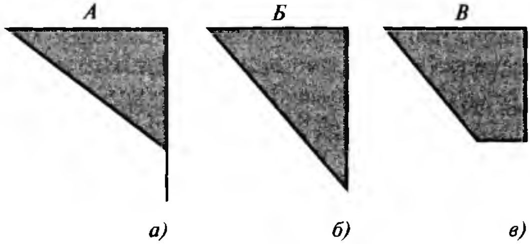

Рис. 9.1

2. Число неизвестных n и число уравнений r совпадают. Расширенная матрица примет треугольный вид (рис. 9.1, *б*):

$$\begin{pmatrix} \dot a_{11} & \dot a_{12} & \dots & \dot a_{1n} & \vert & \dot b_1 \\ 0 & \dot a_{22} & \dots & \dot a_{2n} & \vert & \dot b_2 \\ \dots & \dots & \dots & \dots & \vert & \dots \\ 0 & 0 & \dots & \dot a_{nn} & \vert & \dot b_n \end{pmatrix}.$$

Система уравнений, соответствующая этой матрице, имеет вид

$$\begin{cases} \dot a_{11}x_1 + \dot a_{12}x_2 + \dots + \dot a_{1n}x_n = \dot b_1, \\ \dot a_{22}x_2 + \dots + \dot a_{2n}x_n = \dot b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ \dot a_{nn}x_n = \dot b_n. \end{cases}$$

Из последнего уравнения определяется неизвестная величина x_n. Подставляем ее в предыдущее уравнение с номером $n-1$ и находим x_{n-1}. Продолжая этот процесс, находим неизвестные со все меньшими номерами. Наконец, подставив найденные значения неизвестных $x_2, x_3, ..., x_n$ в первое уравнение, найдем величину x_1. Итак, при $n = r$ система совместна и имеет единственное решение.

3. Число неизвестных n больше числа уравнений r. Вид расширенной матрицы — трапеция (рис. 9.1, *в*). Последнее уравнение содержит переменные $x_r, x_{r+1}, ..., x_n$. Выразим в этом уравнении переменную x_r через остальные неизвестные и подставим в уравнение с номером $r-1$. Найдем переменную x_{r-1}, которая будет выражена через те же неизвестные $x_{r+1}, ..., x_n$. Результат подставим в уравнение с номером $r-2$ и т. д. Таким образом мы можем определить значения переменных $x_r, x_{r-1}, ..., x_1$ через неизвестные $x_{r+1}, x_{r+2}, ..., x_n$:

$$\begin{cases} x_1 = \ddot{a}_{1,r+1}x_{r+1} + \ddot{a}_{1,r+2}x_{r+2} + \ldots + \ddot{a}_{1n}x_n + \ddot{b}_1, \\ x_2 = \ddot{a}_{2,r+1}x_{r+1} + \ddot{a}_{2,r+2}x_{r+2} + \ldots + \ddot{a}_{2n}x_n + \ddot{b}_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_r = \ddot{a}_{r,r+1}x_{r+1} + \ddot{a}_{r,r+2}x_{r+2} + \ldots + \ddot{a}_{r,n}x_n + \ddot{b}_r. \end{cases}$$

Придавая неизвестным $x_{r+1}, ..., x_n$ произвольные значения, получаем бесконечное множество решений системы уравнений.

Пример 9.4. Решить систему линейных уравнений

$$\begin{cases} 3x_1 - 2x_2 + 5x_3 + 4x_4 = 2, \\ 6x_1 - 4x_2 + 4x_3 + 3x_4 = 3, \\ 9x_1 - 6x_2 + 3x_3 + 2x_4 = 4. \end{cases}$$

Решение. Составим расширенную матрицу системы:

$$\left(\begin{array}{cccc|c} 3 & -2 & 5 & 4 & 2 \\ 6 & -4 & 4 & 3 & 3 \\ 9 & -6 & 3 & 2 & 4 \end{array}\right).$$

С помощью элементарных преобразований приведем ее к ступенчатому виду:

$$\left(\begin{array}{cccc|c} 3 & -2 & 5 & 4 & 2 \\ 0 & 0 & -6 & -5 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \text{ или } \left(\begin{array}{cccc|c} 3 & -2 & 5 & 4 & 2 \\ 0 & 0 & -6 & -5 & -1 \end{array}\right).$$

Исходная система эквивалентна следующей системе уравнений:

$$\begin{cases} 3x_1 - 2x_2 + 5x_3 + 4x_4 = 2, \\ \qquad\qquad -6x_3 - 5x_4 = -1. \end{cases}$$

Выразим из второго уравнения x_3 через x_4 и подставим в первое уравнение системы. Тогда

$$x_1 = \frac{1}{18}(12x_2 + x_4 + 7),$$

$$x_3 = \frac{1}{6}(-5x_4 + 1).$$

Придавая произвольные значения неизвестным

$x_2 = a$ и $x_4 = b$,

получим общее решение системы уравнений:

$$x_1 = \frac{1}{18}(12a + b + 7);$$

$$x_2 = a;$$

$$x_3 = \frac{1}{6}(-5b + 1);$$

$$x_4 = b.$$

Решение методом Гаусса представляет собой кропотливый и часто длительный процесс. Когда в конце пути может оказаться, что система не имеет решения, наступает разочарование. Столько сделано работы, и как ничтожен итог. В середине XIX в. Леопольдом Кронекером (1823–1891), профессором Берлинского университета, была найдена та «лакмусовая бумажка», по реакции которой можно судить о наличии или отсутствии решений системы линейных уравнений. Ею оказался ранг.

Теорема 9.3 (теорема Кронекера—Капелли). Линейная система совместна тогда и только тогда, когда ранг матрицы системы равен рангу расширенной матрицы этой системы.

В системе линейных уравнений введем обозначения матрицы коэффициентов

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

и расширенной матрицы

$$A^P = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right),$$

а также обозначения столбцов, составленных из коэффициентов при неизвестных:

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix} = \boldsymbol{a}_1, \quad \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix} = \boldsymbol{a}_2, \quad \dots, \quad \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix} = \boldsymbol{a}_n, \quad \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix} = \boldsymbol{b}.$$

В новых обозначениях система выглядит так:

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \dots + x_n\boldsymbol{a}_n = \boldsymbol{b}.$$

Необходимость. Пусть система совместна, т.е. существуют $x_1 = \lambda_1, x_2 = \lambda_2, \dots, x_n = \lambda_n$ такие, что выполняется равенство

$$\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \dots + \lambda_n\boldsymbol{a}_n = \boldsymbol{b}.$$

Из него следует, что столбец $\boldsymbol{b}$ есть линейная комбинация остальных столбцов матрицы системы. Следовательно, добавление в матрицу коэффициентов при неизвестных столбца свободных членов ранга не меняет: $r(A) = r(A^P)$.

Достаточность. Пусть матрицы A и A^P имеют одинаковый ранг $r(A) = r(A^P)$. Тогда r столбцов матрицы A (пусть это, например, будут первые r столбцов) линейно независимы. Остальные $n - r$ столбцов, а именно $\boldsymbol{a}_{r+1}, \dots, \boldsymbol{a}_n$, а также столбец b являются линейными комбинациями первых r столбцов. В частности, найдутся такие $\lambda_1, \lambda_2, \dots, \lambda_r$, не все равные нулю одновременно, что

$$\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \dots + \lambda_r\boldsymbol{a}_r = \boldsymbol{b}.$$

Расширим это равенство за счет добавления в него слагаемых с коэффициентами $\lambda_{r+1} = 0, \dots, \lambda_n = 0$:

$$\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \dots + \lambda_r\boldsymbol{a}_r + \lambda_{r+1}\boldsymbol{a}_{r+1} + \dots + \lambda_n\boldsymbol{a}_n = \boldsymbol{b}.$$

Но это и означает, что исходная система имеет решения

$$x_1 = \lambda_1, \quad x_2 = \lambda_2, \dots, x_r = \lambda_r, \quad x_{r+1} = \lambda_{r+1} = 0, \dots, x_n = \lambda_n = 0.$$

Однородные системы линейных уравнений

Система линейных уравнений называется однородной, если все свободные члены системы равны нулю:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = 0. \end{cases}$$

В матричном виде систему можно записать так:

$$A \cdot X = O,$$

где $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$; $X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$; $O = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}$.

Свойства однородной системы линейных уравнений

1. Однородная система всегда совместна, так как всегда имеет по крайней мере нулевое решение.

2. Для существования ненулевых решений ранг матрицы коэффициентов должен быть меньше числа переменных: $r < n$, т.е число линейно независимых уравнений должно быть меньше числа переменных. В этом случае $|A| = 0$.

3. Если матрица-столбец (вектор) $\boldsymbol{e} = \begin{pmatrix} x_{11} \\ x_{21} \\ \dots \\ x_{n1} \end{pmatrix}$ есть решение исходной системы, то и столбец $\boldsymbol{e}' = \lambda \cdot \boldsymbol{e} = \begin{pmatrix} \lambda x_{11} \\ \lambda x_{21} \\ \dots \\ \lambda x_{n1} \end{pmatrix}$ также является решением системы.

Пусть $\boldsymbol{e}$ — решение системы. Тогда матричное уравнение при подстановке $X = \boldsymbol{e}'$ обращается в тождество $A \cdot \boldsymbol{e}' = O$.

Действительно, найдем произведение матриц A и $\lambda \cdot \boldsymbol{e}$:

$A \cdot (\lambda \boldsymbol{e}) = \lambda \cdot (A \cdot e) = \lambda \cdot \mathrm{O} = O$.

Отсюда столбец $\boldsymbol{e}' = \lambda \cdot \boldsymbol{e}$ также является решением матричного уравнения.

4. Если матрицы-столбцы $\boldsymbol{e}_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \dots \\ x_{n1} \end{pmatrix}$ и $\boldsymbol{e}_2 = \begin{pmatrix} x_{12} \\ x_{22} \\ \dots \\ x_{n2} \end{pmatrix}$ есть решения исходной системы, т.е. $A \cdot \boldsymbol{e}_1 = O$ и $A \cdot \boldsymbol{e}_2 = O$, то и столбец $\boldsymbol{e} = \lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2$, где λ_1, λ_2 — произвольные числа, также является решением системы $A \cdot \boldsymbol{e} = O$.

Для доказательства перемножим матрицы A и $\boldsymbol{e} = \lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2$:

$A \cdot (\lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2) = A \cdot (\lambda_1 \boldsymbol{e}_1) + A \cdot (\lambda_2 \boldsymbol{e}_2) = \lambda_1 \cdot (A \cdot \boldsymbol{e}_1) + \lambda_2 \cdot (A \cdot \boldsymbol{e}_2) =$
$= \lambda_1 \cdot O + \lambda_2 \cdot O = O$.

Следовательно, всякая линейная комбинация решений однородной системы линейных уравнений также является решением этой системы.

Заметим, что среди решений однородной системы выделяются решения, которые можно назвать главными решениями. Через них выражаются другие решения. Попробуем разобраться в этом и найти эти выделяющиеся, фундаментальные решения.

9.2. ФУНДАМЕНТАЛЬНЫЕ РЕШЕНИЯ ОДНОРОДНОЙ СИСТЕМЫ УРАВНЕНИЙ

Решения $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p$ однородной системы называются *линейно независимыми*, если линейная комбинация этих решений $\lambda_1\boldsymbol{e}_1 + \lambda_2\boldsymbol{e}_2 + \ldots + \lambda_p\boldsymbol{e}_p$ равна нулевому столбцу только при условии $\lambda_1 = \lambda_2 = \ldots = \lambda_p = 0$. Построим матрицу решений, расположив матрицы-столбцы решений по столбцам новой матрицы. В соответствии с теоремой о ранге матрицы ранг новой матрицы будет численно равен числу столбцов новой матрицы, т.е. числу линейно независимых решений системы.

Совокупность линейно независимых решений $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p$ однородной системы уравнений называется *фундаментальной*, если общее решение системы является линейной комбинацией решений $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p$.

Теорема 9.4. Если ранг r матрицы коэффициентов при переменных однородной системы уравнений меньше числа переменных n, то:

1) существует совокупность линейно независимых решений системы;

2) число линейно независимых решений равно $n - r$;

3) любое решение системы можно представить в виде совокупности этих независимых решений, т.е. в виде линейной комбинации фундаментального набора решений.

З а м е ч а н и е 1. Для нахождения множества решений однородной системы достаточно найти какой-нибудь ФНР системы и составить его линейную комбинацию.

З а м е ч а н и е 2. Любая однородная система уравнений с ненулевыми решениями имеет ФНР.

Пример 9.5. Решить систему однородных уравнений, выделив какой-либо ФНР:

$$\begin{cases} 2x_1 - x_2 + x_3 - x_4 = 0, \\ x_1 + 2x_2 - 2x_3 + 3x_4 = 0, \\ 3x_1 + x_2 - x_3 + 2x_4 = 0. \end{cases}$$

Запишем матрицу коэффициентов и, совершая элементарные преобразования со строками, приведем ее к ступенчатому виду:

$$\begin{pmatrix} 2 & -1 & 1 & -1 \\ 1 & 2 & -2 & 3 \\ 3 & 1 & -1 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -2 & 3 \\ 2 & -1 & 1 & -1 \\ 3 & 1 & -1 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -2 & 3 \\ 0 & -5 & 5 & -7 \\ 0 & -5 & 5 & -7 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -2 & 3 \\ 0 & -5 & 5 & -7 \end{pmatrix}.$$

Ранг матрицы $r = 2$. Вернемся к системе уравнений:

$$\begin{cases} x_1 + 2x_2 - 2x_3 + 3x_4 = 0, \\ \quad -5x_2 + 5x_3 - 7x_4 = 0. \end{cases}$$

Возьмем базисными переменными x_1, x_2, тогда свободными останутся x_3, x_4. Найдем x_1, x_2 и запишем решения в удобной для дальнейшей записи форме:

$$\begin{cases} x_1 = 0 \cdot x_3 - \frac{1}{5}x_4, \\ x_2 = 1 \cdot x_3 - \frac{7}{5}x_4, \\ x_3 = 1 \cdot x_3 + 0 \cdot x_4, \\ x_4 = 0 \cdot x_3 + 1 \cdot x_4. \end{cases}$$

В матричном виде решение системы можно записать так:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -\frac{1}{5} \\ -\frac{7}{5} \\ 0 \\ 1 \end{pmatrix} = x_3 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} - \frac{x_4}{5} \begin{pmatrix} 1 \\ 7 \\ 0 \\ -5 \end{pmatrix}.$$

Обозначим коэффициенты перед столбцами в правой части

$$x_3 = c_1, \quad -\frac{x_4}{5} = c_2$$

и запишем общее решение системы еще раз:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = c_1 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 7 \\ 0 \\ -5 \end{pmatrix}, \quad \text{где } c_1, c_2 \in R.$$

Задавая коэффициентам c_1, c_2 произвольные значения, получаем совокупность всех решений системы. Столбцы $\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 1 \\ 7 \\ 0 \\ -5 \end{pmatrix}$ линейно независимы и представляют набор из двух фундаментальных решений (ФНР), через которые выражаются все остальные решения системы.

Глава 10
ВЕКТОРНАЯ АЛГЕБРА

10.1. ВЕКТОРЫ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ. N-МЕРНЫЕ ВЕКТОРЫ

Вектором на плоскости или в пространстве называется направленный отрезок, имеющий начальную и конечную точки. Обычно вектор обозначается строчной буквой с чертой $\bar{a}$ либо буква выделяется жирным шрифтом $\boldsymbol{a}$ или же двумя заглавными буквами, обозначающими начало и конец вектора $\overline{AB}$. Мы будем векторы называть *свободными*, если начальную точку можно выбирать произвольно. Такие векторы можно произвольно переносить параллельно самим себе.

Длиной или *модулем* $|\boldsymbol{a}|$ вектора называется число, равное длине направленного отрезка. Вектор, длина которого равна единице, называется *единичным* вектором, или *ортом*. Вектор, начало и конец которого совпадают, называется *нулевым* и обозначается $\boldsymbol{o}$. Векторы, лежащие на параллельных прямых, называются *коллинеарными*.

1. *Суммой* векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ называется вектор $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$, начало которого совпадает с началом вектора $\boldsymbol{a}$, а конец — с концом вектора $\boldsymbol{b}$ (рис. 10.1). Этот способ построения суммы векторов называется *правилом треугольника*.

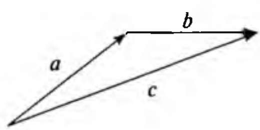

Рис. 10.1

Если на векторах $\boldsymbol{a}$ и $\boldsymbol{b}$ как на сторонах построить параллелограмм, то большая диагональ параллелограмма будет суммой векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. Этот способ построения суммы векторов называется правилом параллелограмма (рис.10.2).

2. *Произведением* числа λ на вектор $\boldsymbol{a}$ называется вектор $\boldsymbol{b}$, имеющий длину $|\boldsymbol{b}| = |\lambda||\boldsymbol{a}|$. Направление вектора $\boldsymbol{b}$ совпадает с направлением $\boldsymbol{a}$, если $\lambda > 0$, и противоположно по направлению, если $\lambda < 0$.

3. Разностью двух векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ называется вектор $\boldsymbol{c} = \boldsymbol{a} + (-1)\boldsymbol{b} = \boldsymbol{a} - \boldsymbol{b}$. На рис. 10.3 вектор $-\boldsymbol{b}$ изображен пунктиром. Сумма

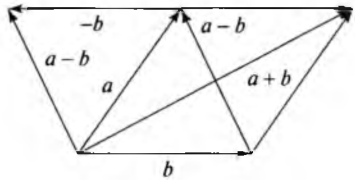

Рис. 10.2

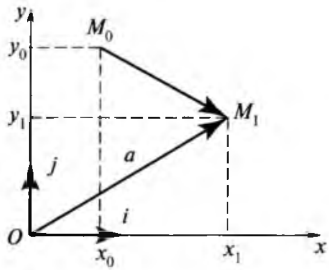

Рис. 10.3

векторов $\boldsymbol{a}$ и $-\boldsymbol{b}$ также изображена пунктиром. Легко видеть, что $\boldsymbol{a} - \boldsymbol{b}$ есть меньшая диагональ параллелограмма.

Укажем основные свойства линейных операций над векторами.

1) $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$;
2) $\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}$;
3) $\boldsymbol{a} + \boldsymbol{o} = \boldsymbol{a}$;
4) $\alpha(\boldsymbol{a} + \boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b}$.

Координаты вектора

Зададим на плоскости точку O и назовем ее *полюсом*. Отложим от точки O как от начала два разнонаправленных неколлинеарных вектора $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ и проведем через полюс две прямые, определяемые этими векторами. Полюс O вместе с прямыми образуют на плоскости так называемую *аффинную систему координат*[1], в которой точка O называется началом координат, а прямые — осями координат. Они называются *осью абсцисс* и *осью ординат* и обозначаются как Ox и Oy. Используются обычно одинаковые масштабы по осям, но они могут и различаться. Аффинную систему координат с прямым углом между осями называют, как правило, *декартовой системой координат*. Сам

[1] Система координат названа аффинной от латинского слова *affinis*, что значит родственная. Термин обусловлен геометрическими преобразованиями фигур, при которых прямые переходят в прямые.

Декарт употреблял только систему координат на плоскости, но в дальнейшем она была расширена на трехмерное пространство добавлением *оси аппликат* (ось Oz), перпендикулярной плоскости Oxy.

Координатами вектора a в аффинной системе координат называются координаты его конечной точки при условии, что начальная точка вектора лежит в начале координат. В дальнейшем по умолчанию будем считать, что все векторы отложены от начала координат.

На рис. 10.3 изображена декартова (прямоугольная) система координат, в которой координатами вектора $\boldsymbol{a}$ на плоскости xOy являются два числа x_1 и y_1, что обычно записывается в виде $\boldsymbol{a} = (x_1, y_1)$. Пусть вдоль осей Ox и Oy направлены векторы единичной длины $\boldsymbol{i}$ и $\boldsymbol{j}$. Тогда произвольный вектор $\boldsymbol{a}$ может быть представлен в виде суммы единичных векторов $\boldsymbol{i}$ и $\boldsymbol{j}$ с коэффициентами x_1 и y_1:

$$\boldsymbol{a} = x_1\boldsymbol{i} + y_1\boldsymbol{j},$$

что соответствует правилу параллелограмма.

Если вектор отложен от точки $M_0(x_0, y_0)$, не совпадающей с началом координат (свободный вектор), до точки $M_1(x_1, y_1)$, то координатами вектора $\overline{M_0M_1}$ будем считать координаты $(x_1 - x_0, y_1 - y_0)$.

Запишем линейные операции над векторами в координатах. Пусть заданы векторы $\boldsymbol{a} = (x_1, y_1)$ и $\boldsymbol{b} = (x_2, y_2)$:

1) $|\boldsymbol{a}| = \sqrt{x_1^2 + y_1^2}$;
2) $\boldsymbol{a} \pm \boldsymbol{b} = (x_1 \pm x_2, y_1 \pm y_2)$;
3) $\lambda \cdot \boldsymbol{a} = (\lambda x_1, \lambda y_1)$.

Проекция вектора

Проекция $\boldsymbol{a}_b$ вектора $\boldsymbol{a}$ на направление, задаваемое вектором $\boldsymbol{b}$, определяется формулой

$$\boldsymbol{a}_b = \boldsymbol{a}\cos\alpha,$$

где α — угол между векторами $\boldsymbol{a}$ и $\boldsymbol{b}$ (рис. 10.4). Проекции $\boldsymbol{a}_x$, $\boldsymbol{a}_y$ на декартовы оси координат представляют собой координаты вектора $\boldsymbol{a}$, умноженные на единичные векторы, направленные вдоль осей:

$$\boldsymbol{a}_x = a_x\boldsymbol{e}_1, \quad \boldsymbol{a}_y = a_y\boldsymbol{e}_2.$$

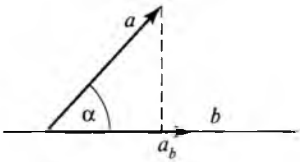

Рис. 10.4

Косоугольные системы координат

При изучении векторных пространств нам придется иметь дело с аффинными системами координат, у которых углы между осями координат отличаются от $\pi/2$. Такие системы координат называются косоугольными. Величины проекций $\boldsymbol{a}'_x$, $\boldsymbol{a}'_y$ вектора $\boldsymbol{a}$ на косоугольные декартовы оси координат будут зависеть от угла между осями. Говорят, например, что $\boldsymbol{a}'_x$ есть проекция вектора $\boldsymbol{a}$ на ось x параллельно оси y (рис. 10.5). Построив векторы $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ как единичные вдоль осей Ox и Oy, получим разложение вектора $\boldsymbol{a}$ по осям с коэффициентами $\boldsymbol{a}'_x$ и $\boldsymbol{a}'_y$:

$$\boldsymbol{a} = \boldsymbol{a}'_x + \boldsymbol{a}'_y = a'_x\boldsymbol{e}_1 + a'_y\boldsymbol{e}_2.$$

По умолчанию в разделе геометрических векторов мы будем ори-

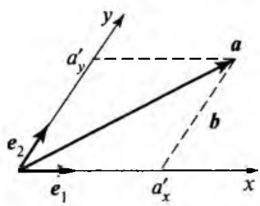

Рис. 10.5

ентироваться на прямоугольную декартову систему координат.

Скалярное произведение векторов

Определение. *Скалярным произведением* двух векторов называется произведение модулей этих векторов, умноженное на косинус угла между векторами и обозначаемое $(\boldsymbol{a}, \boldsymbol{b})$:

$$(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \cos\varphi.$$

Найдем скалярное произведение, выраженное через координаты векторов. На координатной плоскости построим векторы $\boldsymbol{a} = (x_1, y_1)$, $\boldsymbol{b} = (x_2, y_2)$ и $\boldsymbol{c} = \boldsymbol{a} - \boldsymbol{b} = (x_1 - x_2, y_1 - y_2)$. Они образуют треугольник, изображенный на рис. 10.6. По теореме косинусов найдем длину стороны, образуемой вектором $\boldsymbol{c}$:

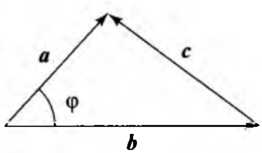

Рис. 10.6

$$|\boldsymbol{c}|^2=|\boldsymbol{a}|^2+|\boldsymbol{b}|^2-2|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi.$$

Отсюда следует

$$|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi=\frac{|\boldsymbol{a}|^2+|\boldsymbol{b}|^2-|\boldsymbol{c}|^2}{2}.$$

Выразим векторы через их координаты:

$$|\boldsymbol{a}|^2=x_1^2+y_1^2,\quad |\boldsymbol{b}|^2=x_2^2+y_2^2,\quad |\boldsymbol{c}|^2=(x_1-x_2)^2+(y_1-y_2)^2.$$

Тогда

$$|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi=\frac{x_1^2+y_1^2+x_2^2+y_2^2-(x_1-x_2)^2-(y_1-y_2)^2}{2}.$$

Раскроем скобки и приведем подобные. Получим

$$(\boldsymbol{a},\boldsymbol{b})=|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi=x_1x_2+y_1y_2,$$

т.е. скалярное произведение двух векторов равно сумме произведений соответствующих координат этих векторов.

Рассмотрим некоторые следствия формулы.

1. В скалярном произведении положим $\boldsymbol{a}=\boldsymbol{b}$. Тогда

$$(\boldsymbol{a},\boldsymbol{b})=(\boldsymbol{a},\boldsymbol{a})=|\boldsymbol{a}||\boldsymbol{a}|\cos 0^0=|\boldsymbol{a}|^2=x_1^2+y_1^2,$$

т.е. скалярный квадрат вектора равен квадрату его длины.

2. Найдем из равенства величину $\cos\varphi$:

$$\cos\varphi=\frac{(\boldsymbol{a},\boldsymbol{b})}{|\boldsymbol{a}||\boldsymbol{b}|}=\frac{x_1x_2+y_1y_2}{\sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}}.$$

3. Пусть вектор $\boldsymbol{b}$ направлен вдоль оси Ox и имеет единичную длину $|\boldsymbol{b}|=1$, т.е. $\boldsymbol{b}=(1, 0)$. Тогда для всякого вектора $\boldsymbol{a}=(x_1, y_1)$ величина

$$\cos\alpha=\frac{(\boldsymbol{a},\boldsymbol{b})}{|\boldsymbol{a}||\boldsymbol{b}|}=\frac{x_1\cdot 1+y_1\cdot 0}{|\boldsymbol{a}|}=\frac{x_1}{\sqrt{x_1^2+y_1^2}}$$

описывает косинус угла между вектором $\boldsymbol{b}$ и положительным направлением оси Ox. Точно так же при направлении единичного вектора $\boldsymbol{b}$ вдоль оси Oy ($\boldsymbol{b}=(0, 1)$) величина

$$\cos\beta=\frac{y_1}{\sqrt{x_1^2+y_1^2}}$$

описывает косинус угла между вектором $\boldsymbol{a}$ и положительным направлением оси Oy (рис. 10.7); $\cos\alpha$ и $\cos\beta$ называются направляющими косинусами вектора $\boldsymbol{a}$.

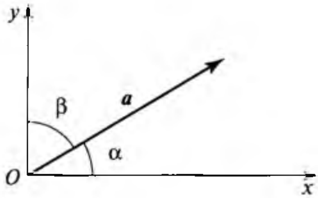

Рис. 10.7

Свойства скалярного произведения

1. $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$.

Справедливость свойства вытекает из формулы скалярного произведения, если учесть четность функции $\cos\varphi$.

2. $(\boldsymbol{a} + \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a}, \boldsymbol{c}) + (\boldsymbol{b}, \boldsymbol{c})$.

Пусть векторы $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ заданы своими координатами $\boldsymbol{a} = (x_1, y_1)$, $\boldsymbol{b} = (x_2, y_2)$, $\boldsymbol{c} = (x_3, y_3)$. Запишем скалярное произведение $(\boldsymbol{a} + \boldsymbol{b}, \boldsymbol{c})$ в координатах:

$$(\boldsymbol{a} + \boldsymbol{b}, \boldsymbol{c}) = (x_1 + x_2)x_3 + (y_1 + y_2)y_3 = (x_1x_3 + y_1y_3) + (x_2x_3 + y_2y_3).$$

Выражения, представленные в скобках, есть соответственно скалярные произведения $(\boldsymbol{a}, \boldsymbol{c})$ и $(\boldsymbol{b}, \boldsymbol{c})$. Отсюда следует свойство.

3. $(\lambda \cdot \boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a}, \lambda \cdot \boldsymbol{b}) = \lambda \cdot (\boldsymbol{a}, \boldsymbol{b})$.

Элементы аналитической геометрии

Рассмотрим задачи описания линейных математических объектов в трехмерном пространстве. Получим уравнения, которые описывают всю совокупность точек этих объектов и которые не справедливы для всех остальных точек пространства. Весьма эффективным является векторный подход. Каждой точке трехмерного пространства ставится в соответствие вектор, начало которого совпадает с началом системы координат, а конец — с рассматриваемой точкой пространства. Тогда бесконечное множество точек пространства будет описано бесконечным множеством соответствующих им векторов, которое можно назвать векторным пространством. Будем рассматривать вектор как матрицу-строку или матрицу-столбец.

Уравнение прямой линии

Положение прямой линии в пространстве будет определено, если задать точку M_0, через которую проведена прямая, и ее направление в пространстве. Точка $M_0(x_0, y_0, z_0)$ задается вектором $\boldsymbol{r}_0 = (x_0, y_0, z_0)$, называемым вектором сдвига, направление — ненулевым вектором $\boldsymbol{e}$. Построим прямую l, параллельную вектору $\boldsymbol{e}$ и проходящую через начало координат (рис. 10.8). Совокупность точек прямой задается уравнением

$$\boldsymbol{r} = \boldsymbol{e}t,$$

где t — некоторая переменная, каждому значению которой соответствует определенная точка на прямой l. Переменная t может принимать любые значения в зависимости от положения точки на прямой. Равенство выполняется для любой точки прямой и нарушается, если точка оказывается вне прямой. Следовательно, равенство выражает свойство, присущее исключительно точкам прямой, и может быть названо *векторным уравнением прямой*. При сдвиге множества точек прямой l на вектор $\boldsymbol{r}_0$ возникнет прямая, проходящая через точку M_0. Ее векторное уравнение имеет вид

$$\boldsymbol{r} = \boldsymbol{r}_0 + \boldsymbol{e}t.$$

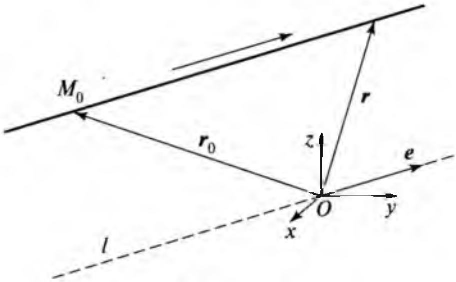

Рис. 10.8

Пусть $\boldsymbol{r} = (x, y, z)$, $\boldsymbol{r}_0 = (x_0, y_0, z_0)$, $\boldsymbol{e} = (x_1, y_1, z_1)$. Расположив координаты векторов по столбцам, получим

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}, \quad \text{где } t \in R.$$

Уравнение называется *уравнением прямой в координатах*. Запишем это равенство как три уравнения:

$$x = tx_1 + x_0, \quad y = ty_1 + y_0, \quad z = tz_1 + z_0.$$

Выразим из каждого уравнения переменную t и приравняем дроби. Получим

$$\frac{x - x_0}{x_1} = \frac{y - y_0}{y_1} = \frac{z - z_0}{z_1}.$$

Уравнение называется *каноническим уравнением прямой в координатах.*

Уравнение прямой можно записать также системой линейных уравнений. Из раздела «Системы линейных уравнений» нам известно, что равенство есть решение неоднородной системы из двух независимых уравнений с тремя переменными

$$\begin{cases} a_{11}x + a_{12}y + a_{13}z = b_1, \\ a_{21}x + a_{22}y + a_{23}z = b_2. \end{cases}$$

Чтобы раскрыть геометрическое содержание уравнений системы, рассмотрим задачу на составление уравнения плоскости.

Уравнение плоскости

Плоскость в трехмерном пространстве можно определить двумя пересекающимися векторами $\boldsymbol{e}_1 = (x_1, y_1, z_1)$, $\boldsymbol{e}_2 = (x_2, y_2, z_2)$ и точкой M_0, через которую проведена плоскость (рис. 10.9).

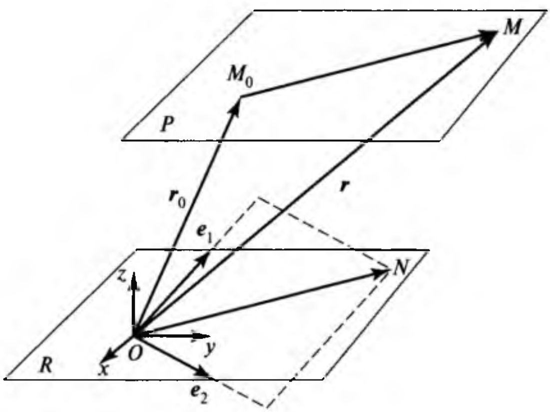

Рис. 10.9

Построим вспомогательную плоскость R на двух пересекающихся векторах $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, проходящую через начало координат. Любая точка N на плоскости может быть задана вектором $\overline{ON}$. Разложим вектор $\overline{ON}$ по правилу параллелограмма по векторам $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$:

$$\overline{ON} = c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2,$$

где c_1, c_2 — некоторые множители. Равенство выполняется для любой точки плоскости и нарушается, если точка оказывается вне плоскости R. Следовательно, равенство выражает свойство, присущее исключительно точкам плоскости, и может быть названо *векторным уравнением плоскости*. Плоскость P, проходящую через точку M_0, получим при сдвиге множества точек плоскости R на вектор $\boldsymbol{r}_0$. Векторное уравнение плоскости P имеет вид

$$\boldsymbol{r} = \boldsymbol{r}_0 + c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2.$$

Напишем уравнение в координатах:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c_1\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + c_2\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix},$$

где $c_1, c_2 \in R$.

Это равенство представляет решение неоднородного уравнения

$$ax + by + cz = d$$

и содержит фундаментальный набор решений однородного уравнения плюс частное решение неоднородного уравнения.

Можно сделать вывод о том, что всякая плоскость в трехмерном пространстве определяется неоднородным уравнением первой степени относительно трех переменных. Верно и обратное утверждение: всякое уравнение первой степени относительно трех переменных описывает плоскость в трехмерном пространстве.

Становится понятным геометрическое содержание уравнений в системе, определяющей линию в пространстве двумя уравнениями. Каждое уравнение описывает плоскость. Их пересечение определяет прямую.

Для составления уравнения плоскости можно использовать вектор, перпендикулярный плоскости. Пусть плоскость проходит через точку M_0, заданную вектором $\boldsymbol{r}_0 = (x_0, y_0, z_0)$ перпендикулярно вектору $\boldsymbol{n} = (a, b, c)$. Проведем вектор $\boldsymbol{r} = (x, y, z)$ в произвольную точку M этой плоскости. Для вектора $\boldsymbol{r} - \boldsymbol{r}_0 = \overline{M_0M}$, лежащего в рассматриваемой плоскости, выполнено условие $\overline{M_0M} \perp \boldsymbol{n}$. Поэтому скалярное произведение векторов равно нулю:

$$(\overline{M_0M}, \boldsymbol{n}) = (\boldsymbol{r} - \boldsymbol{r}_0, \boldsymbol{n}) = 0.$$

В координатной форме

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Раскроем скобки и обозначим $ax_0 + by_0 = cz_0 \equiv d$. Получим неоднородное уравнение первой степени относительно трех переменных, т.е. уравнение плоскости. Вектор $\boldsymbol{n}$ называется в этом случае нормальным.

З а м е ч а н и е 1. Линейные математические объекты трехмерного пространства — прямую и плоскость — можно задать векторным уравнением или системой линейных уравнений (линейным уравнением).

З а м е ч а н и е 2. Если прямая или плоскость проходит через начало координат, векторные уравнения не будут содержать вектора $\boldsymbol{r}_0 = (x_0, y_0, z_0)$, а линейные уравнения окажутся однородными.

N-мерные векторы

Обобщим понятие вектора: *n-мерным вектором* называется математический объект, который состоит из упорядоченной совокупности действительных чисел, называемых *координатами* вектора, и записываемый в виде $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. Название «*n*-мерный вектор» связано с тем, что при $n = 2$ или $n = 3$ совокупность чисел можно интерпретировать как совокупность координат вектора на плоскости или в пространстве.

З а м е ч а н и е. Числа $x_1, x_2, ..., x_n$ отсчитываются от начальной точки отсчета, принятой за нуль. Для геометрических векторов это означает, что начала векторов совпадают с началом координат.

Два *n*-мерных вектора $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ и $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ называются *равными*, если равны все компоненты векторов, т.е. $x_i = y_i$, где $i = 1, 2, ..., n$.

Вводя правила сложения и умножения для *n*-мерных векторов, мы должны уточнить, как следует производить эти действия над совокупностями *n* действительных чисел. Иначе говоря, введем операции над *n*-мерными векторами.

Суммой двух векторов $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ и $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ назовем вектор $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ такой, что $z = (z_1, z_2, ..., z_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$.

Произведением действительного числа λ на вектор $\boldsymbol{x}$ назовем вектор $\boldsymbol{y} = \lambda \cdot \boldsymbol{x}$ такой, что $\boldsymbol{y} = (y_1, y_2, ..., y_n) = (\lambda x_1, \lambda x_2, ..., \lambda x_n)$.

З а м е ч а н и е. Введенные по определению операции над *n*-мерными векторами аналогичны операциям над прямоугольными матрицами. Поэтому *n*-мерные векторы можно рассматривать как матрицы-строки $X = (x_1, x_2, ..., x_n)$ или как матрицы-столбцы $X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$ и совершать над векторами матричные операции.

Линейная зависимость и независимость векторов

Пусть каждый из векторов в наборе $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ есть n-мерный вектор.

О п р е д е л е н и е. Вектор $\boldsymbol{x}$ называется *линейной комбинацией* векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$, если найдутся такие действительные числа $\lambda_1, \lambda_2, \ldots, \lambda_m$, что

$$\boldsymbol{x} = \lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \ldots + \lambda_m\boldsymbol{a}_m.$$

Векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ называются *линейно зависимыми*, если существуют такие числа $\lambda_1, \lambda_2, \ldots, \lambda_m$, не все одновременно равные нулю, что

$$\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \ldots + \lambda_m\boldsymbol{a}_m = \boldsymbol{o}.$$

Если равенство выполняется только при $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, то векторы называются *линейно независимыми*.

Пример 10.1. Даны два неколлинеарных вектора. Доказать, что они линейно независимы.

Решение. Предположим иное: векторы $\boldsymbol{a}_1, \boldsymbol{a}_2$ линейно зависимы. Тогда существуют не равные одновременно нулю числа λ_1, λ_2 такие, что $\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 = \boldsymbol{o}$. Пусть для определенности $\lambda_1 \neq 0$. Разделив обе части равенства на $\lambda_1 \neq 0$, получаем $\boldsymbol{a}_1 = -\dfrac{\lambda_2\boldsymbol{a}_2}{\lambda_1}$. Значит, векторы коллинеарны, что противоречит условию.

Свойства линейной зависимости и независимости векторов

1. Если среди нескольких векторов (набора векторов) один из них есть линейная комбинация части остальных, то весь набор векторов линейно зависим.

2. Если среди набора векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ имеется нулевой вектор, то весь набор векторов линейно зависим.

3. Если векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ линейно независимы и существует вектор $\boldsymbol{x}$, являющийся линейной комбинацией векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$, т.е. $\boldsymbol{x} = \lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \ldots + \lambda_m\boldsymbol{a}_m$, то коэффициенты $\lambda_1, \lambda_2, \ldots, \lambda_m$ определяются по вектору $\boldsymbol{x}$ единственным образом.

Пример 10.2. Являются ли векторы $\boldsymbol{a}_1 = (1, -1, 2, 2)$, $\boldsymbol{a}_2 = (-1, -2, 3, 0)$, $\boldsymbol{a}_3 = (2, 1, -1, 2)$, $\boldsymbol{a}_4 = (1, 0, -2, -1)$ линейно зависимыми? Если да, найти всю совокупность значений коэффициентов, реализующих линейную зависимость.

Решение. Составим векторное равенство $\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \lambda_3\boldsymbol{a}_3 + \lambda_4\boldsymbol{a}_4 = \boldsymbol{o}$. Запишем его в матричном виде, представив векторы как матрицы-столбцы:

$$\lambda_1\begin{pmatrix}1\\-1\\2\\2\end{pmatrix}+\lambda_2\begin{pmatrix}-1\\-2\\3\\0\end{pmatrix}+\lambda_3\begin{pmatrix}2\\1\\-1\\2\end{pmatrix}+\lambda_4\begin{pmatrix}1\\0\\-2\\-1\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}.$$

Равенство есть система линейных однородных уравнений с четырьмя переменными. Составим матрицу из коэффициентов и определим ее ранг:

$$\begin{pmatrix}1&-1&2&1\\-1&-2&1&0\\2&3&-1&-2\\2&0&2&-1\end{pmatrix}\sim\begin{pmatrix}1&-1&2&1\\0&-3&3&1\\0&5&-5&-4\\0&2&-2&-3\end{pmatrix}\sim\begin{pmatrix}1&-1&2&1\\0&-3&3&1\\0&0&0&-7\\0&0&0&-7\end{pmatrix}\sim\begin{pmatrix}1&-1&2&1\\0&-3&3&1\\0&0&0&1\\0&0&0&0\end{pmatrix}.$$

Очевидно, ранг матрицы равен 3. Система имеет кроме нулевого решения $\lambda_1=\lambda_2=\ldots=\lambda_m=0$ бесконечное множество решений. Следовательно, векторы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, $\boldsymbol{a}_4$ линейно зависимы. Далее найдем структуру бесконечного множества решений, для чего продолжим элементарные преобразования со строками матрицы по методу Гаусса — Жордана:

$$\begin{pmatrix}1&-1&2&1\\0&-3&3&1\\0&0&0&1\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&0&1&0\\0&1&-1&0\\0&0&0&1\\0&0&0&0\end{pmatrix}.$$

Отсюда следует

$$\begin{pmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{pmatrix}=c\begin{pmatrix}-1\\1\\1\\0\end{pmatrix},\quad \text{где } c\in R.$$

Решение представляет всю совокупность значений коэффициентов λ_1, λ_2, λ_3, λ_4, реализующих линейную зависимость векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, $\boldsymbol{a}_4$. Например, при $c=1$ имеем $-\boldsymbol{a}_1+\boldsymbol{a}_2+\boldsymbol{a}_3+0\cdot\boldsymbol{a}_4=\boldsymbol{o}$. Среди векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, $\boldsymbol{a}_4$ три являются линейно независимыми, один — линейной комбинацией остальных. Линейно независимыми можно взять тройки векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_4$, или $\boldsymbol{a}_1$, $\boldsymbol{a}_3$, $\boldsymbol{a}_4$, или $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, $\boldsymbol{a}_4$, но нельзя взять $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, поскольку в этом случае вектор $\boldsymbol{a}_4$ невозможно через них выразить.

З а м е ч а н и е. Как видно из примера, вопрос о линейной зависимости векторов сводится к исследованию существования ненулевого решения у линейной однородной системы уравнений.

10.2. ВЕКТОРНОЕ ПРОСТРАНСТВО И ЕГО БАЗИС

Множество W элементов $x, y, z, \ldots$ называется *линейным пространством*, если по некоторому правилу:

а) любым двум элементам x и y из W поставлен в соответствие элемент из W, обозначаемый $x + y$ и называемый суммой элементов x и y;

б) любому элементу x из W и каждому числу λ поставлен в соответствие элемент из W, обозначаемый $\lambda \cdot x$ и называемый произведением числа λ на элемент x, причем справедливы следующие аксиомы:

1) $x + y = y + x$;
2) $(x + y) + z = x + (y + z)$;
3) $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$;
4) $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$;
5) $\lambda \cdot (\mu \cdot x) = (\lambda \cdot \mu) \cdot x$;
6) $1 \cdot x = x$;
7) существует нулевой элемент 0 такой, что $x + 0 = x$ для любого $x \in W$;
8) для каждого элемента x существует противоположный элемент $-x$ такой, что $x + (-x) = 0$.

Назовем эти два правила и восемь аксиом *законами и аксиомами композиции в линейной алгебре.* Элементы любой природы, удовлетворяющие всем этим условиям, по определению образуют линейное пространство. Например, совокупность любых матриц размерами $m \times n$ образует линейное пространство, поскольку для них выполнены оба правила и все аксиомы. Легко проверить, что совокупность геометрических векторов, например, трехмерного пространства также является линейным пространством.

Определение векторного пространства

Пусть для n-мерных векторов выполнены правила сложения и умножения на число, удовлетворяющие аксиомам линейного пространства. Тогда множество всех n-мерных векторов называется *линейным векторным пространством* и обозначается W. Векторное пространство называется *n-мерным*, если среди множества его векторов найдутся n линейно независимых векторов, а любые $n + 1$ векторов уже окажутся зависимыми. Число n называется *размерностью* векторного пространства.

Например, среди бесконечного множества векторов, расположенных в одной плоскости, любые два неколлинеарных вектора являются линейно независимыми. Выберем какие-либо два неколлинеарных вектора $\boldsymbol{a}_1 = (x_1, x_2)$ и $\boldsymbol{a}_2 = (y_1, y_2)$. Добавление третьего вектора $\boldsymbol{a}_3 = (z_1, z_2)$ к выбранным двум делает их линейно зависимыми. Действительно, система уравнений, полученная из векторного равенства $\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \lambda_3\boldsymbol{a}_3 = \mathbf{0}$, будет иметь вид

$$\lambda_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \lambda_2\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \lambda_3\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Матрица коэффициентов $\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix}$ имеет ранг, равный двум, откуда следует, что один из трех столбцов есть линейная комбинация двух других. Следовательно, размерность такого линейного векторного пространства равна двум.

Базис векторного пространства

Упорядоченная совокупность n линейно независимых векторов n-мерного векторного пространства называется *базисом* этого пространства.

Выбранные нами в рассмотренном выше примере в определенном порядке два неколлинеарных вектора составляют базис в двухмерном пространстве. Если векторы поменять местами, они также составят базис этого пространства, но другой. Если выбрать два других неколлинеарных вектора в определенном порядке, на них можно построить свой базис.

В общем случае пусть в n-мерном векторном пространстве нам известны m векторов ($m > n$). Количество способов выбора n линейно независимых векторов из общего числа заданных m не превышает величины

$$C_m^n = \frac{m!}{n!(m-n)!}.$$

Кроме того, выбрав n векторов, можно построить $n!$ упорядоченных совокупностей. Тогда количество базисов в таком n-мерном векторном пространстве может достигать величины

$$N = \frac{m!}{n!(m-n)!} \cdot n! = \frac{m!}{(m-n)!}.$$

Если в векторном пространстве определен базис, другие векторы могут быть выражены через этот базис.

Теорема 10.1. Каждый вектор линейного пространства можно представить и притом единственным образом в виде линейной комбинации векторов базиса.

Пусть векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ образуют базис в n-мерном векторном пространстве. Возьмем произвольный вектор $\boldsymbol{x}$ из этого пространства. Тогда совокупность векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n, \boldsymbol{x}$ линейно зависима, т.е. найдутся такие не равные одновременно нулю числа $\lambda_1, \ldots, \lambda_n, \lambda$, что

$$\lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \ldots \lambda_n\boldsymbol{a}_n + \lambda\boldsymbol{x} = \boldsymbol{o},$$

причем $\lambda \neq 0$. Если бы выполнялось равенство $\lambda = 0$, то хотя бы один из коэффициентов $\lambda_1, \lambda_2, \ldots, \lambda_n$ обязан быть не равным нулю. Но это противоречит определению линейной независимости векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

Разделим обе части уравнения на $\lambda \neq 0$. Получим

$$\boldsymbol{x} = -\frac{\lambda_1}{\lambda}\boldsymbol{a}_1 - \frac{\lambda_2}{\lambda}\boldsymbol{a}_2 - \ldots - \frac{\lambda_n}{\lambda}\boldsymbol{a}_n.$$

Равенство есть линейная комбинация векторов базиса

$$\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n = \sum_{i=1}^{n} x_i\boldsymbol{a}_i,$$

где $x_i = -\dfrac{\lambda_i}{\lambda}$ $i = 1, 2, \ldots, n$. Представление вектора $\boldsymbol{x}$ в виде линейной комбинации является единственным в силу свойства 3 линейной зависимости векторов.

Последнее равенство называется ***разложением вектора x по базису*** $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, а числа $x_1, x_2, \ldots, x_n$ — координатами вектора $\boldsymbol{x}$ в этом базисе. Считая компоненты вектора его координатами, можно представить вектор $\boldsymbol{x}$ набором своих координат $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Таким образом, упорядоченная совокупность действительных чисел (так определяется n-мерный вектор) ***есть набор координат в определенном базисе*** некоторого n-мерного пространства.

Пример 10.3. Векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4, \boldsymbol{a}_5, \boldsymbol{a}_6$ заданы совокупностями действительных чисел:

$$\boldsymbol{a}_1 = (1, 0), \quad \boldsymbol{a}_2 = (0, 1), \quad \boldsymbol{a}_3 = (1, 1), \quad \boldsymbol{a}_4 = (2, 1), \quad \boldsymbol{a}_5 = (1, 2), \quad \boldsymbol{a}_6 = (2, 2).$$

1. Найти размерность и базис линейного пространства, в котором заданы векторы.
2. Задав базис, разложить остальные векторы по этому базису.

Решение. Составим матрицу из координат всех векторов:

$$\begin{pmatrix} 1 & 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 1 & 1 & 2 & 2 \end{pmatrix}.$$

1. Ранг матрицы равен двум. Только два линейно независимых столбца содержатся в этой матрице. Остальные столбцы могут быть представлены как линейные комбинации двух выбранных линейно независимых столбцов. Следовательно, размерность пространства равна двум. Базисными векторами можно выбрать векторы $\boldsymbol{a}_1 = (1, 0)$, $\boldsymbol{a}_2 = (0, 1)$, расположив их в следующем порядке: $\boldsymbol{a}_1, \boldsymbol{a}_2$.

2. Разложим векторы $\boldsymbol{a}_3 = (1, 1)$, $\boldsymbol{a}_4 = (2, 1)$, $\boldsymbol{a}_5 = (1, 2)$, $\boldsymbol{a}_6 = (2, 2)$ по выбранному базису $\boldsymbol{a}_1, \boldsymbol{a}_2$.

Начнем с векторов $\boldsymbol{a}_3$ и $\boldsymbol{a}_4$:

$$\boldsymbol{a}_3 = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2, \quad \boldsymbol{a}_4 = y_1\boldsymbol{a}_1 + y_2\boldsymbol{a}_2.$$

Представим векторы равенства как матрицы-столбцы:

$$\begin{pmatrix}1\\1\end{pmatrix} = x_1\begin{pmatrix}1\\0\end{pmatrix} + x_2\begin{pmatrix}0\\1\end{pmatrix}, \quad \begin{pmatrix}2\\1\end{pmatrix} = y_1\begin{pmatrix}1\\0\end{pmatrix} + y_2\begin{pmatrix}0\\1\end{pmatrix}.$$

В матричной форме

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}, \quad \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix}$$

или в сокращенной матричной форме

$$A \cdot X = B$$

решение матричного уравнения имеет вид

$$X = A^{-1} \cdot B$$

или

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}^{-1} \cdot \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}, \quad \begin{pmatrix}y_1\\y_2\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}^{-1} \cdot \begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix}.$$

Следовательно,

$$\boldsymbol{a}_3 = \boldsymbol{a}_1 + \boldsymbol{a}_2, \quad \boldsymbol{a}_4 = 2\boldsymbol{a}_1 + \boldsymbol{a}_2.$$

Действуя аналогично, для векторов $\boldsymbol{a}_5$ и $\boldsymbol{a}_6$ получим разложения $\boldsymbol{a}_5 = \boldsymbol{a}_1 + 2\boldsymbol{a}_2$, $\boldsymbol{a}_6 = 2\boldsymbol{a}_1 + 2\boldsymbol{a}_2$.

Переход к новому базису

Пусть $A = (\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n)$ и $B = (\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_n)$ — старый и новый базисы линейного n-мерного векторного пространства W. Каждый из векторов нового базиса можно выразить через векторы старого базиса:

$$\begin{cases} \boldsymbol{e}'_1 = a_{11}\boldsymbol{e}_1 + a_{21}\boldsymbol{e}_2 + \ldots + a_{n1}\boldsymbol{e}_n, \\ \boldsymbol{e}'_2 = a_{12}\boldsymbol{e}_1 + a_{22}\boldsymbol{e}_2 + \ldots + a_{n2}\boldsymbol{e}_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ \boldsymbol{e}'_n = a_{1n}\boldsymbol{e}_1 + a_{2n}\boldsymbol{e}_2 + \ldots + a_{nn}\boldsymbol{e}_n, \end{cases}$$

а также представить в матричной форме:

$$(e'_1 \quad e'_2 \quad \dots \quad e'_n) = (e_1 \quad e_2 \quad \dots \quad e_n)\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

или в сокращенной матричной форме

$$B = A \cdot T,$$

где матрица A есть матрица координат старого базиса, расположенных по столбцам; B — матрица координат нового базиса, также расположенного по столбцам.

Матрица

$$T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & .. & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & u_{n2} & \dots & a_{nn} \end{pmatrix}$$

называется *матрицей перехода* от старого базиса e к новому базису e'.

Свойства матрицы перехода

1. Матрица перехода является невырожденной, т.е. $|T| \neq 0$.
2. Матрица перехода от нового базиса к старому имеет вид T^{-1}.
3. Координаты вектора в разных базисах связаны матрицей перехода. Пусть некоторый вектор x имеет координаты $(x_1, x_2, ..., x_n)$ в старом базисе и координаты $(x'_1, x'_2, ..., x'_n)$ в новом базисе. Тогда

$$x = x'_1e'_1 + x'_2e'_2 + ... + x'_ne'_n = x'_1e'_1 + x'_2e'_2 + ... + x'_ne'_n.$$

Подставив в это выражение разложения векторов (e'_1, e'_2, e'_n) по базису (e_1, e_2, e_n), получим

$$x_1e_1 + x_2e_2 + ... + x_ne_n = x'_1(a_{11}e_1 + a_{21}e_2 + ... + a_{n1}e_n) + x'_2(a_{12}e_1 +$$
$$+ a_{22}e_2 + ... + a_{n2}e_n) + ... + x'_n(a_{1n}e_1 + a_{2n}e_2 + ... + a_{nn}e_n).$$

Перенесем все влево и сгруппируем слагаемые с одинаковыми сомножителями e_i:

$$(x_1 - x'_1a_{11} - x'_2a_{12} - ... - x'_na_{1n})e_1 + (x_2 - x'_1a_{21} - x'_2a_{22} - ... - x'_na_{2n})e_2 +$$
$$+ (x_n - x'_1a_{n1} - x'_2a_{n2} - ... - x'_na_{nn})e_n = o.$$

Это равенство выполняется при условии, что все коэффициенты перед e_i равны нулю. Следовательно,

$$\begin{cases} x_1 = x'_1a_{11} + x'_2a_{12} + ... + x'_na_{1n}, \\ x_2 = x'_1a_{21} + x'_2a_{22} + ... + x'_na_{2n}, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ x_n = x'_1a_{n1} + x'_2a_{n2} + ... + x'_na_{nn}. \end{cases}$$

В матричной форме

$$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & .. & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ \dots \\ x'_n \end{pmatrix} \quad \text{или} \quad X = T \cdot X'.$$

Координаты вектора $\boldsymbol{x}$ в новом базисе выражаются через координаты вектора в старом базисе:

$X = T^{-1} \cdot X.$

В развернутой матричной форме

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \dots \\ x'_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & .. & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}^{-1} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

Пример 10.4. Векторы $\boldsymbol{x} = (1, 3, -2)$, $\boldsymbol{e}'_1 = (1, 1, 0)$, $\boldsymbol{e}'_2 = (1, 0, 1)$, $\boldsymbol{e}'_3 = (0, 1, 1)$ заданы своими координатами в старом базисе $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$. Выразить координаты вектора $\boldsymbol{x}$ в новом базисе $\boldsymbol{e}'_1$, $\boldsymbol{e}'_2$, $\boldsymbol{e}'_3$.

Решение. Матрица перехода от старого базиса к новому базису имеет вид

$$T = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Вычисляется обратная матрица:

$$T^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

Координаты вектора $\boldsymbol{x}$ в новом базисе:

$$X' = \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = T^{-1} \cdot X = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 0 \end{pmatrix}.$$

10.3. ЛИНЕЙНЫЕ ПОДПРОСТРАНСТВА

Из множества векторов линейного пространства W выберем некоторую совокупность векторов и обозначим ее V. Пусть для любых векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ из V и любого числа $\lambda \in R$ выполняются следующие условия:

1) $x + y \in V$;

2) $\lambda \cdot x \in V$.

Тогда множество векторов V называется *линейным подпространством* пространства W.

Примеры линейных подпространств

1. Каждое линейное пространство обладает двумя подпространствами: нулевым подпространством и самим пространством. Эти подпространства называют тривиальными.

2. Линейное пространство W^1 векторов на прямой, проходящей через начало координат, имеет два тривиальных подпространства.

3. Линейное пространство W^2 векторов на плоскости (рис. 10.10) имеет, кроме двух тривиальных подпространств, бесконечное множество подпространств V_1, V_2, ... Каждое из них состоит из векторов, которые лежат на прямой, проходящей через начало координат (предполагается, что все векторы отложены от начала координат).

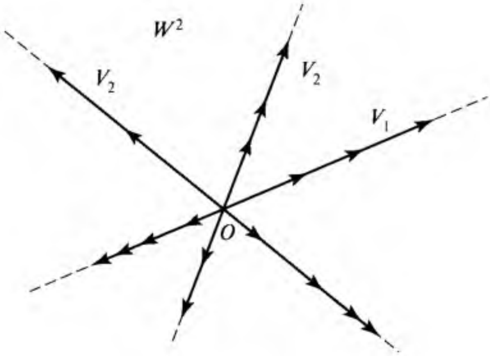

Рис. 10.10

4. В геометрическом пространстве W^3 векторов пространства каждая прямая и каждая плоскость, проходящие через начало координат, определяют линейное подпространство.

Способы задания линейных подпространств

Линейное векторное подпространство задается двумя возможными способами: набором векторов или системами линейных уравнений.

Первый способ. Набор линейно независимых векторов (базис подпространства) тесно связан с понятием линейной оболочки системы векторов.

О п р е д е л е н и е. *Линейной оболочкой* $L(\boldsymbol{x}_1, \boldsymbol{x}_2)$ двух векторов $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$, принадлежащих линейному пространству W, называется совокупность всех линейных комбинаций этих векторов:

$$\boldsymbol{a} = \alpha \cdot \boldsymbol{x}_1 + \beta \cdot \boldsymbol{x}_2, \quad \text{где } \alpha, \beta \in R.$$

Иначе говоря, линейная оболочка состоит из бесконечного множества векторов $\boldsymbol{a}$, представимых в виде линейных комбинаций векторов $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$. На рис. 10.11 построены векторы $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$, а также приведено несколько их линейных комбинаций $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_5$.

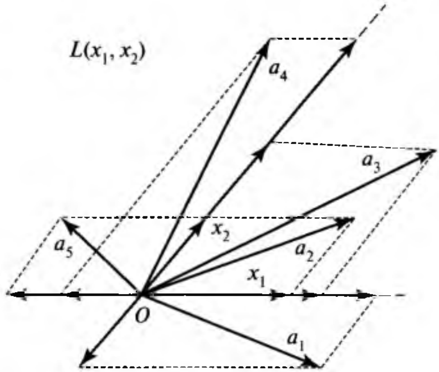

Рис. 10.11

Из определения и свойств линейной оболочки следует, что каждое векторное пространство есть линейная оболочка векторов своего базиса. Поэтому часто в задачах в целях экономии места мы, разыскивая базис подпространства, будем ограничиваться перечислением векторов базиса, не записывая конечным результатом их произвольную линейную комбинацию.

Второй способ. Существует еще один способ задания подпространства — в виде однородной системы линейных уравнений. Рассмотрим однородную линейную систему m уравнений с n переменными

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0, \end{cases}$$

имеющую ненулевые решения. Пусть ранг системы равен r. Она обладает фундаментальным набором решений (ФНР)

$$e_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \dots \\ x_{r1} \\ 1 \\ \dots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} x_{12} \\ x_{22} \\ \dots \\ x_{r2} \\ 0 \\ \dots \\ 0 \end{pmatrix}, \quad \dots, \quad e_{n-r} = \begin{pmatrix} x_{1,n-r} \\ x_{2,n-r} \\ \dots \\ x_{r,n-r} \\ 0 \\ \dots \\ 1 \end{pmatrix},$$

которые линейно независимы. Эти независимые решения, являющиеся совокупностями из n чисел, можно представить как n-мерные линейно независимые векторы. Любое решение системы представляется в виде линейной комбинации ФНР. Если взять векторы $e_1, e_2, \dots, e_{n-r}$ в качестве базиса некоторого линейного векторного подпространства, то все множество решений однородной системы и будет этим векторным *подпространством*, называемым пространством решений однородной системы. Размерность подпространства равна числу независимых векторов, т.е. $n - r$.

Таким образом, векторное подпространство может быть задано как набором векторов, составляющих базис векторного подпространства, так и посредством задания однородной системы линейных уравнений, фундаментальный набор решений которой есть базис линейного векторного подпространства. Переход от задания подпространства в виде набора векторов к заданию в виде однородной системы уравнений и обратно достаточно прост.

Пример 10.5. Линейное подпространство V задано набором линейно независимых векторов $a_1 = (1, -1, 1, 0)$, $a_2 = (1, 1, 0, 1)$. Найти однородную систему линейных уравнений, задающую подпространство V.

Решение. Рассмотрим два способа решения задачи. Введем произвольный вектор $x = (x_1, x_2, x_3, x_4)$, принадлежащий подпространству V. Разложим вектор x по векторам базиса:

$$x = \lambda_1 a_1 + \lambda_2 a_2$$

или в координатном виде:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Первый способ. Использование формы, в которой записывается решение системы однородных уравнений, и представление в этой форме векторов базиса как фундаментальных решений некоторой системы. Запишем равенство в виде решения системы уравнений

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Тогда система имеет вид

$$\begin{cases} x_1 = x_3 + x_4, \\ x_2 = -x_3 + x_4, \\ x_3 = x_3, \\ x_4 = x_4. \end{cases}$$

Окончательно

$$\begin{cases} x_1 = x_3 + x_4, \\ x_2 = -x_3 + x_4. \end{cases}$$

З а м е ч а н и е. Переход к системе уравнений требует наличия в каждом из векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ нулевой координаты. Если ее нет, комбинируя векторы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, такие координаты легко получить.

Второй способ основан на использовании теоремы Кронекера—Капелли. Составим расширенную матрицу из коэффициентов и свободных членов и преобразуем, используя метод Гаусса:

$$\left(\begin{array}{cc|c} 1 & 1 & x_1 \\ -1 & 1 & x_2 \\ 1 & 0 & x_3 \\ 0 & 1 & x_4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & x_1 \\ 0 & 2 & x_1 + x_2 \\ 0 & -1 & -x_1 + x_3 \\ 0 & 1 & x_4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & x_1 \\ 0 & -1 & -x_1 + x_3 \\ 0 & 0 & -2x_1 + 2x_3 + x_1 + x_2 \\ 0 & 0 & -x_1 + x_3 + x_4 \end{array}\right).$$

Система должна иметь решения, поскольку вектор $\boldsymbol{x}$ принадлежит подпространству. Ранги матрицы коэффициентов и расширенной матрицы системы по теореме Кронекера—Капелли должны быть равны. Это выполняется при соблюдении условий

$$\begin{cases} -x_1 + x_2 + 2x_3 = 0, \\ -x_1 + x_3 + x_4 = 0. \end{cases}$$

Полученная система однородных линейных уравнений задает требуемое линейное подпространство. Системы, полученные способами 1 и 2, несколько отличаются друг от друга. От второй системы можно перейти к первой, взяв разность первого и второго уравнений.

З а м е ч а н и е. Продолжив преобразование матрицы по методу Гаусса—Жордана, получим координаты λ_1 и λ_2 вектора $\boldsymbol{x}$ в базисе $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ линейного подпространства V:

$$\left(\begin{array}{cc|c} 1 & 0 & x_3 \\ 0 & 1 & x_1 - x_3 \end{array}\right),$$

т.е. $\lambda_1 = x_3, \quad \lambda_2 = x_1 - x_3.$

Пример 10.6. Линейное подпространство V задано однородной системой линейных уравнений

$$\begin{cases} -x_1 + x_2 + 2x_3 = 0, \\ -x_1 + x_3 + x_4 = 0. \end{cases}$$

Найти набор линейно независимых векторов (базис), задающий линейное подпространство V.

Решение. Найдем фундаментальный набор решений однородной системы. Ранг матрицы коэффициентов уравнений равен 2. Поэтому могут быть найдены две переменные, выраженные через две другие. Положим базисными переменными x_1, x_2. Свободными переменными станут x_3, x_4. Вычтем из первого уравнения второе. Будем иметь

$$\begin{cases} x_1 = x_3 + x_4, \\ x_2 = -x_3 + x_4. \end{cases}$$

Запишем решения в развернутой матричной форме:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Обозначив свободные переменные, стоящие в правой части равенства, $x_3 \equiv \lambda_1$, $x_4 \equiv \lambda_2$, получим

$$\boldsymbol{x} = \lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2,$$

где

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad \boldsymbol{a}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{a}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Любой вектор $\boldsymbol{x}$, координаты которого являются переменными в однородной системе уравнений, представлен как линейная комбинация двух линейно независимых векторов, составляющих ФНР однородной системы. Следовательно, все множество векторов $\boldsymbol{x}$ составляет линейное векторное подпространство с базисом $\boldsymbol{a}_1$, $\boldsymbol{a}_2$. Итак, базис линейного подпространства V: $\boldsymbol{a}_1 = (1, -1, 1, 0)$, $\boldsymbol{a}_2 = (1, 1, 0, 1)$.

Операции с линейными подпространствами

1. Сумма подпространств. *Суммой* $V_1 + V_2$ линейных подпространств V_1 и V_2 линейного пространства W называется совокупность всех векторов $\boldsymbol{a} \in W$, которые можно представить в виде (разложить)

$$\boldsymbol{a} = \boldsymbol{x} + \boldsymbol{y}, \quad \text{где } \boldsymbol{x} \in V_1, \boldsymbol{y} \in V_2.$$

2. Пересечение подпространств. *Пересечением* $V_1 \cap V_2$ линейных подпространств V_1 и V_2 линейного пространства W называется совокупность всех векторов $\boldsymbol{b}$, которые принадлежат одновременно подпространствам V_1 и V_2. На рис. 10.12 пересечению подпространств V_1 и V_2 геометрического пространства W^3 принадлежат векторы $\boldsymbol{b}_1$ и $\boldsymbol{b}_2$.

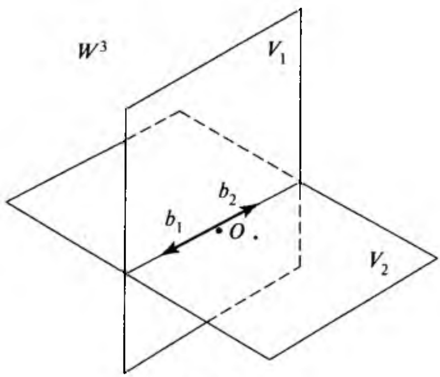

Рис. 10.12

3. Умножение числа на подпространство. Умножение числа $\alpha \neq 0$ на линейное векторное подпространство $V(\boldsymbol{x})$ не изменяет его, т.е. $\alpha \cdot V(\boldsymbol{x}) = V(\boldsymbol{x})$.

4. Сумма подпространства и вектора. Алгебраическая сумма векторного подпространства $V(\boldsymbol{x})$ и отдельного вектора $\boldsymbol{x}_0$, принадлежащего подпространству $V(\boldsymbol{x})$, не изменяет последнего:

$$V(\boldsymbol{x}) + \boldsymbol{x}_0 = V(\boldsymbol{x}), \quad \text{если } \boldsymbol{x}_0 \in V(\boldsymbol{x}).$$

Свойства суммы и пересечения линейных подпространств

1. Сумма и пересечение линейных подпространств являются линейными подпространствами.

2. Суммой двух подпространств $V(\boldsymbol{x})$ и $V(\boldsymbol{x})$ является подпространство V, т.е. $V(\boldsymbol{x}) + V(\boldsymbol{x}) = V(\boldsymbol{x})$.

3. Размерность суммы линейных подпространств равна сумме размерностей подпространств минус размерность их пересечения (формула Грассмана):

$$\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2).$$

Пусть линейные подпространства задаются своими базисами или в виде систем линейных уравнений. Тогда без труда решается задача нахождения базиса или системы уравнений, задающих сумму подпространств или их пересечение. Идеи решений таких задач оформим в виде таблицы.

<table>
<tr><td colspan="3">Задание подпространств V_1 и V_2</td></tr>
<tr><td></td><td>1. своими базисами
V_1: $a_1, a_2, \ldots, a_n$;
V_2: $b_1, b_2, \ldots, b_m$</td><td>2. системами линейных уравнений
V_1: $A \cdot X = O$;
V_2: $B \cdot X = O$</td></tr>
<tr><td colspan="3">Идеи решения</td></tr>
<tr><td>$V_1 + V_2$</td><td>3. Находится ранг системы векторов $\begin{pmatrix} a_1 \\ \ldots \\ a_n \\ b_1 \\ \ldots \\ b_m \end{pmatrix}$, определяются все линейно независимые векторы, которые составляют базис суммы</td><td>4. Находится объединение систем:
$\left[\begin{array}{l} A \cdot X = O, \\ B \cdot X = O \end{array}\right.$</td></tr>
<tr><td>$V_1 \cap V_2$</td><td>5. Используется равенство линейных комбинаций:
$\lambda_1 a_1 + \ldots + \lambda_n a_n = \eta_1 b_1 + \ldots + \eta_m b_m$</td><td>6. Находится пересечение систем:
$\begin{cases} A \cdot X = O, \\ B \cdot X = O \end{cases}$</td></tr>
</table>

Пример 10.7. Линейные подпространства V_1 и V_2 трехмерного векторного пространства W заданы своими базисами векторов:

V_1: $a_1 = (1, 0, 1), a_2 = (1, 1, 1)$;

V_2: $b_1 = (1, -1, -1), b_2 = (0, 0, 1)$.

Найти:

1) системы линейных уравнений, задающие эти подпространства (переход $1 \to 2$);

2) базис $V_1 + V_2$ (переход $1 \to 3$);

3) систему уравнений, задающую $V_1 + V_2$ (переход $1 \to 4$ или $2 \to 4$);

4) базис $V_1 \cap V_2$ (переход $1 \to 5$ или $2 \to 5$);

5) систему уравнений, задающую $V_1 \cap V_2$ (переход $1 \to 6$ или $2 \to 6$).

Решение. 1. Переход 1 → 2 нами был разобран в предыдущем примере. *Ответ*: V_1: $-x_1 + x_3 = 0$; V_2: $x_1 + x_2 = 0$.

2. Составим матрицу из координат векторов и приведем ее к треугольному виду:

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Очевидно, ранг равен 3. Проводя элементарные преобразования, мы не меняли положение строк. Первые три вектора, написанные в координатах по строкам матрицы, являются линейно независимыми и могут составить базис суммы подпространств. Итак, векторы $\boldsymbol{a}_1 = (1, 0, 1)$, $\boldsymbol{a}_2 = (1, 1, 1)$, $\boldsymbol{b}_1 = (1, -1, -1)$ задают подпространство $V_1 + V_2$. Размерности пространств $V_1 + V_2$ и W совпадают. Следовательно, векторы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}_1$ образуют базис всего пространства W.

3. Сумма подпространств $V_1 + V_2$ имеет размерность векторного пространства W и системой уравнений описана быть не может. Для обоснования утверждения рассмотрим произвольный вектор $\boldsymbol{x} = (x_1, x_2, x_3)$, принадлежащий пространству $V_1 + V_2 = W$. Вектор $\boldsymbol{x}$ раскладывается по базису $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}_1$ единственным образом:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1'\boldsymbol{a}_1 + x_2'\boldsymbol{a}_2 + x_3'\boldsymbol{b}_1.$$

Составим расширенную матрицу и воспользуемся преобразованием Гаусса—Жордана:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -x_2 + x_3 \\ 0 & 1 & 0 & 0{,}5(x_1 + 2x_2 - x_3) \\ 0 & 0 & 1 & 0{,}5(x_1 - x_3) \end{array}\right).$$

Получим выражение

$x_1' = -x_2 + x_3,$

$x_2' = 0{,}5(x_1 + 2x_2 - x_3),$

$x_3' = 0{,}5(x_1 - x_3),$

представляющее собой связь координат произвольного вектора $\boldsymbol{x}$ в новом и старом базисах. Таким образом, задача о нахождении системы линейных уравнений, описывающей векторное подпространство, вырождается в задачу нахождения связи координат произвольного вектора в старом и новом базисах.

4. Базис пересечения подпространств $V_1 \cap V_2$. Любой вектор $\boldsymbol{x}$, принадлежащий подпространствам V_1 и V_2, может быть разложен по базисам этих подпространств:

$$\lambda_1\begin{pmatrix}1\\0\\1\end{pmatrix}+\lambda_2\begin{pmatrix}1\\1\\1\end{pmatrix}=\eta_1\begin{pmatrix}1\\-1\\-1\end{pmatrix}+\eta_2\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Решение системы уравнений по методу Гаусса—Жордана

$$\begin{pmatrix}1&1&1&0\\0&1&-1&0\\1&1&-1&1\end{pmatrix}\sim\begin{pmatrix}1&0&0&1\\0&1&0&-1/2\\0&0&1&-1/2\end{pmatrix}$$

дает следующие значения переменных:

$$\lambda_1=\eta_2,\ \ \lambda_2=-\frac{1}{2}\eta_2,\ \ \eta_1=\frac{1}{2}\eta_2.$$

Подставим их в разложение вектора $\boldsymbol{x}$:

$$\eta_2\begin{pmatrix}1\\0\\1\end{pmatrix}-\frac{1}{2}\eta_2\begin{pmatrix}1\\1\\1\end{pmatrix}=\frac{1}{2}\eta_2\begin{pmatrix}1\\-1\\-1\end{pmatrix}+\eta_2\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

При $\eta_2\neq 0$ получим $\boldsymbol{a}_1-\frac{1}{2}\boldsymbol{a}_2=\frac{1}{2}\boldsymbol{b}_1+\boldsymbol{b}_2$ или $2\boldsymbol{a}_1-\boldsymbol{a}_2=\boldsymbol{b}_1+2\boldsymbol{b}_2$. Следовательно, вектор $\boldsymbol{c}=2\boldsymbol{a}_1-\boldsymbol{a}_2=\boldsymbol{b}_1+2\boldsymbol{b}_2$ является общим для подпространств V_1 и V_2 и может быть положен базисом их пересечения. *Ответ*: базис пересечения $V_1\cap V_2$ имеет вид $\boldsymbol{c}=\begin{pmatrix}1\\-1\\1\end{pmatrix}$.

5. Найдем систему уравнений, задающую пересечение подпространств. Ее вид $\begin{cases}-x_1+x_3=0,\\x_1+x_2=0.\end{cases}$ Легко видеть, что фундаментальное решение системы $\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=x_1\begin{pmatrix}1\\-1\\1\end{pmatrix}$ совпадает с вектором $\boldsymbol{c}$ базиса пересечения подпространств.

10.4. ЕВКЛИДОВЫ ПРОСТРАНСТВА

Введенное нами линейное векторное пространство не содержит информации о том, как измерять длины и углы в этом пространстве. Неожиданным, на первый взгляд, является возможность решить этот вопрос, если ввести понятие скалярного произведения.

Пусть любым двум векторам $\boldsymbol{x}$ и $\boldsymbol{y}$ из W ставится в соответствие число, обозначаемое как $(\boldsymbol{x},\boldsymbol{y})$, причем выполняются следующие условия:

1) $(\boldsymbol{x},\boldsymbol{y})=(\boldsymbol{y},\boldsymbol{x})$;

2) $(\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}) = (\boldsymbol{x}, \boldsymbol{z}) + (\boldsymbol{y}, \boldsymbol{z})$;

3) $(\lambda \cdot \boldsymbol{x}, \boldsymbol{y}) = \lambda \cdot (\boldsymbol{x}, \boldsymbol{y})$, где $\lambda \in R$;

4) $(\boldsymbol{x}, \boldsymbol{x}) > 0$, если $\boldsymbol{x}$ — ненулевой вектор; $(\boldsymbol{x}, \boldsymbol{x}) = 0$, если $\boldsymbol{x}$ — нулевой вектор.

Назовем это число *скалярным произведением.*

Пусть векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ задаются набором своих координат в некотором базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$:

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i, \quad \boldsymbol{y} = \sum_{j=1}^{n} y_j \boldsymbol{e}_j.$$

Подстановка этих разложений в скалярное произведение с последующим применением аксиом 2) и 3) дает формулу

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j (\boldsymbol{e}_i, \boldsymbol{e}_j).$$

Скалярное произведение лежит в основе определения евклидова пространства и часто используется. В связи с этим желательно базисы в евклидовом пространстве выбирать так, чтобы процедура вычисления скалярного произведения была простой. Наиболее удачный вариант реализуется, если векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n$ удастся подобрать исходя из условия

$$(\boldsymbol{e}_i, \boldsymbol{e}_j) = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

Скалярное произведение становится равным

$$(\boldsymbol{x}, \boldsymbol{y}) = x_1 y_1 + x_2 y_2 + ... + x_n y_n.$$

Теперь появляется возможность определить *длины и углы* в пространстве. Введем модуль вектора $\boldsymbol{a}$ в векторном пространстве W:

$$|\boldsymbol{a}| = \sqrt{(\boldsymbol{a}, \boldsymbol{a})} = \sqrt{a_1^2 + a_2^2 + ... + a_n^2}.$$

Пусть $\boldsymbol{a} = \boldsymbol{x} - \boldsymbol{y}$, тогда

$$|\boldsymbol{x} - \boldsymbol{y}| = \sqrt{\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y}} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}.$$

В частном случае расстояние между вектором $\boldsymbol{x}$ и нулевым вектором $\boldsymbol{y} = \boldsymbol{o} = (0, 0, ..., 0)$ определяется как норма вектора $\boldsymbol{x}$:

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \rho(\boldsymbol{x}, \boldsymbol{o}) = |\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

Введение скалярного произведения векторов в векторном пространстве W с дополнительным указанием записи его в координатной форме и определением нормы отдельного вектора приводит к появлению пространства, названного *евклидовым*. Будем обозначать его через E.

Норма вектора в евклидовом пространстве для $n = 2$ и $n = 3$ имеет ясный геометрический смысл. Она совпадает с длиной вектора и может быть выражена тем числом, которое мы получаем, приложив линейку к вектору.

Евклидово пространство в трехмерном случае совпадает с физическим пространством. Евклидовым линейное пространство названо по имени древнегреческого математика Евклида, создавшего в III в. до н.э. евклидову геометрию — «первое приближение для описания структуры реального физического пространства».

Введенный нами математический аппарат позволяет описывать евклидовы пространства различных размерностей, рассматривая наше физическое пространство как частный случай и философски осмысливая тот факт, что природа, создавая физический мир, остановилась на цифре три.

Углом между ненулевыми векторами $\boldsymbol{x}$ и $\boldsymbol{y}$ евклидова пространства называется число φ, определяемое из равенства

$$\cos\varphi = \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}, \quad \text{где } 0 \le \varphi \le \pi.$$

Два ненулевых вектора $\boldsymbol{x}$ и $\boldsymbol{y}$ называются *ортогональными*, если их скалярное произведение равно нулю, т. е. $(\boldsymbol{x}, \boldsymbol{y}) = 0$. Из равенства $(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x}| \cdot |\boldsymbol{y}| \cdot \cos\varphi = 0$ следует, что $\cos\varphi = 0$, или угол $\varphi = \frac{\pi}{2}$. Для двух- или трехмерного пространства ортогональность векторов означает, что они взаимно перпендикулярны.

Система векторов $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ называется *ортогональной*, если $(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0$ при $i \ne j$, и *нормированной*, если $|\boldsymbol{e}_i| = 1$ для всех $i = 1, 2, \ldots, n$. Если векторы системы ортогональны и нормированны, они называются *ортонормированными*.

Замечание. Чтобы нормировать ненулевой вектор, необходимо разделить его на норму. Пусть задан вектор $\boldsymbol{x} = (1, -1, 2, 0)$. Его норма $|\boldsymbol{x}| = \sqrt{1^2 + (-1)^2 + 2^2} = \sqrt{6}$. Нормированный вектор имеет вид $\boldsymbol{e} = \frac{\boldsymbol{x}}{|\boldsymbol{x}|} = \left(\frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0\right)$, откуда $|\boldsymbol{e}| = 1$.

Теорема 10.2. Ортонормированная система векторов линейно независима.

Докажем, что ортогональные и нормированные векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ линейно независимы, т.е. докажем, что равенство

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n = o$$

справедливо лишь при $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. Умножив обе части равенства скалярно на вектор e_1, получим

$$\lambda_1(e_1, e_1) + \lambda_2(e_1, e_2) + \ldots + \lambda_n(e_1, e_n) = (e_1, o)$$

или

$$\lambda_1 \cdot 1 + \lambda_2 \cdot 0 + \ldots + \lambda_n \cdot 0 = 0.$$

Отсюда следует $\lambda_1 = 0$. Умножая последовательно равенство скалярно на $e_2, e_3, \ldots, e_n$, будем иметь $\lambda_2 = \ldots = \lambda_n = 0$.

Теорема 10.3. Во всяком n-мерном евклидовом пространстве существует ортонормированный базис.

Доказательство теоремы представляет собой алгоритм последовательного построения ортонормированного базиса по заданному базису $f_1, f_2, \ldots, f_n$, названный *методом ортогонализации* или *процедурой Грама—Шмидта.*

Положим вектор $g_1 = f_1$ и нормируем его: $e_1 = \dfrac{g_1}{|g_1|}$, получив первый вектор ортонормированного базиса. Построим вектор

$$g_2 = f_2 - \alpha_1 \cdot e_1$$

так, чтобы он был ортогонален вектору e_1. Должно выполняться условие $(g_2, e_1) = 0$. Из этого условия найдем α_1. Умножив скалярно равенство на e_1, получим

$$(g_2, e_1) = (f_2 - \alpha_1 \cdot e_1, e_1) = (f_2, e_1) - \alpha_1(e_1, e_1) = 0,$$

откуда $\alpha_1 = (f_2, e_1)$. Тем самым вектор $g_2 = f_2 - (f_2, e_1) \cdot e_1$ станет ортогональным вектору e_1. Тогда вторым вектором ортонормированного базиса станет вектор $e_2 = \dfrac{g_2}{|g_2|}$.

Пользуясь найденными векторами e_1, e_2 и заданным вектором f_3, построим вектор

$$g_3 = f_3 - \beta_1 e_1 - \beta_2 e_2,$$

ортогональный единичным векторам e_1 и e_2, для чего умножим скалярно равенство последовательно на e_1 и e_2 и приравняем нулю:

$$(g_3, e_1) = (f_3 - \beta_1 e_1 - \beta_2 e_2, e_1) = (f_3, e_1) - \beta_1(e_1, e_1) - \beta_2(e_2, e_1) = 0;$$

$$(g_3, e_2) = (f_3 - \beta_1 e_1 - \beta_2 e_2, e_2) = (f_3, e_2) - \beta_1(e_1, e_2) - \beta_2(e_2, e_2) = 0.$$

Поскольку $(e_2, e_1) = 0$, $(e_1, e_1) = (e_2, e_2) = 1$, получим $\beta_1 = (f_3, e_1)$, $\beta_2 = (f_3, e_2)$. Теперь вычислим вектор

$$g_3 = f_3 - (f_3, e_1)e_1 - (f_3, e_2)e_2.$$

Затем нормируем его, сделав третьим вектором ортонормированного базиса $e_3 = \frac{g_3}{|g_3|}$. Продолжая процесс ортогонализации, по заданному базису $f_1, f_2, ..., f_n$ построим ортонормированный базис (или ортобазис) $e_1, e_2, ..., e_n$.

Пример 10.8. Методом ортогонализации построить ортонормированный базис по базису евклидова пространства $f_1 = (1, 1), f_2 = (1, 2)$.

Положим вектор $g_1 = f_1$ и нормируем его:

$$e_1 = \frac{g_1}{|g_1|} = \frac{1}{\sqrt{2}}(1,1) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Построим вектор $g_2 = f_2 - \alpha \cdot e_1$ так, чтобы выполнялось условие $(g_2, e_1) = 0$. Получим

$$(g_2, e_1) = (f_2 - \alpha \cdot e_1, e_1) = (f_2, e_1) - \alpha(e_1, e_1) = 0,$$

откуда

$$\alpha = (f_2, e_1) = \frac{1}{\sqrt{2}} \cdot 1 + \frac{1}{\sqrt{2}} \cdot 2 = \frac{3}{\sqrt{2}}.$$

Вычислим вектор g_2:

$$g_2 = f_2 - \alpha \cdot e_1 = (1,2) - \frac{3}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Нормируем вектор g_2:

$$e_2 = \frac{g_2}{|g_2|} = \frac{1}{\frac{1}{\sqrt{2}}}\left(-\frac{1}{2}, \frac{1}{2}\right) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Векторы e_1 и e_2 образуют ортонормированный базис евклидова пространства.

Проверка:

$$|e_1| = |e_2| = 1, \quad (e_1, e_2) = \frac{1}{\sqrt{2}} \cdot \left(-\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = 0.$$

Ортогональное дополнение

Пусть задано евклидово пространство E. Пусть $V^=$ — некоторое линейное подпространство евклидова пространства E.

Совокупность $V^\perp$ векторов y пространства E, обладающих свойством $(y, x) = 0$, где x — произвольный вектор из $V^=$, называется *ортогональным дополнением* к подпространству $V^=$.

Ортогональное дополнение $V^\perp$ есть линейное подпространство евклидова пространства E. Любой вектор пространства W можно представить, причем единственным образом, в виде суммы векторов из $V^=$ и $V^\perp$:

$$\begin{cases}(\boldsymbol{a}, \boldsymbol{a}_1) = \lambda_1(\boldsymbol{a}_1, \boldsymbol{a}_1) + \lambda_2(\boldsymbol{a}_1, \boldsymbol{a}_2),\\ (\boldsymbol{a}, \boldsymbol{a}_2) = \lambda_1(\boldsymbol{a}_1, \boldsymbol{a}_2) + \lambda_2(\boldsymbol{a}_2, \boldsymbol{a}_2).\end{cases}$$

В численном виде

$$\begin{cases}1 = 2\lambda_1 + 0 \cdot \lambda_2,\\ -1 = 0 \cdot \lambda_1 + 2\lambda_2.\end{cases}$$

Ее решение

$$\lambda_1 = \frac{1}{2}, \lambda_2 = -\frac{1}{2}.$$

Тогда

$$\boldsymbol{x} = \frac{1}{2}\boldsymbol{a}_1 - \frac{1}{2}\boldsymbol{a}_2 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right).$$

Затем находится вектор y.

10.5. СОБСТВЕННЫЕ ВЕКТОРЫ И СОБСТВЕННЫЕ ЗНАЧЕНИЯ МАТРИЦЫ

Определение. Ненулевой вектор $\boldsymbol{x}$ называется собственным вектором матрицы P, если найдется такое число λ, называемое *собственным значением* матрицы, что

$$P \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x}.$$

Не всякая матрица обладает собственными векторами.

Решим задачу нахождения собственных векторов матрицы. Запишем равенство в матричной форме:

$$P \cdot X = \lambda \cdot X.$$

Преобразуем матричное уравнение

$$P \cdot X - \lambda \cdot X = O \quad \text{или} \quad (P - \lambda \cdot E)X = O.$$

Матричное уравнение всегда имеет нулевое решение:

$$X = O = \begin{pmatrix}0\\0\\\dots\\0\end{pmatrix}.$$

Для существования ненулевых решений ранг матрицы коэффициентов должен быть меньше числа переменных ($r < n$), т.е. число линейно независимых уравнений должно быть меньше числа переменных. В этом случае должно быть выполнено условие

$$|P - \lambda \cdot E| = 0.$$

Написав уравнение относительно λ подробнее, будем иметь

$$|P - \lambda \cdot E| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0.$$

Раскроем определитель, получим уравнение n-й степени относительно λ:

$$(-1)^n \lambda^n + \alpha_{n-1}\lambda^{n-1} + \dots + \alpha_1\lambda + \alpha_0 = 0,$$

которое называется *характеристическим уравнением A*. Корни уравнения называются *характеристическими* или *собственными числами* матрицы. Множество всех собственных чисел называется *спектром* этой матрицы. Многочлен левой части уравнения называется *характеристическим многочленом*.

Решив характеристическое уравнение, получаем собственные числа $\lambda_1, \lambda_2, \dots, \lambda_n$. Для каждого найденного собственного значения λ_i найдем векторы матрицы $A - \lambda_i E$. Именно они будут собственными векторами, соответствующими собственному значению λ_i. Другими словами, необходимо решить однородную систему уравнений $(P - \lambda_i \cdot E)X = O$. Ее общее решение дает всю совокупность собственных векторов, отвечающих λ_i.

Общее решение однородной системы, как известно, структурировано. Оно представляет собой линейную комбинацию фундаментального набора линейно независимых решений (векторов).

З а м е ч а н и е 1. Собственные векторы определены с точностью до постоянного множителя.

З а м е ч а н и е 2. Собственные векторы $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ оператора, отвечающие различным собственным значениям $\lambda_1, \lambda_2, \dots, \lambda_n$, линейно независимы.

Теорема 10.4. Матрица в базисе из собственных векторов имеет диагональный вид, где по диагонали стоят собственные числа.

Пусть $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ — собственные векторы линейного оператора $\tilde{P}$ в пространстве R^n, которые возьмем в качестве базиса. Тогда разложение векторов $\tilde{P}(\boldsymbol{e}_1), \tilde{P}(\boldsymbol{e}_2), \dots, \tilde{P}(\boldsymbol{e}_n)$ по базису $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ примет вид

$$\begin{aligned} \tilde{P}(\boldsymbol{e}_1) &= a_{11}\boldsymbol{e}_1 + a_{21}\boldsymbol{e}_2 + \dots + a_{n1}\boldsymbol{e}_n = \lambda_1\boldsymbol{e}_1, \\ \tilde{P}(\boldsymbol{e}_2) &= a_{12}\boldsymbol{e}_1 + a_{22}\boldsymbol{e}_2 + \dots + a_{n2}\boldsymbol{e}_n = \lambda_2\boldsymbol{e}_2, \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ \tilde{P}(\boldsymbol{e}_n) &= a_{1n}\boldsymbol{e}_1 + a_{2n}\boldsymbol{e}_2 + \dots + a_{nn}\boldsymbol{e}_n = \lambda_n\boldsymbol{e}_n. \end{aligned}$$

Отсюда следует, что $a_{ij} = \lambda_i$, если $i = j$, и $a_{ij} = 0$, если $i \neq j$. Поэтому в базисе, составленном из собственных векторов, матрица оператора будет иметь диагональный вид:

$$N = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Замечание. Справедливо и обратное утверждение. Если матрица в некотором базисе имеет диагональный вид, то числа, стоящие по диагонали, есть собственные значения, а имеющийся базис составлен из собственных векторов.

Пример 10.11. Найти собственные значения и собственные векторы матрицы $P = \begin{pmatrix} 3 & -2 \\ -1 & 2 \end{pmatrix}$ в пространстве R^2.

Решение. Составим характеристическое уравнение:

$$|P - \lambda \cdot E| = \begin{vmatrix} 3-\lambda & -2 \\ -1 & 2-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 4 = 0.$$

Из квадратного уравнения найдем собственные значения линейного оператора $\lambda_1 = 1$, $\lambda_2 = 4$. Чтобы найти собственные векторы, решим матричные уравнения

$(P - \lambda_1 \cdot E)$ и $(P - \lambda_2 \cdot E)X = O$.

В развернутом виде

$$\begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} -1 & -2 \\ -1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Соответствующие однородные системы:

$$\begin{cases} 2x_1 - 2x_2 = 0, \\ -x_1 + x_2 = 0 \end{cases} \quad \text{и} \quad \begin{cases} -x_1 - 2x_2 = 0, \\ -x_1 - 2x_2 = 0. \end{cases}$$

Общие решения систем:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c_2\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad \text{где } c_1, c_2 \in R.$$

Таким образом, множество собственных векторов, отвечающих собственным значениям $\lambda_1 = 1$ и $\lambda_2 = 4$, имеет вид $c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix}$; $c_2\begin{pmatrix} -2 \\ 1 \end{pmatrix}$, где $c_1, c_2 \in R$.

Векторы $\boldsymbol{a}_1 = (1, 1)$, $\boldsymbol{a}_2 = (-2, 1)$, например, являются линейно независимыми.

10.6. ВЕКТОРНЫЕ ФУНКЦИИ СКАЛЯРНОГО И ВЕКТОРНОГО АРГУМЕНТА

Предположим, что каждому значению скалярного аргумента x ставится в соответствие n функций $y_1(x)$, $y_2(x)$, ..., $y_n(x)$. Набор этих функций примем координатами вектора $\boldsymbol{y}$:

$$\boldsymbol{y}(x) = (y_1(x), y_2(x), \ldots, y_n(x))$$

и будем считать, что *задана векторная функция (вектор-функция)* $\boldsymbol{y}$ *скалярного аргумента* x.

Пусть $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, ..., $\boldsymbol{e}_n$ — базис n-мерного векторного пространства. Разложим вектор-функцию по векторам базиса:

$$\boldsymbol{y}(x) = y_1(x)\boldsymbol{e}_1 + y_2(x)\boldsymbol{e}_2 + \ldots + y_n(x)\boldsymbol{e}_n.$$

Если вектор-функцию $\boldsymbol{y}(x)$ представить как радиус-вектор $\boldsymbol{r}(x)$, начало которого помещено в начало координат, то конец радиуса-вектора $\boldsymbol{r}(x)$ будет описывать некоторую кривую, называемую *годографом* векторной функции. На рис. 10.13 для евклидова векторного пространства с размерностью, равной трем, построена векторная функция

$$\boldsymbol{r}(x) = (r_1(x), r_2(x), r_3(x)) = \left(\frac{\sin 6x}{x}, \frac{\cos 6x}{x}, x\right),$$

где $x \in [0{,}1;\ 2\pi]$. Перед нами пример годографа, изображающего спираль. Для четырех последовательных значений аргумента $x_0 < x_1 < x_2 < x_3$ показаны радиусы-векторы, выходящие из начала координат.

Дифференцирование векторной функции скалярного аргумента

Пусть для векторной функции $\boldsymbol{y}(x)$ существует предел

$$\lim_{\Delta x \to 0} \frac{\boldsymbol{y}(x_0 + \Delta x) - \boldsymbol{y}(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta \boldsymbol{y}}{\Delta x}.$$

Он называется производной от $\boldsymbol{y}(x)$ по переменной x в точке x_0 и обозначается $\frac{d\boldsymbol{y}}{dx}$. Для координат вектора $\boldsymbol{y}(x)$ это означает по определению, что

$$\frac{d\boldsymbol{y}(x)}{dx} = \lim_{\Delta x \to 0} \frac{\Delta \boldsymbol{y}}{\Delta x} = \left(\lim_{\Delta x \to 0} \frac{\Delta y_1(x)}{\Delta x}, \lim_{\Delta x \to 0} \frac{\Delta y_2}{\Delta x}, \ldots, \lim_{\Delta x \to 0} \frac{\Delta y_n}{\Delta x}\right).$$

Правила дифференцирования векторной функции скалярного аргумента:

1. $\dfrac{d(\boldsymbol{y}(x) + \boldsymbol{z}(x))}{dx} = \dfrac{d\boldsymbol{y}(x)}{dx} + \dfrac{d\boldsymbol{z}(x)}{dx}.$

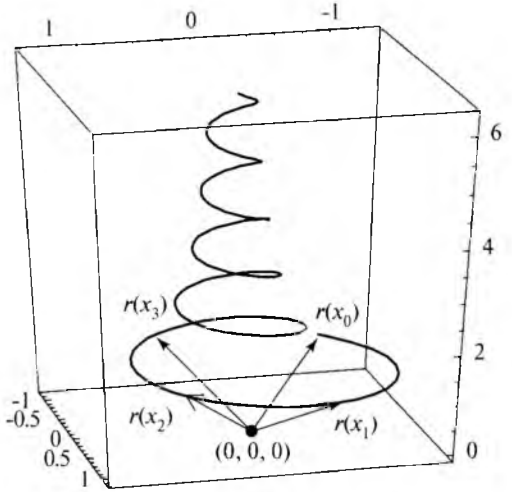

Рис. 10.13

2. $\dfrac{d(\lambda(x)\cdot \boldsymbol{y}(x))}{dx} = \dfrac{d\lambda(x)}{dx}\boldsymbol{y}(x) + \lambda(x)\dfrac{d\boldsymbol{y}(x)}{dx}$, где $\lambda(x)$ — скалярная функция.

3. $\dfrac{d(\boldsymbol{y}(x), \boldsymbol{z}(x))}{dx} = \left(\dfrac{d\boldsymbol{y}(x)}{dx}, \boldsymbol{z}(x)\right) + \left(\boldsymbol{y}(x), \dfrac{d\boldsymbol{z}(x)}{dx}\right)$.

4. $\dfrac{d\boldsymbol{y}(u(x))}{dx} = \dfrac{d\boldsymbol{y}}{du}\cdot\dfrac{du}{dx}$.

Векторные функции векторного аргумента

Рассматривая векторные функции, отойдем от соглашения, что начало любого вектора находится в точке полюса, или в начале координат.

Пусть в некоторой области D заданы n функций от m переменных:

$$y_1 = y_1(x_1, x_2, ..., x_m), \quad y_2 = y_2(x_1, x_2, ..., x_m), \quad y_n = y_n(x_1, x_2, ..., x_m).$$

Другими словами, пусть заданы n функций от векторного аргумента $\boldsymbol{x} = (x_1, x_2, ..., x_m)$:

$$y_1 = y_1(\boldsymbol{x}), \quad y_2 = y_2(\boldsymbol{x}), \quad ..., \quad y_n = (y_n(\boldsymbol{x})).$$

Набор переменных $y_1, y_2, ..., y_n$ можно представить как координаты некоторого нового вектора $\boldsymbol{y}$:

$$\boldsymbol{y} = (y_1, y_2, ..., y_n).$$

Тогда будем говорить, что задана *n-мерная вектор-функция* $\boldsymbol{y}$ *от m-мерного вектора-аргумента* $\boldsymbol{x}$:

$\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), y_2(\boldsymbol{x}), ..., y_n(\boldsymbol{x}))$.

Если в каждой точке $M(x_1, x_2)$ евклидова векторного пространства $\boldsymbol{R}^2$ определена векторная величина

$\boldsymbol{y} = \boldsymbol{y}(M) = y_1(x_1, x_2) \cdot \boldsymbol{e}_1 + y_2(x_1, x_2) \cdot \boldsymbol{e}_2$,

то говорят, что задано *векторное поле* $\boldsymbol{y}$. Координаты $y_1(x_1, x_2)$, $y_2(x_1, x_2)$ образуют две скалярные функции.

Пример векторного поля векторной функции

$\boldsymbol{y} = \boldsymbol{y}(M) = \sin(x_1) \cdot \boldsymbol{e}_1 + \cos(x_2) \cdot \boldsymbol{e}_2$

изображен на рис. 10.14. В каждой точке M с координатами (x_1, x_2), где $0 \le x_1 \le 2\pi$, $0 \le x_2 \le 2\pi$, вычисляются $\sin x_1$ и $\cos x_2$, которые являются координатами вектора $\boldsymbol{y}(M)$. Направления стрелок указывают направления векторов в каждой точке. Длина стрелки пропорциональна величине вектора в данной точке (коэффициент пропорциональности 0.25).

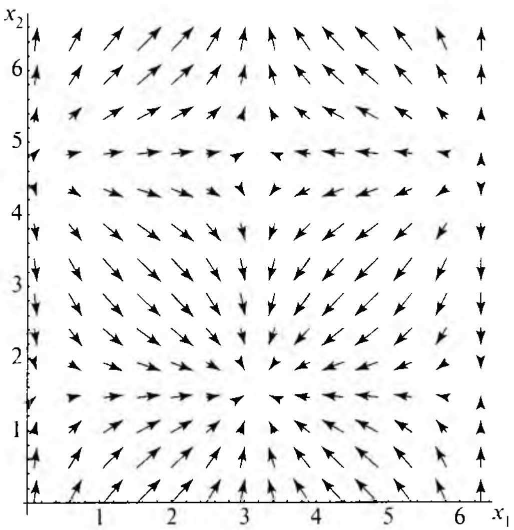

Рис. 10.14

Градиент функции Кобба — Дугласа является полноценной векторной функцией векторного аргумента. Градиент функции $u(x_1, x_2) = x_1^{\frac{1}{2}} x_2^{\frac{1}{2}}$ имеет вид

$$\operatorname{grad} u = \left(\frac{x_2^{\frac{1}{2}}}{2x_1^{\frac{1}{2}}}, \frac{x_1^{\frac{1}{2}}}{2x_2^{\frac{1}{2}}} \right).$$

Векторное поле градиента представлено на рис. 10.15. Длина стрелки пропорциональна величине вектора в данной точке (коэффициент пропорциональности 0,25). Пунктиром изображены векторные линии, сплошной линией — эквипотенциальные линии (линии уровня).

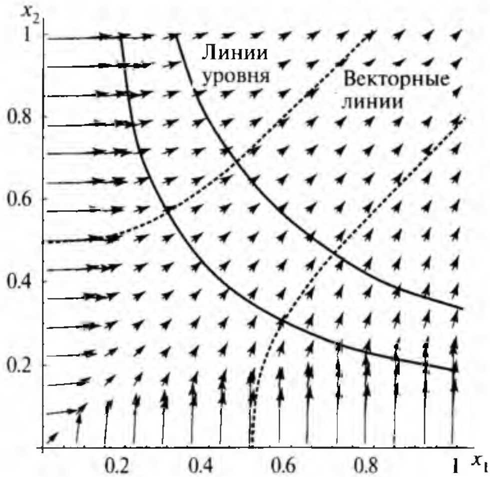

Рис. 10.15

Скалярные функции $y_1(x_1, x_2)$, $y_2(x_1, x_2)$ определяют $\boldsymbol{y}(\boldsymbol{r})$ — векторную функцию векторного аргумента, где $\boldsymbol{r} = \boldsymbol{r}(x_1, x_2)$ — вектор с началом в начале координат и концом в точке M.

Дифференцирование векторной функции векторного аргумента

Предположим, что все функции дифференцируемы в области D.

Производной $\frac{d\boldsymbol{y}}{d\boldsymbol{x}}$ *вектор-функции* $\boldsymbol{y}$ *по вектору-аргументу* $\boldsymbol{x}$ в точке $\boldsymbol{x}_0$ называется предел

$$\lim_{\Delta \boldsymbol{x} \to 0} \frac{\boldsymbol{y}(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - \boldsymbol{y}(\boldsymbol{x}_0)}{\Delta \boldsymbol{x}} = \lim_{\Delta \boldsymbol{x} \to 0} \frac{\Delta \boldsymbol{y}}{\Delta \boldsymbol{x}}.$$

Для координат вектора $\boldsymbol{y}(\boldsymbol{x})$ это означает по определению, что

$$\frac{d\boldsymbol{y}(\boldsymbol{x})}{d\boldsymbol{x}} = \lim_{\Delta x \to 0} \frac{\Delta \boldsymbol{y}}{\Delta \boldsymbol{x}} = \begin{pmatrix} \lim\limits_{\Delta x \to 0} \dfrac{\Delta y_1(\boldsymbol{x})}{\Delta \boldsymbol{x}} \\ \lim\limits_{\Delta x \to 0} \dfrac{\Delta y_2}{\Delta \boldsymbol{x}} \\ \dots \\ \lim\limits_{\Delta x \to 0} \dfrac{\Delta y_n}{\Delta \boldsymbol{x}} \end{pmatrix} = \begin{pmatrix} \dfrac{dy_1}{d\boldsymbol{x}} \\ \dfrac{dy_2}{d\boldsymbol{x}} \\ \dots \\ \dfrac{dy_n}{d\boldsymbol{x}} \end{pmatrix}.$$

В свою очередь, производная скалярной функции векторного аргумента определяется так:

$$\lim_{\Delta \boldsymbol{x} \to 0} \frac{\Delta y_k}{\Delta \boldsymbol{x}} = \left(\lim_{\Delta x_1 \to 0} \frac{\Delta y_k}{\Delta x_1}, \quad \lim_{\Delta x_2 \to 0} \frac{\Delta y_k}{\Delta x_2}, \quad \dots, \quad \lim_{\Delta x_m \to 0} \frac{\Delta y_k}{\Delta x_m} \right),$$

где $k = 1, 2, \dots, n$.

Таким образом, производной от $\boldsymbol{y}(x)$ по переменной $\boldsymbol{x}$ является матрица размерами $n \times m$, составленная из частных производных функций по следующему правилу:

$$\frac{d\boldsymbol{y}}{d\boldsymbol{x}} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \dots & \dfrac{\partial y_1}{\partial x_m} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \dots & \dfrac{\partial y_2}{\partial x_m} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial y_n}{\partial x_1} & \dfrac{\partial y_n}{\partial x_2} & \dots & \dfrac{\partial y_n}{\partial x_m} \end{pmatrix}.$$

Эта матрица получила название матрицы Якоби.

Пример 10.12. Найти матрицу Якоби вектор-функции $\boldsymbol{y}(x) = (2x_1 + 3x_3, -x_2 + 2x_3, 5x_1 - x_2)$.

Решение.

$$\frac{d\boldsymbol{y}}{d\boldsymbol{x}} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_1}{\partial x_3} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \dfrac{\partial y_2}{\partial x_3} \\ \dfrac{\partial y_3}{\partial x_1} & \dfrac{\partial y_3}{\partial x_2} & \dfrac{\partial y_3}{\partial x_3} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 3 \\ 0 & -1 & 2 \\ 5 & -1 & 0 \end{pmatrix}.$$

ЛИТЕРАТУРА

1. *Абчук В.А.* Математика для менеджеров и экономистов. — М.: Изд-во Михайлова, 2002.
2. *Воронцов М.В., Мещеряков Г.П.* Высшая математика для экономистов и менеджеров. — М.: Феникс, 2004.
3. *Ильин В.А., Ким Г.Д.* Линейная алгебра и аналитическая геометрия. — М.: Изд-во Моск. ун-та, 1998.
4. *Краснов М.Л. и др.* Вся высшая математика. Т.1. — М.: Эдиториал: УРСС, 2003.
5. *Красс М.С., Чупрынов Б.П.* Математика для экономистов. СПб.: Питер, 2004.
6. *Кремер Н.Ш.* Высшая математика для экономистов. — М.: ЮНИТИ-ДАНА, 2006.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Раздел IV
ИНСТРУМЕНТАРИЙ ПРИНЯТИЯ РЕШЕНИЯ В УСЛОВИЯХ РИСКА И НЕОПРЕДЕЛЕННОСТИ

Теория вероятностей и **математическая статистика** относятся к числу прикладных математических дисциплин, поскольку возникли из чисто практических потребностей и направлены на решение прикладных задач.

Основная задача математической статистики состоит в разработке методов принятия решений в условиях неопределенности, рекомендаций и выводов на основе анализа статистических данных, научно обоснованного прогнозирования случайных явлений и их взаимосвязей между собой, построения математических моделей реальных экономических ситуаций.

Одной из основных задач теории вероятностей и математической статистики является исследование зависимости между двумя или несколькими случайными величинами. Строгая функциональная зависимость реализуется редко, так как одна или обе величины подвержены случайным факторам. Основная цель изучения зависимостей между случайными величинами заключается в прогнозировании с данной вероятностью области изменения значений одной случайной величины на основании наблюдаемых значений другой случайной величины.

Одной из важнейших сфер приложения теории вероятностей и математической статистики является экономика. В настоящее время невозможно себе представить исследование и прогнозирование экономических явлений без использования эконометрического моделирования, регрессионного анализа, трендовых и сглаживающих моделей и других методов, опирающихся на теорию вероятностей.

Статистические закономерности присущи и централизованно управляемой, и, тем более, децентрализованной экономике. Наличие таких твердо устоявшихся в экономике понятий, как страховой запас, резервные мощности, государственные резервы, финансовые риски и т.п., свидетельствует об этом.

В течение последних 30 лет теория и практика финансов во все большей степени стали опираться на количественные методы математики, статистики и эконометрики. Появился новый раздел литературы, посвященный управлению рисками и производными финансовыми инструментами, такими как опционы, фьючерсы и свопы.

Сегодня реальная экономика немыслима без прогнозирования ожидаемой эффективности инвестиций, выгоды от вложения средств в акции и т.д. Статистические гипотезы, базирующиеся на законах теории вероятностей, и являются основой экономического прогнозирования.

Теории вероятностей и **математической статистики** предоставляет исследователю математический инструментарий, необходимый для построения разного рода экономических моделей и исследования их свойств.

Глава 11
ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

11.1. ПРОСТРАНСТВО ЭЛЕМЕНТАРНЫХ ИСХОДОВ. СЛУЧАЙНЫЕ СОБЫТИЯ

Теория вероятностей — это наука о закономерностях случайных событий. Под *событием* в теории вероятностей понимается всякий факт, который может произойти или не произойти при осуществлении определенного комплекса условий. Каждое такое осуществление называется испытанием, опытом или экспериментом.

События можно подразделить на следующие три вида: достоверные, невозможные и случайные.

Достоверным называется событие, которое обязательно произойдет при испытании.

Невозможным называется событие, которое заведомо не произойдет при испытании.

Случайным называется событие, которое в результате эксперимента может либо произойти, либо не произойти.

Предметом теории вероятностей являются закономерности массовых случайных событий, где под массовостью мы понимаем многократную повторяемость.

Рассмотрим несколько событий.

A — появление герба при бросании монеты.

B — появление трех гербов при трехкратном бросании монеты.

C — попадание в цель при выстреле.

D — выигрыш по билету денежно-вещевой лотереи.

Видим, что каждое из этих событий обладает какой-то степенью возможности. Для того чтобы количественно сравнивать между собой события по степени их возможности, нужно с каждым событием связать определенное число.

Вероятность события есть численная мера степени объективной возможности этого события.

В качестве единицы измерения вероятности принята вероятность достоверного события. Вероятность невозможного события равна нулю, а вероятность любого случайного события обозначается $P(A)$ и изменяется в диапазоне от нуля до единицы: $0 \le P(A) \le 1$.

Пусть проведена серия из n испытаний (n называют *длиной серии*), в каждом из которых может произойти или не произойти событие A. Подсчитаем, сколько раз в этой серии эксперимент заканчивался наступлением события A. Обозначим это число через $n(A)$. Поделив его на общее число n всех повторений эксперимента, получим величину $P_n(A) = \frac{n(A)}{n}$, которая называется *относительной частотой наступления события A*. При небольшом числе экспериментов частота события носит случайный характер и может заметно меняться от одной группы опытов к другой. При увеличении числа экспериментов случайные обстоятельства, свойственные каждому отдельному эксперименту, в массе взаимно погашаются и частота $P_n(A)$ проявляет тенденцию стабилизироваться, приближаясь к некоторой средней величине p.

Этот эмпирический факт *определяет свойство статистической устойчивости частот*.

Теорема 11.1 (Теорема Бернулли). По мере неограниченного увеличения числа однородных и независимых испытаний относительная частота события A неограниченно приближается к некоторой постоянной величине.

Число, к которому приближается относительная частота события при неограниченном увеличении числа экспериментов, и называется вероятностью события A.

Математически данное свойство записывается в виде предела:

$$\lim_{n \to \infty} P_n(A) = P(A).$$

Частота события A отличается от вероятности этого события тем, что вероятность — величина детерминированная, а частота — величина случайная и до опыта неизвестная.

В качестве примера укажем на опыт Бюффона, в котором симметричная монета подбрасывалась 4040 раз, а герб выпадал $m = 2048$ раз. Частота появления герба в данной серии наблюдений равна $m/n = 2048/4040 = 0{,}507$, что близко к интуитивно ожидаемому значению вероятности 0,5.

Для того чтобы формально описать некоторый эксперимент, т.е. создать математическую модель некоторой экономической ситуации, надо указать все возможные варианты исхода, которыми этот эксперимент может закончиться. Считается, что в результате эксперимента должен произойти только один из этих возможных исходов.

Множество Ω всех возможных исходов эксперимента называется *пространством элементарных исходов*, а каждый исход называется *элементарным событием* и обозначается ω.

Если все возможные исходы можно перечислить, то пространство элементарных исходов называется *дискретным* (конечным или счетным): $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ или $\Omega = \{\omega_1, \omega_2, ...\}$.

Рассмотрим несколько примеров экспериментов и соответствующих им пространств элементарных событий.

1. Подбрасывание монеты: $\Omega = \{Г, Р\}$ (Г — выпадение герба, Р — выпадение решки).

2. Выбрасывание одной игральной кости: $\Omega = \{\omega_1 = 1, \omega_2 = 2, \omega_3 = 3, \omega_4 = 4, \omega_5 = 5, \omega_6 = 6\}$.

3. Проверка одного изделия, случайно отобранного из продукции массового производства: $\Omega = \{\omega_1 = \text{годно}, \omega_2 = \text{дефектно}\}$.

Кроме элементарных событий, рассматриваются так называемые *сложные* события. Например, события A_1 — выпадение четного числа очков на игральной кости, A_2 — выпавшее число очков не превзойдет трех — запишутся соответственно $A_1 = \{\omega_2, \omega_4, \omega_6\}$ и $A_2 = \{\omega_1, \omega_2, \omega_3\}$.

Событием A в случае дискретного пространства элементарных событий называется любое подмножество $A = \{\omega_{i1}, \omega_{i2}, ..., \omega_{ik}, ...\}$ пространства элементарных событий $A \subseteq \Omega$.

Осуществление любого из элементарных событий, входящих в A, влечет за собой осуществление события A: обычно говорят, что «событие A произошло», если эксперимент закончился элементарным исходом $\omega \subseteq A$.

Введем некоторые соотношения между событиями.

Если при каждом осуществлении эксперимента, при котором происходит событие A, происходит и событие B, то говорят, что *событие A влечет за собой* событие B, обозначается $A \subset B$. Если событие A влечет за собой событие B и в то же время событие B влечет за собой событие A, то говорят, что события A и B *равносильны*:

$A = B$. Очевидно, что все достоверные события равносильны между собой.

Объединением (*суммой* $A + B$) событий A и B называется событие $C = A \cup B$, которое происходит тогда, когда наступает хотя бы одно из событий — либо A, либо B, либо A и B одновременно.

Пересечением (*произведением* AB) событий A и B называется событие $C = A \cap B$, которое наступает тогда, когда происходит и событие A, и событие B одновременно.

Разностью двух событий A и B называется такое событие $C = A \backslash B$, которое состоит в том, что происходит событие A и не происходит событие B.

Событие $\bar{A}$ называется *дополнительным* (дополнением) к событию A, если оно происходит всякий раз, когда не происходит событие A. События A и $\bar{A}$ называются противоположными событиями, и для них выполняются соотношения $A + \bar{A} = \Omega$, $A\bar{A} = \varnothing$.

Симметрической разностью событий A и B называется событие $C = A \nabla B$, в которое входят те элементарные события, которые входят или в A, или в B, но не входят в их пересечение: $A \nabla B = (A \backslash B) \cup (B \backslash A)$.

Отношения между событиями можно интерпретировать как соотношения между множествами. Для наглядности используют графическую модель, называемую *диаграммой Вьенна* (рис. 11.1).

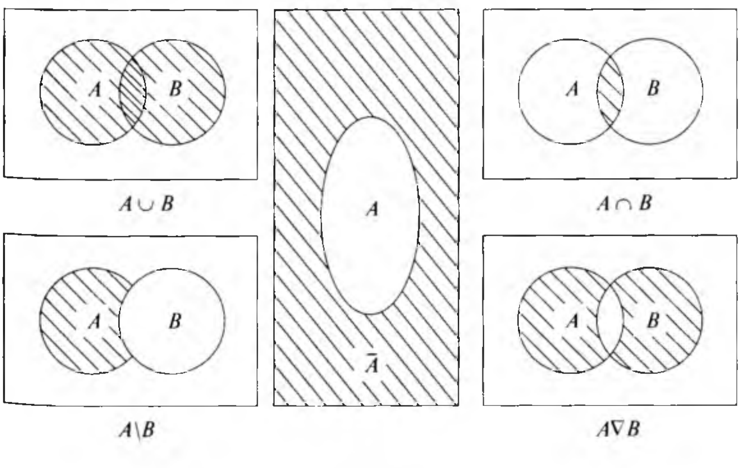

Рис. 11.1

Поскольку все события рассматриваются как подмножества пространства Ω, то все пространство Ω соответствует достоверному событию, а пустое множество $\varnothing$ — невозможному событию.

События A и B называются ***несовместными*** (непересекающимися), если наступление одного из них исключает наступление другого: $A \cap B = \varnothing$.

События $A_1, A_2, \ldots, A_n$ образуют *полную группу событий*, если в результате эксперимента непременно произойдет ровно одно из них, т.е. события несовместны и их объединение является достоверным событием: $\bigcup\limits_{i=1}^{n} A_i = \Omega$, $A_i \cap A_j = \Phi$, $i \neq j$. Например, события A и $\bar{A}$ образуют полную группу.

Несколько событий в данном опыте называются ***равновозможными***, если по условиям симметрии есть основание считать, что ни одно из этих событий не является объективно более возможным, чем другое.

Для любых событий A и B справедливы следующие соотношения.

1. $A \cup \Omega = \Omega, \quad A \cup \varnothing = A, \quad A \cup A = A.$
2. $A \cap \Omega = A, \quad A \cap \varnothing = \varnothing, \quad A \cap A = A.$
3. $\bar{\bar{A}} = A, \quad \bar{\Omega} = \varnothing, \quad \bar{\varnothing} = \Omega.$
4. $\overline{A \cap B} = \bar{A} \cup \bar{B}, \quad \overline{A \cup B} = \bar{A} \cap \bar{B}$ — эти формулы называются принципом *двойственности или формулами де Моргана*.
5. $A \cup B = B \cup A, \quad A \cap B = B \cap A$ — коммутативность операций объединения и пересечения.
6. $A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$ — ассоциативность операций объединения и пересечения.
7. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ — дистрибутивность операции объединения относительно пересечения.
8. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ — дистрибутивность операции пересечения относительно объединения.

Следует отметить, что все действия над событиями можно получить с помощью двух действий — объединения и дополнения или пересечения и дополнения. Например, справедливо соотношение $A \backslash B = A \cap \bar{B}$.

11.2. ВЕРОЯТНОСТЬ В ДИСКРЕТНОМ ПРОСТРАНСТВЕ. КЛАССИЧЕСКАЯ ВЕРОЯТНОСТНАЯ МОДЕЛЬ

Говорят, что в дискретном пространстве элементарных событий $\Omega = \{\omega_1, \omega_2, \ldots\}$ задана *вероятность*, если каждому элементарному событию ω_i поставлено в соответствие неотрицательное число $p_i \geq 0$ такое, что сумма (конечная или бесконечная) вероятностей всех элементарных исходов равна единице: $\sum_i p_i = 1$.

Таким образом, ***вероятность события*** — это числовая функция, определенная на пространстве событий.

Вероятностью любого события A называется сумма вероятностей всех элементарных исходов, входящих в A, т.е. $P(A) = \sum_{\omega_i \in A} P(\omega_i)$.

Из определения следует, что:

1) вероятность события удовлетворяет неравенству $0 \leq P(A) \leq 1$;
2) если Ω — пространство элементарных событий, то $P(\Omega) = 1$;
3) если $\varnothing$ — пустое множество, то $P(\varnothing) = 0$.

Теорема 11.2 (теорема сложения вероятностей несовместных событий). Вероятность объединения двух несовместных событий равна сумме их вероятностей, т.е. если $A \cap B = \varnothing$, *то* $P(A \cup B) = P(A) + P(B)$.

Доказательство проведем для случая конечного числа исходов. Пусть пространство элементарных событий $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ содержит n элементарных исходов, из них n_1 благоприятны событию A и n_2 благоприятны событию B, т.е. $P(A) = \sum_{i:\omega_i \in A}^{n_1} P(\omega_i)$ и $P(B) = \sum_{k:\omega_k \in B}^{n_2} P(\omega_k)$, причем нет исходов, благоприятных одновременно A и B, так как события несовместны. Отсюда следует, что событию $A \cup B$ благоприятны $n_1 + n_2$ исходов и вероятность этого события вычисляется по формуле

$$P(A \cup B) = \sum_{\omega_j \in A \cup B} P(\omega_j) = \sum_{\omega_i \in A} P(\omega_i) + \sum_{\omega_k \in B} P(\omega_k) = P(A) + P(B).$$

Теорема доказана.

Доказательство без труда переносится на случай счетного пространства $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$, когда вместо конечных сумм рассматриваются сходящиеся ряды.

С л е д с т в и я

1. Методом математической индукции эту теорему можно распространить на любое конечное число слагаемых, т.е. если все события A_i несовместны, то

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

2. Если события A_1, A_2, ..., A_n образуют полную группу несовместных событий, то сумма их вероятностей равна единице.

Действительно, так как $\bigcup_{i=1}^{n} A_i = \Omega$, то $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = 1$. Например, так как $\bar{A} \cup A = \Omega$, то $P(\bar{A}) + P(A) = P(\Omega) = 1$, и следовательно, $P(\bar{A}) = 1 - P(A)$.

3. Если $A \subset B$, то $P(A) \leq P(B)$.

Теорема 11.3 (теорема сложения вероятностей произвольных событий).

Для любых событий A и B справедливо равенство $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Доказательство. Если A и B — совместные события, то $A \cup B$ наступит тогда, когда наступит одно из несовместных событий $A \cap B$, $A \cap \bar{B}$ или $\bar{A} \cap B$. По теореме сложения несовместных событий $P(A \cup B) = P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B)$. Событие A наступит, если наступит хотя бы одно из несовместных событий $A \cap B$ или $A \cap \bar{B}$. Тогда вероятность события A по теореме сложения несовместных событий равна $P(A) = P(A \cap B) + P(A \cap \bar{B})$. Аналогично событие B наступит, если наступит хотя бы одно из несовместных событий $A \cap B$ или $\bar{A} \cap B$, и вероятность события B равна $P(B) = P(A \cap B) + P(\bar{A} \cap B)$. Отсюда получаем доказательство теоремы:

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) + P(A \cap B) - P(A \cap B) =$$
$$= P(A) + P(B) - P(A \cap B).$$

Следствия

1. Вероятность пересечения двух событий вычисляется по формуле $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

2. Вероятность суммы любого числа совместных событий вычисляется по формуле включения-исключения:

$$P(A_i \cup \ldots \cup A_n) = \sum_i P(A_i) - \sum_i \sum_j P(A_i \cap A_j) +$$
$$+ \sum_i \sum_j \sum_k P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n).$$

3. Из теорем 1 и 2 для любых событий A и B следует, что $P(A \cup B) \leq P(A) + P(B)$.

Простейшей математической моделью дискретного случайного эксперимента является так называемая классическая модель, в которой пространство элементарных исходов — конечное множество с равновозможными исходами, которые и образуют полную группу событий. Такое вероятностное пространство называется *симметричным*.

Пусть число элементарных событий «симметричного» пространства конечно и равно N, т.е. пространство элементарных событий $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$. Обозначим вероятности элементарных исходов $P(\omega_i) = p_i = p$ для любого $i = 1, 2, ..., N$ и $\sum_{i=1}^{N} P(\omega_i) = \sum_{i=1}^{N} p_i =$ $= pN = 1$. Тогда для любого $i = 1, 2, ..., N$ получаем, что вероятность элементарного исхода равна $p_i = p = \frac{1}{N}$.

Если имеется N равновозможных элементарных исходов, то вероятность любого сложного события A, состоящего из $N(A)$ элементарных исходов, по определению вероятности события равна $P(A) =$ $= \sum_{\omega_i \in A} P(\omega_i) = N(A) \cdot \frac{1}{N} = \frac{N(A)}{N}$, т.е. в рамках классической модели вероятность события определяется как отношение числа элементарных исходов, благоприятных событию A, к общему числу элементарных исходов: $P(A) = \frac{N(A)}{N}$.

Это определение — результат принятия гипотезы о равновероятности элементарных событий.

Пример 11.1. В случае с игральной костью при одном бросании равновозможно выпадение любой из шести граней, на которых нанесены цифры 1, 2, 3, 4, 5, 6. Пространство элементарных событий в этом случае имеет вид $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6)$ и вероятность любого ω_i равна $P(\omega_i) = {}^1/_6$.

Пусть событие A — выпадение четного числа очков $A = \{\omega_2, \omega_4, \omega_6\}$. Тогда $P(A) = {}^3/_6 = {}^1/_2$.

Из доказанного следует, что подсчет вероятностей в случае симметричного пространства сводится к комбинаторным задачам определения числа элементов во множествах A и Ω.

11.3. ЭЛЕМЕНТЫ КОМБИНАТОРНОГО АНАЛИЗА

Теорема 11.4 (основная теорема комбинаторики). Пусть имеется k групп $A_1, A_2, ..., A_k$, причем i-я группа содержит n_i элементов. Тогда общее число N способов, которыми можно выбрать по одному элементу из каждой группы, равно $N = n_1 n_2 ... n_k$.

Эта формула называется основной формулой комбинаторики.

Теорема 11.5 (теорема сложения). Пусть имеется k групп $A_1, A_2, ...,$ A_k, причем i-я группа содержит n_i элементов и при этом любые две группы A_i и A_j не имеют общих элементов. Тогда выбор одного элемента или из A_1, или из A_2, ..., или из A_k можно осуществить $N = n_1 + n_2 + ... + n_k$ способами.

Рассмотрим три типовые ситуации, которые являются следствием общей схемы, изложенной в основной теореме комбинаторики. Будем называть их схемами последовательного выбора с возвращением, последовательного выбора без возвращения и одновременным выбором.

1. **Последовательный выбор с возвращением.**

Пусть имеется некоторая конечная совокупность элементов $\{a_1, a_2, ..., a_n\}$, называемая генеральной совокупностью. Эксперимент состоит в том, что из генеральной совокупности последовательно выбирают k элементов и каждый отобранный элемент перед отбором следующего возвращается в генеральную совокупность. Требуется найти общее число способов, которыми можно произвести последовательную выборку с возвращением k элементов из генеральной совокупности объема n.

Так как каждый раз отобранный элемент перед отбором следующего возвращается в генеральную совокупность, то выбор на каждом шаге производится из совокупности объема n и можно считать, что выбор производится из k групп и все группы состоят из одинакового числа элементов: $n_1 = n_2 = ... = n_k = n$. Тогда в силу основной теоремы комбинаторики число таких способов равно $N = n^k$. Такой выбор называется последовательным выбором с возвращением или *размещениями с повторениями*, их число обозначается $\bar{A}_n^k = n^k$.

2. **Последовательный выбор без возвращения.**

Пусть эксперимент состоит в том, что из генеральной совокупности последовательно выбирают k элементов и отобранный элемент перед отбором следующего в генеральную совокупность не возвращается. Тогда первый элемент выбирается из совокупности объема $n_1 = n$. Так как элемент не возвращается в генеральную совокупность, то следующий элемент выбирается из совокупности, объем которой на один элемент меньше, т.е. второй элемент выбирается уже из совокупности объема $n_2 = n - 1$, третий — из совокупности объема $n_3 = n - 2$ и т.д., k-й элемент выбирается из совокупности объема $n_k = n - (k - 1)$. Число способов, каким можно выбрать последовательно k элементов из генеральной совокупности объема n без возвращения, в силу основной теоремы комбинаторики равно $N = n(n - 1)(n - 2)...(n - (k - 1))$. Это число равно числу размещений из n по k и обозначается $N = A_n^k$.

Размещениями из n элементов по k называются упорядоченные совокупности, которые отличаются друг от друга или своими элементами, или их порядком. Формула для вычисления числа размещений из n элементов по k имеет вид $A_n^k = \dfrac{n!}{(n-k)!}$.

Частным случаем размещений являются перестановки, когда $k = n$.

Перестановками из n элементов называются упорядоченные совокупности, отличающиеся друг от друга только порядком входящих в них элементов. Формула для вычисления числа перестановок из n элементов имеет вид $P_n = n!$.

3. **Одновременный выбор.**

Одновременный неупорядоченный выбор k элементов из генеральной совокупности объема n подразумевает, что мы выбираем элементы без учета порядка. Число таких совокупностей по k элементов из n равно числу сочетаний из n по k и обозначается $N = C_n^k$.

Сочетаниями из n элементов по k называются любые неупорядоченные соединения, отличающиеся друг от друга хотя бы одним элементом. Формула для вычисления числа сочетаний из n элементов по k следующая: $C_n^k = \dfrac{n!}{k!(n-k)!}$.

Если в сочетаниях из n элементов по k некоторые из элементов или все могут оказаться одинаковыми, то такие сочетания называются *сочетаниями с повторениями* из n элементов по k, и число таких сочетаний равно $\bar{C}_n^k = C_{n+k-1}^k$.

Сочетания обладают следующим свойствами.

1. $C_n^0 = 1$.
2. $C_n^m = C_n^{n-m}$.
3. $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$.
4. $C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n$.
5. $A_n^m = P_m C_n^m$.

Пример 11.2. Пусть в урне N шаров, из них M белых и $N - M$ черных. Из урны извлекается выборка объема n. Найти вероятность того, что в этой выборке будет ровно m белых шаров.

Решение. Так как порядок элементов здесь несуществен, то число всех возможных выборок объема n из N элементов равно числу сочетаний C_N^n. Число испытаний, которые благоприятны событию A — появление m белых шаров, остальные $n - m$ — черные, равно $C_M^m C_{N-M}^{n-m}$, и следовательно, искомая вероятность равна $P(A)=\dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$. Описанная ситуация представляет собой пример *«урновой модели»*. Говорят также, что случайное число белых шаров в выборке имеет *гипергеометрическое распределение*.

В общем случае предполагается, что имеется $N = n_1 + n_2 + \ldots + n_k$ различных частиц, причем n_1 частиц — первого типа, n_2 — второго типа, ..., n_k — k-го типа. Случайным образом из этих N частиц выбирается m частиц. Найдем вероятность события A, состоящего в том, что среди выбранных окажется ровно $m_1 < n_1$ частиц первого типа, $m_2 < n_2$ — второго типа, ..., $m_k < n_k$ — k-го типа, так что $m = m_1 + m_2 + \ldots + m_k$.

Поскольку порядок выбора несуществен, то при определении общего числа исходов и числа благоприятных исходов мы должны пользоваться числом сочетаний. Общее число элементарных исходов равно C_N^m. Далее, m_1 частиц первого типа можно выбрать $C_{n_1}^{m_1}$ способами, m_2 частиц второго типа — $C_{n_2}^{m_2}$ способами, ..., m_k частиц k-го типа — $C_{n_k}^{m_k}$ способами. При этом любой выбор частиц определенного типа комбинируют с любыми выборами частиц остальных типов, и следовательно, число благоприятных событию A исходов равно $C_{n_1}^{m_1}C_{n_2}^{m_2}\ldots C_{n_k}^{m_k}$. Поэтому вероятность события A равна

$$P(A) = P(m_1, m_2, \ldots, m_k) = \frac{C_{n_1}^{m_1}C_{n_2}^{m_2}\ldots C_{n_k}^{m_k}}{C_N^m}.$$

Пример 11.3. Найти вероятность того, что в семизначном номере телефона ровно три цифры совпадают, а остальные различны.

Решение. Поскольку каждая из десяти цифр может занять любое из семи мест в числе, то число возможных исходов равно числу размещений с повторением из 10 по 7, т.е. $N = 10^7$. Событие A состоит в том, что три цифры совпадают, а остальные различны. Следовательно, необходимо выбрать пять различных цифр, что можно сделать $n_1 = A_{10}^5$ способами. Здесь использованы размещения, так как это номер телефона и порядок в данном случае существен. К тому же одна из выбранных цифр может занять любые три места из семи. Это можно сделать $n_2 = C_7^3$ способами, так как порядок здесь несуществен по причине того, что размещается одна и та же цифра. Благоприятный исход в силу основной теоремы комбинаторики равен $N(A) = n_1 n_2 = C_7^3 A_{10}^5$. Искомая вероятность равна

$$P(A) = \frac{C_7^3 A_{10}^5}{10^7}.$$

Предположим, что n *неразличимых* частиц распределяются по m ячейкам. Различными и равновозможными считаются распределения частиц по ячейкам, отличающиеся только числом частиц, попавших в каждую ячейку. Такое распределение носит название *статистики Бозе—*

Эйнштейна. Найдем общее число элементарных исходов в статистике Бозе—Эйнштейна. Если считать «белый» элемент частицей, а «черный» — перегородкой, то существует взаимно-однозначное соответствие между способами выбора $m-1$ «черного» элемента и размещениями частиц в статистике Бозе—Эйнштейна. Для этого рассмотрим последовательность из $n+m-1$ элементов и выберем из них $m-1$ «черный» элемент (рис. 11.2).

	1				2	3				$m-1$
○	●	○	○	○	●	●	○	…	○	●
1	2	3	4	5	6	7				$n+m-1$

Рис. 11.2

Так, на рис. 11.2 в первую ячейку попала одна частица, во вторую — три, третья оказалась пустой и т.д., последняя, m-я ячейка также оказалась пустой. Поэтому общее число исходов равно числу сочетаний C_{n+m-1}^{m-1}. Найдем вероятность того, что в фиксированную ячейку попало ровно k частиц (событие A). Заметим, что если в этой фиксированной ячейке уже находится k частиц, то остальные $n-k$ частиц должны быть распределены по оставшимся $m-1$ ячейкам, а это можно сделать $C_{n+m-k-1-1}^{m-1-1} = C_{n+m-k-2}^{m-2}$ способами. Следовательно, искомая вероятность равна

$$P(A) = \frac{C_{n+m-k-2}^{m-2}}{C_{n+m-1}^{m-1}}.$$

В *статистике Ферми—Дирака* n неразличимых частиц распределяются по m ячейкам ($n \le m$), однако в каждой ячейке не может находиться более одной частицы. Число различных элементарных исходов совпадает с числом способов, которыми мы можем выбрать n занятых ячеек из общего числа ячеек m, и так как порядок выбора несущественен, то число способов равно числу сочетаний: $N = C_m^n$. Найдем вероятность того, что заняты k фиксированных ячеек, т.е. событие A — заняты фиксированные k ячеек ($k \le n$). Тогда оставшиеся $m-k$ ячеек должны быть заполнены $n-k$ частицами, а это можно сделать C_{m-k}^{n-k} способами. Поэтому искомая вероятность равна

$$P(A) = \frac{C_{m-k}^{n-k}}{C_m^n}.$$

Предполагая, что n одинаковых частиц распределяются по m ячейкам без ограничений на число попавших в каждую ячейку частиц, получаем статистику *Максвелла—Больцмана*. Поскольку каждая из n частиц может попасть в любую из m ячеек, то общее число элементарных исходов равно m^n. Событие A заключатся в том, что в первую ячейку попало n_1 частиц, во вторую — n_2, ..., в m-ю — n_m частиц ($n_1 + n_2 + ... + n_m = n$). Число благоприятных для события A исходов подсчитаем следующим образом. В первую ячейку могут попасть любые n_1 частиц из имеющихся n частиц первоначально. Это можно сделать $C_n^{n_1}$ способами. Вторую ячейку надо заполнить n_2 частицами из имеющихся $n - n_1$ частиц. Это можно сделать $C_{n-n_1}^{n_2}$ различными способами. Продолжая эту процедуру и используя основную формулу комбинаторики, получаем, что число благоприятных событию A способов равно

$$C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n-n_1-\dots-n_{m-1}}^{n_m} = \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdot \ldots \times$$

$$\times \frac{(n-n_1-\ldots-n_{m-1})!}{n_m!0!} = \frac{n!}{n_1!n_2!\ldots n_m!},$$

т.е. если множество из n элементов разбивается на m групп так, что в первую группу попадает n_1 элементов, во вторую — n_2 элементов, в m-ю группу — n_m элементов, то число таких разбиений равно

$$N_n(n_1, n_2, \ldots, n_m) = \frac{n!}{n_1!n_2!\ldots n_m!}.$$

Таким образом, искомая вероятность равна

$$P(A) = \frac{n!}{n_1!n_2!\ldots n_m!} \frac{1}{m^n}.$$

Статистика Максвелла—Больцмана представляет собой частный случай так называемой полиномиальной схемы.

Полиномиальной схемой называется последовательность независимых испытаний, в каждом из которых возможны $k > 2$ исходов, с вероятностями исходов $p_1, p_2, \ldots, p_k$, $0 < p_i < 1$, $\sum_{i=1}^{k} p_i = 1$. Полиномиальную схему можно трактовать как обобщение статистики Максвелла—Больцмана на случай, когда вероятности попадания каждой частицы в различные ячейки различны. В этом случае вероятностное пространство содержит k^n элементарных событий, а вероятность того, что из n испытаний m_1 закончатся первым исходом, m_2 — вторым исходом, ..., m_k — k-м исходом, равна

$$P_n(m_1, \ldots, m_k) = \frac{n!}{m_1!m_2!\ldots m_k!} p_1^{m_1} p_2^{m_2} \ldots p_k^{m_k}.$$

Полученная формула описывает *полиномиальный закон распределения*.

Пример 11.4. Пять клиентов случайным образом обращаются в пять фирм недвижимости. Найти вероятность того, что ровно в одну фирму никто не обратится.

Решение. Поскольку каждый из пяти клиентов может обратиться в любую из пяти фирм, то общее число исходов равно $N = 5^5$. Число исходов, благоприятных рассматриваемому событию, когда в одну фирму обратятся двое, в одну фирму никто не обратится и в три фирмы обратятся по одному, равно $N_1(3,1,1) = \frac{5!}{3!1!1!}$, т.е. числу способов, какими можно разбить пять фирм на группы: $n_1 = 3$ — куда обратятся по одному, $n_2 = 1$, $n_3 = 1$ — куда или обратятся двое, или никто не обратится. Пятерых клиентов тоже необходимо распределить соответственно по этим фирмам, что можно сделать $N_2(2,1,1,1) = \frac{5!}{2!1!1!1!}$ способами. Следовательно, число благоприятных исходов составляет $m = N_1 N_2$ и искомая вероятность равна $P(A) = \frac{5!}{2!1!1!1!} \cdot \frac{5!}{3!1!1!} \cdot \frac{1}{5^5} = 0{,}384$.

Пример 11.5. Два шахматиста A и B встречались за доской 50 раз, причем 15 раз выиграл A, 10 раз выиграл B и 25 партий закончились вничью. Найти вероятность того, что в матче из 10 партий между этими шахматистами 3 партии выиграет A, 2 партии выиграет B, а 5 партий закончатся вничью.

Решение. Из условия задачи следует, что вероятности различных исходов партии можно оценить как

$$P(A) = \frac{15}{50} = \frac{3}{10}, \quad P(B) = \frac{10}{50} = \frac{1}{5}, \quad P(H) = \frac{25}{50} = \frac{1}{2}.$$

Тогда искомая вероятность равна

$$P_{10}(3,2,5) = \frac{10!}{3!2!5!} \cdot \left(\frac{3}{10}\right)^3 \cdot \left(\frac{1}{5}\right)^2 \cdot \left(\frac{1}{2}\right)^5 = 0{,}085.$$

11.4. УСЛОВНАЯ ВЕРОЯТНОСТЬ. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ

Говорить о вероятности $P(A)$ как о мере возможности появления случайного события A имеет смысл только при осуществлении определенного комплекса условий эксперимента, в рамках которого событие может произойти. При изменении условий эксперимента, вообще говоря, изменится и вероятность.

Поэтому помимо обычной (безусловной) вероятности события A рассматривают так называемую условную вероятность события A, вычисленную при условии, что произошло некоторое событие B.

Условной вероятностью события A при условии, что произошло событие B ($P(B) > 0$), называется число $P(A \mid B)$, равное отношению $P(A|B) = P(AB)/P(B)$.

Аналогично определяется условная вероятность события B: $P(B \mid A) = P(AB)/P(A)$.

Из определения условной вероятности вытекает теорема умножения вероятностей.

Теорема 11.6. Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое из них произошло:

$P(AB) = P(A \mid B)P(B) = P(A)P(B \mid A)$.

Следствие

Вероятность произведения нескольких событий равна произведению вероятностей этих событий, причем вероятность каждого последующего события вычисляется при условии, что все предыдущие имели место:

$$P\left[\bigcap_{i=1}^{n} A_i\right] = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)...P(A_n \mid A_1A_2...A_{n-1}).$$

Для условной вероятности сохраняются все свойства *безусловной вероятности*:

1) $0 \le P(A/B) \le 1$;
2) $P(\bar{A} \mid B) = 1 - P(A \mid B)$;
3) $P(A \mid A) = 1$;
4) $P([A_1 + A_2] \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 \cap A_2 \mid B)$;
5) если $A \cap B = \varnothing$, то $P(A/B) = 0$;
6) если событие $A \cap B \subset C$, то $P(C \mid AB) = 1$;
7) если $A \subset B$, то $P(B \mid A) = 1$.

Доказательства этих свойств вытекают из определения условной вероятности и свойств безусловной вероятности $P(A)$.

Пример 11.6. Известно, что в пятизначном номере телефона все цифры разные. Найти вероятность того, что среди них есть цифры 1 и 2.

Решение. Пусть событие A состоит в том, что в номере телефона есть цифра 1, событие B — в номере телефона есть цифра 2, событие C — все цифры разные. Тогда событие AB — в номере телефона есть цифры и 1, и 2. При вычислении искомой вероятности удобно перейти к противоположному событию. Используя формулы де Моргана, теорему сложения и учитывая свойства условной вероятности, получаем решение в виде следующей очевидной цепочки равенств:

$$P(AB/C) = 1 - P(\overline{AB}/C) = 1 - P((\bar{A} \cup \bar{B})/C) = 1 - (P(\bar{A}/C) + P(\bar{B}/C) -$$

$$- P((\bar{A} \cap \bar{B})/C)) = 1 - \left(\frac{A_9^5}{A_{10}^5} + \frac{A_9^5}{A_{10}^5} - \frac{A_8^5}{A_{10}^5} \right) = \frac{2}{9}.$$

Пусть событие A может произойти при условии появления одного из событий B_1, B_2, ..., B_n, образующих полную группу несовместных событий, и вероятности их до опыта известны. Такие события B_1, B_2, ..., B_n называются *гипотезами*.

Теорема 11.7. Вероятность события A равна сумме произведений вероятностей гипотез $P(B_i)$ на условные вероятности события A при каждой гипотезе B_i:

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A \mid B_i). \tag{11.1}$$

Формула (11.1) называется *формулой полной вероятности*.

Доказательство. Поскольку гипотезы B_i, $i = 1, 2, ..., n$ несовместны, то несовместны и их произведения с событием A, т.е. несовместны комбинации $A \cap B_i$ и $A \cap B_j$ при $i \neq j$. Обозначим $B = \bigcup_{i=1}^{n} B_i$.

Так как B_1, ..., B_n образуют полную группу событий, то событие B — достоверное событие и его вероятность равна единице: $P(B) = 1$. Из теоремы сложения, получаем $P(A \cap B) = P(A) + P(B) - P(A \cup B) =$ $= P(A) + 1 - 1 = P(A)$. Применяя теорему умножения вероятностей для каждого слагаемого, получаем окончательный результат

$$P(A) = P(A \cap B) = P\left(A \cap \left(\bigcup_{i=1}^{n} B_i \right) \right) = P\left(\bigcup_{i=1}^{n} (A \cap B_i) \right) =$$

$$= \sum_{i=1}^{n} P(B_i)P(A/B_i).$$

Возможна другая ситуация: пусть событие A произошло. Требуется найти вероятность того, что событие A произошло именно путем B_i.

Теорема 11.8. Пусть B_1, B_2, ..., B_k — полная группа событий и A — произвольное событие, которое может произойти с одним из них. Тогда для каждого i справедливо равенство

$$P(B_i \mid A) = [P(B_i)P(A \mid B_i)]/P(A). \tag{11.2}$$

Формула (11.2) называется *формулой Байеса.*

Доказательство. По определению условной вероятности

$$P(B_i \mid A) = P(A \cap B_i)/P(A) = [P(B_i)P(A \mid B_i)]/P(A),$$

где $P(A \cap B_i)$ найдена с помощью теоремы умножения, а $P(A)$ находится по формуле полной вероятности. Формула Байеса позволяет производить пересчет вероятностей гипотез B_i с учетом того, что событие A произошло.

Пример 11.7. Фирма имеет три источника поставки комплектующих — фирмы A, B, C. На долю фирмы A приходится 50% общего объема поставок, B — 30% и C — 20%. Из практики известно, что среди поставляемых фирмой A деталей брак составляет 10%, фирмой B — 5% и фирмой C — 6%. Какова вероятность, что взятая наугад и оказавшаяся бракованной деталь получена от фирмы A?

Решение. Пусть событие D — появление бракованной детали. Вероятности гипотез о том, что деталь поставлена фирмами A, B, C, равны соответственно $P(A) = 0{,}5$, $P(B) = 0{,}3$, $P(C) = 0{,}2$. Условные вероятности появления при этом бракованной детали будут равны $P(D/A) = 0{,}1$, $P(D/B) = 0{,}05$, $P(D/C) = 0{,}06$. По формуле Байеса получаем

$$P(A/D) = \frac{P(A)P(D/A)}{P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)} =$$

$$= \frac{0{,}5 \cdot 0{,}1}{0{,}5 \cdot 0{,}1 + 0{,}3 \cdot 0{,}05 + 0{,}2 \cdot 0{,}06} = 0{,}649.$$

Сложные события, представляющие собой серию опытов и комбинаций всех возможных исходов, можно представить с помощью «дерева» вероятностей, на котором отражаются последовательность экспериментов и их результаты (рис. 11.3). Опыты представлены последовательностью кружков, а каждый исход — «ветвью» (линией) от соответствующего кружка. Вероятность соответствующего исхода указана около «ветви». Затем вероятности, стоящие на одном пути, перемножаются, а результаты, полученные для различных путей, складываются.

На рис. 11.3 построена графическая схема для задачи 1 — «дерево» вероятностей, по которой вычисляется вероятность появления годной детали:

$$P(A) = 0{,}5 \cdot 0{,}9 + 0{,}3 \cdot 0{,}95 + 0{,}2 \cdot 0{,}94 = 0{,}923.$$

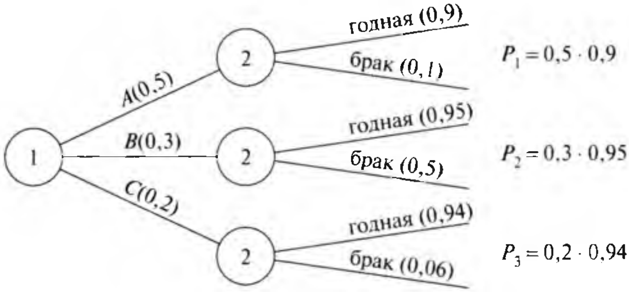

Рис. 11.3

11.5. АКСИОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ. ГЕОМЕТРИЧЕСКАЯ ВЕРОЯТНОСТЬ

Ограниченность «классического» определения вероятностей заложена в предположении равной возможности исходов. Многие реальные случайные эксперименты не укладываются также в рамки дискретной модели с конечным или счетным пространством Ω. Не всегда может помочь и геометрическая интерпретация.

Поэтому в случае бесконечного пространства Ω построение современной теории вероятностей базируется на подходе, предложенном великим русским математиком А.Н. Колмогоровым. Его основная идея заключается в том, что не все подмножества пространства Ω рассматриваются как события. Предполагается, что события — это некоторые подмножества из пространства элементарных событий Ω, совокупность которых замкнута относительно операций конечного или счетного числа объединений и пересечений.

Пусть Ω — произвольное пространство элементарных событий, а $\mathfrak{I}$ — некоторый класс подмножеств множества Ω.

Алгеброй событий $\mathfrak{I}$ называется любая непустая система подмножеств пространства Ω, удовлетворяющая следующим аксиомам:

1) если подмножество A принадлежит $\mathfrak{I}$ (является событием), то его дополнение $\bar{A}$ также принадлежит $\mathfrak{I}$ (также является событием);

2) если подмножества A и B принадлежат $\mathfrak{I}$ (являются событиями), то и их объединение $A \cup B$ принадлежит $\mathfrak{I}$ (также является событием).

Поскольку любую из операций над подмножествами можно получить, используя формулы де Моргана, с помощью только двух операций дополнения и объединения $A \cap B = \overline{\bar{A} \cup \bar{B}}$, $A \setminus B = A \cap \bar{B} =$

$= \overline{\bar{A} \cup \bar{B}}$, то пересечение и разность двух событий также будут событиями: $A \cap B \in \Im$, $A \setminus B \in \Im$ при любых $A \in \Im$, $B \in \Im$. Отсюда следует, что $\Omega \in \Im$ и $\varnothing = \Omega \setminus \Omega \in \Im$ — тоже события.

Алгебра событий $\Im$ называется *σ-алгеброй*, если объединение счетного числа элементов из $\Im$ также является элементом $\Im$, т.е. из того, что $A_n \in \Im$, $n = 1, 2, \ldots$, следует $\bigcup_{n=1}^{\infty} A_n \in \Im$.

Таким образом, σ-алгебру событий $\Im$ можно определить как систему подмножеств пространства элементарных исходов Ω, замкнутую относительно счетного числа теоретико-множественных операций. Тривиальная σ-алгебра событий состоит из полного и пустого множеств: $\Im = \{\varnothing, \Omega\}$. Любая σ-алгебра событий является одновременно и алгеброй событий. Обратное, вообще говоря, неверно, т.е. существуют алгебры событий, не являющиеся σ-алгебрами.

Теперь согласно аксиоматике Колмогорова можно ввести общее понятие вероятности события.

Вероятностью события или **вероятностной мерой** называется числовая функция, заданная на σ-алгебре событий $\Im$, которая каждому событию $A \in \Im$ ставит в соответствие число $P(A)$ так, что выполняются следующие четыре аксиомы.

1. **Аксиома неотрицательности**: $P(A) \geq 0$ для всех $A \in \Im$.
2. **Аксиома нормированности**: $P(\Omega) = 1$.
3. **Аксиома конечной аддитивности**: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ для любых $A_i \cap A_j = \varnothing$, $i \neq j$, и $A_i \in \Im$ для любого $i = 1, 2, \ldots, n$.
4. **Аксиома счетной аддитивности**: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, если события A_i в последовательности $A_1, A_2, \ldots, A_n, \ldots$ несовместны, т.е. для любых $A_i \cap A_j = \varnothing$, $i \neq j$, $A_i \in \Im$ для любого $i = 1, 2, \ldots$

Очевидно, что вероятность, определенная в дискретном вероятностном пространстве условием $P(A) = \sum_{i:\omega_i \in A} P(\omega_i)$, является счетно-аддитивной.

Заметим также, что введенные аксиомы в случае дискретного пространства превращаются в доказуемые утверждения.

Конечно-аддитивная вероятность $P(A_n)$, заданная на σ-алгебре множеств $\Im$, называется *непрерывной*, если для любой убывающей последовательности множеств $A_n \subseteq \Im$, $n = 1, 2, \ldots$ такой, что $A_{n+1} \subseteq A_n$, имеющих пустое пересечение $\bigcap_{n=1}^{\infty} A_n = \varnothing$, имеет место равенство $\lim_{n\to\infty} P(A_n) = 0$.

5. **Аксиома непрерывности.** Если последовательность событий A_1, A_2, ..., A_n,... такова, что каждое последующее влечет за собой предыдущее, а пересечение всех событий A_n пусто, то $P(A_n) \to 0$ при $n \to \infty$.

Теорема 11.9. Аксиома счетной аддитивности равносильна аксиоме непрерывности.

Доказательство. Докажем сначала, что из непрерывности следует счетная аддитивность. Рассмотрим последовательность несовместных событий B_n, и пусть $B = \bigcup_{i=1}^{\infty} B_i$. Введем события вида $A_n = \bigcup_{i=n+1}^{\infty} B_i$, тогда они удовлетворяют условиям аксиомы непрерывности и $P(A_n) \to 0$ при $n \to \infty$. По теореме сложения получаем

$$P(B) = P\left(\bigcup_{i=1}^{n} B_i\right) + P\left(\bigcup_{i=n+1}^{\infty} B_i\right) = \sum_{i=1}^{n} P(B_i) + P(A_n),$$

откуда

$$\sum_{i=1}^{n} P(B_i) = P(B) - P(A_n) \to P(B) = P\left(\bigcup_{i=1}^{\infty} B_i\right), \quad n \to \infty,$$

что и означает счетную аддитивность.

Докажем теперь, что из счетной аддитивности следует непрерывность. Пусть последовательность A_n удовлетворяет условиям аксиомы непрерывности. Введем события $C_n = A_n \setminus A_{n+1}$. Тогда эти события несовместны, причем допустимы представления:

$$A_1 \setminus A_n = \bigcup_{i=1}^{n-1} C_i, \quad A_1 = \bigcup_{i=1}^{\infty} C_i.$$

По теореме сложения получаем

$$P(A_1) = P(A_1 \setminus A_n) + P(A_n) = P\left(\bigcup_{i=1}^{n-1} C_i\right) + P(A_n) = \sum_{i=1}^{n-1} P(C_i) + P(A_n),$$

откуда

$$P(A_n) = P(A_1) - \sum_{i=1}^{n-1} P(C_i) = P\left(\bigcup_{i=1}^{\infty} C_i\right) - \sum_{i=1}^{n-1} P(C_i) \to 0, \quad n \to \infty,$$

так что аксиома непрерывности выполняется.

Тройка (Ω, $\Im$, P), где Ω — пространство элементарных событий; $\Im$ — σ-алгебра подмножеств Ω, называемых событиями; P — вероятностная мера, определенная на событиях, называется *вероятностным пространством.*

Далее будем всюду неявно предполагать, что любые рассматриваемые множества относятся к некоторой σ-алгебре $\Im$, а вероятность удовлетворяет всем необходимым аксиомам.

Пусть рассматривается непрерывная вероятностная модель, т.е. пространство элементарных событий Ω — некоторая область (отрезок, многоугольник, круг, шар и т.д.), имеющая меру $\mu(\Omega)$ (длину, площадь, объем и т.д.) такую, что $0 < \mu(\Omega) < \infty$, и в ней содержится другая подобласть A с мерой $\mu(A)$. Говорят, что точка равномерным образом попадает в пространство Ω (реализуется принцип геометрической вероятности), если вероятность $P(A)$ попадания ее в каждую область A, являющуюся подмножеством пространства Ω, пропорциональна мере этой области $\mu(A)$: $P(A) = \mu(A)/\mu(\Omega)$.

Пример 11.8. Стержень разламывается на две части в случайной точке, равномерно распределенной по длине стержня. Найти вероятность того, что меньший обломок имеет длину, не превосходящую одной трети длины стержня.

Решение. Обозначим длину стержня L, а расстояние точки разлома от одного (например, левого) конца стержня — x. Тогда описанное событие произойдет при условии, если либо $x \le \frac{L}{3}$, либо $x \ge \frac{2L}{3}$. Искомая вероятность равна отношению $P(A) = \dfrac{\frac{L}{3} + \frac{L}{3}}{L} = \dfrac{2}{3}$ (рис. 11.4).

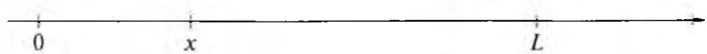

Рис. 11.4

Пример 11.9 (задача о встрече). Два лица A и B условились встретиться в определенном месте между 12 и 13 часами. Пришедший первым ждет другого в течении 20 минут, после чего уходит. Чему равна вероятность встречи лиц A и B, если приход каждого из них может произойти случайно в течение указанного часа и моменты прихода независимы?

*Решение.*Обозначим моменты прихода лица A через x и лица B через y. Для того чтобы встреча произошла, необходимо и достаточно, чтобы $|x - y| \le 20$. Изобразим x и y как координаты на плоскости, в качестве единицы масштаба выберем минуту. Всевозможные исходы представятся точками квадрата со стороной 60, благоприятствующие встрече — расположатся в заштрихованной области. Искомая вероятность равна отношению площади заштрихованной фигуры (рис. 11.5) к площади всего квадрата:

$p = (60^2 - 40^2)/60^2 = 5/9.$

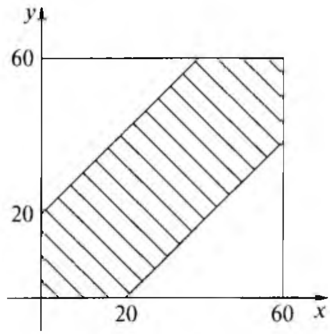

Рис. 11.5

Пример 11.10 (*задача Бюффона*). Плоскость разграфлена параллельными прямыми, отстоящими друг от друга на расстоянии $2a$. На плоскость наудачу бросается игла длиной $2l$ ($l < a$). Найти вероятность того, что игла пересечет какую-нибудь прямую.

Решение. Если игла бросается с достаточной высоты и ее начальное положение случайно, то под словом «наудачу» подразумевается, во-первых, что центр иглы наудачу попадет на отрезок длиной $2a$; во-вторых, угол φ между прямой и иглой равномерно распределен на отрезке $[0; \pi]$; в-третьих, на величину угла не влияет расстояние от центра до прямой. Поэтому изобразим результат бросания точкой с координатами (φ, x), лежащей внутри прямоугольника со сторонами a и π, где x — расстояние от центра иглы до ближайшей прямой. Из рис. 11.6, *а* видно, что пересечение иглы с прямой происходит тогда и только тогда, когда $x < l \cdot \sin\varphi$. Искомая вероятность равна отношению площади заштрихованной области A к площади прямоугольника на рис. 11.6, *б*:

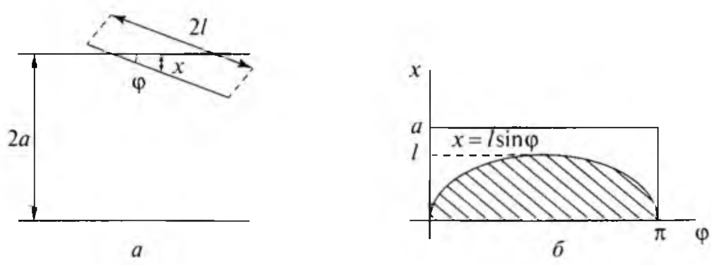

Рис. 11.6

$$P(A)=\frac{1}{\pi a}\int_0^{\pi} l\sin\varphi d\varphi=\frac{2l}{\pi a}.$$

Далее будем всюду неявно предполагать, что любые рассматриваемые множества относятся к некоторой σ-алгебре $\Im$, а вероятность удовлетворяет всем необходимым аксиомам.

11.6. НЕЗАВИСИМОСТЬ СОБЫТИЙ. СХЕМА БЕРНУЛЛИ

Событие A не зависит от события B, если появление события B не меняет значения вероятности события A, т.е. условная вероятность события A равна безусловной вероятности: $P(A \mid B) = P(A)$ в предположении, что $P(B) > 0$. Тогда из теоремы умножения $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$ в предположении, что $P(A) > 0$, получим, что условная вероятность события B равна его безусловной вероятности: $P(B/A) = P(B)$.

Следовательно, события A и B независимы, если появление одного из них не меняет вероятности появления другого.

Из определения независимости двух событий следует теорема.

Теорема 11.10. События A и B независимы, если вероятность произведения событий равна произведению их вероятностей: $P(A \cap B) = P(A)P(B)$.

Это равенство используется как критерий при практической проверке независимости двух событий.

Понятие независимости можно обобщить на любое конечное число событий.

События $A_1, A_2, ..., A_n$ называются *независимыми в совокупности*, если для любого их подмножества $A_{i1}, A_{i2}, ..., A_{ik}$ имеет место равенство $P(A_{i1}A_{i2}...A_{ik}) = P(A_{i1})P(A_{i2})...P(A_{ik})$ для любых k от 1 до n и любых несовпадающих номеров $i_1, i_2, ..., i_k$.

События $A_1, ..., A_n$ называются *попарно независимыми*, если для любых $i \neq j$, $i, j \in \{1, ..., n\}$ события A_i и A_j независимы.

Из определений следует, что из независимости в совокупности следует попарная независимость, но из попарной независимости не следует независимости в совокупности.

Пример 11.11 (пример Бернштейна). На плоскость бросается правильный тетраэдр (треугольная пирамида), три грани которого покрашены в цвета: красный, синий и зеленый, а на четвертую грань нанесены все три цвета. Событие A — при бросании выпала красная грань, событие

B — синяя грань, событие C — зеленая грань. Вероятности этих событий равны между собой и равны

$P(A) = P(B) = P(C) = {}^2/_4 = {}^1/_2$.

Найдем вероятности их попарных произведений:

$P(AB) = {}^1/_4 = {}^1/_2 \cdot {}^1/_2 = P(A)P(B)$; $\quad P(AC) = P(A)P(C) = {}^1/_4 = {}^1/_2 \cdot {}^1/_2$,

$P(BC) = P(B)P(C) = {}^1/_2 \cdot {}^1/_2 = {}^1/_4$.

Отсюда следует, что они попарно независимы. Однако вероятность появления всех трех цветов равна

$P(ABC) = {}^1/_4 \neq P(A)P(B)P(C) = {}^1/_2 \cdot {}^1/_2 \cdot {}^1/_2 = {}^1/_8$,

т.е. вероятность произведения всех трех событий не равна произведению вероятностей этих событий и, следовательно, они зависимы в совокупности.

Пусть некоторый эксперимент повторяется n раз, причем результаты каждого повторения не зависят от исходов предыдущих повторений. Такие серии повторений называют независимыми испытаниями. Частным случаем таких испытаний являются независимые испытания Бернулли, которые характеризуются двумя условиями:

1) результатом каждого испытания является один из двух возможных исходов, называемых соответственно «успехом» или «неуспехом»;

2) вероятность «успеха» в каждом последующем испытании не зависит от результатов предыдущих испытаний и остается постоянной.

Построим вероятностную модель, соответствующую n-кратной серии независимых испытаний Бернулли.

Исходный опыт описывается пространством элементарных исходов $\Omega = \{У, Н\}$, состоящим из двух элементов: У(успех) — появление события A и Н(неуспех) — непоявление события A, а также их вероятностями $P(У) = p$, $P(Н) = q$, $p + q = 1$.

Составной эксперимент, включающий n повторений исходного опыта, задается пространством Ω_n, каждый элемент которого представляет собой упорядоченный n-мерный набор конкретных результатов повторений исходного опыта. Обозначим через B_m событие, состоящее в том, что в n опытах событие A появится ровно m раз. Разложим событие B_m на сумму произведений событий, состоящих в появлении и непоявлении события A в отдельном опыте, при этом обозначим A_i — появление события A в i-м опыте и $\bar{A}_i$ — непоявление A в i-м опыте. Тогда каждый вариант события B_m состоит из m появлений события A и $n - m$ непоявлений события A, т.е.

$$B_m = A_1 A_2 \ldots \bar{A}_{n-1} \bar{A}_n + A_1 \bar{A}_2 \ldots \bar{A}_{n-1} A_n + \ldots + \bar{A}_1 \bar{A}_2 \ldots A_{n-1} A_n.$$

Число всех комбинаций такого рода равно числу способов, какими можно из n элементов одновременно выбрать m элементов, соответствующих m появлениям события A, т.е. числу сочетаний C_n^m. Вероятность каждой такой комбинации (каждого слагаемого) по теореме умножения независимых событий равна $p^m q^{n-m}$, а так как составляющие событие B_m являются несовместными событиями, то согласно теореме сложения несовместных событий $P(B_m) = P_n(m) = C_n^m p^m q^{n-m}$. Итак, доказана следующая теорема.

Теорема 11.11. Если производится серия из n независимых испытаний Бернулли, в каждом из которых событие A появляется с вероятностью p, то вероятность того, что в n испытаниях событие A появится ровно m раз, выражается формулой

$$P_n(m) = C_n^m p^m q^{n-m},$$

где $q = 1 - p$.

Эта формула называется *формулой Бернулли*.

Вероятности $P_n(m)$ называются *биномиальными вероятностями*.

Пример 11.12. Аудитор обнаруживает финансовые нарушения с вероятностью 0,9. Найти вероятность того, что среди четырех фирм-нарушителей будет выявлено больше половины.

Решение. Событие состоит в том, что из четырех фирм-нарушителей будет выявлено три или четыре, т.е. требуется найти вероятности

$$P_4(3) + P_4(4) = C_4^3 0{,}9^3 \cdot 0{,}1 + C_4^4 0.9^4 = 0{,}9^3(0{,}4 + 0{,}9) = 0{,}9477.$$

Число m, при котором биномиальные вероятности $P_n(m)$ достигают своего максимального значения (при фиксированном числе испытаний n), называют *наиболее вероятным (наивероятнейшим) числом успехов*.

Теорема 11.12. Наивероятнейшее число успехов m^* в серии из n независимых испытаний Бернулли (с вероятностью успеха p в одном испытании) определяется соотношением $np - q \le m^* \le np + p$, причем:

1) если число $np - q$ — дробное, то существует одно наивероятнейшее число m^*;

2) если число $np - q$ — целое, то существует два наивероятнейших числа:

$$m^* = np - q, \quad m^* = np + p;$$

3) если np — целое число, то наивероятнейшее число $m^* = np$.

Пример 11.13. Монета подбрасывается три раза. Найти наиболее вероятное число успехов (выпадений герба).

Решение. Возможными значениями для числа успехов в трех рассматриваемых испытаниях являются $m = 0, 1, 2$ или 3. Пусть A_m — событие, состоящее в том, что при трех подбрасываниях монеты герб появляется m раз. По формуле Бернулли легко найти вероятности событий A_m (см. таблицу).

m	0	1	2	3
$P_n(m)$	$^1/_8$	$^3/_8$	$^3/_8$	$^1/_8$

В данном случае числа $(n + 1)p$ и $np - q = (n + 1)p - 1$ целые, так что наиболее вероятными являются два значения числа успехов: $m^* = 1$ и $m^* = 2$.

11.7. ПРИБЛИЖЕННЫЕ ВЫЧИСЛЕНИЯ ВЕРОЯТНОСТЕЙ В СХЕМЕ БЕРНУЛЛИ

При больших значениях n непосредственное нахождение вероятностей $P_n(m)$ по формуле Бернулли сопряжено с трудностями вычислительного порядка. В таких случаях используют различные варианты приближенных формул, основанных на предельных теоремах Пуассона и Муавра—Лапласа.

Приближенная формула Пуассона используется в том случае, когда число испытаний Бернулли (n) велико, а вероятность успеха в отдельном испытании мала ($n \geq 100$, $p < 0,1$, $np < 10$). Тогда для вычисления биномиальных вероятностей $P_n(m)$ используют формулу Пуассона $P_n(m) = \frac{\lambda^m}{m!}e^{-\lambda}$, которая является предельной для формулы Бернулли.

Теорема 11.13 (теорема Пуассона). Пусть число испытаний n в схеме Бернулли велико, а вероятность успеха мала, причем $np \to \lambda =$ $= \text{const}$ при $n \to \infty$. Тогда

$$P_n(m) \to \frac{\lambda^m}{m!}e^{-\lambda}, \quad m = 0, 1, 2, ..., n; \; n \to \infty.$$

Доказательство. В формуле Бернулли после умножения числителя и знаменателя на n^m и некоторых преобразований получаем

$$P_n(m) = C_n^m p^m q^{n-m} = \frac{n(n-1)...(n-m+1)p^m}{m!}(1-p)^{n-m} =$$

$$= \frac{n(n-1)...(n-m+1)n^m}{m!n^m(1-p)^m} p^m \left(1 - \frac{np}{n}\right)^n =$$

$$= \frac{1\cdot\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)}{m!}(np)^m\left(1-\frac{np}{n}\right)^n\frac{1}{(1-p)^m}\to$$

$$\to\frac{\lambda^m}{m!}e^{-\lambda},\quad n\to\infty,\quad p\to 0.$$

Предельные вероятности $P_n(m)\approx\frac{(np)^m e^{-np}}{m!}$ называются *пуассоновскими*. Поскольку при больших n верно $np\approx\lambda$, то можно считать, что $\lambda=np$.

Формула выражает закон распределения Пуассона вероятностей *массовых* (n велико) и *редких* (p мало) явлений. Отсюда название закона Пуассона — *закон редких явлений*. Закон Пуассона широко применяется в теории информации, в теории массового обслуживания при изучении потока событий.

В силу определенной «симметричности» понятий «успех» и «неудача» приближенная формула Пуассона может использоваться в схеме независимых испытаний Бернулли при больших n также и в случае, когда p близко к единице, т.е. при $q<0{,}1$ и $nq<10$:

$$P_n(n-m)=C_n^{n-m}p^{n-m}q^m=C_n^m p^{n-m}q^m\approx\frac{\lambda^m}{m!}e^{-\lambda},\quad \lambda=nq.$$

Пример 11.14. Известно, что процент брака для некоторой детали равен 0,5%. Контролер проверяет 1000 деталей. Какова вероятность обнаружить ровно три бракованные детали? Какова вероятность обнаружить не меньше трех бракованных деталей?

Решение. Имеем 1000 испытаний Бернулли с вероятностью «успеха» $p=0{,}005$. Применяя пуассоновское приближение с $\lambda=np=5$, получаем:

1) $P_{1000}(3)\approx\frac{5^3}{3!}e^{-5}$;

2) $P_{1000}(m\ge 3)=1-P_{1000}(m<3)=1-[P_{1000}(0)+P_{1000}(1)+P_{1000}(2)]\approx$

$\approx 1-\sum\limits_{m=0}^{2}\frac{5^m}{m!}e^{-5}$ — и по таблице значений функции Пуассона находим $P_{1000}(3)\approx 0{,}14$; $P_{1000}(m\ge 3)\approx 0{,}875$.

В силу определенной «симметричности» понятий «успех» и «неудача» приближенная формула Пуассона может использоваться в схеме независимых испытаний Бернулли также и в случае, когда p близко к единице, а q — мало, что эквивалентно переходу к противоположному событию. Тогда

$$P_n(n-m)=C_n^{n-m}p^{n-m}q^m=C_n^m p^{n-m}q^m\approx\frac{(nq)^m}{m!}e^{-nq}.$$

Приближенные формулы Муавра—Лапласа используются в том случае, когда наряду с числом испытаний n велико также значение произведения np.

Теорема 11.14. Если вероятность p появления события A в каждом из независимых испытаний постоянна и отлична от нуля и единицы, то вероятность $P_n(m)$ того, что в n испытаниях событие A появится ровно m раз, приближенно равна $P_n(m) \approx \dfrac{\varphi(x)}{\sqrt{npq}}$, где $\varphi(x) = \dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$; $x = \dfrac{m - np}{\sqrt{npq}}$.

Функция $\varphi(x)$ четная, и для положительных значений x составлена таблица ее значений.

Пример 11.15. Из 100 потенциальных покупателей каждый совершает покупку с вероятностью 0,8. Найти наиболее вероятное число покупателей и его вероятность.

Решение. Из условия задачи следует: $n = 100$; $p = 0{,}8$; $np = 80$; $\sqrt{npq} = 4$. Поскольку $np = 80$ — целое число, то наиболее вероятное значение $m^* = np = 80$. Вероятность

$$P_{100}(80) = \frac{1}{\sqrt{npq}}\varphi(x) = \frac{1}{4}\varphi(0) = \frac{1}{4}0{,}3989 = 0{,}0997,$$

где $\dfrac{m - np}{\sqrt{npq}} = \dfrac{80 - 80}{4} = 0$.

Для вычисления вероятности $P_n(m_1, m_2)$ события, состоящего в том, что число успехов в n независимых испытаниях Бернулли окажется заключенным в пределах от m_1 до m_2, используется следующая теорема.

Теорема 11.15 (интегральная теорема Муавра—Лапласа). Если вероятность p наступления события A в каждом из независимых испытаний постоянна и отлична от нуля и единицы, то вероятность $P_n(m_1, m_2)$ того, что в n испытаниях событие A появится от m_1 до m_2 раз, приближенно равна определенному интегралу

$$P_n(m_1, m_2) \approx \frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{z^2}{2}}\,dz,$$

где $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$, $\quad x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$.

Интеграл $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ называется функцией Лапласа, и для ее значений имеются специальные таблицы при положительных значениях аргумента. Функция Лапласа нечетная, т.е. $\Phi_0(-x) = -\Phi_0(x)$, что следует из равенства

$$\Phi_0(-x) = \frac{1}{\sqrt{2\pi}} \int_0^{-x} e^{-\frac{z^2}{2}} dz = |z = -t, \ \ dz = -dt| = -\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt = -\Phi_0(x).$$

Используя функцию Лапласа, можно записать следующую приближенную формулу:

$$P_n(m_1, m_2) \approx \Phi_0(x_2) - \Phi_0(x_1),$$

где $x_1 = \frac{m_1 - np}{\sqrt{npq}}$; $\quad x_2 = \frac{m_2 - np}{\sqrt{npq}}$.

Пример 11.16. Страховая компания заключила 40 тыс. договоров. Вероятность страхового случая по каждому в течение года составляет 2%. Найти вероятность, что таких случаев будет не более 870.

Решение. По условию задачи $n = 40\,000$, $p = 0{,}02$, $np = 800$, $\sqrt{npq} = 28$. Для вычисления вероятности $P(m \le 870)$ воспользуемся интегральной теоремой Лапласа:

$$P(0 < m \le 870) = \Phi_0(x_2) - \Phi_0(x_1),$$

где $x_1 = \frac{0 - 800}{28} = -28{,}57$ и $x_2 = \frac{870 - 800}{28} = 2{,}5$.

Находим по таблицам значения функции Лапласа:

$$P(0 < m \le 870) = \Phi_0(x_2) - \Phi_0(x_1) = \Phi_0(2{,}5) - \Phi_0(-28{,}57) =$$
$$= 0{,}4938 + 0{,}5 = 0{,}9938.$$

Следствие. Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события равна p, $0 < p < 1$, абсолютная величина отклонения относительной частоты от вероятности появления события не превысит положительного числа ε, приближенно равна удвоенной функции Лапласа при $x = \varepsilon\sqrt{\frac{n}{pq}}$, т.е. $P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$.

Доказательство получим из следующей цепочки очевидных равенств:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = P\left(p - \varepsilon \le \frac{m}{n} \le p + \varepsilon\right) = P(np - n\varepsilon \le m \le np + n\varepsilon) \approx$$

$$\approx \Phi_0\left(\frac{np + n\varepsilon - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{np - n\varepsilon - np}{\sqrt{npq}}\right) =$$

$$= \Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - \Phi_0\left(-\varepsilon\sqrt{\frac{n}{pq}}\right) = 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Пример 11.17. Каждый из 900 посетителей оптового рынка случайным образом обращается в один из 10 ларьков. В каких границах с вероятностью 0,95 лежит число клиентов отдельно взятого ларька?

Решение. Из условия задачи следует: $n = 900$, $p = 0,1$ — и заданная вероятность $P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 0,95$. Отсюда следует, что $\Phi_0(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}) = 0,475$. По таблице находим, что аргумент функции равен $\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} = 1,96$, и получаем значение $\varepsilon = 0,0196$. Тогда из условия $\left|\frac{m}{900} - \frac{1}{10}\right| \le$ $\le 0,0196$ найдем промежуток для m: $|m - 90| \le 17,64$ или $72,36 \le m \le 107,64$, и так как m — число целое, то $73 \le m \le 107$.

Глава 12
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

12.1 СЛУЧАЙНАЯ ВЕЛИЧИНА В ДИСКРЕТНОМ ВЕРОЯТНОСТНОМ ПРОСТРАНСТВЕ

Случайной величиной x в вероятностном пространстве называется любая действительная функция $x = x(\omega)$, $\omega \in \Omega$, определенная на пространстве элементарных событий Ω. В результате опыта случайная величина может принять то или иное значение, причем заранее неизвестно, какое именно. Случайные величины принято обозначать греческими буквами ξ, η, μ и т.д. Примеры случайных величин: курс доллара, цены товаров, прибыль или убытки фирмы, время ожидания транспорта и т.д.

Случайная величина, принимающая конечное или счетное множество значений, называется *дискретной*.

Пример 12.1. Проводится эксперимент — двукратное подбрасывание монеты. Тогда пространство элементарных событий имеет вид Ω = (ГГ, ГР, РГ, РР) = $(\omega_1, \omega_2, \omega_3, \omega_4)$. Если случайная величина $\xi(\omega)$ — число выпадений герба при двукратном бросании монеты, то она принимает значения $\xi(\omega_1) = 2$, $\xi(\omega_2) = \xi(\omega_3) = 1$, $\xi(\omega_4) = 0$. Вероятности этих значений вычисляются по формуле $P(\xi(\omega) \in A) = \sum\limits_{j:\xi(\omega_j)\in A} p_j$. Зная вероятности элементарных исходов, можно вычислить вероятности соответствующих значений случайной величины ξ:

$$P(\xi = 2) = P(\omega_1) = {}^1/_4,\ P(\xi = 1) = P(\omega_2, \omega_3) = {}^1/_2,\ P(\xi = 0) = P(\omega_4) = {}^1/_4.$$

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Простейшей формой закона распределения дискретной случайной величины является ряд распределения.

Рядом распределения дискретной случайной величины ξ называется таблица, где перечислены все возможные значения этой случайной величины и соответствующие им вероятности.

Пусть $x_1, x_2, \ldots, x_n$ — различные значения случайной величины $\xi(\omega)$, заданной на дискретном пространстве Ω, и $p_1, p_2, \ldots, p_n$ — соответствующие вероятности этих значений. Тогда рядом распределения будет таблица вида

ξ	x_1	x_2	...	x_n	
p_i	p_1	p_2	...	p_n	$\sum_{i=1}^{n} p_i = 1$

Графическое изображение ряда распределения называется многоугольником распределения (рис. 12.1).

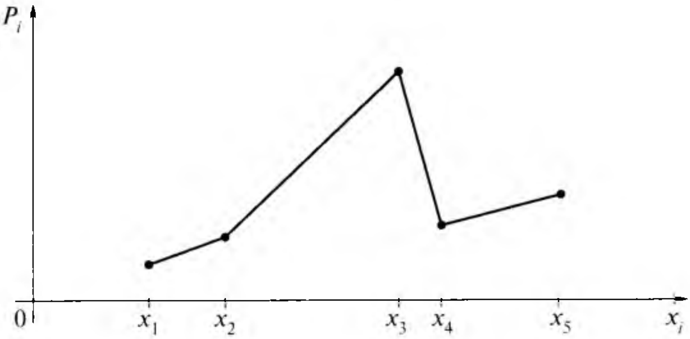

Рис. 12.1

Примеры распределений дискретных случайных величин.

1. Пусть случайная величина $\xi(\omega)$ — число успехов в n испытаниях Бернулли, при этом вероятности успеха и неудачи соответственно равны $P(\text{У}) = p$ и $P(\text{Н}) = 1 - p = q$. Поскольку вероятности m успехов в n испытаниях вычисляются по формуле $P_n(\xi = m) = C_n^m p^m q^{n-m}$, $m = 0, 1, 2, ..., n$, то ряд распределения имеет следующий вид:

$$\begin{bmatrix} 0 & 1 & \dots & m & \dots & n \\ q^n & C_n^1 pq^{n-1} & \dots & C_n^m p^m q^{n-m} & \dots & p^n \end{bmatrix}.$$

Полученная таблица действительно является рядом распределения, так как сумма биномиальных вероятностей равна единице:

$$\sum_{m=0}^{n} P_n(\xi = m) = q^n + C_n^1 pq^{n-1} + C_n^2 p^2 q^{n-2} + \dots + p^n = (1 - p + p)^n = 1.$$

Случайная величина ξ, равная числу «успехов» в n испытаниях Бернулли, имеет биномиальный закон распределения.

2. Распределение Пуассона ($\lambda > 0$) задается формулой $P(\xi = m) = [\lambda^m / m!]e^{-\lambda}$, $m = 0, 1, 2, ...$, и ряд распределения Пуассона имеет вид

$$\begin{bmatrix} 0 & 1 & \ldots & m & \ldots \\ e^{-\lambda} & \lambda e^{-\lambda} & \ldots & \dfrac{\lambda^m}{m!}e^{-\lambda} & \ldots \end{bmatrix},$$

так как сумма вероятностей равна единице:

$$\sum_{m=0}^{\infty} P(\xi = m) = \sum_{m=0}^{\infty} \frac{\lambda^m e^{-\lambda}}{m!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

3. Примерами распределений дискретных случайных величин являются также распределения случайных величин, имеющих:

а) гипергеометрическое распределение, определяемое формулой

$$P(\xi = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m = 0, 1, 2, \ldots, \min(n, M);$$

б) геометрическое распределение, заданное формулой

$$P(\xi = m) = p \cdot q^{m-1}, \quad m = 1, 2, \ldots, \quad 0 < p < 1, \quad q = 1 - p.$$

Геометрическое распределение имеет случайная величина ξ, равная числу испытаний Бернулли до первого «успеха» с вероятностью «успеха» в одном испытании, равной p.

12.2. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ФУНКЦИЯ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Функцией распределения случайной величины ξ называется функция $F_\xi(x)$, определенная для любого действительного x и выражающая вероятность того, что случайная величина ξ примет значение, меньшее x: $F_\xi(x) = P(\xi < x)$.

Функция распределения $F_\xi(x)$ является универсальным законом распределения случайной величины, и ее называют также интегральной функцией распределения или интегральным законом распределения.

Рассмотрим свойства функции распределения.

1. Как любая вероятность, функция распределения неотрицательна и не превосходит единицу, т.е. $0 \le F_\xi(x) \le 1$ для любого $x \in \Re$.

2. Функция распределения — неубывающая функция, т.е. $F_\xi(x_2) \ge F_\xi(x_1)$ для любых $x_2 \ge x_1$.

Доказательство. Представим событие, состоящее в том, что случайная величина примет значение, меньшее x_2, в виде суммы несовместных событий:

$$\{\omega: \xi(\omega) < x_2\} = \{\omega: \xi(\omega) < x_1\} \cup \{\omega: x_1 \le \xi(\omega) < x_2\}.$$

Вероятность суммы несовместных событий равна сумме вероятностей этих событий:

$$F_\xi(x_2) = P(\xi(\omega) < x_2) = P(\xi(\omega) < x_1) + P(x_1 \le \xi(\omega) < x_2) =$$
$$= F(x_1) + P(x_1 \le \xi(\omega) < x_2),$$

и так как вероятность неотрицательна: $P(x_1 \le \xi(\omega) \le x_2) \ge 0$, то получаем неравенство $F_\xi(x_2) \ge F_\xi(x_1)$.

3. Вероятность того, что случайная величина примет значение из полуинтервала $[x_1; x_2)$, равна разности значений функции распределения на концах интервала, т.е. $P(x_1 \le \xi < x_2) = F_\xi(x_2) - F_\xi(x_1)$.

Доказательство вытекает из равенства $F_\xi(x_2) = F_\xi(x_1) + P(x_1 \le$ $\le \xi(\omega) < x_2)$.

4. Справедливы следующие предельные соотношения:

$$\lim_{x\to-\infty} F_\xi(x) = 0; \qquad \lim_{x\to+\infty} F_\xi(x) = 1.$$

Для доказательства достаточно рассмотреть функцию распределения как вероятность $F_\xi(x) = P(\xi < x)$, заданную на числовой прямой.

5. Функция распределения непрерывна слева, т.е. $\lim\limits_{x\to a-0} F_\xi(x) =$ $_\xi($ $= F_\xi(a)$.

6. Справедливо следующее равенство: $P(\xi \ge x) = 1 - F_\xi(x)$.

Действительно, событие $\xi(\omega) \ge x$ является противоположным событию $\xi(\omega) < x$, и следовательно, $P(\xi \ge x) = 1 - P(\xi < x) = 1 - F_\xi(x)$.

Таким образом, каждая функция распределения является неубывающей, непрерывной слева и удовлетворяющей предельным условиям $F(-\infty) = 0$ и $F(+\infty) = 1$. Верно и обратное: каждая функция, удовлетворяющая перечисленным условиям, может рассматриваться как функция распределения некоторой случайной величины.

Функция распределения — универсальный закон распределения случайной величины, и все доказанные свойства функции распределения остаются верными, если пространство элементарных событий не является дискретным.

Функция распределения любой дискретной величины разрывная, возрастает скачками при тех значениях x, которые являются возможными значениями ξ. Величина скачка функции $F_\xi(x)$ в точке x_i равна p_i.

Так, в приведенном в §12.1 примере 1 ряд распределения имеет вид

ξ	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

и функция распределения будет равна $F_\xi(x) = \begin{cases} 0, & x \le 0, \\ \frac{1}{4}, & 0 < x \le 1, \\ \frac{3}{4}, & 1 < x \le 2, \\ 1, & x > 2 \end{cases}$

(рис. 12.2).

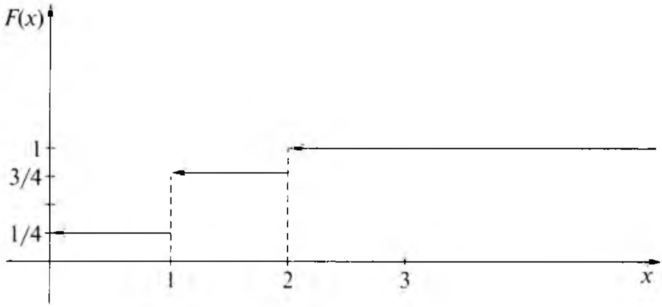

Рис. 12.2

Рассмотрим теперь, когда множество значений непрерывной случайной величины несчетно и обычно представляет собой некоторый промежуток, конечный или бесконечный.

Случайная величина $\xi(\omega)$ называется *непрерывной*, если непрерывна ее функция распределения.

Случайная величина называется *абсолютно непрерывной*, если существует неотрицательная функция $f_\xi(x)$ такая, что при любых x функцию распределения $F_\xi(x)$ можно представить в виде интеграла $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt.$

Функция $f_\xi(x)$ называется *плотностью распределения* вероятностей случайной величины ξ.

Из определения вытекают следующие свойства плотности вероятностей.

1. В точках непрерывности плотность распределения равна производной функции распределения: $f_\xi(x) = F'_\xi(x)$, и так как функция распределения — неубывающая функция, то ее производная неотрицательна: $F'_\xi(x) = f_\xi(x) \ge 0$.

2. Интеграл по всей числовой прямой от плотности распределения вероятностей равен единице.

Доказательство следует из равенства

$$\int_{-\infty}^{\infty} f_\xi(t)dt = \lim_{x\to\infty}\int_{-\infty}^{x} f_\xi(t)dt = \lim_{x\to\infty} F_\xi(x) = 1.$$

3. Плотность распределения определяет закон распределения случайной величины.

Действительно, для любых $x_1 < x_2$

$$P(x_1 \le \xi < x_2) = F_\xi(x_2) - F_\xi(x_1) = \int_{-\infty}^{x_2} f_\xi(t)dt - \int_{-\infty}^{x_1} f_\xi(t)dt = \int_{x_1}^{x_2} f_\xi(t)dt.$$

Вероятность того, что непрерывная случайная величина примет конкретное значение, равна нулю: $P(\xi = a) = 0$.

Представим событие $A = \{\omega\colon \xi = a\}$ в виде произведения $\bigcap_{n=1}^{\infty}(a \le \xi < a + 1/n)$, тогда

$$P(A) = P(\xi = a) = \lim_{n\to\infty} P(a \le \xi < a + 1/n) = \lim_{n\to\infty}\int_{a}^{a+\frac{1}{n}} f_\xi(x)dx = 0.$$

Отсюда получаем, что для непрерывных случайных величин справедливы равенства

$$P(x_1 \le \xi < x_2) = P(x_1 < \xi \le x_2) = P(x_1 \le \xi \le x_2) = P(x_1 < \xi < x_2).$$

График плотности распределения называется *кривой распределения*. Площадь, ограниченная кривой распределения, как следует из свойства 2, равна единице. Тогда геометрически значение функции распределения в точке x_0 есть площадь, ограниченная кривой распределения и осью Ox и лежащая левее точки x_0 (рис. 12.3).

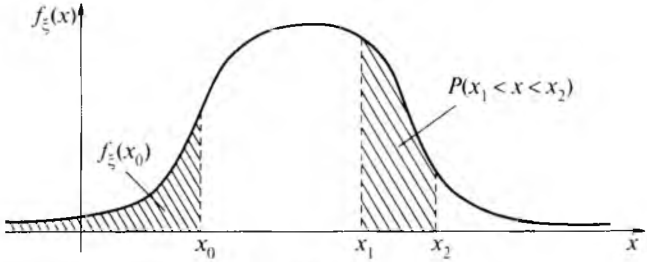

Рис. 12.3

12.3. СЛУЧАЙНЫЙ ВЕКТОР В ДИСКРЕТНОМ ВЕРОЯТНОСТНОМ ПРОСТРАНСТВЕ

Пусть на вероятностном пространстве Ω задано несколько случайных величин $\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega)$, где $\omega \in \Omega$. Такой упорядоченный набор называется ***случайным вектором*** или ***n-мерной случайной величиной*** и обозначается $\xi(\omega) = (\xi_1, \xi_2, ..., \xi_n)$.

Рассмотрим случай $n = 2$, когда на пространстве элементарных исходов заданы две дискретные случайные величины ξ и η, принимающие значения x_i ($i = 1, 2, ...$) и y_j ($j = 1, 2, ...$) соответственно. Упорядоченная пара (ξ, η) называется ***двумерным случайным вектором*** или ***двумерной случайной величиной***. Сами величины ξ и η называются в этом случае ***составляющими*** (или ***компонентами***) случайного вектора.

Геометрически совокупность двух случайных величин можно рассматривать как случайную точку с координатами (ξ, η) на плоскости xOy или как случайный вектор, направленный из начала координат в точку (ξ, η), составляющие которого — случайные величины ξ и η. Совокупность трех случайных величин изображается случайной точкой или случайным вектором в трехмерном пространстве, совокупность n случайных величин — случайной точкой или случайным вектором в пространстве n измерений.

Любое соотношение между возможными значениями случайного вектора и их вероятностями называется ***совместным законом распределения***.

Совместный закон распределения вероятностей дискретных величин ξ и η задается набором вероятностей p_{ij} одновременного осуществления событий $\{\xi = x_i\}$ и $\{\eta = y_j\}$, т.е. $p_{ij} = P\{\xi = x_i, \eta = y_j\}$, и представляется в виде таблицы.

ξ \ η	y_1	y_2	...	y_m
x_1	p_{11}	p_{12}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2m}
...	...	...	...	...
x_n	p_{n1}	p_{n2}	...	p_{nm}

Первый столбец таблицы содержит все возможные значения составляющей ξ, а первая строка — все возможные значения составляющей η. В соответствующей клетке таблицы указана вероятность того, что двумерная случайная величина приняла значение (x_i, y_j).

Поскольку события $\{\omega: \xi(\omega) = x_i, \eta_2(\omega) = y_j\}$ образуют полную группу событий, то сумма вероятностей, помещенных во всех клетках таблицы, равна единице: $\sum_i \sum_j P(\xi = x_i, \eta = y_j) = \sum_i \sum_j p_{ij} = 1$. Такая таблица называется рядом распределения вектора (ξ, η).

Вероятность события типа $\{(\xi, \eta) \in B\}$ — случайная точка (ξ, η) попадает в область B — вычисляется по формуле $P((\xi,\eta) \in B) = $

$= \sum_{(x_i, y_j) \in B} P(\xi = x_i, \eta = y_j)$, где суммирование происходит по всем возможным парам (x_i, y_j) значений случайных величин ξ и η, для которых случайная точка (x_i, y_j) входит в область B.

Частным законом распределения случайной величины ξ называется набор вероятностей событий $\{\xi = x_i\}$. Если задан совместный закон распределения, то частный закон распределения для x можно получить с помощью формулы

$$p_\xi(x_i) = P(\xi = x_i) = \sum_j P(\xi = x_i, \eta = y_j) = \sum_j p_{ij} .$$

Действительно, событие $\{\omega: \xi = x_i\}$ может появиться с одним из событий $\{\xi = x_i, \eta = y_1\}$, $\{\xi = x_i, \eta = y_2\}$, ..., $\{\xi = x_i, \eta = y_m\}$, которые несовместны, и их объединение равно событию $\{\xi = x_i\}$, т.е. $\{\xi = x_i\} = $

$= \bigcup_{j=1}^{m} \{\xi = x_i, \eta = y_j\}$. Отсюда в силу теоремы сложения несовместных событий следует, что $P(\xi = x_i) = \sum_{j=1}^{m} p_{ij}$. Аналогично *частным законом распределения* η называется набор вероятностей событий $\{\eta = y_i\}$, которые также можно вычислить с помощью формулы

$$p_\eta(y_j) = P(\eta = y_j) = \sum_i P(\xi = x_i, \eta = y_j) = \sum_i p_{ij} .$$

Таким образом, распределение каждой случайной величины восстанавливается с помощью совместного закона распределения вероятностей.

Пример 12.2. Совместное распределение пары (ξ, η) задано таблицей в первых трех строках и первых четырех столбцах.

$\eta \backslash \xi$	-1	0	1	p_ξ
-1	1/16	1/8	1/16	1/4
1	3/16	3/8	3/16	3/4
p_η	1/4	1/2	1/4	1

В последней строке и последнем столбце приведены наборы вероятностей, соответствующие частным законам распределения случайных величин.

Частные законы рассчитаны суммированием по строкам и по столбцам.

Случайные величины ξ и η называются *независимыми*, если для любых множеств A и B выполняется условие: $P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B)$, т.е. независимы события $\{\xi \in A\}$ и $\{\eta \in B\}$. Справедлива следующая теорема.

Теорема 12.1. Дискретные случайные величины ξ и η независимы тогда и только тогда, когда события $\{\xi = x_i\}$ и $\{\eta = y_j\}$ независимы для всех значений x_i и y_j.

Пример 12.3. Совместный закон распределения случайных величин ξ и η задан с помощью таблицы.

$\xi \backslash \eta$	1	2
–1	1/16	3/16
0	1/16	3/16
1	1/8	3/8

Вычислить частные законы распределения составляющих величин ξ и η. Определить, зависимы ли они. Вычислить вероятность $P(\xi + \eta \geq 2)$.

Решение. Частное распределение для ξ получается суммированием вероятностей в строках:

$P(\xi = -1) = P(\xi = -1, \eta = 1) + P(\xi = -1, \eta = 2) = 1/16 + 3/16 = 1/4;$

$P(\xi = 0) = P(\xi = 0, \eta = 1) + P(\xi = 0, \eta = 2) = 1/16 + 3/16 = 1/4;$

$P(\xi = 1) = P(\xi = 1, \eta = 1) + P(\xi = 1, \eta = 2) = 1/8 + 3/8 = 1/2.$

Аналогично получается частное распределение для η:

$P(\eta = 1) = 1/16 + 1/16 + 1/8 = 1/4;$

$P(\eta = 2) = 3/16 + 3/16 + 3/8 = 3/4.$

Полученные вероятности можно записать в ту же таблицу напротив соответствующих значений случайных величин.

$\xi \backslash \eta$	1	2	p_ξ
–1	1/16	3/16	1/4
0	1/16	3/16	1/4
1	1/8	3/8	1/2
p_η	1/4	3/4	1

Теперь ответим на вопрос о независимости случайных величин ξ и η. С этой целью для каждой клетки совместного распределения вычислим произведение $P(\xi = x_i, \eta = y_j)$ (т.е. сумм по соответствующей строке и столбцу) и сравним его со значением вероятности $P(\xi = x_i)P(\eta = y_j)$ в этой клетке. Например, в клетке для значений $\xi = -1$ и $\eta = 1$ стоит вероятность $^1/_{16}$, а произведение соответствующих частных вероятностей $^1/_4 \cdot {}^1/_4$ равно $^1/_{16}$, т.е. совпадает с совместной вероятностью. Это условие так же проверяется в оставшихся пяти клетках, и оно оказывается верным во всех. Следовательно, случайные величины ξ и η независимы.

Заметим, что если бы наше условие нарушалось хотя бы в одной клетке, то величины следовало бы признать зависимыми.

Для вычисления вероятности $P(\xi + \eta \geq 2)$ отметим клетки, для которых выполнено условие $\xi + \eta \geq 2$. Таких клеток всего три, и соответствующие вероятности в этих клетках равны $^1/_8$, $^3/_{16}$, $^3/_8$. Их сумма равна $^{11}/_{16}$, это и есть искомая вероятность. Вычисление этой вероятности можно записать так:

$$P(\xi + \eta \geq 2) = P(\xi = 1, \eta = 1) + P(\xi = 0, \eta = 2) + P(\xi = 1, \eta = 2) =$$
$$= {}^1/_8 + {}^3/_{16} + {}^3/_8 = {}^{11}/_{16}.$$

12.4. СОВМЕСТНЫЕ ФУНКЦИЯ И ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОГО ВЕКТОРА

Совместной функцией распределения вероятностей случайных величин ξ_1 и ξ_2 или случайного вектора $\xi = (\xi_1, \xi_2)$ называется функция двух аргументов $F(x_1, x_2)$, равная вероятности $P(\xi_1 < x_1, \xi_2 < x_2) = F(x_1, x_2)$.

Это общее определение верно как в дискретном, так и в непрерывном случае.

Многомерные функции распределения обладают аналогичными свойствами, что и одномерные:

1) $0 \leq F(x_1, x_2) \leq 1$;

2) $F(x_1, x_2)$ есть неубывающая функция по каждому из аргументов;

3) $F(x_1, x_2)$ непрерывна слева по каждому из аргументов;

4) $F(x_1, x_2)$ удовлетворяет соотношениям $F(+\infty, +\infty) = 1$ и $F(x_1, -\infty) = F(-\infty, x_2) = 0$.

В многомерном случае четырех свойств недостаточно, чтобы функция $F(x_1, x_2)$ была функцией распределения. Необходимо, чтобы при любых $a_1 < b_1$ и $a_2 < b_2$ и следующее выражение было неотрицательно: $P(a_1 \leq \xi_1 < b_1, a_2 \leq \xi_2 < b_2) = P[(\xi_1, \xi_2) \in \Pi] \geq 0$.

Геометрически функция распределения $F_\xi(x_1, x_2) = P(\xi_1 < x_1, \xi_2 < x_2)$ определяет вероятность попадания случайной точки в бесконечный угол с вершиной в точке (x_1, x_2) (рис. 12.4).

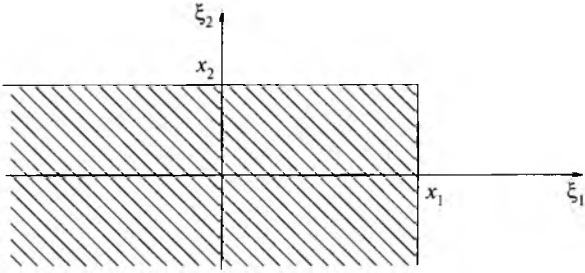

Рис. 12.4

Геометрическая интерпретация функции распределения $F(x_1, x_2)$ позволяет дать простое пояснение предельным свойствам функции распределения: если $x_1 \to -\infty$ (или $x_2 \to -\infty$), правая граница (верхняя граница) бесконечного квадранта неограниченно смещается влево (вниз), то вероятность попадания случайной точки в квадрант стремится к нулю. При $x_1 \to +\infty$ и $x_2 \to +\infty$ бесконечный квадрант превращается во всю плоскость $\xi_1 O \xi_2$ и попадание случайной точки в эту плоскость является достоверным событием.

Если один из аргументов равен $+\infty$, то функция распределения вектора превращается в функцию распределения случайной величины, соответствующей другому аргументу: $F(x_1, +\infty) = F_1(x_1)$, $F(+\infty, x_2) = F_2(x_2)$. Действительно, так как событие $\{\omega\colon \xi_2(\omega) < +\infty\}$ достоверно, то $F(x_1, +\infty)$ определяет вероятность события $\{\omega\colon \xi_1(\omega) <$ $< x_1\}$, т.е. представляет собой функцию распределения составляющей ξ_1. Аналогично $F(+\infty, x_2) = F_2(x_2)$. Функции распределения составляющих $F_1(x_1)$ и $F_2(x_2)$ называются *частными законами распределения* (частными функциями распределения).

Геометрически функция распределения составляющих есть вероятность попадания случайной точки в полуплоскость, ограниченную прямой $x = x_1$, для $F(x_1, +\infty) = F_1(x_1)$ или прямой $x = x_2$ для $F(+\infty, x_2) =$ $= F_2(x_2)$ (рис. 12.5).

Используя геометрическую интерпретацию функции распределения $F_\xi(x_1, x_2)$, можно вычислить вероятность попадания случайного вектора $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega))$ в прямоугольник $\Pi = \{(x_1, x_2)\colon$ $x_1 \in [a_1, b_1), x_2 \in [a_2, b_2)\}$. Действительно (рис. 12.6):

$$P[(\xi_1, \xi_2) \in \Pi] = P(\xi_1 < b_1, \xi_2 < b_2) - P(\xi_1 < b_1, x_2 < a_2) +$$
$$+ P(\xi_1 < a_1, \xi_2 < a_2) - P(\xi_1 < a_1, x_2 < b_2) = F(b_1, b_2) - F(b_1, a_2) -$$
$$- F(a_1, b_2) + F(a_1, a_2).$$

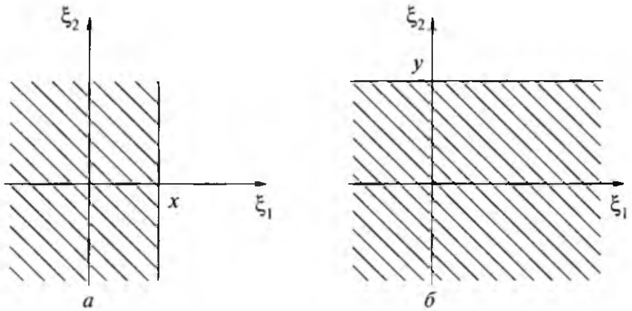

Рис. 12.5

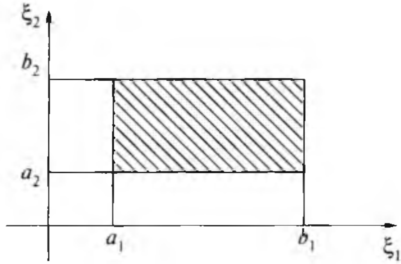

Рис. 12.6

Если вектор $\xi(\omega) = (\xi_1, \xi_2)$ непрерывный, то эта формула верна для любого интервала, полуинтервала или отрезка, так как вероятности точек и прямых в этом случае нулевые.

Условие независимости случайных величин вытекает из следующего утверждения: если случайные величины ξ_1 и ξ_2 независимы, то любые события $(\omega\colon \xi \in B_1)$, $(\omega\colon \xi \in B_2)$, индуцированные случайными величинами ξ_1 и ξ_2, являются независимыми.

Отсюда следует, что если случайные величины ξ_1 и ξ_2 независимы, то совместная функция распределения равна произведению функций распределения составляющих: $F(x_1, x_2) = F_1(x_1)F_2(x_2)$.

Все свойства, доказанные для функции распределения двумерной случайной величины, остаются справедливыми и для функции распределения $F(x_1, x_2, ..., x_n)$ в случае n аргументов.

Случайный вектор $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega))$ называется *непрерывным*, если его функция распределения $F(x_1, x_2)$ непрерывна.

Случайный вектор называется *абсолютно непрерывным*, если существует такая неотрицательная функция $f(x_1, x_2)$, называемая *со-*

вместной плотностью распределения случайных величин ξ_1, ξ_2 или плотностью распределения непрерывного вектора $\xi(\omega)$, что при любых x_1, x_2 имеет место равенство

$$F(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(t_1, t_2)dt_1dt_2.$$

Из определения следует, что в случае непрерывного вектора $\xi(\omega)$ каждая его компонента ξ_k — непрерывная случайная величина. Плотность распределения составляющей вектора ξ_k называется частным законом распределения или ее *частной плотностью*. Поскольку $F_1(x_1) = P(\xi < x_1) = P(\xi_1 < x_1, \xi_2 < \infty) = \int_{-\infty}^{x_1}\int_{-\infty}^{\infty} f(t_1, t_2)dt_1dt_2$, то частная плотность распределения определяется равенством $f_1(t_1) = \int_{-\infty}^{\infty} f(t_1, t_2)dt_2$. Аналогично доказывается и второе равенство: $f_2(t_2) = \int_{-\infty}^{\infty} f(t_1, t_2)dt_1$.

Совместная плотность распределения обладает следующими свойствами:

1) $f(x_1, x_2) \geq 0$;

2) вероятность попадания случайной точки (ξ_1, ξ_2) в какую-нибудь область G определяется равенством $P((\xi_1, \xi_2) \in G) = \int_G f(x_1, x_2)dx_1dx_2$;

3) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1, x_2)dx_1dx_2 = 1$;

4) $f(x_1, x_2) = F''_{x_1x_2}(x_1, x_2)$.

Теорема 12.2. Если непрерывные случайные величины ξ_1 и ξ_2 независимы, то совместная их плотность распределения равна произведению плотностей распределения составляющих: $f(x_1, x_2) = f_1(x_1)f_2(x_2)$.

Справедливо и обратное утверждение: если выполнено равенство $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, то случайные величины ξ_1 и ξ_2 независимы.

Определения и свойства, справедливые для двумерной случайной величины, без ограничения переносятся на многомерный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$.

12.5. ФУНКЦИИ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН. КОМПОЗИЦИЯ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

Пусть $\xi(\omega)$ — случайная величина, определенная в вероятностном пространстве Ω, и $\varphi(x)$ — некоторая функция. Тогда $Y = \varphi(\xi)$ — случайная величина, которая каждому событию $\omega \in \Omega$ ставит в соответствие число $Y(\omega) = \varphi(\xi(\omega))$. Покажем, что, зная закон распределения случайной величины ξ, можно найти закон распределения случайной величины Y.

Если $y = \varphi(x)$ — непрерывная монотонно возрастающая функция, то существует обратная ей (также монотонно возрастающая) функция $x = \varphi^{-1}(y)$ и функцию распределения случайной величины $Y = \varphi(\xi)$ можно представить в виде

$$F_Y(x) = F_\xi(\varphi^{-1}(x)), \text{ так как } F_Y(x) = P(\varphi(\xi) < x) = P(\xi < \varphi^{-1}(x)) = F_\xi(\varphi^{-1}(x)).$$

Если ξ — непрерывная случайная величина с плотностью распределения $f_\xi(x)$ и функция $\varphi(x)$ — дифференцируемая и монотонно возрастающая функция, то

$$f_Y(x) = F_Y'(x) = \frac{dF(\varphi^{-1}(x))}{d\varphi^{-1}(x)} \cdot \frac{d\varphi^{-1}(x)}{dx} = f_\xi(\varphi^{-1}(x))\frac{d\varphi^{-1}(x)}{dx}. \qquad (12.1)$$

Если $y = \varphi(x)$ — дифференцируемая и монотонно убывающая функция, то и обратная ей функция тоже монотонно убывает. Функция распределения случайной величины $Y = \varphi(\xi)$ в этом случае будет равна

$$F_Y(x) = P(\varphi(\xi) < x) = P(\varphi^{-1}(\varphi(\xi)) > \varphi^{-1}(x)) = P(\xi > \varphi^{-1}(x)) = 1 - P(\xi \le \varphi^{-1}(x)).$$

Причем если ξ — непрерывная случайная величина, то, учитывая, что $P(\xi = \varphi^{-1}(x)) = 0$ и $d\varphi^{-1}(x)/dx < 0$, получаем

$$f_Y(x) = (1 - P(\xi \le \varphi^{-1}(x))' = (1 - P(\xi < \varphi^{-1}(x))' =$$

$$= (1 - F_\xi(\varphi^{-1}(x)))' = -\frac{dF_\xi(\varphi^{-1}(x))}{d\varphi^{-1}(x)} \cdot \frac{d\varphi^{-1}(x)}{dx} =$$

$$= -f_\xi(\varphi^{-1}(x))\frac{d\varphi^{-1}(x)}{dx}. \qquad (12.2)$$

Объединяя формулы (12.1) и (12.2), получаем, что если $\varphi(x)$ — монотонная и дифференцируемая функция, а ξ — непрерывная случайная величина с плотностью распределения $f_\xi(x)$, то плотность распределения случайной величины $Y = \varphi(\xi)$ равна

$$f_Y(x) = f_\xi(\varphi^{-1}(x))\left|\frac{d\varphi^{-1}(x)}{dx}\right|.$$

Пример 12.4. Случайная величина ξ равномерно распределена на отрезке $[-1;\ 1]$. Найти плотность распределения случайной величины $\eta = -\ln(2+\xi)$.

Решение. Из условия следует, что $f_\xi(x) = \begin{cases} \frac{1}{2}, & x \in [-1;1], \\ 0, & x \notin [-1;1]. \end{cases}$

Из условия $\eta = -\ln(2+\xi)$ получаем, что $\xi = (e^{-\eta} - 2)$, т.е. $\varphi^{-1}(x) = e^{-x} - 2$, когда $x \in [-\ln 3, 0]$. Тогда плотность распределения случайной величины η будет иметь вид

$$f_\eta(x) = \begin{cases} \frac{1}{2}e^{-x}|-1|, & x \in [-\ln 3; 0], \\ 0, & x \notin [-\ln 3; 0] \end{cases} = \begin{cases} \frac{1}{2}e^{-x}, & x \in [-\ln 3; 0], \\ 0, & x \notin [-\ln 3; 0]. \end{cases}$$

Пусть задан случайный двумерный вектор $\xi = (\xi_1, \xi_2)$ и некоторая функция $Y = \xi_1 + \xi_2$. Зная совместный закон распределения случайных величин (ξ_1, ξ_2), найдем функцию распределения случайной величины Y. По определению получаем

$$F_Y(x) = P\{\xi_1 + \xi_2 < x\} = \iint\limits_{D_x} f(x_1, x_2)dx_1dx_2,$$

где интегрирование производится по множеству $D_x = \{(x_1, x_2): x_1 + x_2 < x\}$, изображенному на рис. 12.7. При фиксированном x_1 переменная x_2 может принимать значения из интервала $(-\infty,\ x - x_1)$, интервал изменения x_1 при этом равен $(-\infty, +\infty)$. Получим

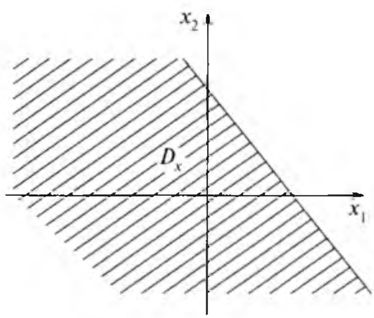

Рис. 12.7

$$F_Y(x) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x-x_1} f_\xi(x_1, x_2)dx_2\right)dx_1.$$

Произведем во внутреннем интеграле замену $x_2 = t - x_1$ и затем, так как двойной интеграл абсолютно сходится, поменяем порядок интегрирования:

$$F_Y(x) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x} f_\xi(x_1, t - x_1)dt\right)dx_1 = \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f_\xi(x_1, t - x_1)dx_1\right)dt.$$

В соответствии с определением функции плотности распределения вероятностей получим, что функция $f_Y(t) = \int_{-\infty}^{\infty} f_\xi(x_1, t - x_1)dx_1$ является плотностью распределения случайной величины $Y = \xi_1 + \xi_2$. Отсюда получается, что

$$f_Y(t) = \int_{-\infty}^{\infty} f_\xi(t - x_2, x_2)dx_2 \quad \text{и} \quad f_Y(t) = \int_{-\infty}^{\infty} f_\xi(x_1, t - x_1)dx_1.$$

Если случайные величины независимы, то $f_\xi(x_1, x_2) = f_1(x_1)f_2(x_2)$. Подставляя последнее равенство в полученные выражения для плотности суммы, приходим к *формуле свертки* ($f_{\xi_1+\xi_2} = f_1 * f_2$) в *композиции для плотностей распределения* $f_1(x_1)$ и $f_2(x_2)$:

$$f_{\xi_1+\xi_2}(t) = \int_{-\infty}^{\infty} f_{\xi_1}(x_1)f_{\xi_2}(t - x_1)dx_1$$

или $f_{\xi_1+\xi_2}(t) = \int_{-\infty}^{\infty} f_{\xi_1}(t - x_2)f_{\xi_2}(x_2)dx_2.$

Формула распределения вероятностей суммы двух независимых дискретных случайных величин имеет вид

$$P(\xi_1 + \xi_2 = x) = P(A_x) = \sum_{i=1}^{\infty} P(\xi_1 = x_i)P(\xi_2 = x - x_i)$$

и называется *сверткой независимых случайных величин с дискретным распределением*.

Глава 13
ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

13.1. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Рассмотрим дискретную случайную величину ξ, принимающую значения $x_1, x_2, ..., x_k$ с вероятностями $p_1, p_2, ..., p_k$. Требуется охарактеризовать каким-то числом положение значений этой случайной величины с учетом того, что эти значения имеют различные вероятности.

Пусть произведено n независимых испытаний и пусть каждое значение $x_1, x_2, ..., x_k$ встречается при этом $n_1, n_2, ..., n_k$ раз, так что $n_1 + n_2 + ... + n_k = n$. Величина

$$\bar{x} = \frac{1}{n}[x_1 n_1 + x_2 n_2 + ... + x_k n_k] = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + ... + x_k \frac{n_k}{n} = \sum_{i=1}^{\infty} \frac{x_i n_i}{n}$$

есть среднее арифметическое из полученных значений. Из свойства статистической устойчивости частот следует, что относительная частота n_i/n события (ω: $\xi = x_i$) приближенно равна вероятности этого события: $n_i/n \approx p_i$. Следовательно, среднее значение случайной величины x приближенно равно вполне определенному числу $M\xi$, зависящему только от распределения вероятностей случайной величины ξ.

Математическим ожиданием (или *средним значением*) дискретной случайной величины ξ называется число, равное сумме произведений всех ее возможных значений на их вероятности: $M\xi = \sum_{i=1}^{\infty} x_i p_i$ в случае абсолютной сходимости этого ряда. Если такой ряд не сходится абсолютно, то говорят, что случайная величина ξ не имеет конечного математического ожидания. Если случайная величина ξ принимает только конечное число значений, то математическое ожидание всегда существует.

Математическим ожиданием непрерывной случайной величины ξ называется интеграл $M\xi = \int_{-\infty}^{\infty} x f_\xi(x)dx$, где $f_\xi(x)$ — плотность распределения ξ, причем интеграл абсолютно сходится (т.е. существует интеграл $\int_{-\infty}^{\infty} |x| f(x)dx$).

Если ξ — случайная величина и $\varphi(x)$ — некоторая функция, то математическое ожидание величины $Y = \varphi(\xi)$ можно вычислить по формуле

$$M\varphi(\xi) = \sum_i \varphi(x_i) p_i \quad \text{или} \quad M\xi = \int_{-\infty}^{\infty} \varphi(x) f_\xi(x) dx$$

при условии, что ряд или интеграл, стоящий справа, абсолютно сходятся.

Под математическим ожиданием n-мерного случайного вектора $\xi = (\xi_1, \xi_2, ..., \xi_n)$ понимается вектор $M\xi = (M\xi_1, M\xi_2, ..., M\xi_n)$.

Свойства математического ожидания следующие.

1. Математическое ожидание — детерминированная величина, значение которой заключено между наименьшим и наибольшим возможными значениями случайной величины.

2. Математическое ожидание постоянной величины равно этой постоянной, т.е. $MC = C$.

Действительно, постоянную величину C можно рассматривать как случайную величину, принимающую значение C с вероятностью 1, тогда $MC = C \cdot 1 = C$.

3. Математическое ожидание суммы случайных величин ξ_1 и ξ_2 равно сумме математических ожиданий: $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$, если существуют конечные математические ожидания $M\xi_1$ и $M\xi_2$.

Доказательство. Используя ранее введенные обозначения, получаем

$$M(\xi + \eta) = \sum_{i,j} (x_i + y_j) p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} =$$
$$= \sum_i x_i p_\xi(x_i) + \sum_j y_j p_\eta(y_j) = M\xi + M\eta.$$

Аналогично доказывается для непрерывной случайной величины.

С л е д с т в и е 1. Математическое ожидание суммы конечного числа случайных величин $\xi_1, \xi_2, ..., \xi_n$ равно сумме математических ожиданий слагаемых: $M\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} M(\xi_i)$.

4. Если ξ_1 и ξ_2 — независимые случайные величины и существуют конечные математические ожидания $M\xi_1$, $M\xi_2$, то математическое ожидание произведения $M\xi_1\xi_2$ равно произведению математических ожиданий: $M\xi_1\xi_2 = (M\xi_1)(M\xi_2)$.

Доказательство следует из равенства

$$M(\xi\eta) = \sum_{i,j} x_i y_j p_{ij} = \sum_{i,j} x_i y_j p_\xi(x_i) p_\eta(y_j) =$$
$$= \sum_i x_i p_\xi(x_i) \cdot \sum_j y_j p_\eta(x_j) = M\xi \cdot M\eta.$$

С л е д с т в и е 2. Постоянную величину можно выносить за знак математического ожидания $M(C\xi) = CM\xi$.

Рассмотрим примеры вычисления математического ожидания.

1. Случайная величина ξ, равная числу успехов в n испытаниях Бернулли, имеет биномиальное распределение, может принимать любое из значений 0, 1, 2, ..., n, и вероятности этих значений определяются формулой Бернулли: $P(\xi = m) = P_n(m) = C_n^m p^m q^{n-m}$, где p — вероятность «успеха», $q = 1 - p$ — вероятность «неуспеха». Математическое ожидание случайной величины ξ можно вычислить по формуле $M\xi = \sum_{m=0}^{n} mP_n(m)$. Однако в данном случае удобнее воспользоваться тем фактом, что математическое ожидание суммы случайных величин равно сумме математических ожиданий этих величин.

Обозначим через ξ_i случайную величину, равную числу успехов в i-м испытании Бернулли. Тогда случайная величина $\xi = \sum_{i=1}^{n} \xi_i$ равна сумме n независимых случайных величин. Следовательно, $M\xi = \sum_{i=1}^{n} M\xi_i = \sum_{i=1}^{n} p = np$. Итак, среднее число успехов в n испытаниях Бернулли равно np.

2. Пуассоновская случайная величина, вероятности которой задаются формулой Пуассона $P_m = \lambda^m e^{-\lambda}/m!$, имеет ряд распределения

$$\begin{pmatrix} 0 & 1 & \dots & m & \dots \\ e^{-\lambda} & \lambda e^{-\lambda} & \dots & \lambda^m e^{-\lambda}/m! & \dots \end{pmatrix}.$$

Тогда математическое ожидание этой случайной величины равно

$$M\xi = \sum_{m=0}^{\infty} \frac{m\lambda^m}{m!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

3. Математическое ожидание геометрически распределенной случайной величины, т.е. числа испытаний по схеме Бернулли до первого «успеха», имеет ряд распределения

$$\begin{pmatrix} 1 & 2 & \dots & m & \dots \\ p & pq & \dots & pq^{m-1} & \dots \end{pmatrix},$$

где $0 < p \le 1$, $\sum_{m=0}^{\infty} pq^m = p\sum_{m=0}^{\infty} q^m = p/(1-q) = 1$. Математическое ожидание равно

$$M\xi = \sum_{m=1}^{\infty} mpq^{m-1} = p\sum_{m=1}^{\infty} mq^{m-1} = p\sum_{m=1}^{\infty} (q^m)' = p\left[\sum_{m=1}^{\infty} q^m\right]' =$$

$$= p[q/(1-q)]' = p/(1-q)^2 = 1/p,$$

т.е. математическое ожидание случайной величины обратно пропорционально вероятности положительного исхода.

Кроме математического ожидания, положение случайной величины характеризуется модой и медианой.

Модой случайной величины называется ее наиболее вероятное значение.

Модой непрерывной величины является точка локального максимума функции плотности распределения. Если многоугольник или кривая распределения имеет один максимум, то распределение называется *унимодальным*, более одного максимума — *мультимодальным*. Распределение, имеющее минимум плотности, называется *антимодальным* (рис. 13.1).

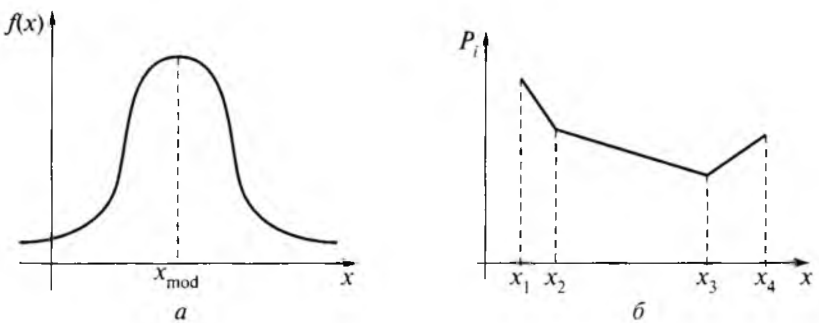

Рис. 13.1

Наивероятнейшим значением называется мода, доставляющая глобальный максимум вероятности для дискретной случайной величины или плотности распределения для непрерывной случайной величины.

Медианой случайной величины ξ называется такое ее значение $\xi = x_{\text{med}}$, что $P(\xi < x_{\text{med}}) = P(\xi > x_{\text{med}}) = 1/2$.

Геометрически медиана — это абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам (рис. 13.2).

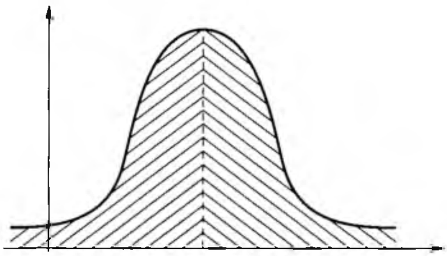

Рис. 13.2

13.2. ДИСПЕРСИЯ, СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Дисперсией случайной величины ξ называется число, равное математическому ожиданию квадрата отклонения случайной величины от ее математического ожидания, и обозначается $D\xi = M(\xi - M\xi)^2$ при условии существования конечного математического ожидания, т.е. дисперсия — это математическое ожидание функции $\varphi(\xi) = (\xi - M\xi)^2$.

Дисперсия характеризует распределение (рассеивание) значений случайной величины около ее математического ожидания.

Из определения следует, что дисперсия имеет размерность квадрата случайной величины. Линейной оценкой рассеивания случайной величины является величина $\sigma_\xi = \sqrt{D\xi}$, которая называется *средним квадратическим* (или *стандартным*) отклонением.

Формулы для вычисления дисперсии следующие:

$$D\xi = \sum_i (x_i - M\xi)^2 p_i$$

для дискретной случайной величины при условии абсолютной сходимости ряда;

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2 f(x)dx$$

для непрерывной случайной величины при условии абсолютной сходимости интеграла.

Свойства дисперсии.

1. Дисперсия — величина детерминированная и всегда неотрицательная.

2. Дисперсия постоянной величины равна нулю:

$DC = M(C - MC)^2 = M(C - C)^2 = M(0) = 0.$

3. Дисперсию можно вычислить по формуле

$$D\xi = M\xi^2 - (M\xi)^2.$$

Доказательство получим из определения дисперсии и свойств математического ожидания $D\xi = M(\xi - M\xi)^2 = M(\xi^2 - 2\xi M\xi + (M\xi)^2) = M\xi^2 - 2(M\xi)^2 + (M\xi)^2 = M\xi^2 - (M\xi)^2$ при условии существования конечных $M\xi^2$ и $M\xi$.

4. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат: $D(C\xi) = C^2D\xi$.

Действительно, $D(C\xi) = M(C\xi)^2 - (M(C\xi))^2 = C^2M\xi^2 - C^2(M\xi)^2 = C^2(M\xi^2 - M\xi)^2) = C^2D\xi$.

5. Дисперсия суммы независимых случайных величин ξ_1 и ξ_2 равна сумме дисперсий этих величин: $D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2$.

Доказательство. Из свойств математического ожидания и свойства 3 получим

$$D(\xi_1 + \xi_2) = M(\xi_1 + \xi_2)^2 - (M(\xi_1 + \xi_2))^2 = M(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2) -$$
$$- (M\xi_1)^2 - (M\xi_2)^2 - 2(M\xi_1)(M\xi_2) = M\xi_1^2 - (M\xi_1)^2 + M\xi_2^2 - (M\xi_2)^2 =$$
$$= D\xi_1 + D\xi_2.$$

Следствия

1. Дисперсия суммы случайной величины и постоянной равна дисперсии случайного слагаемого: $D(\xi + C) = D\xi$.

2. $D(C\xi + B) = C^2D\xi$, если C и B — постоянные величины.

3. Методом математической индукции можно доказать, что если $\xi_1, \xi_2, \ldots, \xi_n$ независимы, то дисперсия их суммы равна сумме их дисперсий: $D(\xi_1 + \xi_2 + \ldots + \xi_n) = D\xi_1 + D\xi_2 + \ldots + D\xi_n$.

Из свойств дисперсии вытекают соответствующие свойства среднего квадратического отклонения случайной величины.

1. $\sigma_\xi = 0$ тогда и только тогда, когда ξ — постоянная величина.

2. Среднее квадратическое отклонение произведения случайной величины ξ на постоянную равно произведению среднего квадратического отклонения случайной величины ξ на модуль постоянной:

$$\sigma_{C\xi} = \sqrt{D(C\xi)} = \sqrt{C^2D\xi} = |C|\sqrt{D\xi} = |C|\sigma_\xi.$$

3. Если ξ и η — независимые случайные величины, то среднее квадратическое отклонение суммы $\xi + \eta$ можно найти по формуле

$$\sigma_{\xi+\eta} = \sqrt{D(\xi + \eta)} = \sqrt{D\xi + D\eta} = \sqrt{\sigma_\xi^2 + \sigma_\eta^2}.$$

4. Среднеквадратическое отклонение среднего арифметического n независимых случайных величин с одинаковой дисперсией σ^2 равно $\dfrac{\sigma}{\sqrt{n}}$.

Действительно, если $\xi_1, \xi_2, ..., \xi_n$ — независимые случайные величины с одинаковой дисперсией $D\xi_i = \sigma^2$, то дисперсия среднего арифметического $\xi = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ этих случайных величин равна $D\xi = = \frac{1}{n^2}\sum_{i=1}^{n} D\xi_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

Случайную величину ξ называют ***нормированной***, если $M\xi = 0$, $D\xi = 1$. Преобразование случайной величины вида $f(\xi) = \frac{\xi - M\xi}{\sigma_\xi}$ называется ***нормированием случайной величины*** ξ.

Рассмотрим примеры вычисления дисперсий.

1. Найдем дисперсию «успеха» в одном испытании Бернулли. Поскольку $M\xi = p$, то ряд распределения для функции $Y = (\xi - M\xi)^2$ будет иметь вид

$$\begin{pmatrix} (0-p)^2 & (1-p)^2 \\ q & p \end{pmatrix},$$

поэтому $D\xi = M(\xi - M\xi)^2 = p^2q + q^2p = pq$.

2. Дисперсия случайной величины ξ, распределенной по биномиальному закону, может быть представлена как дисперсия суммы независимых случайных величин: $\xi = \sum_{i=1}^{n}\xi_i$. Тогда $D\xi = D\left[\sum_{i=1}^{n}\xi_i\right] = = \sum_{i=1}^{n} D\xi_i = npq$, т.е. дисперсия биномиальной случайной величины равна произведению числа испытаний на вероятности положительного и отрицательного исходов.

3. Дисперсия пуассоновской случайной величины равна $D\xi = = M\xi^2 - (M\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$, так как

$$M\xi^2 = \sum_{m=1}^{\infty} m^2 \frac{\lambda^m e^{-\lambda}}{m!} = e^{-\lambda}\sum_{m=1}^{\infty} m(m-1+1)\frac{\lambda^m}{m!} =$$

$$= e^{-\lambda}\left(\sum_{m=1}^{\infty} m(m-1)\frac{\lambda^m}{m!} + \sum_{m=1}^{\infty} m\frac{\lambda^m}{m!}\right) = e^{-\lambda}\lambda^2\sum_{m=2}^{\infty}\frac{\lambda^{m-2}}{(m-2)!} + \lambda e^{-\lambda}e^{\lambda} =$$

$$= \lambda^2 + \lambda.$$

Итак, дисперсия пуассоновской величины равна ее математическому ожиданию, равному параметру λ: $D\xi = M\xi = \lambda$.

4. Для нахождения дисперсии геометрически распределенной случайной величины найдем математическое ожидание ее квадрата:

$$M\xi^2 = \sum_{m=1}^{\infty} m^2 pq^{m-1} = \sum_{m=1}^{\infty} m(m-1+1)pq^{m-1} = \sum_{m=1}^{\infty} m(m-1)pq^{m-1} +$$

$$+\sum_{m=1}^{\infty} mpq^{m-1} = pq\sum_{m=2}^{\infty} m(m-1)q^{m-2} + 1/p = pq\sum_{m=2}^{\infty}(q^m)'' + 1/p =$$

$$= pq\left(\sum_{m=2}^{\infty} q^m\right)'' + 1/p = pq(q^2/(1-q))'' + 1/p = 2pq/(1-q)^3 +$$

$$+1/p = 2q/p^2 + 1/p.$$

Отсюда получаем $D\xi = M\xi^2 - (M\xi)^2 = 2q/p^2 + 1/p - 1/p^2 = q/p^2$.

13.3. КОВАРИАЦИЯ. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Пусть задан двумерный вектор (ξ_1, ξ_2) в вероятностном пространстве Ω.

Ковариацией случайных величин ξ_1 и ξ_2 называется число $\mathrm{cov}(\xi_1, \xi_2)$, равное математическому ожиданию произведения отклонений случайных величин от их математических ожиданий, т.е. $\mathrm{cov}(\xi_1, \xi_2) = M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)]$, в предположении существования математических ожиданий.

Формулы для вычисления ковариации следующие:

$$\mathrm{cov}(\xi_1, \xi_2) = \sum_i \sum_j (x_i - M\xi_1)(y_j - M\xi_2)p_{ij}$$

для дискретных случайных величин;

$$\mathrm{cov}(\xi_1, \xi_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_1 - M\xi_1)(x_2 - M\xi_2)f(x_1, x_2)dx_1dx_2$$

для непрерывных случайных величин.

Из определения ковариации и свойств математического ожидания вытекают следующие свойства ковариации.

1. Ковариация может быть записана в виде

$\mathrm{cov}(\xi_1, \xi_2) = M(\xi_1\xi_2) - (M\xi_1)(M\xi_2)$.

Действительно, справедливы следующие равенства:

$$\mathrm{cov}(\xi_1, \xi_2) = M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)] = M[\xi_1\xi_2 - \xi_2 M\xi_1 - \xi_1 M\xi_2 +$$
$$+ (M\xi_1)(M\xi_2)] = M(\xi_1\xi_2) - (M\xi_2)(M\xi_1) - (M\xi_1)(M\xi_2) + (M\xi_1)(M\xi_2) =$$
$$= M(\xi_1\xi_2) - (M\xi_1)(M\xi_2).$$

Следствия

1. Если случайные величины ξ_1 и ξ_2 имеют конечные дисперсии $D\xi_1$ и $D\xi_2$, то дисперсия суммы этих случайных величин существует и равняется

$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2 + 2\mathrm{cov}(\xi_1, \xi_2).$$

Доказательство. По свойству математического ожидания получаем

$$D(\xi_1 + \xi_2) = M(\xi_1 + \xi_2)^2 - (M(\xi_1 + \xi_2))^2 = M(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2) -$$
$$- (M\xi_1 + M\xi_2)^2 = M\xi_1^2 + 2M\xi_1\xi_2 + M\xi_2^2 - (M\xi_1)^2 - 2M\xi_1 M\xi_2 -$$
$$- (M\xi_2)^2 = D\xi_1 + D\xi_2 + 2\mathrm{cov}(\xi_1, \xi_2).$$

2. Если случайные величины ξ_1 и ξ_2 независимы, то $\mathrm{cov}(\xi_1, \xi_2) = 0$.

Действительно, по свойству математического ожидания для независимых случайных величин получим

$$\mathrm{cov}(\xi_1, \xi_2) = M(\xi_1\xi_2) - (M\xi_1)(M\xi_2) = (M\xi_1)(M\xi_2) - (M\xi_1)(M\xi_2) = 0.$$

Однако обратное, вообще говоря, неверно. Существуют зависимые случайные величины, ковариация которых также равна нулю.

3. Из определения ковариации следует, что $\mathrm{cov}(\xi_1, \xi_2) = \mathrm{cov}(\xi_2, \xi_1)$, так как от перемены мест сомножителей произведение не меняется.

4. Если $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$, то $\mathrm{cov}(\eta_1, \eta_2) = a_1a_2\mathrm{cov}(\xi_1, \xi_2)$.

Доказательство.

$$\mathrm{cov}(\eta_1, \eta_2) = M[(\eta_1 - M\eta_1)(\eta_2 - M\eta_2)] = M[(a_1\xi_1 + b_1 - a_1M\xi_1 - b_1) \times$$
$$\times (a_2\xi_2 + b_2 - a_2M\xi_2 - b_2)] = M(a_1a_2(\xi_1 - M\xi_1)(\xi_2 - M\xi_2) =$$
$$= a_1a_2\mathrm{cov}(\xi_1, \xi_2).$$

5. $\mathrm{cov}(\xi_1 + \xi_2, \eta) = \mathrm{cov}(\xi_1, \eta_1) + \mathrm{cov}(\xi_2, \eta)$; $\mathrm{cov}(\xi, \eta_1 + \eta_2) = \mathrm{cov}(\xi, \eta_1) + \mathrm{cov}(\xi, \eta_2)$.

Воспользуемся свойством сложения математического ожидания:

$$\mathrm{cov}(\xi_1 + \xi_2, \eta) = M(((\xi_1 + \xi_2) - M(\xi_1 + \xi_2))(\eta - M\eta)) =$$
$$= M((\xi_1 - M\xi_1) + (\xi_2 - M\xi_2))(\eta - M\eta)) =$$
$$= M((\xi_1 - M\xi_1)(\eta - M\eta)) + M((\xi_2 - M\xi_2)(\eta - M\eta)) =$$
$$\mathrm{cov}(\xi_1, \eta) + \mathrm{cov}(\xi_2, \eta).$$

Аналогично делается для η_1 и η_2.

6. Очевидно, что $\mathrm{cov}(\xi, \xi) = M(\xi - M\xi)^2 = D\xi$.

7. Справедливо следующее неравенство:

$$-\sqrt{D\xi_1 D\xi_2} \le \mathrm{cov}(\xi_1, \xi_2) \le \sqrt{D\xi_1 D\xi_2}.$$

Для доказательства найдем дисперсию случайной величины $\eta_x = x\xi_1 - \xi_2$, где x — произвольное число. По свойству дисперсии

$D\eta_x = x^2 D\xi_1 - 2x\operatorname{cov}(\xi_1, \xi_2) + D\xi_2$. Как функция от x дисперсия $D\eta_x$ представляет собой квадратичный трехчлен. Но дисперсия любой случайной величины не может быть меньше нуля, так что $D\eta_x \geq 0$ для любого x. И так как $D\xi_1 \geq 0$, то дискриминант должен быть неположительный, т.е.

$$[\operatorname{cov}(\xi_1, \xi_2)]^2 - D\xi_1 D\xi_2 \leq 0 \quad \text{или} \quad |\operatorname{cov}(\xi_1, \xi_2)| \leq \sqrt{D\xi_1 D\xi_2}.$$

8. $\operatorname{cov}(\xi_1, \xi_2) = \pm\sqrt{D\xi_1 D\xi_2}$ тогда и только тогда, когда случайные величины ξ_1 и x_2 линейно зависимы.

Действительно, если дискриминант равен нулю, то уравнение

$$\xi^2 D\xi_1 - 2x\operatorname{cov}(\xi_1, \xi_2) + D\xi_2 = 0$$

имеет решение x_0 и $D\eta_{x_0} = 0$. Это означает, что $\eta_{x_0} \equiv c$ — постоянная величина и случайные величины ξ_1 и ξ_2 связаны линейной функциональной зависимостью $\xi_2 = x_0\xi_1 - c$, причем если коэффициент пропорциональности x_0 положителен, то $\operatorname{cov}(\xi_1, \xi_2) = \sqrt{D\xi_1 D\xi_2}$, а если x_0 отрицателен, то $\operatorname{cov}(\xi_1, \xi_2) = -\sqrt{D\xi_1 D\xi_2}$.

Итак, ковариацию можно считать мерой зависимости случайных величин, так как для независимых случайных величин ковариация равна нулю. Недостатком ковариации является то, что ее размерность совпадает с произведением размерностей случайных величин. Безразмерной характеристикой зависимости является нормированный коэффициент ковариации, который называется *коэффициентом корреляции*.

Коэффициентом корреляции случайных величин ξ_1 и ξ_2 называется число, равное отношению ковариации $\operatorname{cov}(\xi_1, \xi_2)$ к произведению средних квадратических отклонений:

$$\rho(\xi_1, \xi_2) = \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}}.$$

Из свойств ковариации следуют свойства коэффициента корреляции.

1. $\rho(\xi, \xi) = \dfrac{D\xi}{\sqrt{D\xi D\xi}} = 1.$

2. Если ξ_1 и ξ_2 независимы и существуют, $D\xi_1 > 0$ и $D\xi_2 > 0$, то $\rho(\xi_1, \xi_2) = 0$.

3. Если $\eta_1 = a_1\xi_1 + b_1$; $\eta_2 = a_2\xi_2 + b_2$, $a_1 \neq 0$, $a_2 \neq 0$, тогда

$$\rho(\eta_1, \eta_2) = a_1 a_2 \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sqrt{a_1^2 D\xi_1 \cdot a_2^2 D\xi_2}} = \pm\rho(\xi_1, \xi_2).$$

При этом знак плюс надо брать тогда, когда a_1 и a_2 имеют одинаковые знаки, а минус — в противном случае.

4. $-1 \le \rho(\xi_1, \xi_2) \le 1$, что следует из свойства 6 ковариации.

5. $|\rho(\xi_1, \xi_2)| = 1$ тогда и только тогда, когда случайные величины ξ_1 и ξ_2 линейно зависимы: $\xi_2 = a\xi_1 + b$ ($a \ne 0$), что следует из свойства 7 ковариации.

Итак, для независимых случайных величин $\rho = 0$, для линейно связанных $|\rho| = 1$, а в остальных случаях $-1 < \rho < 1$.

Чем ближе $|\rho|$ к единице, тем с большим основанием можно считать, что ξ_1 и ξ_2 находятся в линейной зависимости, т.е. коэффициент корреляции характеризует не всякую зависимость, а только так называемую *линейную вероятностную зависимость*, которая заключается в том, что при возрастании одной случайной величины другая имеет тенденцию изменяться по линейному закону. Можно сказать, что коэффициент корреляции ρ отражает степень линейной зависимости случайных величин. С возрастанием ξ_1 случайная величина ξ_2 имеет тенденцию к увеличению при $\rho > 0$ и к уменьшению при $\rho < 0$. Поэтому при $\rho > 0$ говорят о положительной корреляционной зависимости ξ_1 и ξ_2, при $\rho < 0$ — об отрицательной.

Если $\rho = 0$, то случайные величины называются *некоррелированными*. Из независимости случайных величин вытекает их некоррелированность, но наоборот — не всегда: равенство нулю коэффициента корреляции — необходимое, но недостаточное условие независимости случайных величин. Для нормально распределенных случайных величин свойства независимости и некоррелированности эквивалентны.

13.4. МОМЕНТЫ ВЫСШИХ ПОРЯДКОВ

В качестве других характеристик, описывающих свойства распределения случайной величины, используются начальные и центральные моменты.

Начальным моментом k-го порядка называется математическое ожидание k-й степени случайной величины, которое обозначается $\alpha_k = M\xi^k$.

Очевидно, что $M\xi = \alpha_1$.

Центральным моментом k-го порядка μ_k называется математическое ожидание k-й степени отклонения случайной величины от ее математического ожидания:

$$\mu_k = M(\xi - M\xi)^k.$$

Из определения следует, что $\mu_2 = M(\xi - M\xi)^2 = D\xi$.

Справедливы формулы:

$\mu_0 = 1;$

$\mu_1 = M(\xi - M\xi) = M\xi - M(M\xi) = M\xi - M\xi = 0;$

$\mu_2 = \alpha_2 - a^2;$

$\mu_3 = \alpha_3 - 3a\alpha_2 + 2a^3;$

$\mu_4 = \alpha_4 + 6a^2\alpha_2 - 4a\alpha_3 - 3a^4, \ldots$

Если распределение симметрично относительно математического ожидания, то все центральные моменты нечетного порядка равны нулю.

Величина $\beta_\xi = \mu_3/\sigma^3$ называется *коэффициентом асимметрии* случайной величины ξ. Коэффициент асимметрии характеризует «скошенность» распределения по отношению к математическому ожиданию. Для симметричных распределений $\mu_3 = 0$, поэтому коэффициент асимметрии равен нулю. Для распределений, скошенных влево, $\beta < 0$ (большие отрицательные отклонения определяют знак β), для скошенных вправо $\beta > 0$.

Величина $\gamma_\xi = \mu_4/\sigma^4 - 3$ называется *коэффициентом эксцесса* (или просто эксцессом) случайной величины ξ. Коэффициент эксцесса характеризует «сглаженность» или «крутость» распределения по отношению к нормальному, так как для нормального закона распределения $\mu_4/\sigma^4 = 3$ и, следовательно, $\gamma_\xi = 0$. Для более островершинных распределений $\gamma_\xi > 0$, для менее островершинных $\gamma_\xi < 0$ (рис. 13.3).

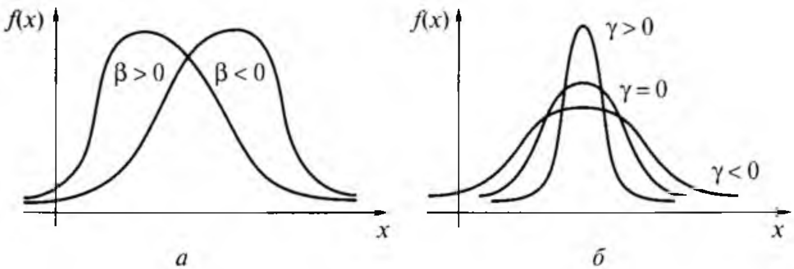

Рис. 13.3

Кроме начальных и центральных моментов, на практике применяются также так называемые *абсолютные моменты*, определяемые формулами

$$\bar{\alpha}'_k = M|\xi|^k, \quad \bar{\mu}'_k = M|\xi - M\xi|^k.$$

Очевидно, что абсолютные моменты четного порядка совпадают с обычными моментами. Чаще других применяется первый абсолютный момент $\mu_1' = M|\xi - M\xi|$, называемый *средним арифметическим отклонением*.

Математическое ожидание, мода, медиана, начальные и центральные моменты являются основными числовыми характеристиками случайной величины. На практике числовыми характеристиками часто пользуются для приближенной замены одного закона распределения другим, но так, чтобы сохранились неизменными несколько важнейших моментов.

Глава 14
ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

14.1. РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ

Непрерывная случайная величина ξ имеет равномерное распределение на отрезке $[a; b]$, если плотность распределения $f_\xi(x)$ сохраняет постоянное значение на этом промежутке:

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a; b], \\ 0, & x \notin [a; b]. \end{cases}$$

Отсюда функция распределения равна

$$F_\xi(x) = \begin{cases} 0, & x \le a, \\ \dfrac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases}$$

Графики функции распределения и плотности распределения вероятностей случайной величины приведены на рис. 14.1, *а*, *б*.

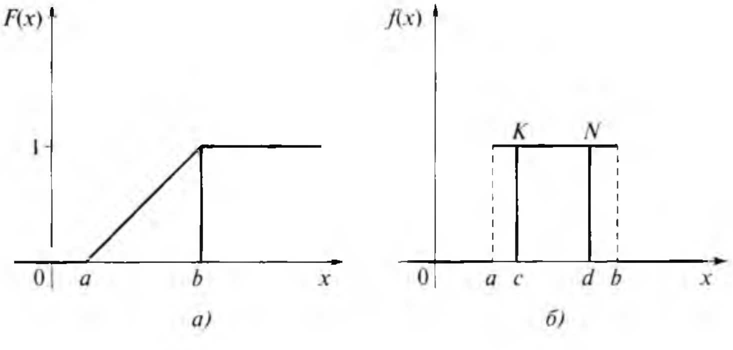

Рис. 14.1

Основные числовые характеристики равномерно распределенной на $[a; b]$ случайной величины ξ: математическое ожидание равно $M\xi = \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2}$, дисперсия равна $D\xi = \int_a^b \frac{\left(x - \frac{a+b}{2}\right)^2}{b-a} dx =$

$= \dfrac{(b-a)^2}{12}$. В силу симметричности равномерного распределения относительно математического ожидания асимметрия равна нулю, $x_{\text{мед}} = M\xi$, а моды равномерное распределение не имеет.

Для определения эксцесса вычислим четвертый центральный момент: $\mu_4 = \dfrac{1}{b-a}\int\limits_a^b \left(x - \dfrac{b+a}{2}\right)^4 dx = \dfrac{(b-a)^4}{80}$. Отсюда эксцесс равен $\gamma_\xi = \dfrac{\mu_4}{\sigma^4} - 3 = -1{,}2$.

Вероятность попадания случайной величины ξ на отрезок $[c; d]$ равна $P(c < \xi < d) = F(d) - F(c) = \dfrac{c-d}{b-a}$, зависит только от длины отрезка и не зависит от того, где этот отрезок расположен; геометрически это площадь прямоугольника $cKNd$.

14.2. ЭКСПОНЕНЦИАЛЬНОЕ (ПОКАЗАТЕЛЬНОЕ) РАСПРЕДЕЛЕНИЕ

Непрерывная случайная величина ξ, принимающая неотрицательные значения, имеет показательное распределение с параметром $\lambda > 0$, если плотность распределения вероятностей случайной величины равна

$$f_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Функция показательного распределения имеет вид

$$F_\xi(x) = \int\limits_0^x \lambda e^{-\lambda t} dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Графики функции распределения и плотности распределения экспоненциальной случайной величины приведены на рис. 14.2, *а*, *б*.

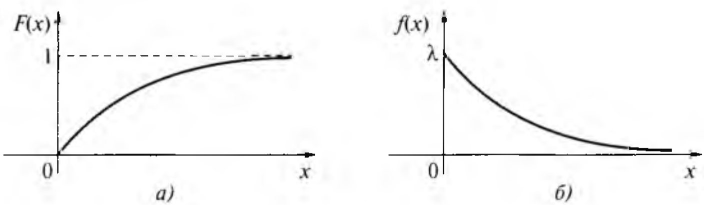

Рис. 14.2

Математическое ожидание экспоненциальной случайной величины равно

$$M\xi = \int_0^\infty t\lambda e^{-\lambda t}dt = |\lambda t = z, dz = \lambda dt| = \frac{1}{\lambda}\int_0^\infty ze^{-z}dz =$$

$$= \frac{1}{\lambda}\left(-ze^{-z}\Big|_0^\infty + \int_0^\infty e^{-z}dz\right) = \frac{1}{\lambda}.$$

Дисперсия случайной величины ξ, распределенной по показательному закону, равна

$$D\xi = M\xi^2 - (M\xi)^2 = \int_0^\infty t^2\lambda e^{-\lambda t}dt - \frac{1}{\lambda^2} = -\frac{1}{\lambda^2}\int_0^\infty z^2e^{-z}dz - \frac{1}{\lambda^2} =$$

$$= 2\frac{1}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Можно показать, что асимметрия $\beta_\xi = 2$, эксцесс $\gamma_\xi = 6$, $x_{\text{mod}} = 0$, $x_{\text{med}} = \frac{1}{\lambda}\ln 2$.

Случайная величина, распределенная по показательному закону, получила широкое применение в теории массового обслуживания.

14.3. РАСПРЕДЕЛЕНИЕ ЛАПЛАСА

Распределение Лапласа задается функцией плотности (двусторонняя показательная плотность, рис.14.3) $f_\xi(x) = \frac{\lambda}{2}e^{-\lambda|x|}, -\infty < x < \infty$.

Функция распределения $F(x) = \begin{cases} \frac{1}{2}e^{-\lambda x}, & x < 0, \\ 1 - \frac{1}{2}e^{-\lambda x}, & x \geq 0. \end{cases}$

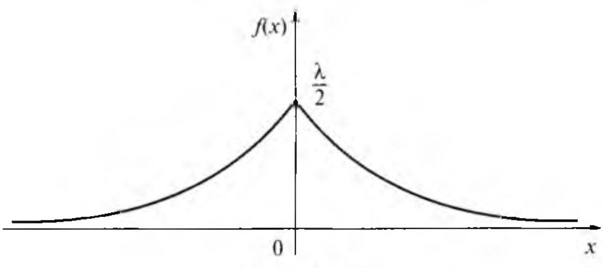

Рис. 14.3

Функция плотности распределения четная, и поэтому математическое ожидание $M\xi = 0$. В силу симметрии имеем $M\xi = X_{\text{med}} = X_{\text{mod}} = 0$, $\beta_\xi = 0$.

Дисперсия в два раза больше дисперсии случайной величины, распределенной по показательному закону:

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\lambda}{2}\int_{-\infty}^{\infty} x^2 e^{-\lambda|x|}dx = \lambda\int_0^{\infty} x^2 e^{-\lambda x}dx = \frac{2}{\lambda^2}.$$

Эксцесс равен $\gamma_\xi = 3$.

Симметричная унимодальная функция плотности этого закона с острым максимумом в точке нуль иногда используется для описания распределений остаточных случайных компонент (ошибок) в моделях регрессионного типа.

14.4. РАСПРЕДЕЛЕНИЕ ВЕЙБУЛЛА

Случайная величина ξ распределена по закону Вейбулла, если она имеет плотностью распределения функцию

$$f_\xi(x) = \begin{cases} \lambda\alpha x^{\alpha-1}e^{-\lambda x^\alpha}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Функция распределения имеет вид

$$F_\xi(x) = \begin{cases} 1 - e^{-\lambda x^\alpha}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Графики плотности и функции распределения Вейбулла приведены на рис. 14.4, *а*, *б*.

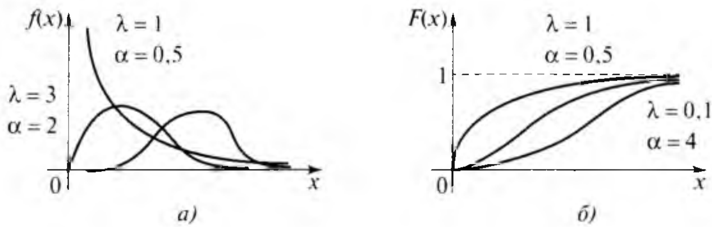

Рис. 14.4

Случайная величина, распределенная по закону Вейбулла, описывает время безотказной работы многих технических устройств, т.е. характеризует длительность жизни элемента сложной системы или индивидуума. В задачах данного профиля важной характеристикой является *интенсивность отказа* (*коэффициент смертности*) $\lambda(t)$ исследуемых элементов возраста t, определяемая соотношением $\lambda(t)=\dfrac{f_\xi(t)}{1-F_\xi(t)}$. Для распределения Вейбулла этот показатель принимает достаточно простой вид степенной функции: $\lambda(t)=\lambda\alpha t^{\alpha-1}$.

Если $\alpha=1$, то распределение Вейбулла превращается в показательное распределение, а если $\alpha=2$ — в так называемое распределение Рэлея.

Основные числовые характеристики распределения Вейбулла:

- математическое ожидание $M\xi=\lambda^{-\frac{1}{\alpha}}\Gamma\left(1+\dfrac{1}{\alpha}\right)$;
- дисперсия $D\xi=\lambda^{-\frac{2}{\alpha}}\left[\Gamma\left(1+\dfrac{2}{\alpha}\right)-\Gamma^2\left(1+\dfrac{1}{\alpha}\right)\right]$;
- мода — $x_{\text{mod}}=\begin{cases}0, & \alpha\le 1;\\ \lambda^{-\frac{1}{\alpha}}\left(1-\dfrac{1}{\alpha}\right)^{\frac{1}{\alpha}}, & \alpha\ge 1;\end{cases}$
- медиана — $x_{\text{med}}=\left(\dfrac{\ln 2}{\lambda}\right)^{\frac{1}{\alpha}}$;
- начальный момент k-го порядка $\alpha_k=M\xi^k=\lambda^{-\frac{k}{\alpha}}\Gamma\left(1+\dfrac{k}{\alpha}\right)$, где $\Gamma(x)$ — гамма-функция (или интеграл Эйлера второго рода: $\Gamma(\gamma)=\int\limits_0^\infty x^{\gamma-1}e^{-x}dx$, которая обладает следующими свойствами.

  1. $\Gamma(\gamma+1)=\gamma\Gamma(\gamma)$.
  2. $\Gamma(n)=(n-1)!$ для целых n.

14.5. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Нормальное распределение, или распределение Гаусса, занимает в теории вероятностей центральное место, поскольку при определенных условиях сумма достаточно большого числа случайных величин имеет распределение, близкое к нормальному распределению.

Непрерывная случайная величина называется распределенной по нормальному закону с параметрами a и σ^2, если ее плотность распределения равна

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Множество случайных величин, распределенных по нормальному закону с параметрами a и σ^2, обозначается через $\xi \in N(a, \sigma^2)$.

График плотности вероятностей нормального распределения называется кривой Гаусса (рис. 14.5).

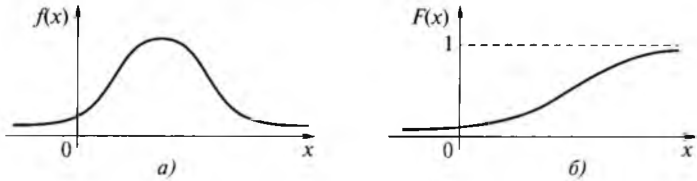

Рис. 14.5

Используя интеграл Эйлера — Пуассона

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \qquad \text{или} \qquad \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} = \sqrt{2\pi},$$

легко найти математическое ожидание и дисперсию нормально распределенной случайной величины:

$$M\xi = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left| \frac{x-a}{\sigma\sqrt{2}} = t, dt = \frac{dx}{\sigma\sqrt{2}} \right| =$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (a + t\sigma\sqrt{2})e^{-t^2} dt = \frac{a}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt + \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt =$$

$$= \frac{a}{\sqrt{\pi}} \sqrt{\pi} = a.$$

Аналогично вычисляется дисперсия $D\xi = \sigma^2$.

Таким образом, плотность нормального распределения зависит от двух параметров: $a = M\xi$ $\sigma^2 = D\xi$ — математического ожидания и дисперсии, которые полностью определяют нормальный закон распределения. Плотность нормального распределения симметрична относительно $x = a$, поэтому $x_{\text{med}} = x_{\text{mod}} = M\xi = a$.

Если $a = 0$ и $\sigma^2 = 1$, то нормальное распределение называется *стандартным* и класс таких распределений обозначается $N(0, 1)$.

Плотность стандартной нормальной случайной величиной равна $\varphi_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Функция стандартного нормального распределения имеет вид

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \Phi_0(x),$$

т.е выражается через функцию Лапласа $\Phi_0(x)$, и называется нормальной функцией распределения. Тогда функция распределения нормальной случайной величины с параметрами a и σ^2: $\xi \in N(a, \sigma^2)$ имеет вид

$$F_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt = \frac{1}{2} + \Phi_0\left(\frac{x-a}{\sigma}\right).$$

Отсюда, используя таблицы функции Лапласа, легко вычислить вероятность попадания нормально распределенной случайной величины на отрезок $[c_1; c_2]$:

$$P(c_1 \le x \le c_2) = F(c_2) - F(c_1) = \Phi_0((c_2 - a)/\sigma) - \Phi_0((c_1 - a)/\sigma).$$

Теоретически случайная величина, распределенная по нормальному закону, отлична от нуля в любой точке числовой прямой. Однако практически все возможные значения нормально распределенной случайной величины сосредоточены на отрезке $[a - 3\sigma; a + 3\sigma]$. Действительно,

$$P(a - 3\sigma < \xi < a + 3\sigma) = \Phi_0[(a + 3\sigma - a)/\sigma] - \Phi_0[(a - 3\sigma - a)/\sigma] =$$
$$= 2\Phi_0(3) = 0{,}9973.$$

Вероятность попасть вне этого отрезка не более 0,0027.

Отсюда правило трех сигм: *если случайная величина распределена по нормальному закону, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.*

Важными свойствами нормального распределения являются следующие.

1. Линейное преобразование стандартной нормальной случайной величины является случайной величиной, распределенной по нормальному закону.

Доказательство. Пусть случайная величина $\eta = \sigma\xi + a$, где случайная величина ξ имеет стандартное нормальное распределение

$\xi \in N(0, 1)$, и плотность ее распределения будет $f_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

Найдем плотность распределения η по формуле

$$f_\eta(x) = f_\xi(\varphi^{-1}(x)) \left| \frac{d\varphi^{-1}(x)}{dx} \right|.$$

Из условия $y = \sigma x + a$ получим, что $x = \frac{y - a}{\sigma}$ и $x' = \frac{1}{\sigma}$.

Тогда

$$f_\eta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \left| \frac{1}{\sigma} \right| = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \text{ т.е. } \eta \in N(a, \sigma^2).$$

Итак, любую нормальную случайную величину с параметрами a и σ^2 можно считать линейным преобразованием стандартной нормальной случайной величины $\xi \in N(0, 1)$.

Следствие. Если $\eta = A\xi + B$, где $\xi \in N(a, \sigma^2)$, то $\eta \in N(Aa + B, A^2\sigma^2)$.

Действительно, пусть $\eta = A\xi + B$ и $\xi \in N(a, \sigma^2)$. Тогда можно записать, что $\xi = \sigma\xi_1 + a$, где $\xi_1 \in N(0, 1)$. Отсюда следует $\eta = A(\sigma\xi_1 + a) + B = A\sigma\xi_1 + Aa + B$, и в силу доказанного свойства $\eta \in N(Aa + B, A^2\sigma^2)$.

2. Сумма двух независимых нормально распределенных случайных величин имеет нормальный закон распределения. При этом их математические ожидания и дисперсии складываются.

Для центральных моментов любого порядка нормально распределенной случайной величины можно вывести рекуррентное соотношение $\mu_k = (k - 1)\sigma^2\mu_{k-2}$, позволяющее выражать моменты высших порядков через моменты низших порядков. Так как $\mu_1 = 0$, то все нечетные моменты нормального распределения равны нулю. Для четных моментов получаем следующие выражения:

$$\mu_0 = 1, \quad \mu_2 = \sigma^2, \quad \mu_4 = 3\sigma^4, \ldots, \mu_{2k} = (2k - 1)!!\sigma^{2k}.$$

Отсюда *асимметрия* равна $\beta_\xi = \mu_3/\sigma^3 = 0$, *эксцесс* равен $\gamma_\xi = \mu_4/\sigma^4 - 3 = 0$.

14.6. ЛОГАРИФМИЧЕСКИ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Неотрицательная случайная величина ξ называется логарифмически нормально распределенной, если ее логарифм $\eta = \ln\xi$ подчинен нормальному закону распределения.

Из определения вытекает, что ξ может быть представлена как показательная функция $\xi = e^\eta$ от нормально распределенной случайной величины: $\eta \in N(a, \sigma^2)$. Найдем плотность распределения случайной величины ξ, имеющей логарифмически нормальное распределение.

Пусть $\eta = \ln\xi$, $\eta \in N(\ln a, \sigma^2)$, тогда, подставляя в формулу плотности нормального распределения $f_\eta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\ln a)^2}{2\sigma^2}}$ функцию $\varphi^{-1}(x) = \ln x$ и ее производную $\frac{d\varphi^{-1}(x)}{dx} = \frac{1}{x}$, получим

$$f_\xi(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \ln a)^2}{2\sigma^2}}.$$

Графики плотности логарифмически нормального распределения при различных значениях параметров приведены на рис. 14.6.

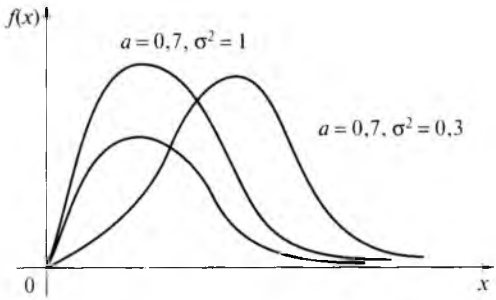

Рис. 14.6

Функция логарифмически нормального распределения имеет вид

$$F_\xi(x) = P(\xi < x) = P(\ln\xi < \ln x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\ln x} e^{-\frac{(t-\ln a)^2}{2\sigma^2}} dt.$$

Для описания логарифмически нормального распределения используется различная параметризация, и соответственно, по-разному выражаются математическое ожидание и дисперсия.

Если $\xi = ae^\eta$, $\eta \in N(0, \sigma^2)$, то математическое ожидание $M\xi = ae^{\frac{\sigma^2}{2}}$ и дисперсия $D\xi = a^2e^{\sigma^2}(e^{\sigma^2} - 1)$.

Если $\xi = e^\eta$ и $\eta \in N(a, \sigma^2)$, что будем обозначать $\xi \in \exp\{N(a, \sigma^2)\}$, то:

- математическое ожидание $M\xi = Ae^{\frac{\sigma^2}{2}}$;

- дисперсия $D\xi = A^2 e^{\sigma^2}(e^{\sigma^2} - 1)$;
- мода $x_{\mathrm{mod}} = Ae^{-\sigma^2}$;
- медиана $x_{\mathrm{med}} = A$, где $A = e^a$;
- асимметрия $\beta_\xi = (e^{\sigma^2} - 1)^{\frac{1}{2}}(e^{\sigma^2} + 2)$;
- эксцесс $\nu_\xi = (e^{\sigma^2} - 1)(e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6)$.

Из формул видно, что асимметрия и эксцесс логарифмически нормального распределения всегда положительны и тем ближе к нулю, чем ближе к нулю σ. Мода и медиана стремятся к слиянию по мере стремления к нулю величины σ.

Значения логарифмически нормальной случайной величины образуются как «случайные искажения» некоторого «истинного значения» A, которое является не средним значением, а медианой. Значения логарифмически нормальной случайной величины оказываются характерными для многих конкретных физических и социально-экономических ситуаций (размеры и вес частиц, образующихся при дроблении; заработная плата работника; доход семьи; размеры космических образований; долговечность изделия, работающего в режиме износа и старения, и др.).

14.7. РАСПРЕДЕЛЕНИЕ ПАРЕТО

В различных задачах прикладной статистики часто встречаются так называемые усеченные распределения. Такие распределения описывают вероятностные закономерности в неполных («усеченных») генеральных совокупностях, т.е. в таких совокупностях, из которых заранее изъяты все элементы с количественным признаком, превышающим некоторый заданный уровень c_0 (или, наоборот, меньшим, чем c_0). Например, налоговые органы интересуются распределением доходов тех лиц, годовой доход которых превосходит некоторый порог c_0, установленный законами о налогообложении. Эти распределения оказываются приближенно совпадающими с распределением Парето.

Распределение Парето задается функцией распределения вида $F_\xi(x) = P(\xi < x) = 1 - \left(\dfrac{c_0}{x}\right)^\alpha$ и плотностью распределения вероятностей $f_\xi(x) = \dfrac{\alpha}{c_0}\left(\dfrac{c_0}{x}\right)^{\alpha+1}$.

В этих формулах $\alpha > 0$, а $x > c_0$, т.е. область возможных значений случайной величины x есть полупрямая $(c_0, +\infty)$. Функция плотности имеет вид монотонно убывающей кривой, выходящей из точки $(c_0, \alpha/c_0)$ (рис. 14.7).

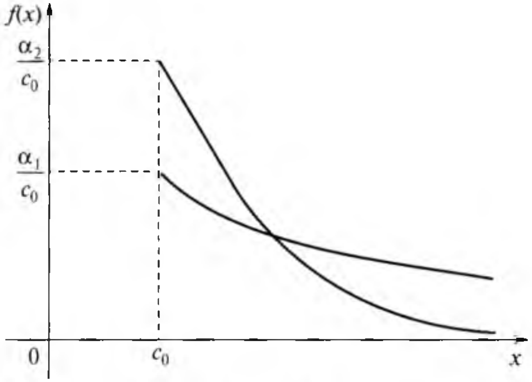

Рис. 14.7

Основные числовые характеристики этого распределения существуют не всегда, а лишь при соблюдении определенных требований к значению параметра α:

- математическое ожидание $M\xi = \dfrac{\alpha c_0}{\alpha - 1}$ при $\alpha > 1$;
- мода $x_{\text{mod}} = c_0$;
- медиана $x_{\text{med}} = c_0 2^{\frac{1}{\alpha}}$;
- дисперсия $D\xi = \dfrac{\alpha c_0^2}{(\alpha - 1)^2(\alpha - 2)}$ существует при $\alpha > 2$;
- начальный момент k-го порядка $M\xi^k = \dfrac{\alpha c_0^k}{\alpha - k}$ существует при $\alpha > k$.

14.8. РАСПРЕДЕЛЕНИЕ КОШИ

Этот закон весьма специфичен, так как ни один из его моментов положительного порядка (в том числе математическое ожидание) не существует. Распределение Коши унимодально, симметрично относительно своего модального значения и задается *функцией плотности распределения*

$$f_\xi(x) = \frac{c}{\pi(c^2 + (x - \alpha)^2)}, \quad -\infty < x < +\infty,$$

где $c > 0$ — *параметр масштаба* и α — *параметр сдвига* (*центр группировки*), определяющий одновременно значение моды и медианы. Функция распределения задается формулой

$$F_\xi(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg}\frac{x - \alpha}{c}.$$

Стандартное распределение Коши соответствует случаю $\alpha = 0$, $c = 1$.

Отметим два важных свойства (самовоспроизводимости) распределения Коши.

1. Если случайная величина ξ имеет распределение Коши с параметрами c и α, то любая линейная функция $b_0 + b_1\xi$ имеет распределение того же типа с параметрами $c' = |b_1|c$ и $\alpha' = b_1\alpha + b_0$.

2. Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и имеют одно и то же распределение Коши, то среднее арифметическое $\bar{\xi} = (\xi_1 + \xi_2 + \ldots + \xi_n)/n$ имеет то же распределение Коши, что и ξ_i.

14.9. ПРОИЗВОДЯЩАЯ ФУНКЦИЯ МОМЕНТОВ

Производящая функция моментов случайной величины ξ — это функция от параметра t, которая равна математическому ожиданию функции $e^{t\xi}$: $m_\xi(t) = Me^{t\xi}$.

Наиболее важным является то, что производящая функция $m_\xi(t)$ содержит в себе сведения о всех начальных моментах («производит» моменты). Действительно, производная от производящей функции в точке нуль равна первому начальному моменту: $m'_\xi(0) = M\xi e^{t\xi}(0) = M\xi = \alpha_1$. В общем случае для любого k получаем $m_\xi^{(k)}(0) = M\xi^k e^{t\xi}(0) = \alpha_k$, т.е. k-я производная от производящей функции при $t = 0$ равна начальному моменту k-го порядка, а любой центральный момент можно выразить через начальный момент.

Свойства производящей функции следующие.

1. $m_{c\xi+a}(t) = Me^{t(c\xi+a)} = m_\xi(ct)e^{at}$.

2. Производящая функция суммы независимых случайных величин равна произведению производящих функций этих величин.

Действительно, пусть $\xi = \sum_{i=1}^{n} \xi_i$, тогда

$$m_\xi(t) = M\exp\left(\sum_{i=1}^{n} t\xi_i\right) = M\prod_{i=1}^{n}(\exp(t\xi_i)) = \prod_{i=1}^{n} m_{\xi_i}(t).$$

По производящей функции можно также определить функцию распределения, содержащую все сведения о случайной величине. В этом смысле производящая функция и функция распределения являются эквивалентными обобщающими характеристиками случайной величины

Примеры.

1. Производящая функция моментов показательного распределения с параметром λ описывается формулой $m_\xi(t) = \dfrac{\lambda}{\lambda - t}$ при $t < \lambda$ (в противном случае производящей функции не существует).

2. Производящая функция моментов нормального распределения имеет вид $m_\xi(t) = e^{at + \frac{\sigma t}{2}}$ и существует при любом значении параметра t.

14.10. УСЛОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ. ФУНКЦИЯ РЕГРЕССИИ

Пусть двумерная случайная величина (ξ, η) задана в вероятностном пространстве $(\Omega, \mathfrak{I}, P)$ совместной функцией распределения $F_{\xi\eta}(x, y)$ и пусть дискретные случайные величины ξ и η принимают значения x_i $(i = 1, 2, ...)$ и y_j $(j = 1, 2, ...)$ соответственно.

Условной функцией распределения $F_{\xi|\eta}(\xi \mid \eta = y_j)$ дискретной случайной величины ξ при условии $\eta = y_j$ называется условная вероятность события $\{\omega: \xi < x\}$ при условии, что имело место событие $\{\omega: \eta = y_j\}$, т.е.

$$F_{\xi|\eta}(x \mid \eta = y_j) = P(\xi < x \mid \eta = y_j) = \frac{P(\xi < x_i, \eta = y_j)}{P(\eta = y_j)}$$

при условии, что $P(\eta = y_j) > 0$.

В общем случае *условную функцию распределения* случайной величины ξ при условии $\eta = y$ также естественно было бы определить формулой

$$F_{\xi|\eta}(x \mid \eta = y) = \frac{P(\xi < x, \eta = y)}{P(\eta = y)}.$$

Однако поскольку для непрерывной случайной величины $P(\eta = y) = 0$, то вместо события $\{\omega: \eta = y\}$ рассматривают событие $\{\omega: y \le \eta < y + \Delta y\}$:

$$P(\xi < x/y \le \eta < y+\Delta y) = \frac{P(\xi < x, y \le \eta < y+\Delta y)}{P(y \le \eta < y+\Delta y)} =$$

$$= \frac{F_{\xi\eta}(x, y+\Delta y) - F_{\xi\eta}(x, y)}{F_{\eta}(y+\Delta y) - F_{\eta}(y)},$$

и условной функцией распределения $F_{\xi|\eta}(x/\eta = y)$ непрерывной случайной величины ξ называется предел

$$F_{\xi|\eta}(x/\eta = y) = \lim_{\Delta y \to 0} P(\xi < x/y \le \eta < y+\Delta y).$$

Такой предел всегда существует, но не однозначен. Если в выражении для $P(\xi < x/y \le \eta < y+\Delta y)$ поделить числитель и знаменатель на Δy и перейти к пределу при $\Delta y \to 0$, то

$$F_{\xi|\eta}(x/\eta = y) = \frac{\dfrac{\partial F_{\xi\eta}(x, y)}{\partial y}}{f_{\eta}(y)},$$

где $F_{\xi\eta}(x, y) = P(\xi < x, \eta < y)$.

Условная функция распределения обладает всеми свойствами, которые присущи обычной (безусловной) функции распределения.

Если вектор (ξ, η) представляет собой двумерную абсолютно непрерывную случайную величину с совместной плотностью распределения $f_{\xi\eta}(x, y)$, то

$$F_{\xi|\eta}(x \mid \eta = y) = \int_{-\infty}^{x} \frac{f_{\xi\eta}(u, y)}{f_{\eta}(y)} du$$

и условная плотность распределения случайной величины ξ при условии $\eta = y$ равна $f_{\xi|\eta}(x \mid \eta = y) = \dfrac{\partial F_{\xi|\eta}(x/\eta = y)}{\partial x} = \dfrac{f_{\xi\eta}(x, y)}{f_{\eta}(y)}$.

Условным математическим ожиданием дискретной случайной величины ξ при условии $\eta = y_i$ называется число

$$M(\xi \mid \eta = y_j) = \sum_i x_i P_{\xi|\eta}(x_i \mid y_j).$$

В случае непрерывной случайной величины ξ условное математическое ожидания определяется формулой

$$M(\xi \mid \eta = y) = \int_{-\infty}^{\infty} x f_{\xi|\eta}(x/\eta) dx,$$

где $f_{\xi|\eta}(x \mid \eta = y) = f(x, y)/f_{\eta}(y)$ — условная плотность распределения случайной величины ξ при условии $\eta = y$.

Условное математическое ожидание описывает среднее значение случайной величины ξ при условии, что случайная величина η приняла значение y.

Из определения следует, что условное математическое ожидание случайной величины ξ является функцией $M(\xi \mid \eta = y) = \varphi_{\xi|\eta}(y)$ от случайной величины η и, следовательно, тоже случайная величина. Область определения функции $\varphi(y)$ совпадает с областью значений случайной величины η, т.е. можно сказать, что зависимость поведения «в среднем» случайной величины ξ от значения случайной величины η характеризуется функцией $\varphi(y) = M(\xi \mid \eta = y)$.

Функция $\varphi(y) = M(\xi \mid \eta = y)$ называется *функцией регрессии* ξ по η, а ее график называется *линией регрессии* ξ по η.

Линия регрессии дает наглядное изображение зависимости «в среднем» случайной величины ξ от значений случайной величины η.

Аналогично значение условного математического ожидания случайной величины η при условии $\xi = x$ определяется формулой

$$M(\eta/\xi = x) = \int_{-\infty}^{\infty} y f_{\eta}(y/\xi = x)dy$$

и называется функцией регрессии случайной величины η по ξ, а ее график — линией регрессии η по ξ.

Условное математическое ожидание $M(\xi \mid \eta)$ обладает следующими свойствами.

1. $M(c \mid \eta) \equiv c$.
2. $M(a\xi + b \mid \eta) = aM(\xi \mid \eta) + b$.
3. $M(\xi_1 + \xi_2 \mid \eta) = M(\xi_1 \mid \eta) + M(\xi_2 \mid \eta)$. Эти свойства аналогичны свойствам безусловного математического ожидания.
4. $M(\xi_1\xi_2 \mid \eta) = M(\xi_1 \mid \eta)M(\xi_2 \mid \eta)$, при этом условие независимости случайных величин ξ_1 и ξ_2 надо заменить требованием, которое называется условной независимостью случайных величин ξ_1 и ξ_2 при условии η.
5. $M\xi = M[M(\xi \mid \eta)]$.
6. $M[f(\xi)h(\eta) \mid \eta] = h(\eta)M[f(\xi) \mid \eta]$, где $f(\xi)$ и $h(\eta)$ — произвольные функции от случайных величин ξ и η.
7. $M(\xi \mid \eta) \equiv M(\xi)$, если ξ и η независимы.

Пример 14.1. Двумерная случайная величина задана рядом распределения:

$\eta \backslash \xi$	1	3	4	8	P_{η}
3	0,15	0,06	0,25	0,04	

$\eta \backslash \xi$	1	3	4	8	P_η
6	0,30	0,10	0,03	0,07	
P_ξ					

Найти условное математическое ожидание составляющей η при $\xi = 1$.

Решение. Условное математическое ожидание равно

$M(\eta/\xi = x_1) = y_1 P(y_1/x_1) + y_2 P(y_2/x_1)$.

Из условия задачи найдем распределение составляющих η и ξ (последний столбец и последняя строка таблицы).

$\eta \backslash \xi$	1	3	4	8	P_η
3	0,15	0,06	0,25	0,04	0,50
6	0,30	0,10	0,03	0,07	0,50
P_ξ	0,45	0,16	0,28	0,11	

Так как $P(x_1) = P(x_1, y_1) + P(x_1, y_2) = 0,15 + 0,30 = 0,45$, то условные вероятности найдем по формулам

$$P(y_1/x_1) = \frac{P(x_1, y_1)}{P(x_1)} = \frac{0,15}{0,45} = \frac{1}{3}, \qquad P(y_2/x_1) = \frac{P(x_1, y_2)}{P(x_1)} = \frac{0,30}{0,45} = \frac{2}{3},$$

и искомое условное математическое ожидание будет равно

$$M(\eta/\xi = 1) = 3 \cdot \frac{1}{3} + 6 \cdot \frac{2}{3} = 5.$$

Пример 14.2. Предположим, что в соответствии с принципом геометрической вероятности мы бросаем точку случайным образом в круг радиуса R с центром в начале координат. Найти условную плотность распределения случайной величины ξ — абсциссы точки падения при условии, что ордината η приняла значение y.

Решение. Поскольку точка не может попасть за пределы круга, то $f(x, y) = 0$ при $x^2 + y^2 > R^2$. Для каждой области внутри круга вероятность попадания пропорциональна площади этой области. Поэтому $f(x, y) = A$ при $x^2 + y^2 \le R^2$, т.е. плотность внутри круга постоянна. Определим константу A:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = \iint\limits_{x^2+y^2 \le R^2} Adxdy = \pi A R^2 = 1.$$

Отсюда $A = 1/\pi R^2$ и плотность совместного распределения

$$f(x, y) = \begin{cases} 0, & x^2 + y^2 > R^2, \\ \dfrac{1}{\pi R^2}, & x^2 + y^2 \le R^2 \end{cases} = \begin{cases} 0, & |x| > \sqrt{R^2 - y^2}, \\ \dfrac{1}{\pi R^2}, & |x| \le \sqrt{R^2 - y^2}. \end{cases}$$

Частные плотности распределения случайных величин ξ и η соответственно равны:

$$f_\eta(y) = \begin{cases} 0, & |y| > R, \\ \dfrac{2\sqrt{R^2 - y^2}}{\pi R^2}, & |y| \le R, \end{cases} \qquad f_\xi(x) = \begin{cases} 0, & |x| > R, \\ \dfrac{2\sqrt{R^2 - x^2}}{\pi R^2}, & |x| \le R, \end{cases}$$

и условная плотность распределения ξ при условии, что $\eta = y$, равна

$$f_\xi(x/\eta = y) = \begin{cases} 0, & |x| > \sqrt{R^2 - y^2}, \\ \dfrac{1}{2\sqrt{R^2 - y^2}}, & |x| \le \sqrt{R^2 - y^2} \end{cases} \quad \text{при } |y| \le R.$$

Таким образом, случайная величина ξ при условии $\eta = y$ равномерно распределена на отрезке $\left[-\sqrt{R^2 - y^2};\ +\sqrt{R^2 - y^2}\right]$. Интересно отметить, что условная плотность распределения случайной величины ξ при условии $\eta = y$ равномерна, в то время как безусловная плотность ξ таковой не является. И в этом примере случайные величины ξ и η зависимы между собой.

Глава 15
ЗАКОН БОЛЬШИХ ЧИСЕЛ

15.1 НЕРАВЕНСТВО МАРКОВА. НЕРАВЕНСТВО ЧЕБЫШЕВА

Для любой случайной величины ξ показателем разброса ее значений вокруг математического ожидания является дисперсия. Однако, с точки зрения исследователя, разброс естественнее характеризовать вероятностью $P(|\xi - M\xi| < \varepsilon)$ отклонения случайной величины ξ от ее математического ожидания.

Неравенство Маркова. Для любой случайной величины ξ вероятность события $\{\omega: |\xi| \geq \varepsilon\}$ не превосходит произведения $1/\varepsilon$ на математическое ожидание модуля случайной величины: $P(|\xi| \geq \varepsilon) \leq \dfrac{M|\xi|}{\varepsilon}$.

Заметим, что событие $\{\omega: |\xi| \geq \varepsilon\}$ равносильно событию $\{\omega: \xi^2 \geq \varepsilon^2\}$. Используя неравенство Маркова, можно получить оценку события $\{\omega: |\xi| \geq \varepsilon\}$ через второй начальный момент: $P(|\xi| \geq \varepsilon) = P(\xi^2 \geq \varepsilon^2) \leq \dfrac{M\xi^2}{\varepsilon^2}$.

Неравенство Чебышева. Для каждой случайной величины ξ, имеющей дисперсию $D\xi = \sigma^2$, и для любого $\varepsilon > 0$ справедливо неравенство $P(|\xi - M\xi| \geq \varepsilon) \leq \dfrac{D\xi}{\varepsilon^2}$.

Доказательство. Применим следствие неравенства Маркова к случайной величине $\eta = |\xi - M\xi|$, равной отклонению случайной величины ξ от ее математического ожидания. Вероятность события $\{\omega: |\eta| \geq \varepsilon\}$ удовлетворяет неравенству $P(|\eta| \geq \varepsilon) \leq \dfrac{M\eta^2}{\varepsilon^2}$ или, заменяя η его выражением через ξ, получим

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{M(\xi - M\xi)^2}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

Неравенство Чебышева справедливо для любой случайной величины, имеющей дисперсию, и оценка вероятности в нем не зависит от закона распределения случайной величины ξ. Применять неравенство Чебышева имеет смысл только тогда, когда $\sigma < \varepsilon$, в противном случае оно дает тривиальную оценку.

Следствие. Для любого $\varepsilon > 0$ и любой случайной величины ξ, имеющей дисперсию $D\xi = \sigma^2$, справедливо неравенство $P(|\xi - M\xi| < \varepsilon) \geq 1 - \dfrac{D\xi}{\varepsilon^2}$.

Неравенство Колмогорова. Если независимые случайные величины $\xi_1, \xi_2, ..., \xi_n$ имеют математические ожидания $M\xi_i$ и дисперсии $D\xi_i$, то для любого $\varepsilon > 0$ верно

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}M\xi_i\right| < \varepsilon\right) \geq 1 - \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}D\xi_i.$$

Пример 15.1. В 400 испытаниях Бернулли вероятность успеха в каждом испытании равна 0,8. С помощью неравенства Чебышева оценить вероятность того, что разница между числом успехов в этих испытаниях и средним числом успехов будет меньше 20.

Решение. Число успехов в этих испытаниях распределено по закону Бернулли, поэтому среднее число успехов равно $M\xi = np = 400 \cdot 0{,}8 = 320$, а дисперсия $D\xi = npq = 400 \cdot 0{,}8 \cdot 0{,}2 = 64$. Тогда в силу неравенства Чебышева имеем

$$P(|\xi - 320| < 20) \geq 1 - \frac{D\xi}{20^2} = 1 - \frac{64}{400} = 0{,}84.$$

Вычислим эту же вероятность с помощью приближенной (интегральной) формулы Муавра—Лапласа (см. гл. 4):

$$P(|\xi - 320| < 20) = P(|\xi - np| < \varepsilon) = P\left(-\frac{\varepsilon}{\sqrt{npq}} < \frac{\xi - np}{\sqrt{npq}} < \frac{\varepsilon}{\sqrt{npq}}\right) =$$

$$= 2\Phi_0\left(\frac{\varepsilon}{\sqrt{npq}}\right) = 2\Phi_0\left(\frac{20}{\sqrt{64}}\right) = 2\Phi_0(2{,}5) = 2 \cdot 0{,}4938 = 0{,}9876.$$

Последнее вычисление показывает, что неравенство Чебышева дает довольно грубые оценки вероятностей.

15.2. ЗАКОН БОЛЬШИХ ЧИСЕЛ В ФОРМЕ ЧЕБЫШЕВА

Теорема 15.1. Пусть $\xi_1, \xi_2, ..., \xi_n, ...$ — последовательность независимых случайных величин, дисперсии которых ограничены сверху одним и тем же числом c: $D\xi_i \leq c$, $i = 1, 2, ...$. Тогда для любого $\varepsilon > 0$ справедливо равенство

$$\lim_{n\to\infty}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}M\xi_i\right| < \varepsilon\right) = 1.$$

Доказательство. Поскольку $D\xi_i \leq c$, $i = 1, 2, ...$, то $\frac{1}{n}\sum_{i=1}^{n}D\xi_i \leq c$, и в силу следствия из неравенства Чебышева получаем

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}M\xi_i\right| < \varepsilon\right) \geq 1 - \frac{\sum_{i=1}^{n}D\xi_i}{n^2\varepsilon^2};$$

и так как вероятность любого события не больше единицы, то

$$1-\frac{c}{n\varepsilon^2}\le P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}\sum_{i=1}^{n}M\xi_i\right|<\varepsilon\right)\le 1.$$

Переходя к пределу в полученном неравенстве при $n\to\infty$, получаем доказательство теоремы.

Следствия

1. Если $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — последовательность попарно независимых одинаково распределенных случайных величин с математическим ожиданием a и дисперсией σ^2, то для любого $\varepsilon > 0$ справедливо равенство $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i-a\right|<\varepsilon\right)=1.$

2. Закон больших чисел для числа успехов в независимых испытаниях Бернулли называется **теоремой Бернулли**.

Теорема 15.2. Если μ_n — число успехов в n независимых испытаниях Бернулли и p — вероятность успеха в одном испытании, то

$$\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)=1.$$

Доказательство следует из неравенства Чебышева, если μ_n представить в виде суммы случайных величин $\mu_n=\sum\limits_{i=1}^{n}\xi_i$, где ξ_i — число успехов в i-м испытании.

3. **Теорема 15.3.** Если $\bar{\mu}_n$ — число успехов в n независимых испытаниях, в каждом из которых успех появляется с вероятностью p_i, $i = 1, 2, \ldots$, то для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n}-\frac{1}{n}\sum_{i=1}^{n}p_i\right|<\varepsilon\right)=1.$$

Закон больших чисел в случае различных p_i называют **теоремой Пуассона**.

15.3. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА (ЦПТ)

Смысл закона больших чисел заключается в том, что при осреднении большого числа случайных слагаемых все меньше ощущается характерный для случайных величин неконтролируемый разброс их значений и в пределе при $n\to\infty$ этот разброс исчезает вовсе или, как принято говорить, случайная величина вырождается в неслучайную.

Однако при конечном числе слагаемых случайный разброс остается и возникает вопрос исследования характера этого разброса. Фундаментальный результат в этом направлении, известный как центральная предельная теорема, заключается в том, что для широкого класса независимых случайных величин предельный закон распределения их нормированной суммы вне зависимости от типа распределения слагаемых стремится к нормальному закону распределения.

Теорема 15.4 (центральная предельная теорема). Пусть ξ_1, ξ_2, ..., ξ_n, ... — независимые одинаково распределенные случайные величины с $M\xi_i = a$ и $D\xi_i = \sigma^2$, $i = 1, 2, ..., n$, ... Тогда для любого числа x верно

$$\lim_{n\to\infty} P\left(\frac{\xi_1 + \xi_2 + ... + \xi_n - na}{\sigma\sqrt{n}} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}dy.$$

Из ЦПТ следует, что вероятность попадания суммы $S_n = \xi_1 + ... + \xi_n$ в любой интервал имеет следующий предел:

$$\lim_{n\to\infty} P\left(c_1 < \frac{S_n - na}{\sigma\sqrt{n}} < c_2\right) = \frac{1}{\sqrt{2\pi}}\int_{c_1}^{c_2} e^{-\frac{y^2}{2}}dy = \Phi(c_2) - \Phi(c_1) =$$

$$= \Phi_0(c_2) - \Phi_0(c_1),$$

откуда следует приближенная формула

$$P(A < S_n < B) \approx \Phi_0\left(\frac{B - na}{\sigma\sqrt{n}}\right) - \Phi_0\left(\frac{A - na}{\sigma\sqrt{n}}\right),$$

которой имеет смысл пользоваться, если A и B не очень далеки от na.

В общем случае сходимость случайной последовательности X_n к пределу (случайной величине) X вида

$$\lim_{n\to\infty} P(X_n < x) = P(X < x)$$

в точках непрерывности функции распределения X называется *сходимостью по распределению* и обозначается так: $X_n \xrightarrow{d} X$.

Из сходимости по вероятности следует сходимость по распределению, обратное верно не всегда (но верно в случае сходимости к константе).

Пример 15.2. Пусть ξ_1, ξ_2, ..., ξ_i, ... — последовательность случайных величин, удовлетворяющая условиям предыдущего примера. В этом случае сумма $S_n = \xi_1 + \xi_2 + ... + \xi_n$ есть число успехов m в n испытаниях Бернулли. Из ЦПТ следует, что

$$\lim_{n\to\infty} P\left(c_1 < \frac{m-np}{\sqrt{npq}} < c_2\right) = \frac{1}{\sqrt{2\pi}}\int_{c_1}^{c_2} e^{-\frac{x^2}{2}}dx = \Phi_0(c_2) - \Phi_0(c_1),$$

где $\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{y^2}{2}}dy$ — функция Лапласа.

Тогда вероятность того, что число успехов заключено между m_1 и m_2, равна

$$P(m_1 \le m \le m_2) = P\left(\frac{m_1-np}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}} \le \frac{m_2-np}{\sqrt{npq}}\right) \approx$$

$$\approx \Phi_0\left(\frac{m_2-np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{m_1-np}{\sqrt{npq}}\right).$$

Эта формула следует из интегральной теоремы Муавра—Лапласа.

Пример 15.3. В продукции цеха детали отличного качества составляют 50%. Детали укладываются в коробки по 200 шт. в каждой. Какова вероятность того, что число деталей отличного качества в коробке отличается от 100 не более чем на 5?

Решение. Пусть ξ_i — случайное число деталей отличного качества в i-й коробке, тогда при $n = 200$, $p = q = {}^1/_2$ получим

$$P(95 \le m \le 105) = P\left(-\frac{5}{\sqrt{50}} \le \frac{m-np}{\sqrt{npq}} \le \frac{5}{\sqrt{50}}\right) \approx$$

$$\approx \Phi_0(0{,}71) - \Phi_0(-0{,}71) \approx 0{,}52.$$

Пример 15.4. Используя условия примера 15.3, указать, в каких границах с вероятностью 0,997 находится число деталей отличного качества в коробке.

Решение. По таблице функции Лапласа при условии $P\left(\left|\frac{m-np}{\sqrt{npq}}\right| \le u\right) \approx$ $\approx 0{,}997$ находим $u = 3$, и следовательно, S_n лежит в пределах $np \pm 3\sqrt{npq}$, т.е. число деталей отличного качества в коробке с вероятностью 0,997 находится в пределах 100 ± 21.

Пример 15.5. Используя условия примера 15.3, определить, сколько деталей надо взять, чтобы с вероятностью, не меньшей 0,99, можно было утверждать, что число деталей отличного качества среди них не менее 100.

Решение. Обозначим $u = \frac{100-np}{\sqrt{npq}}$. Используя нормальное приближение, получаем

$$P(m \ge 100) = P\left(\frac{m-np}{\sqrt{npq}} \ge u\right) \approx 1 - \Phi(u) = \frac{1}{2} - \Phi_0(u) \ge 0{,}99..$$

Отсюда $\Phi_0(u) \le -0{,}49$, а из таблицы значений функции Лапласа, находящейся в приложении, получаем неравенство $u \le -2{,}32$. Обозначив $x = \sqrt{n} > 0$, с учетом $p = q = {}^1/_2$ приходим к квадратному неравенству $x^2 - 2{,}3x - 200 \ge 0$, решая которое получаем $n \ge 236$.

Можно предложить и другой метод, а именно: пусть ξ_i — число деталей, которые пришлось перебрать, чтобы найти i-ю деталь отличного качества (включая ее саму). Случайные величины имеют геометрическое распределение с параметром $p = {}^1/_2$. Можем вычислить $M\xi = 1/p = 2$, $D\xi = (1-p) : : p^2 = 2$. Используя ЦПТ, получаем неравенство

$$P(S_{100} \le n) = \Phi\left(\frac{n - 100 \cdot 2}{\sqrt{2}\sqrt{100}}\right) = \frac{1}{2} + \Phi_0\left(\frac{n - 200}{14{,}14}\right) \ge 0{,}99,$$

откуда следует $n \ge 200 + 14{,}14 \cdot 2{,}32 = 232{,}8$, или, округляя, $n \ge 234$.

Результаты получаются близкие, но первый метод более точен и потому здесь предпочтительней. Вторым методом лучше пользоваться, если нужно определить границы, в которых лежит неизвестное число деталей.

Пример 15.6. Доходы жителей города имеют математическое ожидание 10 тыс. руб. и среднее квадратическое отклонение 2 тыс. руб. (в месяц). Найти вероятность того, что средний доход 100 случайно выбранных жителей составит от 9,5 до 10,5 тыс. руб.

Решение. Переформулируем условие задачи для суммарного дохода: он должен составлять от 950 до 1050 тыс. руб. Используя ЦПТ, получаем

$$P(950 < S_{100} < 1050) = \Phi_0\left(\frac{1050 - 100 \cdot 10}{2\sqrt{100}}\right) - \Phi_0\left(\frac{950 - 100 \cdot 10}{2\sqrt{100}}\right) =$$

$$= 2\Phi_0(2{,}5) = 0{,}9876.$$

ЛИТЕРАТУРА

1. *Бочаров П.П., Печинкин А.В.* Теория вероятностей. Математическая статистика. — М. Гардарика, 1998.
2. *Гнеденко Б.В.* Курс теории вероятностей. — М.: Наука, 1960.
3. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1998.
4. *Грачева М.В., Фадеева Л.Н., Черемных Ю.Н.* Количественные методы в экономических исследованиях. — М.: ЮНИТИ, 2004.
5. *Крамер Гаральд.* Математические методы статистики. — М.: Мир, 1975.
6. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
7. *Фадеева Л.Н.* Теория вероятностей и математическая статистика. — М.: Эксмо, 2006.
8. *Фадеева Л.Н., Лебедев А.В.* Теория вероятностей и математическая статистика. — М.: Эксмо, 2010.
9. *Чистяков В.П.* Курс теории вероятностей. — М.: Наука, 1982.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Часть 2
КОЛИЧЕСТВЕННЫЙ АНАЛИЗ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

Раздел V
АНАЛИЗ СТАТИСТИЧЕСКИХ МЕТОДОВ И МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

Глава 16
ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

16.1. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ СОВОКУПНОСТИ

Математическая статистика — это наука, которая, основываясь на методах теории вероятностей, занимается систематизацией и обработкой статистических данных для получения научных и практических выводов. *Статистическими данными* называются сведения о множестве объектов, обладающих теми или иными признаками.

Группа объектов, объединенных по некоторому качественному или количественному признаку, называется *статистической совокупностью*. Объекты, входящие в совокупность, называются ее *элементами*, а общее число элементов совокупности — ее *объемом*.

Генеральной совокупностью называется множество объектов произвольной природы, обладающих признаками, доступными для наблюдения и количественного измерения.

Например, в случае социально-экономических исследований это может быть население какого-то города, региона или страны, а измеряемыми признаками — доходы, расходы или объем сбережений отдельно взятого человека. Если какой-то признак имеет качественный характер (например, пол, национальность, социальное положение, род деятельности и т.п.), но принадлежит к конечному множеству вариантов, он может быть также закодирован числом (как это часто делают в анкетах).

Объекты, входящие в генеральную совокупность, называются ее *элементами*, а их общее число — ее *объемом*.

Пусть число элементов генеральной совокупности равно N. Примем каждый из них за элементарный исход ω некоторого вероятностного пространства Ω и припишем всем исходам одинаковую вероятность $1/N$. Тогда соответствие между объектами и значениями какого-либо их признака задает случайную величину $\xi = \xi(\omega)$ как функцию на вероятностном пространстве (согласно аксиоматике А.Н. Колмогорова). Числовые характеристики введенной формально случайной величины отражают важные свойства совокупности исследуемых объектов. В частности, ее математическое ожидание $M\xi$ равно среднему значению признака, а функция распределения $F_{\xi}(x) = P(\xi < x)$ показывает долю объектов, для которых значение признака меньше x. Например, в социально-экономических исследованиях нас могут интересовать средний доход на душу населения, доля людей с доходами меньше прожиточного минимума и т.п.

Распределение ξ называют *распределением генеральной совокупности*.

Будем последовательно извлекать из генеральной совокупности ее элементы, выбирая их случайным образом (наудачу), измерять и записывать значения некоторого признака для них: $x_1, x_2, \ldots, x_n$. Эти значения называются *наблюдениями* (признака), их набор — *выборкой*, а число сделанных наблюдений — *объемом выборки n*. Понятно, что наблюдения представляют собой случайные величины в силу случайности нашего выбора объекта. Они заданы уже на другом вероятностном пространстве — на множестве всех вариантов выбора n элементов из генеральной совокупности. Полученные данные называют *наблюдениями случайной величины* ξ, а также говорят, что случайная величина ξ «принимает значения» $x_1, x_2, \ldots, x_n$.

Выборка называется *повторной*, если отобранный объект перед выбором следующего возвращается в генеральную совокупность. *Выборка* называется *бесповторной*, если отобранный объект в генеральную совокупность не возвращается. Для конечной генеральной совокупности случайный отбор без возвращения на каждом шаге приводит к зависимости отдельных наблюдений, случайный равновозможный выбор с возвращением — к независимости наблюдений. На практике, как правило, имеют дело с бесповторными выборками. Тем не менее, когда объем генеральной совокупности N во много раз больше, чем объем выборки n (например, в сотни или тысячи раз), зависимостью наблюдений можно пренебречь, а генеральную совокупность считать бесконечной.

Таким образом, в математической статистике наблюдения x_1, x_2, ..., x_n моделируются независимыми случайными величинами с одинаковым распределением — тем же, что и у исходной случайной величины ξ. При этом допускается непрерывное распределение ξ (как удобная для расчетов модель, представляющая собой приближение к действительности). Например, говорят о *нормальной генеральной совокупности*, имея в виду нормальное распределение ξ, и т.п.

Иногда рассматриваются по-разному распределенные наблюдения (например, в случае неравноточных измерений), а иногда — и зависимые наблюдения, но все эти случаи оговариваются особо.

Бывают ситуации, когда затруднительно или невозможно описать объекты, чьи признаки наблюдаются. Речь идет об измерении какой-то величины, которая меняет свое значение случайным образом от одного наблюдения к другому. Это могут быть, например, колебания в ценах акций, курсах валют, случайные ошибки измерения и т.п. В качестве «объекта» выступает некоторое стечение обстоятельств, которые в принципе могли сложиться различным образом. В этих случаях, тем не менее, все равно применяют изложенную выше статистическую модель, хотя «выбор» здесь осуществляется без личного участия исследователя, какими-то внешними силами.

Как уже было отмечено, случайная величина ξ имеет определенную функцию распределения $F_\xi(x)$ и другие числовые характеристики, которые будем называть *теоретическими* в отличие от *выборочных*, которые определяются по наблюдениям.

Ряд наблюдений, упорядоченных по возрастанию, называют *вариационным рядом.* Его члены обычно обозначают $x_{(1)}$, $x_{(2)}$, ..., $x_{(n)}$. Наименьшее и наибольшее значения (минимум и максимум) обозначают $x_{\min}$ и $x_{\max}$ и называют их *крайними членами вариационного ряда*. Различные значения признака, появившиеся в процессе наблюдения, называют *вариантами.* Варианты, как и наблюдения, обозначаются через x_1, x_2, ..., x_n.

Когда наблюдается дискретная случайная величина, она может принимать одни и те же значения много раз. Поэтому для экономии места и времени каждое значение записывают только один раз с указанием, сколько всего раз оно появилось. Число n_i, показывающее, сколько раз появилось значение x_i в n наблюдениях, называют *частотой* данного значения, а отношение $w_i = n_i/n$ — *относительной частотой*. Число k различных значений в n наблюдениях всегда конечно, и $k \le n$. Очевидно, имеют место равенства: $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} w_i = 1$.

Результаты можно записать в таблицу.

x_1	x_2	...	x_k
w_1	w_2	...	w_k

В случае непрерывной случайной величины на практике часто применяют *группировку*. Это означает, что весь интервал наблюдаемых значений разбивается на k частичных интервалов $[c_0; c_1)$, $[c_1; c_2)$, ..., $[c_{k-1}; c_k]$ равной длины h и затем подсчитываются числа попаданий наблюдений в эти интервалы, которые принимают за частоты n_i (для некоторой новой, уже дискретной случайной величины). В качестве новых значений вариант x_i обычно берутся середины интервалов (либо в таблице указываются сами интервалы). Группировка может применяться и в случае дискретных случайных величин, если шаг, с которым меняются их значения, кажется нам слишком мелким.

Согласно формуле Стерджеса рекомендуемое число интервалов разбиения выборки $k \approx 1 + \log_2 n$, а длины частичных интервалов $h = (x_{\max} - x_{\min})/k$. Предполагается, что при этом весь интервал имеет вид $[x_{\min}, x_{\max}]$.

Набор вариант x_i (или частичных интервалов) и их относительных частот w_i называют ***статистическим рядом***. Графически статистические ряды могут быть представлены в виде полигона, гистограммы или графика накопленных частот.

Полигоном частот называют ломаную линию, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), ..., (x_k, n_k). *Полигоном относительных частот* называют ломаную, отрезки которой соединяют точки (x_1, w_1), (x_2, w_2), ..., (x_k, w_k). Полигоны обычно служат для изображения выборки в случае дискретных случайных величин (рис. 16.1).

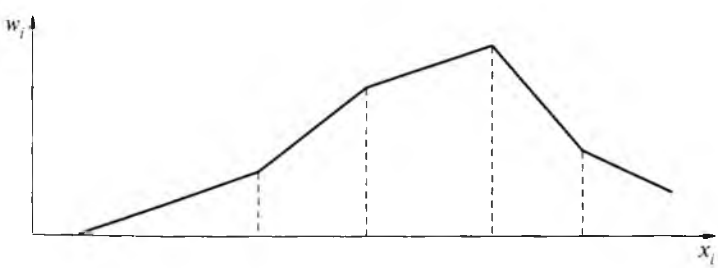

Рис. 16.1

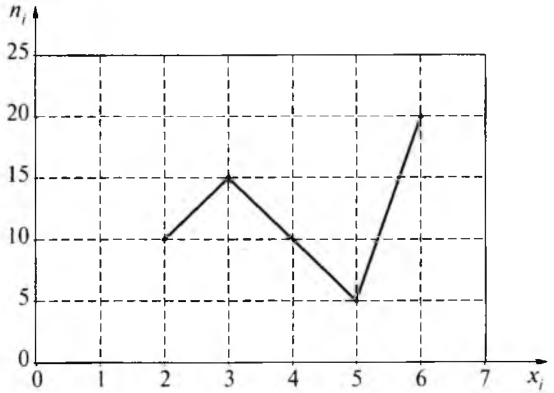

Рис. 16.2

Пример 16.1. Построить полигон частот по данному распределению выборки.

x_i	2	3	5	6
n_i	10	15	5	20

Решение. Отложим на оси абсцисс варианты x_i, а на оси ординат — соответствующие им частоты n_i, затем соединим последовательно точки (x_i, n_i) (рис. 16.2).

Гистограммой относительных частот[1] (или просто гистограммой) называется ступенчатая фигура, состоящая из прямоугольников, основанием которых служат частичные интервалы длиною h, а высоты равны w_i/h. Гистограмма обычно служит для изображения выборки в случае непрерывных случайных величин. Площадь гистограммы равна единице.

Поэтому гистограмму можно рассматривать как график *эмпирической (выборочной) плотности распределения* $p_n(x)$. Если у теоретического распределения существует конечная плотность, то эмпирическая плотность является некоторым приближением теоретической. В этом и состоит практическая польза гистограммы.

При построении гистограмм в реальных исследованиях следует понимать, что формула Стерджеса (как и любая другая) для числа интервалов разбиения k дает лишь рекомендацию, а не строгое правило. Поскольку гистограммы, как правило, строятся не вручную, а на компьютере, исследователь легко может варьировать параметры

[1] На практике гистограммами также называют ступенчатые фигуры с высотами w_i.

гистограммы (нижнюю и верхнюю границы интервала, число частичных интервалов) и в конечном счете выбрать тот вариант, при котором, по его мнению, график выглядит лучше всего.

Графиком накопленных частот называется фигура, строящаяся аналогично гистограмме, с той разницей, что для расчета высот прямоугольников берутся не простые, а *накопленные относительные частоты*, т.е. величины $w_i^C = \sum_{j=1}^{i} w_j$. Эти величины не убывают, и таким образом, график накопленных частот имеет вид ступенчатой «лестницы» (от 0 до 1).

График накопленных частот и эмпирическая функция распределения на практике используются для приближения теоретической функции распределения.

Пример 16.2. Анализируется выборка из 100 малых предприятий региона. Цель обследования — измерение коэффициента соотношения заемных и собственных средств (x_i) на каждом i-м предприятии. Результаты представлены в табл. 16.1 [2].

Таблица 16.1

Коэффициенты соотношений заемных и собственных средств предприятий

5,56	5,45	5,48	5,45	5,39	5,37	5,46	5,59	5,61	5,31
5,46	5,61	5,11	5,41	5,31	5,57	5,33	5,11	5,54	5,43
5,34	5,53	5,46	5,41	5,48	5,39	5,11	5,42	5,48	5,49
5,36	5,40	5,45	5,49	5,68	5,51	5,50	5,68	5,21	5,38
5,58	5,47	5,46	5,19	5,60	5,63	5,48	5,27	5,22	5,37
5,33	5,49	5,50	5,54	5,40	5.58	5,42	5,29	5,05	5,79
5,79	5,65	5,70	5,71	5,85	5,44	5,47	5,48	5,47	5,55
5,67	5,71	5,73	5,05	5,35	5,72	5,49	5,61	5,57	5,69
5,54	5,39	5,32	5,21	5,73	5,59	5,38	5,25	5,26	5,81
5,27	5,64	5,20	5,23	5,33	5,37	5,24	5,55	5,60	5,51

Построить гистограмму и график накопленных частот.

Решение. Построим группированный ряд наблюдений:

1) определим в выборке $x_{min} = 5,05$ и $x_{max} = 5,85$;

2) разобьем весь диапазон $[x_{min}; x_{max}]$ на k равных интервалов: $k \approx 1 + \log_2 100 = 7,62$; $k = 8$, отсюда длина интервала

$$h = \frac{x_{max} - x_{min}}{k} = \frac{5,85 - 5,05}{8} = 0,1.$$

На рис. 16.3 и 16.4, построенных по данным табл. 16.2, представлены гистограмма и график накопленных частот. Кривые соответствуют плотности и функции нормального распределения, «подобранного» к данным.

Таблица 16.2

Сгруппированный ряд наблюдений

Номер интервала	Интервал	Середина интервала x_i	w_i	w_i^C	$p_n(x)$
1	5,05–5,15	5,1	0,05	0,05	0,5
2	5,15–5,25	5,2	0,08	0,13	0,8
3	5,25–5,35	5,3	0,12	0,25	1,2
4	5,35–5,45	5,4	0,20	0,45	2,0
5	5,45–5,55	5,5	0,26	0,71	2,6
6	5,55–5,65	5,6	0,15	0,86	1,5
7	5,65–5,75	5,7	0,10	0,96	1,0
8	5,75–5,85	5,8	0,04	1,00	0,4

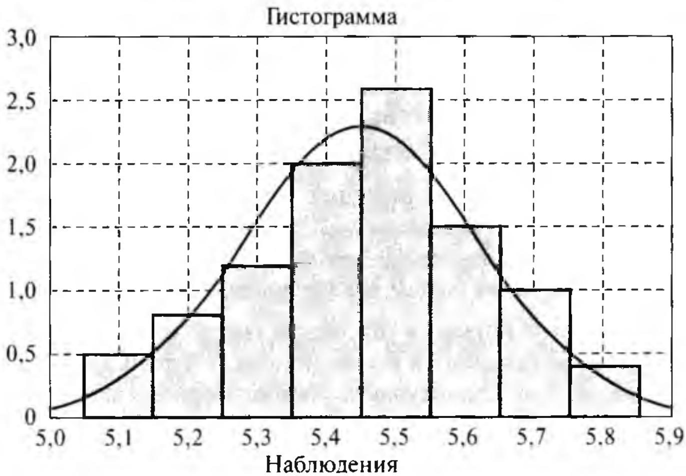

Рис. 16.3

16.2. ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Эмпирической функцией распределения (или *функцией распределения выборки*) называют функцию $F_n(x)$, определяющую для каждого числа x относительную частоту события $\xi < x$, т.е. $F_n(x) = n_x/n$, где n_x — число наблюдений, меньших x; n — объем выборки.

Из определения следует, что значения эмпирической функции распределения при каждом x являются случайными величинами.

Эмпирическая функция распределения обладает всеми свойствами обычной функции распределения, а также некоторыми специфическими.

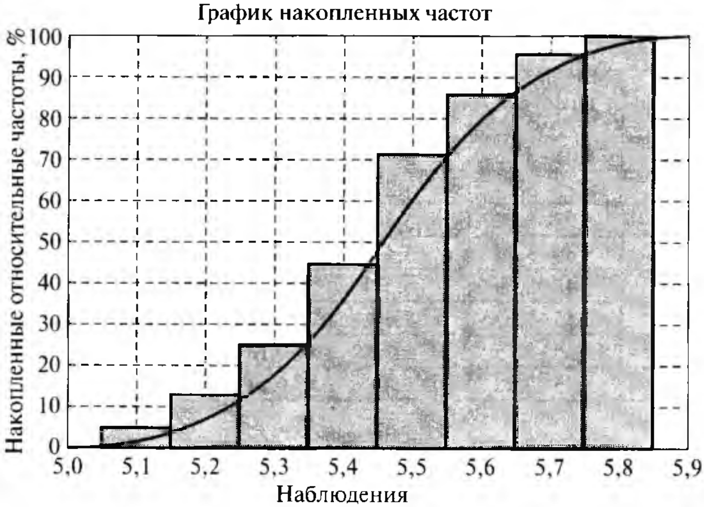

Рис. 16.4

1. $0 \le F_n(x) \le 1$.
2. $F_n(x)$ — неубывающая функция.
3. $F_n(x)$ непрерывна слева.
4. $F_n(x) = 0$ при $x \le x_{\min}$ и $F_n(x) = 1$ при $x > x_{\max}$.
5. $P(\lim\limits_{n\to\infty} \sup|F_n(x) - F(x)| = 0) = 1$ (теорема Гливенко—Кантелли).

Смысл теоремы Гливенко–Кантелли заключается в том, что при увеличении объема выборки n у эмпирической функции распределения исчезают свойства случайности и она приближается к теоретической функции распределения $F(x)$.

График эмпирической функции распределения есть неубывающая ступенчатая кривая со скачками, равными $1/n$ в точках вариационного ряда (если значения не совпадают). Если n_i точек вариационного ряда совпадают и равны x_i, то скачок в точке x_i равен n_i/n.

Пример 16.3. По таблице наблюдений случайной величины найти эмпирическую функцию распределения и построить ее график.

ξ_i	2	3	5
n_i	75	20	5
w_i	0,75	0,2	0,05

Решение. Эмпирическая функция распределения имеет вид

$$F_n(x) = \begin{cases} 0, & x \le 2, \\ 0{,}75, & 2 < x \le 3, \\ 0{,}95, & 3 < x \le 5, \\ 1, & x > 5. \end{cases}$$

Ее график представлен на рис. 16.5.

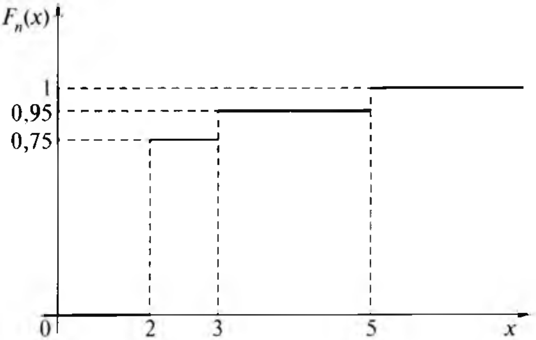

Рис. 16.5

Пример 16.4. Пусть $x_1, x_2, ..., x_n$ — выборка независимых наблюдений из непрерывной генеральной совокупности с функцией распределения $F(x)$ и плотностью распределения $f(x)$. Найти функции распределения и плотности распределения крайних членов вариационного ряда: $x_{\min}$ и $x_{\max}$.

Решение. Из определения функции распределения следует, что

$$F_{x_{\max}}(y) = P(x_{\max} < y) = P(x_1 < y, x_2 < y, ..., x_n < y) = P^n(x_i < y) = F_\xi^n(y).$$

Тогда

$$f_{x_{\max}}(y) = [F_{x_{\max}}(y)]' = [F_\xi^n(y)]' = nF_\xi^{n-1}(y) \cdot F_\xi'(y) = nF_\xi^{n-1}(y) \cdot f_\xi(y).$$

Аналогично

$$F_{x_{\min}}(y) = P(x_{\min} < y) = 1 - P(x_{\min} \ge y) = 1 - P(x_1 \ge y, x_2 \ge y, ..., x_n \ge y) =$$
$$= 1 - [1 - F_\xi(y)]^n.$$

Отсюда получаем функцию плотности

$$f_{x_{\min}}(y) = -n[1 - F_\xi(y)]^{n-1} \cdot (-F_\xi'(y)) = n[1 - F_\xi(y)]^{n-1} \cdot f_\xi(y).$$

Для непрерывной случайной величины лучшую оценку плотности распределения $f(x)$ дает полигон относительных частот.

Глава 17
ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

17.1. ВЫБОРОЧНЫЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

По данным выборки, используя формулы для дискретных случайных величин, можно вычислить приближенные значения числовых характеристик случайной величины: математическое ожидание, дисперсию, начальные и центральные моменты, которые называются *выборочными характеристиками*.

Пусть $x_1, x_2, ..., x_n$ — выборка из генеральной совокупности. Тогда:

1) *выборочное среднее* $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;

2) *выборочная дисперсия* $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$; $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$;

3) *выборочные начальные моменты k-го порядка* $\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$;

4) *выборочные центральные моменты k-го порядка* $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$.

На практике используются также *выборочная мода* x_{mod}, равная значению варианты с наибольшей частотой, и *выборочная медиана* x_{med}, равная значению варианты, стоящей в середине вариационного ряда (либо полусумме двух вариант, стоящих в середине ряда, при четном числе наблюдений). *Коэффициентом вариации* вариационного ряда называется процентное отношение выборочного среднего квадратического отклонения к средней арифметической: $\hat{V} = \frac{\hat{\sigma}}{\bar{x}}100\%$. Выборочным *коэффициентом асимметрии* называется число $\hat{\beta} = \frac{\hat{\mu}_3}{s^3}$, и выборочным *эксцессом* вариационного ряда называется число $\hat{\nu} = \frac{\hat{\mu}_4}{s^4} - 3$.

Все эти характеристики не совпадают с соответствующими характеристиками генеральной совокупности, поскольку являются случайными величинами. Распределение этих случайных величин однозначно определяется распределением генеральной совокупности.

Итак, все выборочные характеристики являются функциями наблюдений. Любая функция $\hat{\theta}_n = U(x_1, x_2, ..., x_n)$ выборочных наблюдений $x_1, x_2, ..., x_n$ называется *статистикой*. Статистика $\hat{\theta}_n$, используемая в качестве приближенного значения неизвестного параметра θ, называется *статистической оценкой* этого параметра.

Для нахождения оценок параметров распределения из исследуемой генеральной совокупности производится выборка $x_1, x_2, ..., x_n$ объема n. Чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, желательно, чтобы оценки были несмещенными, состоятельными и эффективными.

Оценка $\hat{\theta}_n$ называется *состоятельной*, если при неограниченном увеличении объема выборки она сходится по вероятности к оцениваемому параметру: $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$ для любого $\varepsilon > 0$.

Свойство состоятельности оценки — асимптотическое свойство. Оно может проявляться при столь больших объемах выборки, которых на практике не имеется.

Оценка $\hat{\theta}_n$ называется *несмещенной* (оценкой без систематической ошибки), если ее математическое ожидание при любом n равно оцениваемому параметру: $M\hat{\theta}_n = \theta$.

Несмещенность оценки характеризует ее «доасимптотические» свойства, т.е. является показателем ее «хороших» свойств при любом конечном объеме выборки.

Дисперсия любой несмещенной оценки параметра θ удовлетворяет неравенству Рао—Фреше—Крамера:

$$D(\hat{\theta}_n) \geq -\frac{1}{nM\left(\dfrac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}\right)},$$

где n — объем выборки, а $f(x, \theta)$ — плотность распределения случайной величины.

Оценка называется *эффективной*, если она имеет минимальную дисперсию в определенном классе оценок. Для эффективной оценки в неравенстве Рао—Фреше—Крамера достигается равенство. В дальнейшем будут рассмотрены различные виды эффективности оценок.

Пример 17.1. Найти выборочное среднее по выборке объема $n = 20$.

x_i	2560	2600	2620	2650	2700
N_i	2	3	10	4	1

Решение. Для упрощения расчетов перейдем к условным вариантам $u_i = x_i - 2620$.

u_i	−60	−20	0	30	80
n_i	2	3	10	4	1

Тогда $\bar{u} = [2 \cdot (-60) + 3 \cdot (-20) + 10 \cdot 0 + 1 \cdot 80]/20 = 1$ и $\bar{x} = 2620 + \bar{u} = 2621$.

Замечание. В качестве числа, которое вычитается при переходе к условным вариантам (*условный нуль*), обычно выбирается или варианта, стоящая в середине ряда, или та, для которой частота максимальна (выборочная мода). В данном примере они совпадают.

Пример 17.2. Найти неисправленную выборочную дисперсию по выборке объема $n = 50$.

x_i	18,4	18,9	19,3	19,6
n_i	5	10	20	15

Решение. Перейдем к условным вариантам $u_i = 10(x_i - 19{,}3)$. Тогда $Du_i = D(10x_i - 193) = 100Dx_i$ и

u_i	−9	−4	0	3
n_i	5	10	20	15

Найдем выборочную дисперсию для новой варианты u_i:

$$\hat{\sigma}^2(u) = \frac{1}{n}\left(\sum_{i=1}^{4} n_i u_i^2\right) - \left(\sum_{i=1}^{4} \frac{n_i u_i}{n}\right)^2 =$$

$$= \frac{1}{50}(5 \cdot (-9)^2 + 10 \cdot (-4)^2 + 20 \cdot 0^2 + 15 \cdot 3^2) - \left[\frac{1}{50}(-45 - 40 + 0 + 45)\right]^2 = 13{,}36.$$

Тогда, переходя к первоначальной варианте x_i: $\hat{\sigma}^2(x) = \hat{\sigma}^2(u)/100 = 13{,}36/100 = 0{,}1336$.

17.2. СТАТИСТИЧЕСКАЯ УСТОЙЧИВОСТЬ И СВОЙСТВА ОСНОВНЫХ ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК

Покажем, что числовые характеристики выборки сходятся по вероятности к числовым характеристикам генеральной совокупности и, следовательно, мало отличаются от истинных значений числовых характеристик генеральной совокупности (для больших выборок).

Пусть генеральная совокупность ξ имеет характеристики $M\xi = a$, $D\xi = \sigma^2$, $M\xi^k = \alpha_k$, $M(\xi - M\xi)^k = \mu_k$, а соответствующие выборочные характеристики будут $\bar{x}$, $\hat{\sigma}^2$, $\hat{\alpha}_k$, $\hat{\mu}_k$. Докажем, что выборочное среднее и выборочные начальные моменты $\hat{\alpha}_k$ сходятся по вероятности к теоретическим начальным моментам α_k.

Для доказательства воспользуемся неравенством Чебышева $P(|\eta - M\eta| \geq \varepsilon) \leq \dfrac{D\eta}{\varepsilon^2}$, справедливым для любой случайной величины η и любого фиксированного $\varepsilon > 0$.

Вычислим математическое ожидание и дисперсию выборочного среднего:

$$M\bar{x} = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Mx_i = \frac{na}{n} = a, \quad \text{так как } Mx_i = M\xi.$$

Поскольку x_i — независимые случайные величины, то дисперсия суммы этих случайных величин равна сумме их дисперсий $Dx_i = D\xi = \sigma^2$:

$$D\bar{x} = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0 \quad \text{при } n \to \infty.$$

Так как вероятность неотрицательна, то в силу неравенства Чебышева для любого фиксированного $\varepsilon > 0$

$$0 \leq P(|\bar{x} - a| \geq \varepsilon) \leq \frac{D\bar{x}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{при } n \to \infty.$$

Итак, среднее арифметическое выборки $\bar{x}$ сходится по вероятности к математическому ожиданию случайной величины, т.е. является несмещенной и состоятельной оценкой математического ожидания случайной величины, распределенной по любому закону. Аналогично можно доказать состоятельность оценок начальных моментов любого порядка.

Утверждение о сходимости выборочного среднего к среднему значению генеральной совокупности справедливо при более слабых предположениях.

Теорема 17.1 (теорема Хинчина). Если $\xi_1, \xi_2, ..., \xi_n$ — независимые, одинаково распределенные случайные величины и существует $M\xi_i = a$, тогда $\bar{\xi} = \dfrac{1}{n}\sum_{i=1}^{n} \xi_i$ сходится по вероятности к a.

С л е д с т в и я

1. Если $x_1, x_2, ..., x_n$ — выборка из некоторой генеральной совокупности и существует $Mx_i = a$, $i = 1, 2, ..., n$, то $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ сходится по вероятности к a: $\bar{x} \xrightarrow{p} a$.

2. Если $\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ и существует $Mx_i^k = a_k$, то $\hat{a}_k \xrightarrow{p} a_k$.

Покажем, что выборочная дисперсия $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ является смещенной оценкой дисперсии $D\xi = \sigma^2$ генеральной совокупности, и найдем несмещенную оценку. Предварительно докажем следующую теорему.

Теорема 17.2. Для любого b справедливо равенство $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - (\bar{x} - b)^2$.

Доказательство вытекает из следующей цепочки равенств:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}[(x_i - b) - (\bar{x} - b)]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[(x_i - b)^2 - 2(x_i - b)(\bar{x} - b) + (\bar{x} - b)^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - 2(\bar{x} - b)\frac{1}{n}\sum_{i=1}^{n}(x_i - b) + \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - b)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - \frac{\bar{x} - b}{n}\left[2\sum_{i=1}^{n} x_i - 2nb - n\bar{x} + nb\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - \frac{\bar{x} - b}{n}\left[2\sum_{i=1}^{n} x_i - nb - \sum_{i=1}^{n} x_i\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - (\bar{x} - b)\left[\frac{1}{n}\sum_{i=1}^{n} x_i - b\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - (\bar{x} - b)^2.$$

С л е д с т в и е. Если $M\xi = a$ есть средняя генеральной совокупности, то $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - (\bar{x} - a)^2$.

Вычислим математическое ожидание выборочной дисперсии:

$$M\hat{\sigma}^2 = M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2\right] - M(\bar{x} - a)^2 = \frac{1}{n}\sum_{i=1}^{n} M(x_i - a)^2 - D\bar{x} =$$

$$= \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Отсюда следует, что $\hat{\sigma}^2$ является смещенной оценкой дисперсии. Введем случайную величину $s^2 = \frac{n}{n-1}\hat{\sigma}^2$.

Поскольку

$$Ms^2 = M\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \frac{n}{n-1}M(\hat{\sigma}^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2,$$

то несмещенной оценкой дисперсии будет статистика $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Оценку $\hat{\sigma}^2$ (при получении которой сумма делится на n) называют смещенной или неисправленной выборочной дисперсией, а оценку s^2 (при получении которой сумма делится на $n - 1$) называют несмещенной или исправленной выборочной дисперсией. Выборочным средним квадратическим отклонением (исправленным или неисправленным) называют положительный корень из соответствующей выборочной дисперсии. Эта оценка теоретического среднего квадратического отклонения $\sigma = \sqrt{D\xi}$ в обоих случаях является смещенной.

Закон больших чисел неприменим к центрированным величинам, так как после центрирования они становятся зависимыми. Для доказательства сходимости по вероятности выборочных центральных моментов к соответствующим теоретическим моментам случайной величины используем следующую теорему.

Теорема 17.3 (теорема Слуцкого). Пусть функция $f(x, y)$ непрерывна в точке (a, b) и случайные последовательности x_n и y_n сходятся по вероятности соответственно $x_n \xrightarrow{p} a$, $y_n \xrightarrow{p} b$. Тогда функция $f(x_n, y_n)$ сходится по вероятности к $f(a, b)$.

Предположим, что существует дисперсия генеральной совокупности $D\xi = \sigma^2$. Докажем, что выборочная дисперсия сходится по вероятности к дисперсии генеральной совокупности: $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. Выборочная дисперсия равна

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2.$$

Рассмотрим функцию $f(x, y) = x - y^2$. Эта функция непрерывна по совокупности переменных всюду. Положим $y_n = \bar{x}$, и $x_n = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \hat{a}_2$. В силу следствия 1 из теоремы Хинчина имеем $y_n \xrightarrow{p} a$. Вычислим математическое ожидание:

$$M\hat{a}_2 = M\left(\frac{1}{n}\sum_{i=1}^{n}[(x_i - a) + a]^2\right) =$$

$$= M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 + 2a\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - a)\right) + a^2\right] = \sigma^2 + a^2.$$

Из следствия 2 теоремы Хинчина следует, что $x_n = \hat{a}_2 \xrightarrow{p} a_2 = \sigma^2 + a^2$, тогда в силу теоремы Слуцкого

$$\hat{\sigma}^2 = f(x_n, y_n) \xrightarrow{p} f(\sigma^2 + a^2, a) = \sigma^2 + a^2 - a^2 = \sigma^2.$$

Итак, доказали, что выборочная дисперсия сходится по вероятности к дисперсии генеральной совокупности: $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Этот общий подход позволяет доказывать сходимость по вероятности всех выборочных центральных моментов $\hat{\mu}_k \xrightarrow{p} \mu_k$, если существуют теоретические центральные моменты генеральной совокупности $\mu_k = M(\xi - a)^k$. Действительно, из сходимости по вероятности случайных величин $\xi_i(n)$ к некоторым постоянным числам a_i следует в силу теоремы Слуцкого сходимость по вероятности любой рациональной функции $\varphi(\xi_1(n), \xi_2(n), ..., \xi_k(n))$ к ее значению в точке $(a_1, a_2, ..., a_k)$, т.е. если $\xi_i(n) \xrightarrow{p} a_i$, то $\varphi(\xi_1(n), \xi_2(n), ..., \xi_k(n)) \xrightarrow{p} \varphi(a_1, a_2, ..., a_k)$ при $n \to \infty$.

Так как все выборочные центральные моменты (а также их некоторые модификации — асимметрия, эксцесс и т.п.) являются рациональными функциями от наблюдений, то они сходятся по вероятности к соответствующим теоретическим значениям, т.е. являются состоятельными оценками центральных моментов.

Справедлива следующая теорема.

Теорема 17.4. Если $\hat{\theta}_n$ — несмещенная оценка параметра θ и $D(\hat{\theta}_n) \to 0$ при $n \to \infty$, то оценка состоятельная.

Доказательство. Пусть $\hat{\theta}_n$ — несмещенная оценка параметра θ, т.е. $M(\hat{\theta}_n) = \theta$. Тогда для любого $\varepsilon > 0$ из неравенства Чебышева получаем $P(|\hat{\theta}_n - \theta| < \varepsilon) \geq 1 - \frac{D\hat{\theta}_n}{\varepsilon^2}$. По условию $D\hat{\theta}_n \to 0$ при $n \to \infty$, и следовательно, при каждом фиксированном $\varepsilon > 0$ разность $1 - \frac{D\hat{\theta}}{\varepsilon^2} \to 1$ при $n \to \infty$, а так как вероятность не может быть больше единицы, то $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$. Теорема доказана.

17.3. АСИМПТОТИЧЕСКИ НОРМАЛЬНЫЙ ХАРАКТЕР ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК. ЭФФЕКТИВНОСТЬ ОЦЕНОК

Свойства выборки объема n зависят от распределения генеральной совокупности, но по мере увеличения n ($n \to \infty$) выборочные характеристики начинают вести себя одинаковым образом независимо от специфики генеральной совокупности. Поэтому характер поведения выборочных характеристик следует рассматривать в двух вариантах: при фиксированном n (ограниченном объеме выборки) и при $n \to \infty$ (асимптотические свойства выборки). При фиксированном n свойства выборок будут различны для разных типов генеральной совокупности (нормальной, экспоненциальной, равномерной, пуассоновской и т.д.). В условиях асимптотики ($n \to \infty$) общий характер поведения числовых характеристик практически не зависит от типа анализируемой генеральной совокупности.

Последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ называется *асимптотически нормальной*, если существуют числовые последовательности $A_1, A_2, \ldots$ и $B_1, B_2, \ldots$ ($B_i > 0$ для всех i) такие, что

$$P\left(\frac{\xi_i - A_i}{\sqrt{B_i}} < x\right) \to \Phi(x) \quad \text{при } n \to \infty,$$

где $\Phi(x)$ — функция стандартного нормального распределения.

Числа A_i и B_i называются параметрами асимптотически нормально распределенной случайной величины ξ_i. То, что последовательность $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ асимптотически нормальная, записывается в виде $\xi_i \xrightarrow{\in} N(A_i, B_i)$. Легко видеть, что $\xi_i \xrightarrow{\in} N(A_i, B_i)$ тогда и только тогда, когда $\frac{\xi_i - A_i}{\sqrt{B_i}} \xrightarrow{\in} N(0, 1)$.

Используя введенный термин, **центральную предельную теорему** можно сформулировать следующим образом.

Пусть $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ — независимые одинаково распределенные случайные величины с конечными моментами первого и второго порядков: $M\eta_i = a$; $D\eta_i = \sigma^2$. Тогда если

$$S_n = \eta_1 + \eta_2 + \ldots + \eta_n, \quad \text{то } S_n \xrightarrow{\in} N(na, n\sigma^2),\ n \to \infty.$$

Теорема 17.5. Если распределение генеральной совокупности имеет конечные математическое ожидание a и дисперсию σ^2, то при $n \to \infty$ основные выборочные характеристики являются асимптотически нормальными:

1. $\bar{x}_n \xrightarrow{\in} N\left(a, \dfrac{\sigma^2}{n}\right)$;
2. $\hat{\sigma}^2 \xrightarrow{\in} N\left(\dfrac{(n-1)\sigma^2}{n}, \dfrac{2(n-1)\sigma^4}{n^2}\right)$;
3. $s^2 \xrightarrow{\in} N\left(\sigma^2, \dfrac{\mu_4 - \sigma^4}{n}\right)$;
4. $F_n(x) \xrightarrow{\in} N\left(F(x); \dfrac{F(x)(1 - F(x))}{n}\right)$.

Другими словами, при больших объемах выборки все основные выборочные характеристики можно считать практически нормально распределенными.

Для одних и тех же параметров распределения существует бесконечно много различных несмещенных и состоятельных оценок. Поэтому важной задачей является сравнение их между собой и поиск наилучшей среди них. Естественным критерием такого поиска является дисперсия как мера разброса значений случайной величины вокруг его среднего значения. Наилучшей оценкой является та, чья дисперсия минимальна. Однако для смещенной оценки $\hat{\theta}_n$ дисперсия служит мерой близости не к оцениваемому параметру θ, а к математическому ожиданию этой оценки $M\hat{\theta}_n$. Поэтому естественно искать оценки с наименьшей дисперсией только среди несмещенных оценок.

Рассмотрим класс несмещенных оценок $\hat{\theta}_n$ скалярного параметра θ, от которого зависит функция плотности вероятностей $f(x, \theta)$ исследуемой генеральной совокупности.

Информацией Фишера о неизвестном параметре θ, содержащейся в одном из независимых наблюдений случайной величины ξ, называется величина

$$I(\theta) = M\left(\frac{\partial \ln p(\xi, \theta)}{\partial \theta}\right)^2,$$

где в качестве $p(x, \theta)$ берется либо плотность $f(x, \theta)$ в точке x (для непрерывных случайных величин), либо вероятность $P(x, \theta)$ принять случайной величиной значение x (для дискретных случайных величин). Говорят, что величина $I(\theta)$ определяет количество информации Фишера.

Теорема 17.6 (теорема Рао—Фреше—Крамера). Пусть функция плотности $f(x, \theta)$ удовлетворяет следующим условиям регулярности:

1) область $G_n = \{x: f(x, \theta) > 0\}$ всех возможных значений случайной величины, для которых $f(x, \theta) \neq 0$, не зависит от параметра θ;

2) в формулах $\int\limits_{-\infty}^{+\infty} xf(x,\theta)dx \equiv M\xi$ и $\int\limits_{-\infty}^{+\infty} f(x,\theta)dx \equiv 1$ допустимо дифференцирование по параметру θ под знаком интеграла;

3) информация Фишера $I(\theta)$ конечна и положительна.

Тогда для произвольной несмещенной оценки $\hat{\theta}_n$, построенной по выборке объема n, выполняется неравенство (Рао—Фреше—Крамера)

$$D\hat{\theta}_n \geq \frac{1}{nI(\theta)}.$$

Теорема верна и в дискретном случае, если в условии 2) заменить интегралы на суммы (по всем возможным значениям случайной величины).

Заметим, что информацию Фишера можно также представить в виде

$$I(\theta) = D\left(\frac{\partial \ln p(\xi,\theta)}{\partial\theta}\right) \quad \text{и} \quad I(\theta) = -M\left(\frac{\partial^2 \ln p(\xi,\theta)}{\partial^2\theta}\right).$$

Пример 17.3. Подсчитать количество информации о математическом ожидании, содержащееся в одном наблюдении нормально распределенной случайной величины.

Решение. Пусть случайная величина x подчинена нормальному закону $N(a, \sigma^2)$ с плотностью $f(x, a, \sigma^2)$, в котором среднее значение $a = \theta$ — неизвестный параметр, а дисперсия известна. Поскольку плотность нормального закона распределения равна

$$f(x,a,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}},$$

то

$$\ln f(x,a,\sigma^2) = \frac{-(x-a)^2}{2\sigma^2} - \ln\sigma\sqrt{2\pi}.$$

Отсюда получим $\dfrac{\partial \ln f(x,a,\sigma^2)}{\partial a} = \dfrac{x-a}{\sigma^2}.$

Поэтому количество информации в одном наблюдении равно

$$I(a) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, a, \sigma^2)}{\partial a} \right)^2 f(x, a, \sigma^2) dx =$$

$$= \int_{-\infty}^{\infty} \frac{(x-a)^2}{\sigma^4} f(x, a, \sigma^2) dx = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2},$$

т.е. чем больше дисперсия σ^2 случайной величины, тем меньше информации заключено о величине ее среднего значения в одном наблюдении.

Обозначим правую часть неравенства Рао—Фреше—Крамера через $\Delta_n = 1/(nI(\theta))$, которая является нижней гранью всех возможных дисперсий оценок.

Эффективностью несмещенной оценки $\hat{\theta}_n$ называется отношение минимально возможного значения дисперсии оценки в классе всех несмещенных оценок параметра θ к дисперсии рассматриваемой оценки:

$$e(\hat{\theta}_n) = \frac{\Delta_n}{D\hat{\theta}_n} = \frac{1}{nI(\theta)D\hat{\theta}_n}.$$

Из определения следует, что эффективность любой несмещенной оценки удовлетворяет неравенству $0 \le e(\hat{\theta}_n) \le 1$, и чем ближе она к единице, тем лучше оценка.

Несмещенная оценка $\hat{\theta}_n$ называется *эффективной*, если $e(\hat{\theta}_n) = 1$.

Асимптотической эффективностью оценки называется предел $e_0(\hat{\theta}) = \lim\limits_{n\to\infty} e(\hat{\theta}_n)$, если он существует.

Оценку $\hat{\theta}_n^*$ называют *асимптотически эффективной*, если $e_0(\hat{\theta})^* = 1$.

Пример 17.4. Пусть выборка $x_1, x_2, ..., x_n$ произведена из генеральной совокупности с равномерным на интервале $(0; \theta)$ распределением. Проверить на эффективность оценку $\hat{\theta}_n = \frac{n+1}{n} x_{max}$ для неизвестного параметра θ.

Решение. Функция распределения $F_{max}(x)$ максимума x_{max} задается формулой $F_{max}(x) = \left(\frac{x}{\theta}\right)^n$ на отрезке $0 \le x \le \theta$ (см. §1.3). Следовательно, функция плотности величины x_{max} имеет вид $f_{max}(x) = \frac{nx^{n-1}}{\theta^n}$. Отсюда получаем

$$M\hat{\theta}_n = \frac{n+1}{n} \int_0^{\theta} nx \frac{x^{n-1}}{\theta^n} dx = \theta.$$

Значит, оценка $\hat{\theta}_n$ несмещенная. Найдем дисперсию этой оценки:

$$M\hat{\theta}_n^2 = \left(\frac{n+1}{n}\right)^2 \int_0^\theta nx^2 \frac{x^{n-1}}{\theta^n} dx = \frac{(n+1)^2}{n(n+2)}\theta^2,$$

$$D\hat{\theta}_n^2 = M\hat{\theta}_n^2 - (M\hat{\theta}_n)^2 = \frac{\theta^2}{n(n+2)}.$$

Видим, что дисперсия оценки $\hat{\theta}_n$ при $n \to \infty$ убывает как вторая степень бесконечно малой величины $\left(\frac{1}{n}\right)^2$. Такая оценка оказалась лучше эффективной, поскольку дисперсия эффективной оценки имеет порядок убывания только $\frac{1}{n}$. Разгадка парадокса в том, что для данного семейства не выполнены условия теоремы Рао—Фреше—Крамера. Подобные оценки называют *сверхэффективными*.

17.4. ОЦЕНКА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ПО НЕРАВНОТОЧНЫМ НАБЛЮДЕНИЯМ

Для доказательства эффективности среднего арифметического $\bar{x}_n$ как оценки математического ожидания (в классе линейных несмещенных оценок) в общем случае рассмотрим оценку математического ожидания по неравноточным наблюдениям.

Пусть результат i-го выборочного испытания описывается случайной величиной ξ_i с параметрами $M\xi_i = a$, $D\xi_i = \sigma_i^2$ и испытания независимы друг от друга. Наилучшую в статистическом смысле оценку математического ожидания будем искать в классе линейных оценок в виде $\hat{a}_n = \sum_{i=1}^n C_i\xi_i = \hat{a}_n(C)$. Из требования несмещенности оценки из равенства $M\hat{a}_n = M\sum_{i=1}^n C_i\xi_i = \sum_{i=1}^n C_i(M\xi_i) = a\sum_{i=1}^n C_i$ получаем, что оценка будет несмещенной, если $\sum_{i=1}^n C_i = 1$.

По определению оценка будет эффективной, если она имеет минимальную дисперсию. Определим коэффициенты C_i так, чтобы дисперсия была минимальна, при условии, что $\sum_{i=1}^n C_i = 1$. Вычислим дисперсию $D\hat{a}_n$: поскольку случайные величины ξ_i независимы, дисперсия суммы в этом случае равна сумме дисперсий:

$$D\hat{a}_n = D\left[\sum_{i=1}^n C_i\xi_i\right] = \sum_{i=1}^n C_i^2(D\xi_i) = \sum_{i=1}^n C_i^2\sigma_i^2.$$

Задача свелась к нахождению таких коэффициентов C_i, при которых функция $f(C) = \sum_{i=1}^{n} C_i^2 \sigma_i^2$ имеет минимум при условии $g(C) = \sum_{i=1}^{n} C_i - 1 = 0$. Это задача на условный экстремум. Функция Лагранжа имеет вид $L(C) = f(C) - \lambda(C)$. Для определения коэффициентов C_i приравниваем нулю производные:

$$\begin{cases} \dfrac{\partial L}{\partial C_i} = 2C_i\sigma_i^2 - \lambda = 0, \\ \dfrac{\partial L}{\partial \lambda} = -g(C) = -\left(\sum_{i=1}^{n} C_i - 1\right) = 0, \end{cases} \qquad \begin{cases} C_i = \dfrac{\lambda}{2\sigma_i^2}, \quad i = 1, 2, \ldots, n, \\ \sum_{i=1}^{n} C_i = 1, \end{cases}$$

$$\begin{cases} C_i = \dfrac{\lambda}{2\sigma_i^2}, \\ \dfrac{\lambda}{2}\sum_{i=1}^{n} \dfrac{1}{\sigma_i^2} = 1. \end{cases}$$

Отсюда $C_i = \dfrac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$ для любого $i = 1, 2, \ldots, n$.

Итак, эффективная линейная оценка математического ожидания по неравноточным наблюдениям получается тогда, когда наблюдаемые значения входят в оценку обратно пропорционально своим дисперсиям: $\hat{a}_n = \dfrac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\xi_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$. Это означает, что менее точные наблюдения входят с меньшим весом, а более точные — с большим.

Если дисперсии всех наблюдений равны $\sigma_i = \sigma$ для любого $i = 1, 2, \ldots, n$ и если $\xi_i = x_i$ есть элементы выборки из генеральной совокупности, то $C_i = 1/n$ для любого $i = 1, 2, \ldots, n$ и $\hat{a}_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, т.е. среднее арифметическое выборки является эффективной оценкой математического ожидания в классе линейных функций при любом распределении.

Глава 18
МЕТОДЫ ПОСТРОЕНИЯ ОЦЕНОК

18.1. МЕТОД МОМЕНТОВ

Метод моментов, предложенный английским статистиком Карлом Пирсоном в 1894 г., заключается в приравнивании определенного числа выборочных моментов к соответствующим теоретическим, которые являются функциями неизвестных параметров θ_1, θ_2, ..., θ_k. Рассматривая количество моментов, равное числу k неизвестных параметров, подлежащих определению, и решая полученные уравнения относительно этих параметров, получаем искомые оценки. Иначе говоря, оценки параметров θ_1, θ_2, ..., θ_k являются решениями систем уравнений

$$\alpha_i(\theta_1, \theta_2, \ldots, \theta_k) = \hat{\alpha}_i \quad \text{или} \quad \mu_i(\theta_1, \theta_2, \ldots, \theta_k) = \hat{\mu}_i$$

для некоторых $i = i_1, i_2, \ldots, i_k$.

Метод моментов содержит неопределенность, поскольку можно получить уравнения для неизвестных параметров, используя как начальные, так и центральные моменты, а также некоторые их модификации типа асимметрии или эксцесса.

Пример 18.1. Функция $p(x) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & x > 0, \\ 0, & x \le 0 \end{cases}$ задает плотность распределения Рэлея.

Найдем оценку параметра θ, приравнивая начальные выборочные и теоретические моменты. Первый начальный момент имеет вид $\alpha_1 = \sqrt{\pi\theta/2}$.

Приравнивая, получаем первую оценку параметра:

$$\sqrt{\pi\hat{\theta}_n/2} = \bar{x},$$

откуда $\hat{\theta}_n = \dfrac{2\bar{x}^2}{\pi}$.

Приравнивая вторые начальные моменты, можем получить другую оценку:

$$\hat{\theta}_n = \frac{1}{2n}\sum_{i=1}^{n} x_i^2,$$

а из уравнения, которое возникает при использовании второго центрального момента (дисперсии), — третью оценку: $\hat{\theta}_n = \dfrac{2s^2}{4-\pi}$.

Рекомендуется для нахождения оценки одного параметра брать первый момент, для двух — первые два момента и т.п. Однако этот подход годится не всегда. Он не проходит, например, если некоторые моменты равны нулю или не зависят от нужных параметров.

Оценки, полученные методом моментов, часто оказываются смещенными. К достоинствам метода моментов следует отнести его простую вычислительную реализацию, а также то, что оценки являются функциями от выборочных моментов.

В силу теоремы Слуцкого любая непрерывная функция от выборочных моментов сходится по вероятности к постоянной, получаемой подстановкой в эту функцию теоретических моментов, если они существуют и если получаемая таким образом постоянная конечна. Для определенности рассмотрим функцию $H(\alpha_1, \alpha_2)$ от двух моментов (начальных или центральных), хотя результат можно обобщить на любое конечное число аргументов.

Теорема 18.1 (теорема Крамера). Пусть в некоторой окрестности точки (α_1, α_2) функция $H(\alpha_1, \alpha_2)$ непрерывна и имеет непрерывные частные производные первого и второго порядка. Обозначим

$$H_1 = \frac{\partial H(\alpha_1, \alpha_2)}{\partial \alpha_1}, \qquad H_2 = \frac{\partial H(\alpha_1, \alpha_2)}{\partial \alpha_2} \qquad \text{и} \qquad H_0 = H(\alpha_1, \alpha_2).$$

Тогда случайная величина $H(\alpha_1, \alpha_2)$ асимптотически нормальна при $n \to \infty$ со следующими параметрами:

$$H(\hat{\alpha}_1, \hat{\alpha}_2) \xrightarrow{\in} N(H_0,\ D\hat{\alpha}_1 H_1^2 + 2H_1 H_2 \operatorname{cov}(\hat{\alpha}_1, \hat{\alpha}_2) + D\hat{\alpha}_2 H_2^2).$$

На практике часто возникает необходимость оценить не сами параметры распределения (которые представляют собой некие абстракции), а определенные экономически значимые показатели, зависящие от этих параметров функционально: $G = g(\theta_1, \theta_2, ..., \theta_k)$. Самый простой (хотя и не самый точный) способ такого оценивания — подставить полученные оценки в соответствующую функцию: $\hat{G} = g(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)$.

Если распределение определяется одним параметром, то для построения оценки один теоретический момент приравнивают одному эмпирическому моменту того же порядка (как правило, первого).

Пример 18.2. Случайная величина x (число появлений события A в m независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром p. Ниже приведено эмпирическое распределение числа появлений события в 10 опытах по 5 испытаний в каждом (в первой строке указано число x_i появлений события A в одном опыте; во второй строке указана частота n_i — количество опытов, в которых наблюдалось столько появлений события A):

x_i	0	1	2	3	4
n_i	5	2	1	1	1

Найти методом моментов точечную оценку параметра p биномиального распределения. Оценить вероятность $p_0 = P(\xi = 0)$.

Решение. Математическое ожидание биномиального распределения известно: $M\xi = mp$. Приравняв математическое ожидание к выборочному среднему, получим уравнение: $m\hat{p} = \bar{x}$, откуда $\hat{p} = \bar{x}/m$. Для рассматриваемого примера имеем:

$\bar{x} = (0 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 1)/10 = 1,1;$

$\hat{p} = 1,1/5 = 0,22;\ \hat{p}_0 = (1 - 0,22)^5 \approx 0,29.$

Если распределение определяется двумя параметрами, то для построения их оценок два теоретических момента приравнивают двум соответствующим эмпирическим моментам тех же порядков (как правило, первым двум).

Пример 18.3. Случайная величина ξ (отклонение контролируемого размера изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами a и σ. Ниже приведена таблица наблюдаемых отклонений от номинала, подвергнутых группировке, для $n = 200$ изделий. В первой строке указаны середины интервалов отклонений x_i (мм); во второй строке приведена частота n_i — число наблюдений, попадающих в данный интервал.

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки неизвестных параметров a и σ нормального распределения. Оценить долю изделий с отклонением менее 1,5 мм в генеральной совокупности, используя нормальное приближение.

Решение. Для нахождения двух неизвестных параметров необходимо два уравнения. Получаем их, приравнивая теоретические среднее и дисперсию выборочному среднему и исправленной выборочной дисперсии соответственно. Находим величины $\bar{x}$ и $\hat{\sigma}^2$ по данным выборки: $\bar{x} \approx 1,27$; $\sigma^2 \approx 0,25$. Таким образом, получаем: $a^* \approx 1,27$ (мм); $\sigma^* \approx 0,5$ (мм).

Оценим долю изделий с отклонением менее 1,5 мм как вероятность для нормальной ξ:

$$\hat{p} = \hat{P}(\xi < 1,5) = \Phi\left(\frac{1,5 - 1,27}{0,5}\right) \approx 0,68.$$

З а м е ч а н и е. Эту долю можно оценить также непосредственно по таблице: поскольку значение 1,5 мм делит соответствующий интервал пополам, получается, что 122 из 200 изделий имеют отклонение меньше заданного, что дает близкую оценку 0,61.

Отметим, что если между полученными таким образом оценками наблюдается сильное расхождение, это может означать, что исходное предположение о законе распределения случайной величины неверно (т.е. не согласуется с реальными данными).

Пример 18.4. Предполагается, что выполнение некоторой работы занимает случайное время с распределением Симпсона на отрезке $[a; b]$. Хронометраж 20 испытаний дал среднее время работы 30 мин и исправленную выборочную дисперсию 24 мин^2. Оценить параметры a и b методом моментов. Оценить, за какое время работа будет выполняться с вероятностью 98%.

Решение. Для распределения Симпсона (плотность которого имеет вид равнобедренного треугольника с основанием на заданном отрезке) имеем

$$M\xi = \frac{a+b}{2}; \quad D\xi = \frac{(b-a)^2}{24}.$$

Параметры распределения можно выразить через математическое ожидание и дисперсию:

$$a = M\xi - \sqrt{6D\xi};$$

$$b = M\xi + \sqrt{6D\xi}.$$

Подставляя вместо теоретических моментов выборочные, получаем оценки:

$$\hat{a} = \bar{x} - \sqrt{6s^2}; \quad \hat{b} = \bar{x} + \sqrt{6s^2},$$

откуда $\hat{a} = 18$ (мин), $\hat{b} = 42$ (мин).

Функция распределения Симпсона имеет вид

$$F(x) = \begin{cases} 0, & x < a; \\ 2(x-a)^2/(b-a)^2, & a \le x \le (a+b)/2; \\ 1 - 2(b-x)^2/(b-a)^2, & (a+b)/2 \le x \le b; \\ 1, & x > b. \end{cases}$$

Решая уравнение $F(x) = 0{,}98$, находим искомое время $T = b - 0{,}1(b-a)$.

Подставляя полученные оценки в формулу вместо теоретических параметров, получаем $\hat{T} = 42 - 0{,}1(42 - 18) = 39{,}6$ (мин).

18.2. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Одним из основных методов получения оценок параметров генеральной совокупности по данным выборки является *метод максимального правдоподобия*, предложенный Р. Фишером. Основу метода составляет функция правдоподобия, выражающая плотность вероятности (либо вероятность) совместного появления результатов выборки $x_1, x_2, ..., x_n$. Совместное распределение этих величин задается в виде произведения частных распределений (поскольку предполагается, что наблюдения независимы), и следовательно, *функция правдоподобия* имеет вид

$$L(x_1, x_2, ..., x_n, \theta) = \prod_{i=1}^{n} P(x_i, \theta)$$

в случае дискретного распределением, заданного вероятностями $P(x, \theta)$, и

$$L(x_1, x_2, ..., x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

в случае непрерывного распределения с плотностью $f(x, \theta)$.

Из определения функции правдоподобия следует, что чем вероятнее (правдоподобнее) оказывается набор значений $(x_1, x_2, ..., x_n)$ случайной величины ξ при фиксированном θ, тем больше значение функции правдоподобия. Поэтому в качестве оценки неизвестного параметра θ принимается такое значение $\hat{\theta}_n$, которое максимизирует функцию правдоподобия.

Во многих случаях поиск оценки упрощается, если максимизировать не саму функцию правдоподобия, а ее логарифм, потому что максимумы $L(X, \theta)$ и $\ln L(X, \theta)$ достигаются при одном и том же значении параметра θ. Функцию $l_n(x_1, ..., x_n, \theta) = \ln L(X, \theta)$ называют *логарифмической функцией правдоподобия*.

Используется следующий алгоритм для отыскания оценки параметра θ: решают уравнение или систему уравнений правдоподобия, получаемых приравниванием производной (частных производных) по параметру (параметрам) θ нулю: $\frac{\partial \ln L}{\partial \theta} = 0$. Затем отбирают то решение, которое соответствует именно максимуму функции $\ln L$, т.е. вторая производная в данной точке должна быть отрицательной: $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$. Иногда функция правдоподобия имеет несколько максимумов и приходится искать наибольший среди них.

Бывают случаи, когда алгоритм не действует, поскольку функция правдоподобия достигает максимума не во внутренней точке, а на границе некоторой области, либо когда она недифференцируема в точке максимума. Такие случаи называют *нерегулярными*.

Достаточные *условия регулярности* (в одномерном случае) следующие:

1) в интервале всех возможных значений параметра θ плотность $f(x,\ \theta)$ трижды дифференцируема по θ, причем $\left|\frac{\partial^3 f}{\partial\theta^3}\right| \le H(x)$, где $MH(\xi) \le C$, C не зависит от θ;

2) тождество $\int\limits_{-\infty}^{+\infty} f(x,\theta)dx \equiv 1$ можно дважды дифференцировать под знаком интеграла по параметру θ;

3) информация Фишера $I(\theta) = M\left(\frac{\partial \ln f(\xi,\theta)}{\partial\theta}\right)^2$ конечна и положительна.

Теорема 18.2. Если выполнены условия регулярности 1–3, то:

1) решение $\hat{\theta}_n$ уравнения правдоподобия существует;

2) $\hat{\theta}_n$ — состоятельная оценка параметра θ;

3) распределение оценки $\hat{\theta}_n$ асимптотически нормально с параметрами θ и $\frac{1}{nI(\theta)}$;

4) оценка максимального правдоподобия асимптотически эффективна.

Если для оценки максимального правдоподобия выполнены условия теоремы Рао—Фреше—Крамера (см. § 11.3) , то справедливо утверждение:

Теорема 18.3. Если эффективная (по Рао—Фреше—Крамеру) оценка существует, то она является оценкой максимального правдоподобия.

Однако это не означает, что любая оценка максимального правдоподобия эффективна. Но если оценка максимального правдоподобия оказывается неэффективна, это значит, что эффективных оценок вообще нет, хотя при этом могут существовать оценки с дисперсией, сколь угодно близкой к Δ_n.

Следует также отметить, что метод максимального правдоподобия иногда дает те же оценки, что и метод моментов, а иногда — другие.

Бывает, что ни один из этих методов не дает хороших оценок и приходится использовать другие методы.

А. Дискретные распределения

Пример 18.5. Найти методом максимального правдоподобия оценку вероятности «успеха» θ в схеме испытаний Бернулли.

Решение. Рассмотрим случайную величину

$$\xi = \begin{cases} 1, & \text{в случае «успеха»}, \\ 0, & \text{в случае «неудачи»}, \end{cases}$$

тогда функция вероятностей случайной величины ξ запишется в виде

$$P(x, \theta) = \begin{cases} \theta, & x = 1, \\ 1 - \theta, & x = 0. \end{cases}$$

Логарифмическая функция правдоподобия для одного испытания будет иметь вид

$$l_1(\theta) = \ln P(x, \theta) = \begin{cases} \ln\theta, & x = 1, \\ \ln(1 - \theta), & x = 0. \end{cases}$$

Для n испытаний и m «успехов» в n испытаниях получаем

$$l_n(\theta) = m\ln\theta + (n - m)\ln(1 - \theta).$$

Отсюда $\ln(\theta) = \frac{m}{\theta} - \frac{n-m}{1-\theta} = 0$ и $\hat{\theta}_n = \frac{m}{n}$. Проверим знак второй производной:

$$l_n''(\theta) = -\frac{m}{\theta^2} - \frac{n-m}{(1-\theta)^2} < 0.$$

Таким образом, относительная частота появления события является оценкой вероятности «успеха» в одном испытании Бернулли, найденной методом максимального правдоподобия. Поскольку $M\hat{\theta}_n = \theta$, то оценка $\hat{\theta}_n$ является несмещенной оценкой вероятности.

Пример 18.6. Найти методом максимального правдоподобия оценку параметра λ распределения Пуассона.

Решение. Выпишем функцию правдоподобия:

$$L = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\ldots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} = \frac{1}{x_1!x_2!\ldots x_n!}\lambda^{x_1+x_2+\ldots+x_n}e^{-n\lambda}.$$

Найдем точку максимума логарифмической функции правдоподобия, для чего приравняем к нулю ее первую производную по λ:

$$\ln L = \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - n\lambda - \ln(x_1!x_2!\ldots x_n!), \qquad \frac{d\ln L}{d\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0.$$

Получаем $\lambda^* = \bar{x}$. Убедимся, что полученное значение λ является точкой максимума. Для этого найдем вторую производную и проверим ее знак в точке λ^*: $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i$. Если в последнее уравнение подставить $\lambda^* = \bar{x}$, то вторая производная будет отрицательна, значит, $\bar{x}$ является точкой максимума.

Б. Непрерывные распределения

Пример 18.7. Методом максимального правдоподобия найти оценки параметров a и σ нормального распределения, плотность которого

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Решение. Выпишем функцию правдоподобия:

$$L(x_1, x_2, \ldots, x_n, a, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right).$$

Логарифмическая функция правдоподобия имеет вид

$$\ln L = -n\ln\sigma - n\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2.$$

Найдем точку максимума, решив систему из двух уравнений, получающихся путем приравнивания первых двух частных производных по неизвестным параметрам к нулю:

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i - a) = 0, \\ \dfrac{\partial \ln L}{\partial \sigma} = -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\sum\limits_{i=1}^{n}(x_i - a)^2 = 0, \end{cases} \Rightarrow \begin{cases} \sum\limits_{i=1}^{n}(x_i - a) = 0, \\ \dfrac{n}{\sigma} = \dfrac{\sum\limits_{i=1}^{n}(x_i - a)^2}{\sigma^3}, \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \hat{a} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i, \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2. \end{cases}$$

Проверим, является ли точка $(\hat{a}, \hat{\sigma}^2)$ точкой максимума функции правдоподобия:

$$\begin{cases} \dfrac{\partial^2 \ln L}{\partial a^2} = -\dfrac{n}{\sigma^2}, \\ \dfrac{\partial^2 L}{\partial \sigma^2} = \dfrac{n}{\sigma^2} - \dfrac{3}{\sigma^4}\sum\limits_{i=1}^{n}(x_i - a)^2, \\ \dfrac{\partial^2 \ln L}{\partial a \partial \sigma} = -\dfrac{2}{\sigma^3}\sum\limits_{i=1}^{n}(x_i - a). \end{cases}$$

Как известно из математического анализа, чтобы функция $f(x_1, x_2, ..., x_n)$ достигала максимума в некоторой точке, достаточно, чтобы матрица второго дифференциала функции d^2f в этой точке была отрицательно определена. По критерию Сильвестра для этого необходимо и достаточно, чтобы ее главные миноры чередовались по знаку, а именно $\Delta_1 < 0$, $\Delta_2 > 0$. Рассмотрим матрицу производных:

$$A = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i - a) \\ -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i - a) & \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i - a)^2 \end{pmatrix}.$$

Отсюда

$$\Delta_1 = -\frac{n}{\sigma^2} < 0;$$

$$\Delta_2 = -\frac{n^2}{\sigma^4} + \frac{3n}{\sigma^6}\sum_{i=1}^{n}(x_i - a)^2 - \frac{4}{\sigma^6}\left(\sum_{i=1}^{n}(x_i - a)\right)^2 = -\frac{n^2}{\sigma^4} + \frac{3n^2}{\sigma^4} - 0 = \frac{2n^2}{\sigma^4} > 0.$$

Следовательно, точка $(\hat{a}, \hat{\sigma}^2)$ — действительно точка максимума и полученные оценки являются оценками максимального правдоподобия.

Пример 18.8. Найти оценку максимального правдоподобия для параметра сдвига θ распределения Коши, заданного плотностью $f(x) = \frac{1}{\pi(1 + (x - \theta)^2)}$, по выборке из двух наблюдений, если: а) $x_1 = -1$, $x_2 = 1$; б) $x_1 = -2$, $x_2 = 2$.

Решение. Функция правдоподобия для двух наблюдений имеет вид

$$L(\theta) = \frac{1}{\pi(1 + (x_1 - \theta)^2)} \cdot \frac{1}{\pi(1 + (x_2 - \theta)^2)}.$$

Введем функцию $R(\theta) = \frac{1}{\pi^2 L(\theta)}$. Тогда задача максимизации функции правдоподобия эквивалентна задаче минимизации $R(\theta)$:

а) если $x_1 = -1$, $x_2 = 1$, то $R(\theta) = (1 + (1 + \theta)^2)(1 + (\theta - 1)^2) = \theta^4 + 4$. Функция R достигает минимума в точке $\theta = 0$, так что это и есть оценка максимального правдоподобия;

б) если $x_1 = -2$, $x_2 = 2$, то $R(\theta) = (1 + (2 + \theta)^2)(1 + (\theta - 2)^2) = \theta^4 - 6\theta^2 + 25$. В этом случае производная имеет три нуля: в точках $\theta = 0$ и $\theta = \pm\sqrt{3}$. При этом точка $\theta = 0$ оказывается точкой максимума функции R. Точкам $\theta = \pm\sqrt{3}$ соответствует минимум функции R, причем в обеих этих точках величина R одинакова. Таким образом, оба значения $\pm\sqrt{3}$ являются в данном случае оценками максимального правдоподобия.

З а м е ч а н и е. Ни метод моментов, ни метод максимального правдоподобия не могут дать хороших оценок для параметра сдвига распределения Коши. Тем не менее, существует простая оценка для него — выборочная медиана: $\hat{\theta} = x_{\text{med}}$, поскольку $F(\theta) = {}^1/_2$.

В. Нерегулярные случаи

Пример 18.9. Найти методом максимального правдоподобия оценки параметров a и b равномерного закона распределения

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a; b], \\ 0, & x \notin [a; b]. \end{cases}$$

Решение. Заметим, что в данном случае не выполняются условия регулярности, поскольку плотность не только недифференцируема, но даже разрывна (как функция от a и b).

Выпишем функцию правдоподобия для равномерного распределения:

$$L(x_1, x_2, \ldots, x_n) = \frac{1}{(b-a)^n}, \text{ если все } x_i \in [a; b],$$

$$L(x_1, x_2, \ldots, x_n) = 0 \text{ — иначе.}$$

Если брать частные производные и приравнивать их к нулю, то легко убедиться, что полученные уравнения правдоподобия не имеют решений. В данном случае нужно действовать другим способом.

Условие, что все наблюдения принадлежат отрезку $[a; b]$, можно выразить через неравенства для крайних членов вариационного ряда: $a \le x_{\min}$, $b \ge x_{\max}$. При фиксированном значении a функция правдоподобия убывает по b при $b \ge x_{\max}$ и, следовательно, принимает максимальное значение при $b = x_{\max}$. При фиксированном значении b функция правдоподобия возрастает по a при $a \le x_{\min}$ и, следовательно, принимает максимальное значение при $a = x_{\min}$. Таким образом, оценками максимального правдоподобия будут крайние члены вариационного ряда: $\hat{a} = x_{\min}$ и $\hat{b} = x_{\max}$.

Пример 18.10. Построить оценку методом максимального правдоподобия параметра сдвига θ для распределения Лапласа с плотностью

$$f(x) = \frac{1}{2}\exp(-|x-\theta|), \quad -\infty < x < \infty.$$

Решение. Функция правдоподобия имеет вид

$$L(\theta) = \frac{1}{2^n}\exp\left(-\sum_{i=1}^{n}|x_i - \theta|\right).$$

Логарифмируя, получаем:

$$\ln L(\theta) = -n\ln 2 - \sum_{i=1}^{n}|x_i - \theta|.$$

Заметим, что эта функция недифференцируема во всех точках x_1, x_2, ..., x_n, а в остальных точках производная имеет вид

$$\frac{d\ln L}{d\theta}=\sum_{i=1}^{n}\begin{cases}+1, & \theta < x_i,\\ -1, & \theta > x_i.\end{cases}$$

Отсюда следует, что функция правдоподобия возрастает, если слева от значения θ находится меньше членов вариационного ряда, чем справа, и убывает в противном случае. Следовательно, максимума она достигает посередине вариационного ряда. Если $n = 2k + 1$, то это происходит в точке $x_{(k)}$. Если $n = 2k$, то функция постоянна на интервале $(x_{(k)}; x_{(k+1)})$, где принимает наибольшее значение, и в качестве оценки можно взять середину этого интервала. Таким образом, оценкой максимального правдоподобия оказывается выборочная медиана: $\hat{\theta} = x_{\text{med}}$.

18.3. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Метод наименьших квадратов особенно эффективен при нахождении оценок параметров функциональной зависимости между переменными, значения которых определяются из эксперимента.

Пусть Y — некоторый экономический показатель, объективный закон которого описывается функциональной зависимостью $Y = \varphi(X, \theta)$, где $\theta = (\theta_1, \theta_2, ..., \theta_k)$ — параметр; X — многомерная неслучайная переменная. Пусть в результате i-го наблюдения мы получили значение y_i функции $\varphi(x_i, \theta)$ со случайной ошибкой ε_i, т.е. $y_i = \varphi(x_i, \theta) + \varepsilon_i$.

Требуется по наблюдениям (x_1, y_1), ..., (x_n, y_n) оценить значения параметров $\theta_1, \theta_2, ..., \theta_k$.

Если в результате опыта получено n пар значений (x_i, y_i), где x_i — значение аргумента, а y_i — значение функции, то параметры $(\theta_1, \theta_2, ..., \theta_k)$ аппроксимирующей функции выбираются так, чтобы обратилась в минимум сумма

$$S=\sum_{i=1}^{n}[y_i+\varphi(x_i,\theta)]^2.$$

Метод получения оценок, основанный на минимизации суммы квадратов отклонений измерений от теоретических значений, называют *методом наименьших квадратов*.

Пусть измерения независимы. Ошибки ε_i — суммарный результат действия многих независимых причин и неучтенных факторов, которые при достаточно большом n в среднем нивелируют друг друга и, следовательно, их математическое ожидание можно считать равным нулю: $M\varepsilon_i = 0$. Предположим, что ошибки ε_i имеют одинаковую дисперсию, равную $D\varepsilon_i = \sigma^2$. В отличие от предыдущих схем оценивания в данном случае мы не обязаны задаваться общим видом закона распределения ошибок ε_i, однако можно предположить, что распределение ε_i нормальное: $\varepsilon_i \in N(0, \sigma^2)$. Величина y_i как сумма постоянной величины $\varphi(x_i, \theta)$ и случайной величины ε_i является случайной величиной, и ее распределение также нормальное: $y_i \in N(\varphi(x_i, \theta), \sigma^2)$. Плотность распределения y_i будет иметь вид

$$f(y_i, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \varphi(x_i, \theta))^2}{2\sigma^2}}.$$

Функция правдоподобия для наблюдаемых значений $x_1, x_2, \ldots, x_n$ будет равна

$$L(Y, \theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \varphi(x_i, \theta)]^2}.$$

Отсюда функция правдоподобия $L(\theta)$ при изменении θ имеет максимум тогда и только тогда, когда статистика

$$T_n(\theta) = \sum_{1}^{n}[y_i - \varphi(x_i, \theta)]^2$$

достигает минимума.

Задача свелась к задаче математического программирования: найти такое значение $\hat{\theta}_n$, которое минимизировало бы квадратичную форму

$$T_n(\theta_n) = \min_{\theta_1, \ldots, \theta_k} \sum_{i=1}^{n}[y_i - \varphi(x_i, \theta)]^2,$$

где $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ — случайная выборка. Так как $\hat{\theta}_n = (\theta_1, \theta_2, \ldots, \theta_k)$ является точкой минимума статистики $T_n(\theta)$, то, приравнивая нулю частные производные,

$$\frac{\partial T_n(\theta_n)}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, k,$$

получаем систему нормальных уравнений, решения которых и являются оценками $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \theta_n)$ неизвестных параметров, найденных по методу наименьших квадратов.

Система нормальных уравнений всегда имеет решение, так как положительный квадратичный многочлен всегда достигает минимума. Однако решение не обязательно является однозначным. Может случиться, что нормальные уравнения однозначно разрешимы лишь для некоторых определенных линейных комбинаций параметров $\theta_1, \theta_2, \ldots, \theta_k$, а относительно самих параметров однозначного решения нет. Такие линейные комбинации индийский статистик Рао назвал *допускающими оценку*.

Теорема 18.4 (теорема Гаусса — Маркова). Если случайные остатки ε_i имеют нулевые средние значения и одинаковые дисперсии: $E\varepsilon_i = 0$, $D\varepsilon_i = \sigma^2$, не зависящие от номера i и параметра θ, и остатки некоррелированы, а функция $\varphi(x, \theta)$ непрерывна и дифференцируема по всем параметрам $\theta_1, \theta_2, \ldots, \theta_k$, то оценки, полученные методом наименьших квадратов, являются состоятельными, асимптотически несмещенными, асимптотически нормальными и асимптотически эффективными.

Метод наименьших квадратов получил самое широкое распространение в практике статистических исследований, так как, во-первых, не требует знания закона распределения выборочных данных, во-вторых, достаточно хорошо разработан в плане вычислительной реализации.

Линейной регрессией называется сведение наблюдаемой на опыте зависимости некоторой переменной (зависимой или объясняемой) от одной или более других переменных (независимых или объясняющих) к линейной (в предположении, что строгая линейная зависимость между ними нарушается случайными ошибками). Для проведения линейной регрессии используется метод наименьших квадратов.

В простейшем случае речь идет о двух переменных. Пусть x — независимая переменная, y — зависимая и между ними существует следующая связь:

$$y_i = a + bx_i + \varepsilon_i,$$

где a и b — числовые коэффициенты; ε_i — случайные ошибки, $M\varepsilon_i = 0$ и $D\varepsilon_i < \infty$. Задача состоит в том, чтобы по имеющимся наблюдениям $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ построить оценки для a и b.

Согласно методу наименьших квадратов необходимо решить следующую задачу:

$$T = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \to \min.$$

Решим задачу, вычислив частные производные суммы квадратов по каждому из коэффициентов и приравняв эти производные к нулю. Получим систему *нормальных уравнений*:

$$\begin{cases} \dfrac{\partial T}{\partial a} = -2\sum\limits_{i=1}^{n}(y_i - a - bx_i) = 0, \\ \dfrac{\partial T}{\partial b} = -2\sum\limits_{i=1}^{n}(y_i - a - bx_i)x_i = 0. \end{cases}$$

Решая данную систему относительно параметров a и b, получим оценки:

$$\hat{b} = \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}; \quad \hat{a} = \bar{y} - \hat{b}\bar{x}.$$

Уравнение вида $\hat{y} = \hat{a} + \hat{b}x$ называется *уравнением линейной регрессии*, а получаемые из него значения $\hat{y}_i = \hat{a} + \hat{b}x_i$ называются *предсказанными значениями* в отличие от *наблюдаемых значений* y_i. Заметим, что уравнение линейной регрессии часто бывает удобно записать в виде $\hat{y} = \bar{y} + \hat{b}(x - \bar{x})$. Соответствующая прямая всегда проходит через точку выборочных средних $(\bar{x}, \bar{y})$. Числитель и знаменатель в формуле оценки параметра b можно вычислять по следующим эквивалентным формулам:

$$\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}; \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2.$$

Здесь также можно перейти к условным вариантам $u_i = x_i - c$, $v_i = y_i - d$, и оценка b от этого не изменится.

В некоторых случаях, например когда величина y (по смыслу) представляет собой некую долю от x, рассматривают более простой тип линейной зависимости:

$$y_i = bx_i + \varepsilon_i,$$

т.е. полагают $a = 0$. В этом случае оценка b методом наименьших квадратов имеет вид

$$\hat{b} = \frac{\sum\limits_{i=1}^{n}y_i}{\sum\limits_{i=1}^{n}x_i}.$$

Важным и практически значимым результатом линейной регрессии является то, что она позволяет «предсказывать» значения зависимой переменной даже для таких значений независимых, которые не наблюдались реально. Таким образом, например, можно строить прогнозы на будущее.

Пример 18.11. Затраты x на развитие производства и y — величина годовой прибыли фирмы в течение пяти лет представлены в условных единицах следующей таблицей:

x	6	3	7	5	10
y	33	27	32	28	42

На величину прибыли влияют случайные факторы. Предполагается, что имеет место линейная зависимость $y_i = a + bx_i + \varepsilon_i$ между затратами x и прибылью y. Среднее значение ε_i равно нулю, и дисперсия конечна. Каждый год случайное влияние не коррелировано с предыдущими годами. Оценить параметры a и b. Оценить годовую прибыль в том случае, если на развитие производства будет затрачено 12 у.е.

Решение. Перейдем к условным вариантам $u_i = x_i - 6$, $v_i = y_i - 33$.

u	0	–3	1	–1	4
v	0	–6	–1	–5	9

Получаем $\bar{u} = (0 - 3 + 1 - 1 + 4)/5 = 0{,}2$; $\bar{v} = (0 - 6 - 1 - 5 + 9)/5 = -0{,}6$.
Отсюда $\bar{x} = 6{,}2$; $\bar{y} = 32{,}4$.
Далее вычисляем следующие суммы:

$$\sum_{i=1}^{n} u_i v_i - n\bar{u}\bar{v} = ((-3)\cdot(-6) + 1\cdot(-1) + (-1)\cdot(-5) + 4\cdot 9) - 5\cdot 0{,}2\cdot(-0{,}6) =$$

$$= 18 - 1 + 5 + 36 + 0{,}6 = 58{,}6;$$

$$\sum_{i=1}^{n} u_i^2 - n\bar{u}^2 = (0 + (-3)^2 + 1^2 + (-1)^2 + 4^2) - 5\cdot 0{,}2^2 =$$

$$= 9 + 1 + 1 + 16 - 0{,}2 = 26{,}8.$$

Получаем $\hat{b} = 58{,}6/26{,}8 \approx 2{,}187$; $\hat{a} = 32{,}4 - 2{,}187 \cdot 6{,}2 \approx 18{,}843$. Имеем $\hat{y}(12) \approx 45$ (у.е.).

На рис. 18.1 представлены данные задачи и прямая линейной регрессии.

Замечание. На самом деле по столь небольшому числу точек нельзя делать серьезных выводов на будущее.

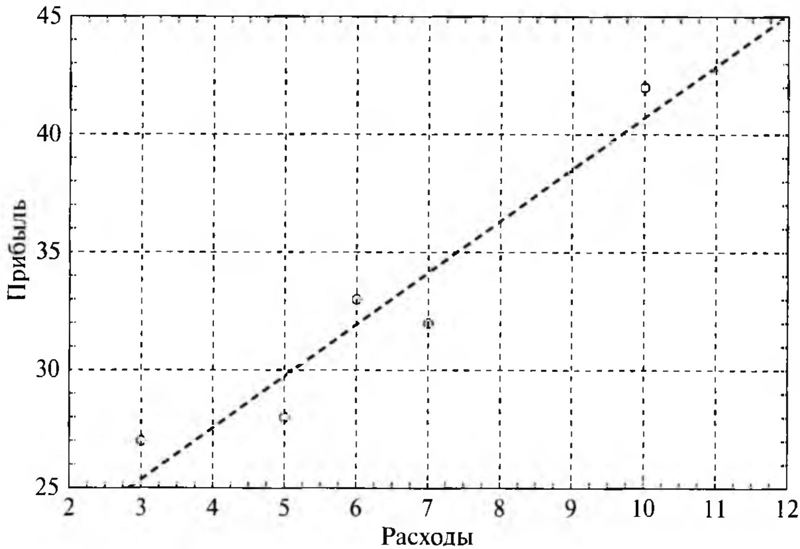

Рис. 18.1. Зависимость между расходами на развитие и прибылью

Пример 18.12. В таблице представлены данные о годовых доходах и расходах на личное потребление (в долл.) для 10 семей.

Годовой доход	*Расходы на личное потребление*
2508	2406
2572	2464
2408	2336
2522	2281
2700	2641
2531	2385
2390	2297
2595	2416
2524	2460
2685	2448

Провести линейную регрессию расходов по доходам в виде $y = bx$. Оценить параметр b. Оценить величину расходов для семьи с годовым доходом 2500 долл.

Решение. Суммируем доходы x_i, расходы y_i и делим одно на другое. Получаем значение параметра:

$$\hat{b} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{24\,134}{25\,435} \approx 0{,}949.$$

Для семьи с доходом 2500 долл. получаем оценку расходов в 2372 долл.

На рис. 18.2 представлены данные задачи и прямая линейной регрессии.

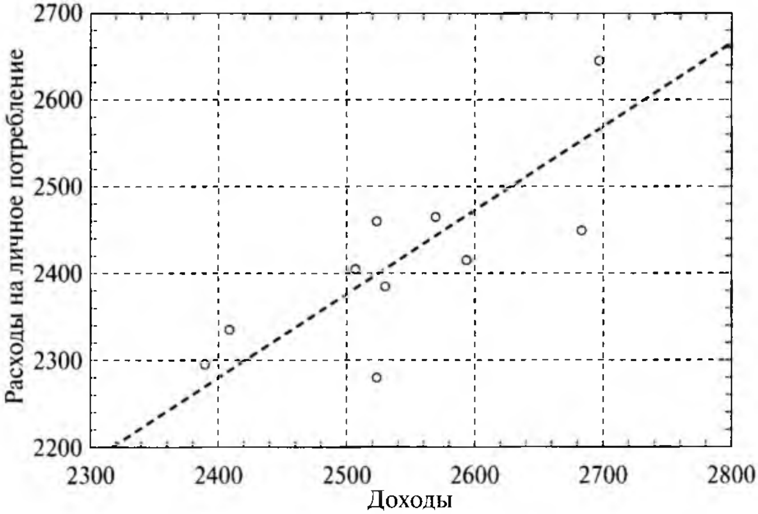

Рис. 18.2. Зависимость расходов и доходов

Глава 19
ФУНКЦИИ И РАСПРЕДЕЛЕНИЯ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

19.1. БЕТА- И ГАММА-ФУНКЦИИ

Бета-функцией или интегралом Эйлера *первого рода* называется интеграл вида $B(a,b)=\int\limits_0^1 x^{a-1}(1-x)^{b-1}dx$, где параметры $a>0$, $b>0$.

Введенный интеграл сходится для любых положительных значений параметров.

Свойства бета-функции следующие.

1. $B(a,b)=B(b,a)$.
2. $B(a,b)=\dfrac{b-1}{a+b-1}B(a,b-1)$, $B(1,\ 1)=1$.

Отсюда, в частности, следует, что $B(m,n)=\dfrac{(n-1)!(m-1)!}{(n+m-1)!}$.

3. $B(a,b)=\int\limits_0^\infty \dfrac{y^{a-1}}{(1+y)^{a+b}}dy$.

В частности, $B(a,1-a)=\int\limits_0^\infty \dfrac{y^{a-1}}{1+y}dy=\dfrac{\pi}{\sin a\pi}$.

Гамма-функцией или интегралом Эйлера *второго рода* называется функция вида $\Gamma(\alpha)=\int\limits_0^\infty x^{\alpha-1}e^{-x}dx$, где интеграл сходится для любого значения параметра $a>0$.

Свойства гамма-функции следующие.

1. $\Gamma(a+1)=a\Gamma(a)$, откуда следует $\Gamma(a+n)=(a+n-1)(a+n-2)\dots(a+1)\Gamma(a)$.
2. $\Gamma(1)=1$.
3. $\Gamma(n+1)=n!$, если $n\in N$.
4. $\Gamma(^1/_2)=\sqrt{\pi}$.
5. $B(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ — формула связи гамма- и бета-функций.
6. $\Gamma(a)\Gamma(1-a)=\dfrac{\pi}{\sin a\pi}$ (*формула дополнения*).
7. $\Gamma(x+1)\sim\left(\dfrac{x}{e}\right)^x\sqrt{2\pi x}$, $x\to\infty$ (*формула Стирлинга*).

Поскольку гамма-функция не выражается через элементарные функции, ее значения табулированы. Обычно в таблицах представлены значения $\Gamma(x)$ для $1 \le x \le 2$, и этого достаточно для вычисления функции при любых $x > 0$ (с помощью свойства 1).

Гамма-функция стремится к бесконечности при $x \to 0$ и при $x \to +\infty$. На рис. 19.1 представлен график гамма-функции на отрезке [0,01; 5].

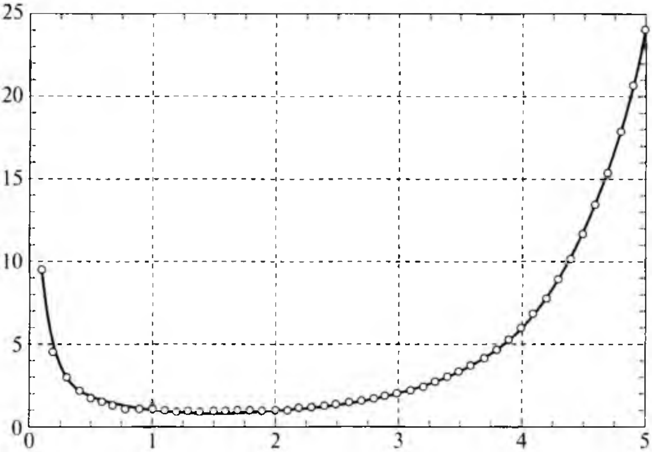

Рис. 19.1. График гамма-функции

Пример 19.1

1. $\Gamma(3/2) = \Gamma(1 + 1/2) = \dfrac{1}{2}\Gamma\left(\dfrac{1}{2}\right) = \dfrac{\sqrt{\pi}}{2}$.
2. Пользуясь свойствами гамма-функции, можно вычислить:

$$B\left(\frac{7}{2}, \frac{3}{2}\right) = \frac{\Gamma\left(\frac{7}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{7}{2} + \frac{3}{2}\right)} = \frac{\frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right)}{\Gamma(5)} = \frac{5\cdot 3\cdot 1}{4!\cdot 2^4}\pi = \frac{5\pi}{128}.$$

Пример 19.2. Доказать формулу, связывающую бета и гамма-функции:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

Решение. По определению бета-функция равна интегралу

$$B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}dx.$$

Сделаем замену переменных, положив $y = \frac{x}{1-x}$, тогда $x = \frac{y}{1+y}$ и

$$B(a,b) = \int_0^{+\infty} \frac{y^{a-1}}{(1+y)^{a-1}}\left(1 - \frac{y}{1+y}\right)^{b-1} \cdot \frac{1}{(1+y)^2}dy = \int_0^{+\infty} \frac{y^{a-1}}{(1+y)^{a+b}}dy.$$

Для доказательства равенства преобразуем гамма-функцию, сделав в интегральном выражении для $\Gamma(\alpha)$ замену $x = ty$, тогда

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx = \int_0^{+\infty} t^{\alpha-1}y^{\alpha-1}e^{-ty}tdy.$$

Отсюда следует равенство

$$\frac{\Gamma(\alpha)}{t^{\alpha}} = \int_0^{+\infty} y^{\alpha-1}e^{-y}dy.$$

Заменим в полученном равенстве t на $t+1$ и положим $\alpha = a+b$. Получим

$$\frac{\Gamma(a+b)}{(1+t)^{a+b}} = \int_0^{+\infty} y^{(a+b)-1}e^{-(1+t)y}dy.$$

Умножив это равенство на t^{a-1} и проинтегрировав по t от 0 до $+\infty$, получим следующее равенство:

$$\Gamma(a+b)\cdot\int_0^{+\infty}\frac{t^{a-1}}{(1+t)^{a+b}}dt = \int_0^{+\infty} t^{a-1}dt\int_0^{+\infty} y^{(a+b)-1}e^{-y(1+t)}dy;$$

$$\Gamma(a+b)\cdot B(a,b) = \int_0^{+\infty} y^{(a+b)-1}e^{-y}dy\int_0^{+\infty} t^{a-1}e^{-ty}dt = \int_0^{+\infty} y^{(a+b)-1}e^{-y}dy\cdot\frac{\Gamma(a)}{y^a} =$$

$$= \Gamma(a)\int_0^{+\infty} y^{b-1}e^{-y}dy = \Gamma(a)\Gamma(b).$$

19.2. КВАНТИЛИ, ПРОЦЕНТНЫЕ И КРИТИЧЕСКИЕ ТОЧКИ

В математической статистике при использовании различных распределений помимо их основных числовых характеристик важную роль играет определение границ, в которые значения случайных величин попадают с заданной вероятностью.

Квантилью уровня p или *p-квантилью* непрерывной случайной величины ξ с функцией распределения $F(x)$ называется такое возможное значение x_p этой случайной величины, для которого вероятность события $\xi < x_p$ равна заданной величине p: $P(\xi < x_p) = p$, $0 < p < 1$.

Геометрически x_p есть такое значение случайной величины ξ, при котором площадь криволинейной трапеции, ограниченная графиком плотности распределения и осью абсцисс и лежащая левее x_p, равна p.

Процентной точкой уровня q или *q%-ной точкой* (при $0 \le q \le 100$) для непрерывной случайной величины ξ с функцией распределения $F(x)$ называется такое значение v_q случайной величины, что вероятность события $\xi \ge v_q$ равна $q/100$, т.е. $1 - F(v_q) = P(\xi \ge v_q) = q/100$ (рис. 19.2).

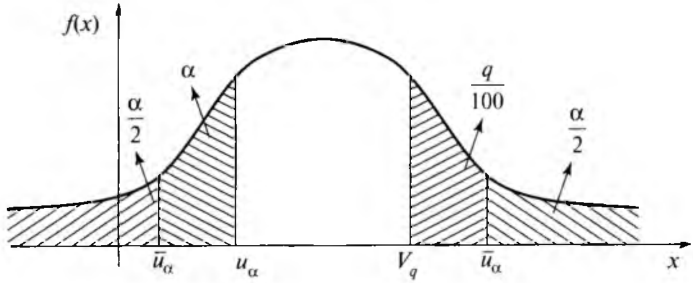

Рис. 19.2

Геометрически q%-ная точка — это значение случайной величины, при котором площадь криволинейной трапеции, ограниченной графиком плотности распределения, осью абсцисс и лежащая правее v_q, равна $q/100$.

На рис. 19.3 показана квантиль уровня 0,8 (20%-ная точка) для стандартного нормального распределения (на графиках плотности и функции распределения).

К понятию процентной точки близко понятие *критической точки*, широко используемое в задачах проверки гипотез.

Критические точки для заданного распределения определяют границы, за пределы которых случайная величина выходит достаточно редко. Например, если рассматриваются большие положительные значения случайной величины ξ, то критическая точка $t_{кр}$ может быть определена из условия $P(\xi > t_{кр}) = \alpha$, где α мало. Если же рассматриваются значения, большие по абсолютной величине (положительные или отрицательные), то можно определить критическую точку из условия $P(|\xi| > t_{кр}) = \alpha$.

Конкретные значения критических точек для различных распределений и уровней значимости можно найти в таблицах.

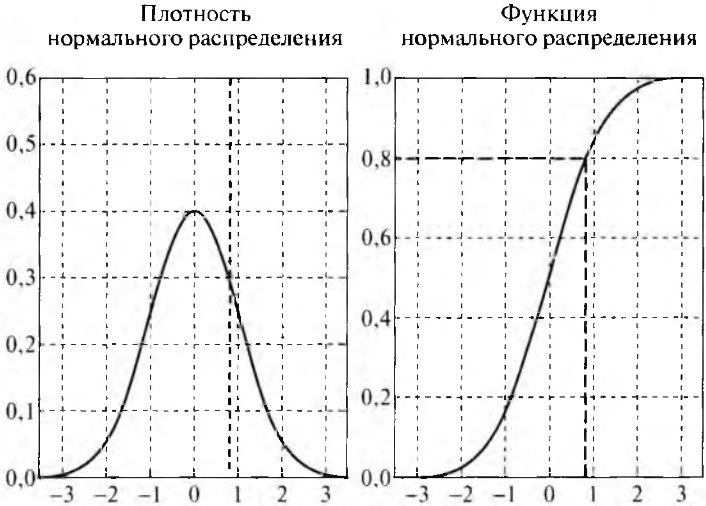

Рис. 19.3

19.3. РАСПРЕДЕЛЕНИЕ ХИ-КВАДРАТ (ЗАКОН ПИРСОНА)

Распределением хи-квадрат с n степенями свободы (обозначается χ_n^2) называется распределение суммы квадратов n независимых случайных величин со стандартным нормальным распределением, т.е. $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$, если $\xi_i \in N(0, 1)$.

Такое же распределение будет иметь величина $\chi_n^2 = \sum_{i=1}^{n} \frac{(\xi_i - a)^2}{\sigma^2}$, если $\xi_i \in N(a, \sigma^2)$.

Плотность распределения случайной величины χ_n^2 с n степенями свободы имеет вид

$$f_{\chi^2}(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Функция распределения

$$F(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \displaystyle\int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Основные числовые характеристики распределения χ_n^2 следующие:

- математическое ожидание $M\chi_n^2 = n$;
- мода $X_{\text{mod}} = n - 2$ при $n > 2$;
- дисперсия $D\chi_n^2 = 2n$;
- асимметрия $\beta_{\chi^2} = \sqrt{\dfrac{8}{n}}$;
- эксцесс $\gamma_{\chi^2} = \dfrac{12}{n}$.

Плотность распределения χ_n^2 зависит от одного параметра n — *числа степеней свободы*. При $n \le 2$ функция плотности убывает, а при $n > 2$ имеет единственный максимум в точке $X_{\text{mod}} = n - 2$. С ростом числа степеней свободы n распределение χ_n^2 приближается к нормальному закону распределения со средним n и дисперсией $2n$ (в смысле асимптотической нормальности).

Вид кривой распределения зависит от числа степеней свободы (рис. 19.4).

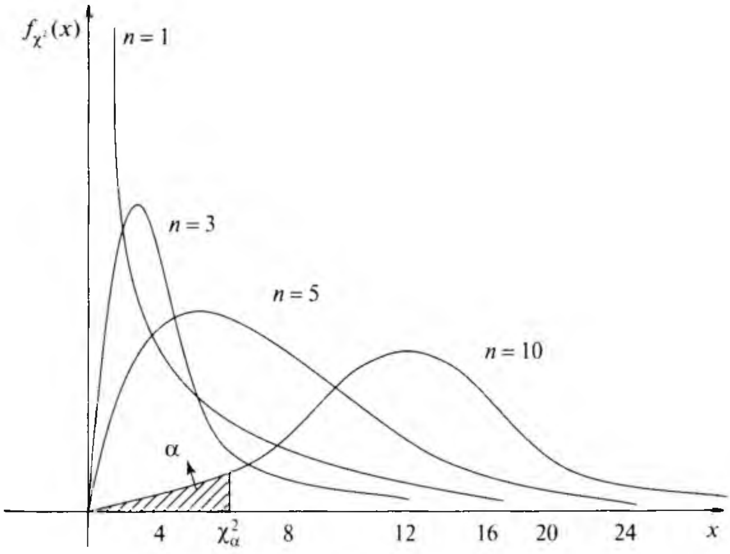

Рис. 19.4

На рис. 19.5 представлен график плотности распределения хи-квадрат с пятью степенями свободы.

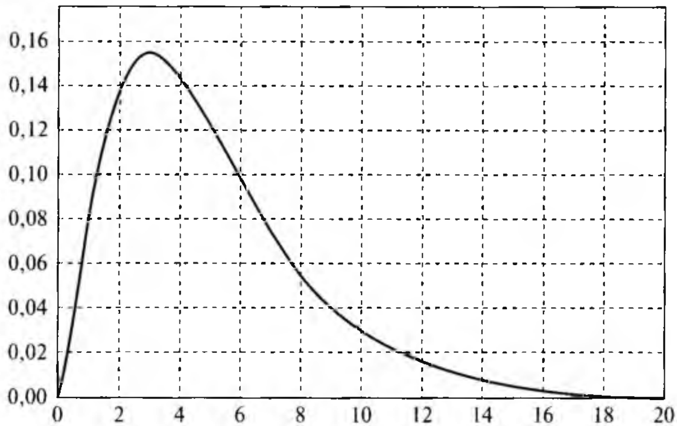

Рис. 19.5. Плотность распределения хи-квадрат

Значение квантиля $\chi^2_{\alpha,n}$, соответствующего заданному уровню значимости α, определяется из уравнения $P(\chi^2 < \chi^2_{\alpha,n}) = \alpha$. Геометрически квантиль $\chi^2_{\alpha,n}$ есть такое значение, при котором площадь заштрихованной криволинейной трапеции равна α. Таблица квантилей χ^2-распределения содержит процентные точки только для $n < 30$, так как для $n > 30$ квантили $\chi^2_{\alpha,n}$ можно определить с помощью таблиц нормального закона распределения.

19.4. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Пусть $n + 1$ случайных величин $\xi_0, \xi_1, \ldots, \xi_n$ независимы и имеют стандартное нормальное распределение: $\xi_i \in N(0, 1)$. Пусть $\eta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}$. Случайная величина $t_n = \frac{\xi_0}{\eta}$ называется *безразмерной дробью Стьюдента*, а ее распределение — *распределением Стьюдента с n степенями свободы.*

Плотность распределения t_n имеет вид

$$f_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Плотность распределения Стьюдента зависит от одного параметра — числа степеней свободы n. При $n = 1$ распределение Стьюдента совпадает с распределением Коши. С ростом числа степеней свободы n распределение Стьюдента сходится к стандартному нормальному распределению (при $n \to \infty$).

Основные числовые характеристики:

- мода и медиана равны нулю, т.е. распределение унимодально и симметрично относительно точки $x = 0$; математическое ожидание существует при $n > 1$ и также равно нулю;
- дисперсия $Dt_n = \dfrac{n}{n-2}$ и существует только при $n > 2$;
- асимметрия $\beta = 0$;
- эксцесс $\gamma_n = \dfrac{6}{n-4}$ и существует только при $n > 4$.

На рис. 19.6 представлен график плотности распределения Стьюдента с тремя степенями свободы.

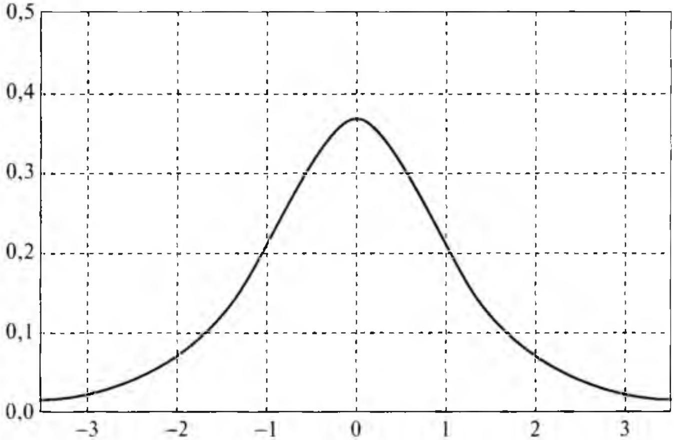

Рис. 19.6. Плотность распределения Стьюдента

Квантили распределения Стьюдента находятся из уравнения

$$P\left(|t_n| > t_{\frac{\alpha}{2},n}\right) = 2\int\limits_{t_{\frac{\alpha}{2},n}}^{\infty} f(t)dt = \alpha.$$

Геометрически квантиль $t_{\frac{\alpha}{2},n}$ есть такое значение случайной величины t_n, что суммарная площадь заштрихованных криволинейных трапеций равна α (рис. 19.7).

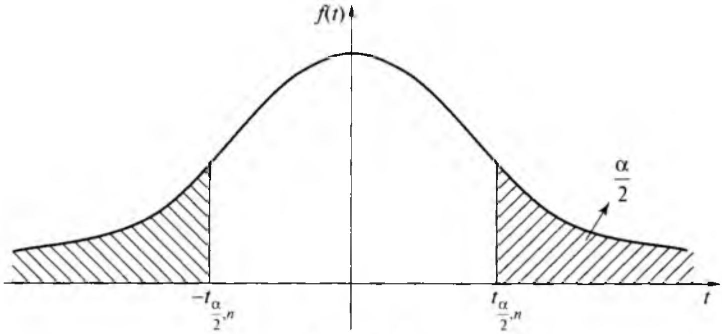

Рис. 19.7

19.5. РАСПРЕДЕЛЕНИЕ ФИШЕРА

Если χ_n^2 и χ_m^2 — независимые случайные величины, распределенные по закону χ^2 с числами степеней свободы n и m соответственно, то случайная величина

$$F(n,m)=\frac{\dfrac{\chi_n^2}{n}}{\dfrac{\chi_m^2}{m}}$$

имеет распределение, которое называют *распределением Фишера–Снедекора* (F-распределением) с числами степеней свободы n и m или *распределением дисперсионного отношения*.

Функция распределения и функция плотности имеют вид:

$$F_{n,m}(x)=\frac{\Gamma\left(\dfrac{n+m}{2}\right)n^{\frac{n}{2}}m^{\frac{m}{2}}}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\int_0^x\frac{u^{\frac{n}{2}-1}}{(nu+m)^{\frac{n+m}{2}}}du;$$

$$f_{n,m}(x)=\frac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\cdot\frac{n^{\frac{n}{2}}m^{\frac{m}{2}}x^{\frac{n}{2}-1}}{(nx+m)^{\frac{n+m}{2}}},\quad x>0.$$

Основные числовые характеристики:

- математическое ожидание $M_F = \dfrac{m}{m-2}$ существует только при $m > 2$;
- дисперсия $D_F = \dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ существует только при $m > 4$;
- мода $X_{\text{mod}} = \dfrac{m(n-2)}{n(m+2)}$ при $n > 1$;
- асимметрия $\beta_F = \dfrac{(2n+m-2)\sqrt{8(m-4)}}{(m-6)\sqrt{n(n+m-2)}}$ при $m > 6$.
- эксцесс $\gamma_F = \dfrac{3(m-6)\left(2+\frac{1}{2}\beta_F^2\right)}{m-8} - 3$ при $m > 8$.

Общий вид графика плотности распределения Фишера при различных значениях параметров приведен на рис. 19.8.

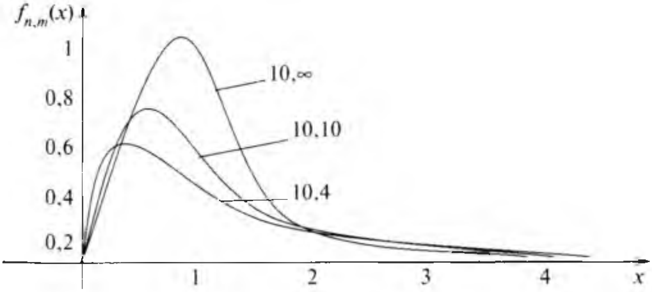

Рис. 19.8

На рис. 19.9 представлен график плотности распределения Фишера с числами степеней свободы $n = 10$ и $m = 15$.

19.6. ГАММА-РАСПРЕДЕЛЕНИЕ

Двухпараметрический закон гамма-распределения случайной величины $\xi = \gamma(\alpha, \lambda)$ описывается функцией плотности

$$f_\gamma(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

где $\Gamma(\alpha)$ — гамма-функция Эйлера; $\alpha > 0$ — параметр «формы»; $\lambda > 0$ — параметр «масштаба». Если $\lambda = 1$, случайная величина $\gamma(\alpha, 1)$

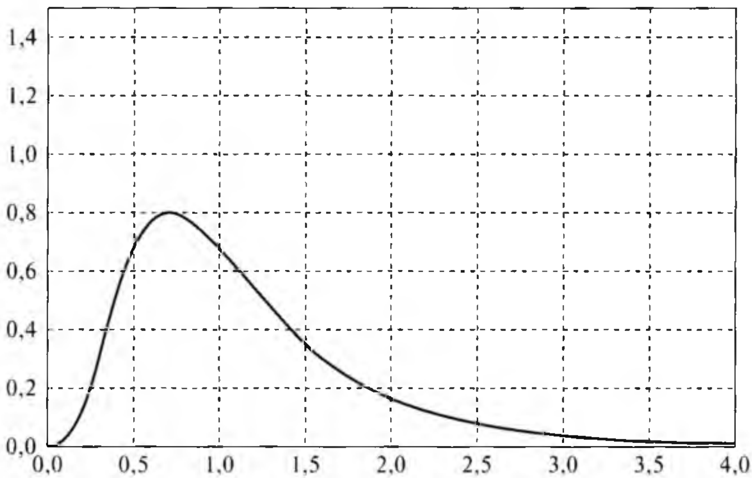

Рис. 19.9. Плотность распределения Фишера при $n = 10$ и $m = 15$

зависит от одного параметра α и подчинена *однопараметрическому закону гамма-распределения.*

Основные свойства гамма-распределения следующие.

1. Если случайная величина имеет гамма-распределение $\xi \in \gamma(\alpha, \lambda)$, то $c\xi \in \gamma(\alpha, \lambda/c)$, в частности $\lambda\xi \in \gamma(\alpha, 1)$.

2. Сумма любого числа независимых гамма-распределенных случайных величин с одинаковым параметром масштаба λ и параметрами формы $\alpha_1, \alpha_2, \ldots, \alpha_n$ также подчиняется гамма-распределению с параметрами $\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_n$ и λ.

3. Распределение χ_n^2 является частным случаем гамма-распределения с параметрами $\alpha = \dfrac{n}{2}, \lambda = \dfrac{1}{2}$, т.е. $\chi_n^2 = \gamma\left(\dfrac{n}{2}, \dfrac{1}{2}\right)$.

Основные числовые характеристики случайной величины $\gamma(\alpha, \lambda)$ следующие:

- математическое ожидание $M_\gamma = \dfrac{\alpha}{\lambda}$;
- мода $X_{\text{mod}} = \dfrac{\alpha - 1}{\lambda}$ при $\alpha \geq 1$;
- дисперсия $D_\gamma = \dfrac{\alpha}{\lambda^2}$;
- асимметрия $\beta_\gamma = \dfrac{2}{\sqrt{\alpha}}$;
- эксцесс $\gamma_\gamma = \dfrac{6}{\alpha}$.

На рис. 19.10 представлен график плотности гамма-распределения при $\alpha = 3$ и $\lambda = 1$.

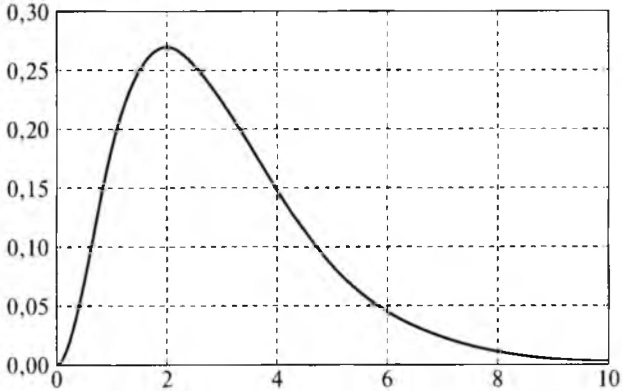

Рис. 19.10. Плотность гамма-распределения при $\alpha = 3$ и $\lambda = 1$

Общий вид графика плотности гамма-распределения при различных значениях параметров приведен на рис. 19.11.

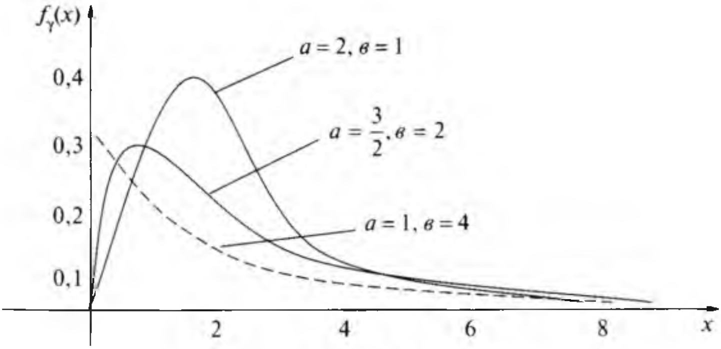

Рис. 19.11

Гамма-распределение иногда используют при моделировании реальных ситуаций. Например, с его помощью описывается распределение доходов или сбережений населения в некоторых случаях.

Пример 19.3. Пусть случайная величина ξ имеет плотность распределения вида

$$f(x, \alpha) = \begin{cases} Cx^{\alpha-1}e^{-x}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

Найти константу C, математическое ожидание $M\xi$ и дисперсию $D\xi$.

Решение. По свойству плотности распределения имеем

$$\int_{-\infty}^{+\infty} f(x)dx = C\int_{0}^{+\infty} x^{\alpha-1}e^{-x}dx = C\Gamma(\alpha) = 1, \quad \alpha > 0.$$

Отсюда получаем, что $C = \dfrac{1}{\Gamma(\alpha)}$. Вычислим математическое ожидание:

$$M\xi = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{+\infty} \frac{1}{\Gamma(\alpha)}x^{\alpha}e^{-x}dx = \frac{1}{\Gamma(\alpha)}\int_{0}^{+\infty} x^{(\alpha+1)-1}e^{-x}dx = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha.$$

Для нахождения дисперсии вычислим второй начальный момент:

$$M\xi^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{1}{\Gamma(\alpha)}\int_{0}^{+\infty} x^{(\alpha+2)-1}e^{-x}dx = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} =$$

$$= \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha(\alpha+1).$$

Тогда дисперсия равна

$D\xi = M\xi^2 - (M\xi)^2 = \alpha(\alpha+1) - \alpha^2 = \alpha.$

Пример 19.4. Пусть случайная величина η подчиняется стандартному нормальному закону распределения: $\eta \in N(0, 1)$. Найти функцию плотности случайной величины $\xi = \dfrac{\eta^2}{2}$. К какому параметрическому семейству распределений относится ξ?

Решение. Так как случайная величина $\eta \in N(0, 1)$, то ее плотность распределения $f_\eta(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Найдем функцию распределения случайной величины ξ:

$$F_\xi(x) = P(\xi < x) = P\left(\frac{\eta^2}{2} < x\right) = P(\eta^2 < 2x) = P(-\sqrt{2x} < \eta < \sqrt{2x}) =$$

$$= \int_{-\sqrt{2x}}^{\sqrt{2x}} p_\eta(t)dt = \frac{2}{\sqrt{2\pi}}\int_{0}^{\sqrt{2x}} \exp\left(-\frac{t^2}{2}\right)dt =$$

$$= \left|\frac{t^2}{2} = u; t^2 = 2u; 2tdt = 2du; dt = \frac{du}{t} = \frac{du}{\sqrt{2u}}\right| = \frac{2}{\sqrt{2\pi}}\int_{0}^{x} e^{-u}\frac{du}{\sqrt{2u}} =$$

$$= \frac{1}{\sqrt{\pi}}\int_{0}^{x} u^{-\frac{1}{2}}e^{-u}du.$$

Отсюда получаем плотность распределения случайной величины ξ в виде

$$f_\xi(x) = \frac{1}{\sqrt{\pi}}x^{-\frac{1}{2}}e^{-x},$$

а эта функция является плотностью однопараметрического гамма-распределения с параметром, равным $\alpha = \frac{1}{2}$, т.е. $\gamma\left(\frac{1}{2}, 1\right) = \gamma\left(\frac{1}{2}\right)$.

19.7. БЕТА-РАСПРЕДЕЛЕНИЕ

Случайная величина $\xi \in \beta(a, b)$, подчиняющаяся закону *бета-распределения* (*B-распределение*), с параметрами $a > 0$ и $b > 0$ имеет плотность распределения

$$f_\xi(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & 0 \le x \le 1, \\ 0, & x \notin [0;1]. \end{cases}$$

Основные свойства бета-распределения следующие.

1. Если $\xi_1 \in \gamma(a_1, \lambda)$ и $\xi_2 \in \gamma(\alpha_2, \lambda)$ — две независимые гамма-распределенные случайные величины, то отношение $\eta = \dfrac{\xi_1}{\xi_1 + \xi_2}$ имеет бета-распределение с параметрами α_1 и α_2:

$\eta \in \beta(\alpha_1, \alpha_2)$.

2. Случайная величина $\beta(1, 1)$ распределена равномерно на отрезке [0; 1];

3. Функция распределения квадрата стьюдентовской величины t_m^2 связана с функцией распределения случайной величины β соотношением

$$F_{t_m^2}(x) = F_{\beta\left(\frac{1}{2},\frac{m}{2}\right)}\left(\frac{x}{m+x}\right).$$

4. Функция распределения случайной величины $F(n, m)$ связана с функцией распределения случайной величины β соотношением

$$F_{n,m}(x) = F_{\beta\left(\frac{n}{2},\frac{m}{2}\right)}\left(\frac{nx}{m+nx}\right).$$

5. Между функцией распределения случайной величины β и биномиальным распределением существует соотношение

$$F_{\beta(m,n-m+1)}(x) = \sum_{k=m}^{n} C_n^k x^k (1-x)^{n-k}.$$

6. Имеет место симметрия: $p_{\beta(a,b)}(x) = p_{\beta(b,a)}(1-x)$ и $F_{\beta(a,b)}(x) = 1 - F_{\beta(b,a)}(1-x)$.

Основные числовые характеристики случайной величины $\beta(a, b)$:

- математическое ожидание (среднее) $M_{\beta(a,b)} = \dfrac{a}{a+b}$;
- мода $X_{\text{mod}} = \dfrac{a-1}{a+b-2}$ при $a > 1$, $b > 1$;
- дисперсия $D_{\beta(a,b)} = \dfrac{ab}{(a+b)^2(a+b+1)}$;
- асимметрия $\beta_\beta = \dfrac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$;
- эксцесс $\gamma_\beta = \dfrac{3(a+b+1)[2(a+b)^2 + ab(a+b-6)]}{ab(a+b+2)(a+b+3)} - 3$.

Бета-распределение используется для описания некоторых реальных распределений, сосредоточенных на отрезке [0; 1], например для описания распределения величин субъективных вероятностей, полученных в ходе экспертного опроса.

На рис. 19.12 представлен график плотности бета-распределения с $a = 3$, $b = 5$.

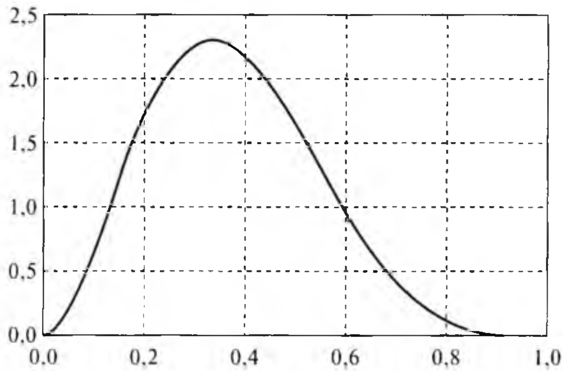

Рис. 19.12. Плотность бета-распределения при $a = 3$, $b = 5$

19.8. ПРИЛОЖЕНИЯ РАСПРЕДЕЛЕНИЙ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ. ТЕОРЕМА ФИШЕРА

При построении вероятностно-статистических моделей в экономике важным является исследование свойств конечной выборки при фиксированном n, когда выборка $x_1, x_2, \ldots, x_n$ сделана из нормальной генеральной совокупности $N(a, \sigma^2)$. Справедлива теорема Фишера, описывающая доасимптотические свойства оценок.

Теорема 19.1 (теорема Фишера). Пусть $\bar{x}$ и s^2 — выборочное среднее и исправленная выборочная дисперсия, построенные по выборке $x_1, x_2, \ldots, x_n$ из нормальной генеральной совокупности $N(a, \sigma^2)$. Тогда при любом фиксированном объеме выборки n их совместный закон распределения описывается следующим образом:

1) $\bar{x} \in N\left(a, \frac{\sigma^2}{n}\right)$;

2) $\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$;

3) случайные величины $\bar{x}$ и s^2 независимы.

Следствие 1. $Ds^2 = \frac{2\sigma^4}{n-1}$.

Следствие 2. Случайная величина

$$T = \frac{\bar{x} - a}{s}\sqrt{n}$$

распределена по закону Стьюдента с $n - 1$ степенями свободы.

Далее, предположим, что случайная выборка $x_1, x_2, \ldots, x_n$ произведена из нормальной генеральной совокупности $N(a_1, \sigma_1^2)$, а выборка $y_1, y_2, \ldots, y_m$ — из нормальной генеральной совокупности $N(a_2, \sigma_2^2)$, причем эти выборки независимы. Обозначим их выборочные средние $\bar{x}$ и $\bar{y}$, а исправленные выборочные дисперсии — s_1^2 и s_2^2 соответственно.

Следствие 3. $\bar{x} - \bar{y} \in N\left(a_1 - a_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$.

В частности, если обе выборки взяты из одной и той же генеральной совокупности ($a_1 = a_2$, $\sigma_1 = \sigma_2$), то

$$\bar{x} - \bar{y} \in N\left(0, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

Следствие 4. При одинаковых дисперсиях $\sigma_1^2 = \sigma_2^2 = \sigma^2$ случайная величина

$$T = \frac{(\bar{x} - \bar{y}) - (a_1 - a_2)}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)s_1^2 + (m-1)s_2^2}{n + m - 2}}}$$

имеет распределение Стьюдента с $(n + m - 2)$ степенями свободы.

Следствие 5. При одинаковых дисперсиях $\sigma_1^2 = \sigma_2^2 = \sigma^2$ случайная величина

$$F = \frac{s_1^2}{s_2^2}$$

распределена по закону Фишера с $n-1$ и $m-1$ степенями свободы.

Теорема Фишера и ее следствия применяются при исследовании различных свойств оценок, построении доверительных интервалов и проверке гипотез.

Пример 19.5. Доказать асимптотическую эффективность исправленной выборочной дисперсии как оценки теоретической дисперсии в случае нормального распределения.

Решение. Для того чтобы доказать эффективность оценки, найдем информацию Фишера. Выпишем функцию плотности нормального распределения

$$f_\xi(x,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и прологарифмируем ее:

$$\ln f_\xi(x,\sigma^2)=\ln\frac{1}{\sigma\sqrt{2\pi}}-\frac{(x-a)^2}{2\sigma^2}=-\ln\sqrt{2\pi}-\frac{1}{2}\ln\sigma^2-\frac{(x-a)^2}{2\sigma^2}.$$

Продифференцируем по параметру $\theta=\sigma^2$:

$$\frac{d\ln f_\xi(x,\theta)}{d\theta}=-\frac{1}{2\theta}+\frac{(x-a)^2}{2\theta^2};$$

$$\left(\frac{d\ln f_\xi(x,\theta)}{d\theta}\right)^2=\frac{1}{4\theta^4}((x-a)^2-\theta)^2.$$

Тогда информация Фишера равна

$$I(\theta)=\frac{1}{4\theta^4}M((\xi-a)^2-\theta)^2=\frac{1}{4\theta^4}D(\xi-a)^2=\frac{1}{4\theta^4}\theta^2D\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2=$$

$$=\frac{1}{4\theta^2}D\chi_1^2=\frac{1}{2\theta^2}=\frac{1}{2\sigma^4},$$

так как случайная величина $\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2=\chi_1^2$ имеет распределение хи-квадрат с одной степенью свободы и $D\chi_1^2=2$. Отсюда получим $\Delta_n=\frac{2\sigma^4}{n}$. С другой стороны, $Ds^2=\frac{2\sigma^4}{n-1}$, и в результате имеем $\frac{\Delta_n}{Ds^2}=\frac{2\sigma^4}{n\cdot\frac{2\sigma^4}{n-1}}\to 1$ при $n\to\infty$. Таким образом, в случае нормального распределения исправленная выборочная дисперсия является асимптотически эффективной оценкой теоретической дисперсии.

Глава 20
ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

20.1. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ. ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ

Точечная оценка неизвестного параметра, найденная по выборке объема n, не указывает, какую ошибку мы допускаем, принимая вместо точного значения параметра θ его приближенное значение. Поэтому строятся интервальные оценки, указывающие границы, в которых может быть заключено истинное значение параметра.

Доверительным интервалом или *интервальной оценкой* называется интервал $(\hat{\theta}_1, \hat{\theta}_2)$, который покрывает неизвестный параметр θ с заданной *доверительной вероятностью* $0 < \gamma < 1$ (ее называют также *надежностью* доверительного интервала). Часто доверительный интервал может быть представлен в виде $(\hat{\theta} - \delta, \hat{\theta} + \delta)$, тогда величина δ (половина длины интервала) называется *точностью* оценки (точностью доверительного интервала). При заданном значении γ точность δ зависит от объема выборки n. Понятно, что чем меньше длина доверительного интервала, тем точнее оценка. Границы доверительного интервала не должны явно зависеть от неизвестного значения параметра.

Однако сама интервальная оценка, как правило, конструируется вокруг точечной оценки, вид которой определятся законом распределения случайной величины, который сам зависит от неизвестного параметра, т.е. границы доверительного интервала зависят от значения неизвестного параметра и, следовательно, пользоваться такими границами нельзя.

Существуют два основных метода построения доверительных интервалов.

1. Классический метод состоит в искусственном подборе пары статистик $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \ldots, x_n)$, $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \ldots, x_n)$ таких, что $\hat{\theta}_1 < \hat{\theta}_2$ при любых $x_1, x_2, \ldots, x_n$, θ, и $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma$.

2. Другой метод основан на асимптотических свойствах оценок (как правило, асимптотической нормальности). Поэтому такой

метод дает приближенные результаты и пригоден только при достаточно больших объемах выборок.

В обоих случаях бывает, что границы доверительных интервалов, построенных формально из каких-то теоретических соображений, выходят за рамки возможного (например, становятся отрицательными для положительных по смыслу величин). В таких случаях их «округляют» до разумных пределов.

20.2. ТОЧНЫЕ ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

Точные доверительные интервалы строятся, как правило, в предположении нормальности данных. Следует понимать, что реальные данные, на основании которых мы строим эти интервалы, могут совсем не выглядеть нормальными (например, это целые положительные числа, в то время как нормальное распределение непрерывно и рассредоточено по всей действительной прямой). Тем не менее широкое практическое применение описываемых методов дает неплохие результаты (это объясняется, в частности, асимптотической нормальностью оценок), и такое несоответствие не должно нас смущать в дальнейшем.

Предположим, что наблюдается случайная величина $\xi \in N(a, \sigma^2)$. Для параметров строятся следующие точные доверительные интервалы.

1. Для неизвестного среднего a при известной дисперсии σ^2:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}} u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

2. Для неизвестного среднего a при неизвестной дисперсии σ^2:

$$\bar{x} - \frac{s}{\sqrt{n}} t_\gamma < a < \bar{x} + \frac{s}{\sqrt{n}} t_\gamma,$$

где t_γ — критическая точка распределения Стьюдента (для двусторонней области) с $n - 1$ степенями свободы и уровнем значимости $\alpha = 1 - \gamma$.

3. Для неизвестной дисперсии σ^2:

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}},$$

где $\chi^2_{\frac{\alpha}{2}, n-1}$ — критические точки χ^2-распределения с $n - 1$ степенями свободы и соответствующими уровнями значимости $\alpha = 1 - \gamma$.

Доказательство.

1. По теореме Фишера (см. § 12.8) имеем $\bar{x} \in N\left(a, \frac{\sigma^2}{n}\right)$. Случайная величина $\eta = \frac{\bar{x} - a}{\sigma}$ тогда имеет стандартное нормальное распределение $N(0, 1)$, и

$P(|\eta| < u) = P(-u < \eta < u) = \Phi(u) - \Phi(-u) = \Phi_0(u) - \Phi_0(-u) = 2\Phi_0(u).$

Выбирая u_γ из соотношения $\Phi_0(u_\gamma) = \gamma/2$, получаем, что $P(|\eta| < u_\gamma) = \gamma$. Таким образом, с вероятностью γ верно $\left|\frac{\bar{x} - a}{\sigma}\right| < u_\gamma$, откуда следует

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}} u_\gamma .$$

2. По следствию 2 из теоремы Фишера (см. § 12.8) статистика $T = \frac{\bar{x} - a}{s}\sqrt{n}$ имеет распределение Стьюдента с $n - 1$ степенями свободы. Заметим, что критическая точка распределения Стьюдента (для двусторонней области) определяется таким образом, чтобы выполнялось $P(|T| > t_{\text{кр}}) = \alpha$. Выбирая $\alpha = 1 - \gamma$, получаем $P(|T| < t_{\text{кр}}) = \gamma$. Таким образом, с вероятностью γ верно $\left|\frac{\bar{x} - a}{s}\right| < t_\gamma$, откуда следует

$$\bar{x} - \frac{s}{\sqrt{n}} t_\gamma < a < \bar{x} + \frac{s}{\sqrt{n}} t_\gamma .$$

3. По теореме Фишера имеем $\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$. Заметим, что критические точки распределения хи-квадрат $\chi^2_{p,n-1}$ определяются из условия $P(\chi^2 > \chi^2_{p,n-1}) = p$. Выбирая $\alpha = 1 - \gamma$, получаем

$$P\left(\chi^2_{1-\frac{\alpha}{2},n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2},n-1}\right) = P\left(\chi^2 > \chi^2_{1-\frac{\alpha}{2},n-1}\right) - P\left(\chi^2 > \chi^2_{\frac{\alpha}{2},n-1}\right) =$$

$$= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha = \gamma.$$

Таким образом, с вероятностью γ верно

$$\chi^2_{1-\frac{\alpha}{2},n-1} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2},n-1},$$

откуда следует

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}.$$

Можно также по выборке $x_1, x_2, ..., x_n$ построить доверительный интервал для следующего, $(n + 1)$-го наблюдения (т.е. определить границы, в которых оно будет лежать с заданной вероятностью γ). А именно имеем

$$\bar{x} - st_\gamma\sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_\gamma\sqrt{1 + \frac{1}{n}}.$$

Понятно, что это может быть полезно в качестве прогноза на будущее.

Доказательство. Заметим, что случайные величины x_{n+1} и $\bar{x}$ независимы, причем $x_{n+1} \in N(a, \sigma^2)$ и $\bar{x} \in N(a, \sigma^2/n)$, следовательно, $x_{n+1} - \bar{x} \in N\left(0, \left(1 + \frac{1}{n}\right)\sigma^2\right)$. Тогда величина $\dfrac{x_{n+1} - \bar{x}}{\sigma\sqrt{1 + \frac{1}{n}}}$ имеет стандартное нормальное распределение $N(0, 1)$.

Введем статистику

$$T^* = \frac{\dfrac{x_{n+1} - \bar{x}}{\sigma\sqrt{1 + \frac{1}{n}}}}{\sqrt{\dfrac{(n-1)s^2}{(n-1)\sigma^2}}}.$$

Поскольку $\dfrac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$, то статистика T^* имеет распределение Стьюдента с $n - 1$ степенями свободы. Рассуждая, как ранее в доказательстве 2, получаем, что с вероятностью γ верно $\left|\dfrac{x_{n+1} - \bar{x}}{s\sqrt{1 + \frac{1}{n}}}\right| < t_\gamma$, откуда следует

$$\bar{x} - st_\gamma\sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_\gamma\sqrt{1 + \frac{1}{n}}.$$

Пример 20.1. Найти минимальный объем выборки, при котором с надежностью 0,925 точность оценки математического ожидания нормально распределенной случайной величины (по выборочному среднему $\bar{x}$) равна $\delta = 0,2$, если известно среднее квадратическое отклонение $\sigma = 1,5$.

Решение. Формула, определяющая точность оценки математического ожидания по выборочному среднему, выглядит следующим образом: $\delta = \dfrac{u_\gamma\sigma}{\sqrt{n}}$. Отсюда следует $n = \dfrac{u_\gamma^2\sigma^2}{\delta^2}$ (при этом обычно n округляют в бóльшую сторону для надежности). По таблице функции Лапласа находим u_γ для данного примера, учитывая, что функция принимает значение $\Phi_0(u_\gamma) = 0{,}925/2 = 0{,}4625$. Таким образом, $u_\gamma = 1{,}78$. Подставив данные задачи, получим искомый объем выборки:

$$n = \frac{1{,}78^2 \cdot 1{,}5^2}{0{,}2^2} \approx 178{,}22.$$

Берем округленно $n = 179$.

Пример 20.2. Из генеральной совокупности извлечена выборка объема $n = 12$:

Варианта x_i	–0,5	–0,4	–0,2	0	0,2	0,6	0,8	1	1,2	1,5
Частота n_i	1	2	1	1	1	1	1	1	2	1

Оценить с надежностью 0,95 математическое ожидание a нормально распределенной случайной величины с помощью доверительного интервала.

Решение. Найдем выборочное среднее $\bar{x}$ и исправленное выборочное среднее квадратическое отклонение s. Пусть $u_i = 10x_i$, тогда

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{10} n_i x_i = 4{,}2; \quad \bar{x} = \frac{\bar{u}}{10} = 0{,}42;$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{10} n_i x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{10} n_i x_i\right)\right] \approx 0{,}52.$$

Находим для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $n - 1 = 11$ по таблице распределения Стьюдента критическую точку $t_\gamma = 2{,}23$ и определяем границы доверительного интервала:

$$\bar{x} - \frac{t_\gamma s}{\sqrt{n}} = 0{,}42 - \frac{2{,}23 \cdot 0{,}72}{\sqrt{12}} = -0{,}04;$$

$$\bar{x} + \frac{t_\gamma s}{\sqrt{n}} = 0{,}42 + \frac{2{,}23 \cdot 0{,}72}{\sqrt{12}} = 0{,}88.$$

Таким образом, искомый доверительный интервал $-0{,}04 < a < 0{,}88$.

Пример 20.3. Для отрасли, включающей 1200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что исправленное среднее квадратическое отклонение для числа работающих на фирме составляет $s = 25$ (человек). Построить 90%-ный доверительный интервал для среднего квадратического отклонения числа работающих на фирме по всей отрасли.

Решение. Доверительный интервал для параметра σ имеет вид

$$s\sqrt{\frac{(n-1)}{\chi^2_{\frac{\alpha}{2},n-1}}} < \sigma < s\sqrt{\frac{(n-1)}{\chi^2_{1-\frac{\alpha}{2},n-1}}},$$

где $\chi^2_{\frac{\alpha}{2},n-1}$ и $\chi^2_{1-\frac{\alpha}{2},n-1}$ находят по таблице критических точек распределения хи-квадрат. По таблице определяем для данного примера $\chi^2_{0,05;18} = 28,9$; $\chi^2_{0,95;18} = 9,39$. Подставив в формулу необходимые величины, получаем искомый доверительный интервал:

$$25\sqrt{18/28,9} < \sigma < 25\sqrt{18/9,39},\text{ откуда } 19,74 < \sigma < 34,61 \text{ (человек).}$$

Пример 20.4. За последние пять лет годовой рост цены актива *A* составлял в среднем 20% со средним квадратическим отклонением (исправленным) 5%. Построить доверительный интервал с вероятностью 95% для цены актива в конце следующего года, если в начале года она равна 100 д.е.

Решение. Рассмотрим величины относительного прироста цены актива за год. Будем пользоваться нормальным приближением[1]. Применяем формулу

$$\bar{x} - st_\gamma\sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_\gamma\sqrt{1 + \frac{1}{n}},$$

где t_γ находим из таблицы критических точек распределения Стьюдента (для двусторонней области): $t_\gamma = t_{кр}(0,05;\ 4) = 2,78$.

Получаем $0,2 - 0,05\cdot 2,78\sqrt{1,2} < x_6 < 0,2 + 0,05\cdot 2,78\sqrt{1,2}$, откуда $0,05 < x_6 < 0,35$. Таким образом, цена актива в следующем году составит от 105 до 135 д.е.

Помимо случаев построения доверительных интервалов для параметров одной выборки, иногда рассматривают и случай двух выборок. Например, когда имеются две выборки $x_1, x_2, \ldots, x_n$ и $y_1, y_2, \ldots, y_m$ из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$ соответственно и надо построить доверительный интервал для разности средних.

1. При известных дисперсиях:

$$\bar{x} - \bar{y} - \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}u_\gamma < a_1 - a_2 < \bar{x} - \bar{y} + \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

Доказательство основано на следствии 3 теоремы Фишера (см. §12.8).

2. При неизвестных, но равных дисперсиях:

$$\bar{x} - \bar{y} - \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}\sqrt{\frac{n+m}{nm}}t_\gamma < a_1 - a_2 <$$

$$< \bar{x} - \bar{y} + \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}\sqrt{\frac{n+m}{nm}}t_\gamma,$$

[1] Такое приближение и соответствующие оценки являются довольно грубыми. На практике распределение относительного прироста цены обычно далеко от нормального и, к сожалению, не описывается ни одной из классических формул.

где t_γ — критическая точка распределения Стьюдента (для двусторонней области) с $n+m-2$ степенями свободы и уровнем значимости $\alpha = 1-\gamma$.

Доказательство основано на следствии 4 теоремы Фишера.

20.3. АСИМПТОТИЧЕСКИЕ ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

Асимптотическим доверительным интервалом при оценивании параметра θ называется такой интервал (θ_1, θ_2), что $P(\theta_1 < \theta < \theta_2) \to \gamma$ (при $n \to \infty$).

Предположим, мы решили учитывать тот факт, что наши наблюдения имеют распределение, отличное от нормального. Пусть это распределение зависит только от одного параметра θ, для которого надо построить доверительный интервал, а также известно, что оценка $\hat{\theta}$ асимптотически нормальна и верно $\hat{\theta} \xrightarrow{\in} N(\theta, \sigma^2(\theta)/n)$ при $n \to \infty$. Приведем два метода построения асимптотических доверительных интервалов.

1. Подстановка оценки параметра в формулу для дисперсии. Получаем

$$\hat{\theta} - \frac{\sigma(\hat{\theta})}{\sqrt{n}} u_\gamma < \theta < \hat{\theta} + \frac{\sigma(\hat{\theta})}{\sqrt{n}} u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

2. Использование функционального преобразования. Определим функцию $g(u) = \int \frac{du}{\sigma(u)}$, тогда верно $g(\hat{\theta}) \xrightarrow{\in} N(g(\theta), 1/n)$ при $n \to \infty$.

Следовательно, можно использовать асимптотическое неравенство

$$g(\hat{\theta}) - \frac{u_\gamma}{\sqrt{n}} < g(\theta) < g(\hat{\theta}) + \frac{u_\gamma}{\sqrt{n}}.$$

Решая его относительно θ, получаем асимптотический доверительный интервал.

Например, доверительный интервал для вероятности «успеха» p в n испытаниях Бернулли обычно строится первым методом. Поскольку $M\xi = p$ и $D\xi = p(1-p)$, то из ЦПТ получаем $w \xrightarrow{\in} N(p, p(1-p)/n)$, и доверительный интервал имеет вид

$$w - \sqrt{\frac{w(1-w)}{n}} u_\gamma < p < w + \sqrt{\frac{w(1-w)}{n}} u_\gamma,$$

где w — относительная частота события.

Асимптотические доверительные интервалы рекомендуется применять, когда объем выборки достаточно велик (порядка сотен и более).

Доказательство (обоснование метода функционального преобразования). Если оценка $\hat{\theta}$ асимптотически нормальна и верно $\hat{\theta} \xrightarrow{\in} N(\theta, \sigma^2(\theta)/n)$, то для любой функции g, гладкой в окрестности θ, верно $g(\hat{\theta}) \xrightarrow{\in} N(g(\theta), (g'(\theta))^2\sigma^2(\theta)/n)$.

Чтобы дисперсия не зависела от θ, необходимо функцию g подобрать так, чтобы выражение $\sigma(\theta)g'(\theta)$ было постоянным, например положив $\sigma(\theta)g'(\theta) = 1$.

Задача свелась к выбору такой функции g, чтобы она являлась решением дифференциального уравнения $g'(\theta) = 1/\sigma(\theta)$. Решение этого уравнения имеет вид $g(\theta) = \int \frac{d\theta}{\sigma(\theta)}$, при этом произвольная постоянная в неопределенном интеграле выбирается из соображений простоты окончательного выражения и обычно полагается равной нулю. Тогда получается $g(\hat{\theta}) \xrightarrow{\in} N(g(\theta), 1/n)$ при $n \to \infty$.

Задача свелась к построению доверительного интервала для математического ожидания нормально распределенной случайной величины при известной дисперсии с заданной надежностью γ. Согласно ранее доказанному такой интервал имеет вид

$$g(\hat{\theta}_n) - \frac{1}{\sqrt{n}}u_\gamma < g(\theta) < g(\hat{\theta}_n) + \frac{1}{\sqrt{n}}u_\gamma.$$

Пример 20.5. Произведено 300 испытаний, в каждом из которых неизвестная вероятность p появления события A постоянна. Событие A появилось в 250 испытаниях. Найти доверительный интервал, покрывающий неизвестную вероятность p с надежностью 0,95.

Решение. Число испытаний $n = 300$ достаточно велико, поэтому можем воспользоваться следующими формулами для границ доверительного интервала:

$$p_1 = w - \sqrt{\frac{w(1-w)}{n}}u_\gamma; \quad p_2 = w + \sqrt{\frac{w(1-w)}{n}}u_\gamma.$$

Значение u_γ находим из соотношения $\Phi_0(u_\gamma) = \gamma/2 = 0{,}475$ по таблице функции Лапласа, в данном случае $u_\gamma = 1{,}96$. Относительная частота события A составляет $w = 250/300 \approx 0{,}83$. Подставим это значение w в формулы для p_1 и p_2:

$$p_1 = 0{,}83 - 1{,}96\sqrt{\frac{0{,}83 \cdot 0{,}17}{300}} \approx 0{,}79; \quad p_2 = 0{,}83 + 1{,}96\sqrt{\frac{0{,}83 \cdot 0{,}17}{300}} \approx 0{,}87.$$

Итак, получаем искомый доверительный интервал: $0{,}79 < p < 0{,}87$.

Пример 20.6. Построить доверительный интервал для вероятности «успеха» в испытаниях Бернулли методом функционального преобразования.

Решение. Необходимо построить функцию $g(u) = \int \frac{du}{\sigma(u)}$, где $\sigma(p) = \sqrt{p(1-p)}$.

Для рассматриваемого случая имеем

$$g(p) = \int \frac{dp}{\sqrt{p(1-p)}} = \left| p = t^2, dp = 2tdt \right| = 2\int \frac{tdt}{\sqrt{t^2(1-t^2)}} =$$

$$= 2\int \frac{dt}{\sqrt{1-t^2}} = 2\arcsin t = 2\arcsin\sqrt{p}.$$

С учетом вида функции $g(p)$ асимптотическое неравенство примет вид

$$2\arcsin\sqrt{w} - \frac{u_\gamma}{\sqrt{n}} < 2\arcsin\sqrt{p} < 2\arcsin\sqrt{w} + \frac{u_\gamma}{\sqrt{n}}.$$

Для функции $y = g(x) = 2\arcsin\sqrt{x}$ при $0 \le x \le 1$ обратной функцией будет $x = g^{-1}(y) = \sin^2(y/2)$, где $0 \le y \le \pi$. Поэтому, если выполняются неравенства

$$2\arcsin\sqrt{w} - \frac{u_\gamma}{\sqrt{n}} \ge 0, \qquad 2\arcsin\sqrt{w} + \frac{u_\gamma}{\sqrt{n}} \le \pi,$$

то, применив обратное преобразование g^{-1}, получим доверительный интервал:

$$\sin^2\left(\arcsin\sqrt{w} - \frac{u_\gamma}{2\sqrt{n}}\right) < p < \sin^2\left(\arcsin\sqrt{w} + \frac{u_\gamma}{2\sqrt{n}}\right).$$

З а м е ч а н и е. Очевидно, первым способом получается гораздо более простой и удобный в применении доверительный интервал, хотя и менее точный.

Пример 20.7. Построить асимптотический доверительный интервал для параметра λ показательного закона распределения (двумя способами).

Решение. В данном случае удобно перейти к новому параметру $\theta = 1/\lambda$. Эффективной оценкой для него является выборочное среднее: $\hat{\theta} = \bar{x}$. Имеем $M\hat{\theta} = \theta$, $D\hat{\theta} = \theta^2/n$, откуда $\sigma(\theta) = \theta$.

Первым способом, подставляя оценку в формулу для дисперсии, получаем

$$\bar{x} - \frac{\bar{x}}{\sqrt{n}}u_\gamma < \theta < \bar{x} + \frac{\bar{x}}{\sqrt{n}}u_\gamma.$$

Возвращаясь к параметру $\lambda = 1/\theta$, приходим к неравенству

$$\frac{1}{\bar{x}}\left(1+\frac{u_\gamma}{\sqrt{n}}\right)^{-1} < \lambda < \frac{1}{\bar{x}}\left(1-\frac{u_\gamma}{\sqrt{n}}\right)^{-1}.$$

Вторым способом, определив функцию $g(u) = \int \frac{du}{u} = \ln u$, получаем асимптотическое неравенство

$$\ln \bar{x} - \frac{u_\gamma}{\sqrt{n}} < \ln\theta < \ln\bar{x} + \frac{u_\gamma}{\sqrt{n}},$$

откуда следует

$$\bar{x}\exp\left(-\frac{u_\gamma}{\sqrt{n}}\right) < \theta < \bar{x}\exp\left(\frac{u_\gamma}{\sqrt{n}}\right)$$

и

$$\frac{1}{\bar{x}}\exp\left(-\frac{u_\gamma}{\sqrt{n}}\right) < \lambda < \frac{1}{\bar{x}}\exp\left(\frac{u_\gamma}{\sqrt{n}}\right).$$

Заметим, что по ходу решения задачи мы получили доверительные интервалы для математического ожидания θ показательного распределения. Данный результат имеет и самостоятельную ценность.

В дальнейших задачах асимптотические интервалы по умолчанию строятся методом функционального преобразования (кроме задач на вероятности «успехов»).

20.4. ИНТЕРВАЛЬНАЯ ОЦЕНКА КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Пусть $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ — независимые наблюдения над двумерной нормальной случайной величиной. Построим асимптотический доверительный интервал для коэффициента корреляции ρ, соответствующий надежности γ.

Точечной оценкой методом моментов для коэффициента корреляции является выборочный коэффициент корреляции

$$\hat{\rho}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]}}.$$

Известно, что оценка асимптотически нормальна со следующими параметрами:

$$\hat{\rho}_n \xrightarrow{\in} N\left(\rho, \frac{(1-\rho^2)^2}{n}\right), \quad n\to\infty.$$

Отсюда следует, что $\sigma(\rho) = 1 - \rho^2$, и

$$g(\rho) = \int \frac{d\rho}{1-\rho^2} = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \operatorname{arth}\rho.$$

Получено так называемое *z-преобразование Фишера* для коэффициента корреляции. Это преобразование хорошо исследовано, и для него известны следующие соотношения:

$$M(\operatorname{arth}\hat{\rho}_n) \approx \operatorname{arth}\rho + \frac{\rho}{2(n-1)}; \quad D(\operatorname{arth}\hat{\rho}_n) = \frac{1}{n-3}.$$

Построим доверительный интервал для $\operatorname{arth}\rho$. Заменив n на $n-3$ и пренебрегая в математическом ожидании при достаточно большом n величиной $\frac{\rho}{2(n-1)}$, получаем

$$P\left(\operatorname{arth}\hat{\rho}_n - \frac{1}{\sqrt{n-3}}u_\gamma < \operatorname{arth}\rho < \operatorname{arth}\hat{\rho}_n + \frac{1}{\sqrt{n-3}}u_\gamma\right) = \gamma.$$

Функция $\operatorname{th}x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (гиперболический тангенс), обратная функции $\operatorname{arth}x$, существует, однозначна и монотонно возрастает, поэтому после преобразования доверительный интервал имеет вид

$$\operatorname{th}\left(\operatorname{arth}\hat{\rho}_n - \frac{1}{\sqrt{n-3}}u_\gamma\right) < \rho < \operatorname{th}\left(\operatorname{arth}\hat{\rho}_n + \frac{1}{\sqrt{n-3}}u_\gamma\right),$$

где u_γ определяется соотношением $\Phi_0(u_\gamma) = \gamma/2$.

Найденный приближенный доверительный интервал настолько мало отличается от истинного, что может применяться уже для выборок объема $n \geq 10$.

Глава 21
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

21.1. СТАТИСТИЧЕСКАЯ ГИПОТЕЗА. КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ

Статистической гипотезой называется любое предположение относительно генеральной совокупности. Гипотеза называется *параметрической*, если в ней содержится некоторое утверждение о параметрах распределения случайной величины (когда сам закон распределения считается известным), и *непараметрической* в иных случаях. В этой главе мы будем иметь дело с параметрическими гипотезами.

Нулевой (*основной*) гипотезой H_0 называется первоначальное предположение относительно генеральной совокупности, т.е. гипотеза, которая утверждает, что различия между сравниваемыми величинами отсутствует, а наблюдаемые отклонения объясняются случайными колебаниями выборки.

Альтернативной (*конкурирующей*) гипотезой H_1 называется гипотеза, которая противоречит основной гипотезе H_0 и которую мы принимаем, если отвергаем основную гипотезу.

Случайная величина K, построенная по наблюдениям для проверки нулевой гипотезы, называется *статистикой критерия*.

Схема построения критерия такова: все выборочное пространство делится на две взаимодополняющие области — область S отклонения основной гипотезы H_0 и область $\bar{S}$ принятия этой гипотезы. Область S, при попадании в которую выборочной точки $x = (x_1, x_2, ..., x_n)$ основная гипотеза отклоняется, называется *критической*. Область $\bar{S}$ называют *областью допустимых значений*.

При проверке гипотез могут быть ошибки двух типов.

Ошибка первого рода состоит в том, что основная гипотеза отвергается, хотя на самом деле она верна. Ее вероятность обозначают обычно α.

Ошибка второго рода состоит в том, что основная гипотеза принимается, хотя на самом деле она неверна. Ее вероятность обозначают обычно β.

Часто вероятности ошибок называют просто ошибками (первого и второго рода) для краткости.

Вероятность α совершить ошибку первого рода называют также *уровнем значимости* или *размером* критерия. Вероятность $1 - \beta$ *не* совершить ошибку второго рода называют *мощностью* критерия.

Критерий называется ***наиболее мощным***, если из всех возможных критериев с заданным уровнем значимости a он обладает наибольшей мощностью.

Заметим, что можно установить связь между задачами проверки гипотез и задачами построения доверительных интервалов (см. гл. 14). Например, пусть построен доверительный интервал $(\hat{\theta}_1, \hat{\theta}_2)$ для параметра θ с надежностью γ. Тем самым утверждается, что истинное значение параметра лежит в интервале с вероятностью γ, а вне этого интервала — с малой вероятностью $\alpha = 1 - \gamma$. Таким образом, если мы проверяем гипотезу H_0: $\theta = \theta_0$ против какой-либо альтернативной гипотезы, то в качестве критерия можно взять $S = \{x: \theta_0 \notin (\hat{\theta}_1, \hat{\theta}_2)\}$. Уровень значимости в этом случае будет равен α. Однако такой критерий не обязательно будет оптимальным.

Основным методом построения наиболее мощных статистических критериев (по теореме Неймана—Пирсона) является *метод отношения правдоподобия*, суть которого заключается в следующем: пусть ξ — непрерывная случайная величина, имеющая плотность распределения $f_0(x)$ при условии истинности нулевой гипотезы H_0 и $f_1(x)$ при выполнении гипотезы H_1. Функции правдоподобия в точке x соответственно равны:

$$L_0(x) = f_0(x_1)f_0(x_2)...f_0(x_n); \qquad L_1(x) = f_1(x_1)f_1(x_2)...f_1(x_n).$$

О предпочтительности гипотезы H_0 перед H_1 можно судить по отношению правдоподобия L_1/L_0 ($L_0 \neq 0$): чем правдоподобнее выборка в условиях истинности гипотезы H_0 по сравнению с H_1, тем меньше L_1 по сравнению с L_0 и тем меньше отношение L_1/L_0.

Теорема 20.1 (Неймана—Пирсона). Критическая область S наиболее мощного критерия имеет вид

$$S = \left\{x: L_0(x) = 0 \cup \frac{L_1(x)}{L_0(x)} > C, L_0(x) \neq 0\right\},$$

где константа $C = C(\alpha)$ является решением уравнения[1]

$$P\left(\frac{L_1(x)}{L_0(x)} > C \mid H_0\right) = \alpha.$$

Подобный метод построения критической области, использующий отношение правдоподобия, дает нам *критерий отношения*

[1] Бывает, что данное уравнение не имеет решения; и в таких случаях используются так называемые рандомизированные критерии, которые мы здесь рассматривать не будем.

правдоподобия. В дискретном случае построение проводится аналогично (только вместо плотностей берутся вероятности).

Статистика критерия здесь имеет вид $K = L_1/L_0$ (при $L_0 = 0$ полагаем $K = +\infty$), тогда критическая область $S = \{K > C\}$, а область допустимых значений $\bar{S} = \{K \le C\}$.

Пример 21.1. Пусть случайная величина $\xi \in N(a, \sigma^2)$, причем значение параметра a неизвестно, а дисперсия σ^2 известна. Требуется на уровне значимости α проверить нулевую гипотезу H_0: $a = a_0$, если альтернативная гипотеза H_1: $a = a_1 > a_0$. Построить критерий отношения правдоподобия. Вычислить объем выборки n, необходимый для достижения ошибок второго рода α и β.

Решение. Если верна гипотеза H_0, т.е. $\xi \in N(a_0, \sigma^2)$, то функция правдоподобия в точке $x = (x_1, x_2, ..., x_n)$ равна

$$L_0(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum\limits_{i=1}^{n}(x_i - a_0)^2}{2\sigma^2}\right).$$

Если же верна гипотеза H_1, т.е. $\xi \in N(a_1, \sigma^2)$, то функция правдоподобия равна

$$L_1(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum\limits_{i=1}^{n}(x_i - a_1)^2}{2\sigma^2}\right).$$

Отношение правдоподобия имеет вид

$$\frac{L_1}{L_0} = \exp\left(\frac{(a_1 - a_0)(2\bar{x} - a_1 - a_0)n}{2\sigma^2}\right).$$

Так как $a_1 > a_0$, то это отношение является монотонно возрастающей функцией от $\bar{x}$, и поскольку $L_0(x) \neq 0$, то неравенство $L_1/L_0 > C$ равносильно неравенству $\bar{x} > \bar{C}$, где C и $\bar{C}$ — некоторые константы. Поэтому критическая область имеет вид $S = \{x: \bar{x} > \bar{C}\}$, где $P(\bar{x} > \bar{C} \mid H_0) = \alpha$.

При условии истинности нулевой гипотезы H_0 имеем $\bar{x} \in N\left(a_0, \frac{\sigma^2}{n}\right)$, поэтому

$$\alpha = P(\bar{x} > \bar{C} \mid H_0) = 1 - \Phi\left(\frac{\bar{C} - a_0}{\sigma}\sqrt{n}\right) = \frac{1}{2} - \Phi_0\left(\frac{\bar{C} - a_0}{\sigma}\sqrt{n}\right).$$

Отсюда $\Phi_0\left(\frac{\bar{C} - a_0}{\sigma}\sqrt{n}\right) = \frac{1}{2} - \alpha.$

Обозначим через u_α решение уравнения $\Phi_0(u_\alpha) = 1/2 - \alpha$, тогда $\bar{C} = a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$. Заметим, что величина u_α является квантилью уровня $1 - \alpha$ для стандартного нормального распределения и выступает здесь в качестве *критической точки*. Значение ее можно найти по таблице функции Лапласа.

Итак, наиболее мощным критерием проверки гипотезы H_0: $a = a_0$ при альтернативной H_1: $a = a_1 > a_0$ оказывается следующий:

- если $\bar{x} < a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$, то H_0 принимается;
- если $\bar{x} > a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$, то H_0 отклоняется (и принимается H_1).

По определению ошибки второго рода имеем

$$\beta = P(\bar{x} \le \bar{C} \mid H_1) = \Phi\left(\frac{\bar{C} - a_1}{\sigma}\sqrt{n}\right).$$

Отсюда $\Phi\left(\frac{a_1 - \bar{C}}{\sigma}\sqrt{n}\right) = 1 - \beta.$

Получаем, что должно выполняться

$$\bar{C} = a_1 - u_\beta \frac{\sigma}{\sqrt{n}} = a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}.$$

Решая уравнение относительно n, получаем

$$n = \frac{(u_\alpha + u_\beta)^2}{(a_1 - a_0)^2}\sigma^2.$$

Полученное значение обычно округляется до целого в большую сторону для уменьшения вероятностей ошибок. Мощность критерия в данном случае равняется

$$1 - \beta = \Phi\left(\frac{a_1 - \bar{C}}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{a_1 - a_0}{\sigma}\sqrt{n} - u_\alpha\right).$$

Пример 21.2. Крупная торговая фирма желает открыть в новом районе города филиал. Известно, что фирма будет работать прибыльно, если средний месячный доход жителей района превышает 400 долл. Также известно, что среднее квадратическое отклонение дохода σ составляет 20 долл. Проводится выборочное обследование населения по величине доходов, чтобы принять решение об открытии филиала.

1. Определить правило принятия решения, с помощью которого, основываясь на выборке $n = 100$ (человек) и уровне значимости $\alpha = 0{,}05$, можно установить, что филиал будет работать прибыльно.

2. Рассчитать вероятность того, что при применении правила принятия решения, полученного при ответе на вопрос пункта 1, будет совершена ошибка второго рода, если в действительности средний доход достигает 406 долл.

3. Считая альтернативное значение среднего месячного дохода жителей равным 410 долл., рассчитать объем выборки, при котором ошибка первого рода не превысит 2,5%, а ошибка второго рода не превысит 5%.

Решение.

1. Фирма не откроет филиал, если средний доход жителей не превысит 400 долл. Поэтому будем считать, что H_0: $a = a_0 = 400$, а H_1: $a = a_1 > 400$. Значение дисперсии σ^2 дохода известно. Находим u_α по таблице функции Лапласа исходя из равенства $\Phi_0(u_\alpha) = 1/2 - \alpha$. Так как $a_0 = 400$ и $u_\alpha = 1{,}65$, то H_1 принимают и, следовательно, филиал открывают, если средний месячный доход 100 жителей

$$\bar{x} > 400 + \frac{20}{\sqrt{100}} \cdot 1{,}65 = 403{,}3.$$

2. Вероятность ошибки второго рода β можно легко найти, зная мощность критерия, которая равна $1 - \beta$. Получаем

$$1 - \beta = \Phi\left(\frac{406 - 400}{20}\sqrt{100} - 1{,}65\right) = \Phi(1{,}35) = \frac{1}{2} + \Phi_0(1{,}35) \approx 0{,}91.$$

Отсюда вероятность ошибки второго рода $\beta \approx 0{,}09$.

3. Используем формулу для необходимого объема выборки. Для этого найдем квантили u_α и u_β с помощью таблицы функции Лапласа и получим

$$n \geq \frac{(1{,}95 + 1{,}65)^2}{100} \cdot 20^2 = 51{,}84, \text{ откуда } n = 52.$$

З а м е ч а н и е. Приведенная формула может давать слишком малые значения n, так как основывается на строго нормальном распределении. На практике n должно быть достаточно велико, чтобы пользоваться асимптотической нормальностью оценок.

21.2. ПРОВЕРКА ГИПОТЕЗ ДЛЯ ОДНОЙ ВЫБОРКИ

Рассмотрим простые методы проверки параметрических гипотез в случае нормального распределения (которые являются формально точными), а также гипотезы о вероятности «успеха» в испытаниях Бернулли (на основе асимптотической нормальности). Как и ранее, нас не будет смущать тот факт, что реальные данные, на основании которых проверяются гипотезы, могут совсем не выглядеть нормальными (например, это целые положительные числа, в то время как нормальное распределение непрерывно и рассредоточено по всей действительной прямой). Тем не менее широкое практическое применение описываемых методов дает неплохие результаты (это объясняется, в частности, асимптотической нормальностью оценок).

Следующие три типа гипотез проверяются для нормальных данных: $\xi \in N(a, \sigma^2)$.

1. **Гипотезы о неизвестном среднем a при известной дисперсии σ^2.**

Основная гипотеза H_0: $a = a_0$, альтернативная гипотеза H_1 может быть трех видов: а) $a \neq a_0$; б) $a > a_0$; в) $a < a_0$. Во всех трех случаях для проверки используется статистика критерия

$$U = \frac{\bar{x} - a_0}{\sigma}\sqrt{n}.$$

В случае а) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = (1 - \alpha)/2$. Если $|U| < u_{кр}$, гипотеза H_0 принимается; если $|U| > u_{кр}$ — отвергается. Таким образом, в данном случае имеет место *двустороння критическая область*.

В случаях б) и в) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = 1/2 - a$.

В случае б), если $U < u_{кр}$, гипотеза H_0 принимается; если $U > u_{кр}$ — отвергается.

В случае в), если $U > -u_{кр}$, гипотеза H_0 принимается, если $U < -u_{кр}$ — отвергается.

Здесь имеют место *односторонние критические области* (*правосторонняя* и *левосторонняя* соответственно).

Замечание. Этим методом можно пользоваться и в случае неизвестной дисперсии при больших объемах выборки (порядка сотен), когда оценку дисперсии можно принять за ее точное значение.

2. Гипотезы о неизвестном среднем a при неизвестной дисперсии σ^2.

Основная гипотеза H_0: $a = a_0$, альтернативная гипотеза H_1 может быть трех видов: а) $a \neq a_0$; б) $a > a_0$; в) $a < a_0$. Во всех трех случаях для проверки используется статистика критерия

$$T = \frac{\bar{x} - a_0}{s}\sqrt{n}.$$

Для проверки берутся критические точки $t_{кр}$ распределения Стьюдента с $n - 1$ степенями свободы и уровнем значимости α, причем в случае а) — для *двусторонней* критической области, в случаях б) и в) — для *односторонней* критической области.

В случае а), если $|T| < t_{кр}$, гипотеза H_0 принимается; если $|T| > t_{кр}$ — отвергается.

В случае б), если $T < t_{кр}$, гипотеза H_0 принимается, если $T > t_{кр}$ — отвергается.

В случае в), если $T > -u_{кр}$, гипотеза H_0 принимается, если $T < -u_{кр}$ — отвергается.

3. Гипотезы о неизвестной дисперсии σ^2.

Обычно предполагается, что хотя дисперсия неизвестна, но дана ее несмещенная оценка s^2. Основная гипотеза H_0: $\sigma^2 = \sigma_0^2$, альтернативная гипотеза H_1 может быть трех видов: а) $\sigma^2 \neq \sigma_0^2$; б) $\sigma^2 > \sigma_0^2$;

в) $\sigma^2 < \sigma_0^2$. Во всех трех случаях для проверки используется статистика критерия

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}.$$

Для проверки берутся критические точки распределения хи-квадрат с $n-1$ степенями свободы (и различными уровнями значимости).

В случае а), если $\chi^2_{1-\alpha/2;n-1} < \chi^2 < \chi^2_{\alpha/2;n-1}$, гипотеза H_0 принимается; иначе она отвергается. Заметим, что границы области *несимметричны* относительно оценки s^2.

В случае б), если $\chi^2 < \chi^2_{\alpha;n-1}$, гипотеза H_0 принимается; иначе отвергается.

В случае в), если $\chi^2 > \chi^2_{1-\alpha;n-1}$, гипотеза H_0 принимается; иначе отвергается.

Следующая гипотеза проверяется приближенно на основе асимптотической нормальности оценки.

4. **Гипотеза о неизвестной вероятности «успеха» в испытаниях Бернулли.**

Основная гипотеза H_0: $p = p_0$, альтернативная гипотеза H_1 может быть трех видов: а) $p \neq p_0$; б) $p > p_0$; в) $p < p_0$. Во всех трех случаях для проверки используется статистика критерия

$$U = \frac{w - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n},$$

где w — относительная частота «успехов» в n наблюдениях.

Далее критические точки и области для проверки выбираются так же, как при проверке гипотезы о неизвестном среднем при известной дисперсии.

Замечание. Этим методом можно пользоваться только при больших объемах выборки (порядка нескольких десятков или сотен).

Описанные выше критерии проверки гипотез можно представить в виде следующей таблицы (табл. 21.1).

Пример 21.3. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 5$ извлечена выборка объема $n = 100$ и по ней найдено выборочное среднее 26,5. Требуется на уровне значимости 0,05 проверить гипотезу H_0: $a = a_0 = 25$ против альтернативной гипотезы H_1: $a \neq 25$. Изменится ли результат, если мы заменим альтернативную гипотезу на H_1: $a > 25$?

Решение. Найдем значение статистики критерия:

Таблица 21.1

H_0	Предположение	Статистика критерия	H_1	Область принятия H_0
$a = a_0$	σ^2 известно	$U = \frac{\bar{x} - a_0}{\sigma}\sqrt{n}$	$a > a_0$	$U < u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$a < a_0$	$U > -u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$a \neq a_0$	$\lvert U\rvert < u_{кр}$, $\Phi_0(u_{кр}) = (1 - \alpha)/2$
$a = a_0$	σ^2 неизвестно	$T = \frac{\bar{x} - a_0}{s}\sqrt{n}$	$a > a_0$	$T < t_{кр}(\alpha, n-1)$ для односторонней области
			$a < a_0$	$T > -t_{кр}(\alpha, n-1)$ для односторонней области
			$a \neq a_0$	$\lvert T\rvert < t_{кр}(\alpha, n-1)$ для двусторонней области
$\sigma^2 = \sigma_0^2$	a неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$	$\chi^2 < \chi^2_{\alpha;n-1}$
			$\sigma^2 < \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha;n-1}$
			$\sigma^2 \neq \sigma_0^2$	$\chi^2_{1-\alpha/2;n-1} < \chi^2 < \chi^2_{\alpha/2;n-1}$
$p = p_0$	n порядка нескольких десятков или сотен	$U = \frac{w - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$, где $w = m/n$	$p > p_0$	$U < u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$p < p_0$	$U > -u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$p \neq p_0$	$\lvert U\rvert < u_{кр}$, $\Phi_0(u_{кр}) = (1 - \alpha)/2$

$$U = \frac{26{,}5 - 25}{5}\sqrt{100} = 3.$$

При проверке гипотезы H_1: $a \neq 25$ из соотношения $\Phi_0(u_{кр}) = 0{,}95/2 = 0{,}475$ находим $u_{кр} = 1{,}96$, и $|U| > u_{кр}$, так что основная гипотеза отвергается. При проверке гипотезы H_1: $a > 25$ из соотношения $\Phi_0(u_{кр}) = 0{,}45$ находим $u_{кр} = 1{,}65$, и $U > u_{кр}$, так что основная гипотеза отвергается. В обоих случаях результат одинаков.

Пример 21.4. По выборке объема $n = 16$, извлеченной из нормальной генеральной совокупности, найдены выборочное среднее, равное 12,4, и исправленное среднее квадратическое отклонение, равное 1,2. Требуется при уровне значимости 0,05 проверить нулевую гипотезу H_0: $a = 11{,}8$ при конкурирующей гипотезе H_1: $a \neq 11{,}8$.

Решение. Найдем наблюдаемое значение статистики критерия:

$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{s} = \frac{(12{,}4 - 11{,}8)\cdot 4}{1{,}2} = 2.$$

Так как конкурирующая гипотеза имеет вид $a \neq a_0$, то искомая критическая область двусторонняя. По таблице критических точек распределения Стьюдента найдем по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $n - 1 = 15$ критическую точку $t_{кр} = t_{кр}(0{,}05;\ 15) = 2{,}13$.

В силу того что $|T| < t_{кр}$, у нас нет оснований отвергнуть нулевую гипотезу.

Пример 21.5. Точность работы станка-автомата проверяется по дисперсии размеров изделий, которая не должна превышать $\sigma_0^2 = 0{,}01$ (мм²). По выборке из 25 изделий получена исправленная выборочная дисперсия $\sigma^2 = 0{,}02$ (мм²). На уровне значимости 0,05 проверить, обеспечивает ли станок необходимую точность.

Решение. Найдем значение статистики критерия:

$$\chi^2 = \frac{24 \cdot 0{,}02}{0{,}01} = 48.$$

Альтернативной гипотезой в данном случае является H_1: $\sigma^2 > \sigma_0^2$.

По таблице находим критическую точку распределения хи-квадрат: $\chi^2_{0{,}05;24} = 36{,}4$.

Поскольку 48 > 36,4, то основная гипотеза отвергается. Следовательно, станок не обеспечивает необходимой точности.

Пример 21.6. Из нормальной генеральной совокупности извлечена выборка объема $n = 31$. В следующей таблице представлены сгруппированные данные:

Варианты x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
Частоты n_i	1	3	7	10	6	3	1

Требуется на уровне значимости 0,05 проверить нулевую гипотезу H_0: $\sigma^2 = 0{,}18$, приняв в качестве конкурирующей гипотезы H_1: $\sigma^2 > 0{,}18$.

Решение. Перейдем к условным вариантам $u_i = 10x_i - 110$.

u_i	−9	−7	−4	2	5	8	10
n_i	1	3	7	10	6	3	1

Получаем:

$$s_u^2 = \frac{1}{n-1}\left[\sum_{i=1}^{7} n_i u_i^2 - \frac{1}{n}\left(\sum_{i=1}^{7} n_i u_i\right)^2\right] = \frac{1}{30}\left(822 - \frac{26^2}{31}\right) \approx 26{,}67;$$

$$s_x^2 = \frac{s_u^2}{100} \approx 0{,}27; \quad \chi^2 = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{30 \cdot 0{,}27}{0{,}18} = 45.$$

По таблице критических точек распределения хи-квадрат по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = n - 1 = 30$ находим $\chi^2_{\text{кр}}(0{,}05;\ 30) = 43{,}8$.

Поскольку 45 > 43,8, основная гипотеза H_0 отвергается.

Пример 21.7. Партия изделий принимается, если доля брака составляет не более 2%. Среди случайно отобранных 500 изделий оказалось 13 бракованных. Следует ли принять партию (на уровне значимости 0,05)?

Решение. Относительная частота брака составляет $w = 13/500 = 0{,}026$. Найдем значение статистики критерия:

$$U = \frac{0,026 - 0,02}{\sqrt{0,02 \cdot 0,98}}\sqrt{500} \approx 0,96.$$

Альтернативной гипотезой в данном случае является H_1: $p > p_0$.

Из соотношения $\Phi_0(u_{кр}) = 0,45$ находим $u_{кр} = 1,65$ и получаем $U < u_{кр}$, так что основная гипотеза принимается. Таким образом, партию изделий можно принять.

Пример 21.8. Торговец утверждает, что он получает заказы в среднем по крайней мере от 30% предполагаемых клиентов. Можно ли при 5%-ном уровне значимости согласиться с этим утверждением, если торговец получил заказы от 20 из 100 случайно отобранных потенциальных клиентов?

Решение. В данном случае нулевая гипотеза будет выглядеть как H_0: $p = p_0 = 0,3$, а конкурирующая — как H_1: $p < 0,3$. Найдем значение статистики критерия, учитывая, что относительная частота для данной задачи равна $w = 20/100 = 0,2$:

$$U = \frac{(w - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} = \frac{(0,2 - 0,3) \cdot 10}{\sqrt{0,3 \cdot 0,7}} \approx -2,18.$$

Из соотношения $\Phi_0(u_{кр}) = 0,45$ находим $u_{кр} = 1,65$, и $U < -u_{кр}$, так что нулевая гипотеза отвергается и с утверждением торговца согласиться нельзя.

21.3. ПРОВЕРКА ГИПОТЕЗ ДЛЯ ДВУХ ВЫБОРОК

Зависимые выборки: парные наблюдения. Под случаем «зависимых выборок» обычно имеют в виду ситуацию, когда речь идет об одном и том же наборе объектов до и после какого-либо воздействия на них. Предполагается, что воздействие может повлиять на признаки, сдвинув их средние значения в большую или меньшую сторону, и это необходимо проверить.

Вначале признаки объектов принимают значения x_i, после воздействия — значения y_i. Такие наблюдения называются *парными*. Вычислим их разности $d_i = y_i - x_i$. Тогда ставится следующая задача: по наблюдениям $d_1, d_2, \ldots, d_n$ проверить гипотезу о равенстве нулю генерального среднего (H_0: $a_d = 0$) при неизвестной дисперсии σ_d^2. Предполагается, что случайные изменения признаков распределены нормально. Тогда гипотезу можно проверить, как это описано в предыдущем параграфе.

Пример 21.9. Физическая подготовка девяти спортсменов была проверена при поступлении в спортивную школу, а затем после недели тренировок. Итоги проверки в баллах оказались следующими (в первой строке указано число баллов, полученных каждым спортсменом при поступлении в школу; во второй строке — после обучения):

x_i	76	71	57	49	70	69	26	65	59
y_i	81	85	52	52	70	63	33	83	62

Требуется на уровне значимости 0,05 проверить, существенно ли улучшилась физическая подготовка спортсменов (используя нормальное приближение).

Решение. Вычислим разности $d_i = y_i - x_i$:

d_i	5	14	–5	3	0	-6	7	18	3

Получаем:

$$\sum_{i=1}^{9} d_i = 5 + 14 - 5 + 3 - 6 + 7 + 18 + 3 = 39;$$

$$\sum_{i=1}^{9} d_i^2 = 25 + 196 + 25 + 9 + 36 + 49 + 324 + 9 = 673.$$

Отсюда находим выборочное среднее и исправленное среднее квадратическое отклонение:

$$\bar{d} = \frac{1}{9}\sum_{i=1}^{9} d_i = \frac{39}{9} \approx 4{,}33;$$

$$s_d = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} d_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2\right]} = \sqrt{\frac{673 - 169}{8}} \approx 7{,}94.$$

Проверяем гипотезу H_0: $a_d = 0$ против H_1: $a_d > 0$.

Найдем значение статистики критерия:

$$T = \frac{\bar{d}\sqrt{n}}{s_d} = \frac{4{,}33 \cdot 3}{7{,}94} \approx 1{,}47.$$

По таблице критических точек распределения Стьюдента для односторонней области по уровню значимости 0,05 и числу степеней свободы $k = n - 1 = 8$ определяем $t_{кр}(0{,}05;\, 8) = 1{,}86$. Поскольку $T < t_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Таким образом, нельзя утверждать, что подготовка спортсменов существенно улучшилась.

Независимые выборки. Пусть имеются две независимые выборки: $x_1, x_2, \ldots, x_n$ и $y_1, y_2, \ldots, y_m$, имеющие нормальное распределение с параметрами (a_x, σ_x^2) и (a_y, σ_y^2) соответственно. Обычно ставится задача проверки их однородности, т.е. равенства обоих параметров, либо надо проверить равенство параметров по отдельности.

1. **Гипотеза о равенстве дисперсий двух выборок.**

Предположим, что известны исправленные выборочные дисперсии для обоих выборок — s_x^2 и s_y^2. Проверяем гипотезу H_0: $\sigma_x^2 = \sigma_y^2$. Альтернативная гипотеза H_1 может быть трех видов: а) $\sigma_x^2 \neq \sigma_y^2$; б) $\sigma_x^2 > \sigma_y^2$; в) $\sigma_x^2 < \sigma_y^2$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно.

В случае а) делят большую выборочную дисперсию на меньшую:

$$F = \frac{s_{\max}^2}{s_{\min}^2}.$$

Обозначим через $n_{\min}$ объем выборки с меньшей выборочной дисперсией и через $n_{\max}$ — с большей. По таблице для распределения Фишера находим критическую точку с уровнем значимости $\alpha/2$ и числами степеней свободы $n_{\max} - 1$ и $n_{\min} - 1$. Если $F < F_{\text{кр}}$, то основная гипотеза принимается, иначе — отвергается.

В случае б) делят первую выборочную дисперсию на вторую:

$$F = \frac{s_x^2}{s_y^2}.$$

По таблице для распределения Фишера находим критическую точку с уровнем значимости α и числами степеней свободы $n - 1$ и $m - 1$. Если $F < F_{\text{кр}}$, то основная гипотеза принимается, иначе отвергается.

2. **Гипотеза о равенстве средних при известных дисперсиях.**

Проверяем гипотезу H_0: $a_x = a_y$. Альтернативная гипотеза H_1 может быть трех видов: а) $a_x \neq a_y$; б) $a_x > a_y$; в) $a_x < a_y$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно. Во всех случаях вычисляют статистику критерия:

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}}.$$

В случае а) критическая точка $u_{\text{кр}}$ выбирается из условия $\Phi_0(u_{\text{кр}}) = (1 - \alpha)/2$. Если $|U| < u_{\text{кр}}$, гипотеза H_0 принимается; если $|U| > u_{\text{кр}}$ — отвергается.

В случае б) критическая точка $u_{\text{кр}}$ выбирается из условия $\Phi_0(u_{\text{кр}}) = 1/2 - \alpha$. Если $U < u_{\text{кр}}$, то гипотеза H_0 принимается; если $U > u_{\text{кр}}$ — отвергается.

З а м е ч а н и е. Гипотеза о средних обычно проверяется таким образом и в случае неизвестных дисперсий для больших выборок (объемом порядка сотен), когда оценки дисперсий можно принять за их точные значения.

3. **Гипотеза о равенстве средних при неизвестных равных дисперсиях.**

Проверяем гипотезу H_0: $a_x = a_y$. Альтернативная гипотеза H_1 может быть трех видов: а) $a_x \neq a_y$; б) $a_x > a_y$; в) $a_x < a_y$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно. Во всех случаях вычисляют статистику критерия:

$$T = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

где $s^2 = \dfrac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2}$.

Величина s^2 является объединенной оценкой дисперсии (общей для выборок).

Эту же формулу можно представить в виде

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Для проверки берутся критические точки $t_{кр}$ распределения Стьюдента с $n + m - 2$ степенями свободы и уровнем значимости α, причем в случае а) — для двусторонней критической области, в случае б) — для односторонней критической области.

В случае а), если $|T| < t_{кр}$, то гипотеза H_0 принимается, если $|T| > t_{кр}$ — отвергается.

В случае б), если $T < t_{кр}$, то гипотеза H_0 принимается, если $T > t_{кр}$ — отвергается.

З а м е ч а н и е. Поскольку для проверки гипотезы требуется равенство дисперсий у двух выборок, то сначала необходимо проверить гипотезу о равенстве дисперсий. В противном случае данный метод применять нельзя.

4. **Гипотеза о равенстве вероятностей «успеха» в двух сериях испытаний Бернулли.**

Гипотеза проверяется на основе асимптотической нормальности относительных частот, так что данный метод может применяться только при больших объемах выборок (порядка сотен). Пусть в одной серии из n_1 испытаний получили m_1 «успехов», в другой серии из n_2 испытаний получили m_2 «успехов». Проверяем гипотезу H_0: $p_1 = p_2$.

Альтернативная гипотеза может быть трех видов: а) $p_1 \neq p_2$; б) $p_1 > p_2$; в) $p_1 < p_2$, однако случай в) сводится к б) перестановкой индексов и не будет рассматриваться отдельно.

Во всех случаях вычисляют статистику критерия:

$$U = \frac{w_1 - w_2}{\sqrt{w(1-w)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

где $w = \frac{m_1 + m_2}{n_1 + n_2}$.

В случае а) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = (1 - \alpha)/2$. Если $|U| < u_{кр}$, гипотеза H_0 принимается, если $|U| > u_{кр}$ — отвергается.

В случае б) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$. Если $U < u_{кр}$, то гипотеза H_0 принимается, если $U > u_{кр}$ — отвергается.

Описанные выше критерии проверки гипотез можно представить в виде следующей таблицы (табл. 21.2).

Пример 21.10. По выборке объема $n = 30$ найден средний вес изделий, равный 130 г, изготовленных на первом станке; по выборке объема $m = 40$ найден средней вес изделий, равный 125 г, изготовленных на втором станке. Генеральные дисперсии известны: $\sigma_x^2 = 60$ г^2, $\sigma_y^2 = 80$ г^2. Требуется на уровне значимости 0,05 проверить нулевую гипотезу H_0: $a_x = a_y$ при конкурирующей гипотезе H_1: $a_x \neq a_y$. Предполагается, что случайные величины распределены нормально и выборки независимы.

Решение. Найдем значение статистики критерия:

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} = \frac{130 - 125}{\sqrt{\frac{60}{30} + \frac{80}{40}}} = 2{,}5.$$

По таблице функции Лапласа найдем критическую точку из равенства $\Phi(u_{кр}) = (1 - \alpha)/2 = 0{,}475$, получаем $u_{кр} = 1{,}96$. Так как $|U| > u_{кр}$, то H_0 отвергается. Таким образом, нельзя утверждать, что средние значения веса изделий двух станков совпадают.

Пример 21.11. По двум независимым выборкам, объемы которых $n = 9$ и $m = 16$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии $s_x^2 = 34{,}02$ и $\sigma_y^2 = 12{,}15$. На уровне значимости 0,01 проверить нулевую гипотезу H_0: $\sigma_x^2 = \sigma_y^2$ против конкурирующей гипотезы H_1: $\sigma_x^2 > \sigma_y^2$.

Решение. Рассчитаем значение статистики критерия:

Таблица 21.2

H_0	Предположение	Статистика критерия	H_1	Область принятия H_0
$a_x = a_y$	σ_x^2 и σ_y^2 известны	$U = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}}$	$a_x > a_y$	$U < u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$a_x < a_y$	$U > -u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$a_x \neq a_y$	$\lvert U\rvert < u_{кр}$, $\Phi_0(u_{кр}) = (1 - \alpha)/2$
$a_x = a_y$	σ_x^2 и σ_y^2 неизвестны, но равны	$T = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$, где $s^2 = \dfrac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$	$a_x > a_y$	$T < t_{кр}(\alpha, n+m-2)$ для односторонней области
			$a_x < a_y$	$T > -t_{кр}(\alpha, n+m-2)$ для односторонней области
			$a_x \neq a_y$	$\lvert T\rvert < t_{кр}(\alpha, n+m-2)$ для двусторонней области
$\sigma_x^2 = \sigma_y^2$	a_x и a_y неизвестны	$F = \dfrac{s_x^2}{s_y^2}$, где $s_x^2 > s_y^2$	$\sigma_x^2 > \sigma_y^2$	$F < F_{кр}(\alpha, n-1, m-1)$
			$\sigma_x^2 \neq \sigma_y^2$	$F < F_{кр}(\alpha/2, n-1, m-1)$
$p_1 = p_2$	n_1 и n_2 порядка сотен	$U = \dfrac{w_1 - w_2}{\sqrt{w(1-w)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, где $w = \dfrac{m_1 + m_2}{n_1 + n_2}$	$p_1 > p_2$	$U < u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$p_1 < p_2$	$U > -u_{кр}$, $\Phi_0(u_{кр}) = 1/2 - \alpha$
			$p_1 \neq p_2$	$\lvert U\rvert < u_{кр}$, $\Phi_0(u_{кр}) = (1 - \alpha)/2$

$$F = \frac{s_x^2}{s_y^2} = \frac{34{,}02}{12{,}15} = 2{,}8.$$

Числа степеней свободы $k_1 = n - 1 = 8$, $k_2 = m - 1 = 15$. По таблице критических точек распределения Фишера—Снедекора по заданному уровню значимости $\alpha = 0{,}01$ и числам степеней свободы находим $F_{кр}(0{,}01;\ 8;\ 15) = 4$. Поскольку $F < F_{кр}$, нулевая гипотеза принимается.

Пример 21.12. Реклама утверждает, что из двух типов пластиковых карт — «Русский Экспресс» и «Супер-Понт» их пользователи предпочитают первый. С целью проверки этого утверждения были обследованы среднемесячные платежи $n = 16$ обладателей «Русского Экспресса» и $m = 11$ обладателей «Супер-Понта». Выяснилось, что платежи по картам «Русский Экспресс» составляют в среднем 563 долл. с исправленным средним квадратическим отклонением 178 долл., а по картам «Супер-Понт» — в среднем 485 долл. с исправленным средним квадратическим отклонением 196 долл.

Предварительный анализ законов распределения месячных расходов как среди обладателей «Русского Экспресса», так и среди обладателей «Супер-Понта» показал, что они достаточно хорошо описываются нормальным приближением.

Проверить утверждение рекламы на уровне значимости 10%.

Решение. В данном случае речь идет о проверке гипотезы о средних при неизвестных дисперсиях (объемы выборок малы). Поэтому прежде всего необходимо проверить гипотезу о равенстве дисперсий и лишь затем двигаться дальше. Имеем

$$F = \frac{s^2_{\max}}{s^2_{\min}} = \frac{196^2}{178^2} = \frac{38416}{31684} \approx 1{,}21.$$

По таблице критических точек распределения Фишера—Снедекора по уровню значимости $\alpha/2 = 0{,}05$ и числам степеней свободы $k_1 = n_{\max} - 1 = 10$ и $k_2 = n_{\min} - 1 = 15$ найдем критическую точку $F_{\text{кр}} = 2{,}55$. Так как $1{,}21 < 2{,}55$, то мы принимаем гипотезу о равенстве дисперсий двух выборок.

Теперь мы можем воспользоваться критерием Стьюдента для проверки гипотезы о равенстве средних. Имеем

$$s^2 = \frac{10 \cdot 38\,416 + 15 \cdot 31\,684}{25} = 34\,376{,}8; \quad s \approx 185{,}4.$$

Вычисление статистики критерия дает

$$T = \frac{563 - 485}{185{,}4\sqrt{1/11 + 1/16}} \approx 1{,}074.$$

Из таблиц критических точек распределения Стьюдента (для односторонней области) по уровню значимости $\alpha = 0{,}1$ и числу степеней свободы 25 находим $t_{\text{кр}} = 1{,}32$.

Поскольку $T < t_{\text{кр}}$, то принимается основная гипотеза (о равенстве средних). Таким образом, утверждение рекламы не подтверждается имеющимися данными.

Пример 21.13. В партии из 500 деталей, изготовленных первым станком-автоматом, оказалось 60 нестандартных; из 600 деталей второго станка — 42 нестандартных. На уровне значимости $\alpha = 0{,}01$ проверить нулевую гипотезу H_0: $p_1 = p_2$ о равенстве вероятностей изготовления нестандартной детали обоими станками против конкурирующей гипотезы H_1: $p_1 \neq p_2$.

Решение. Имеем $w_1 = 60/500 = 0{,}12$; $w_2 = 42/600 = 0{,}07$; $w = (60 + 42)$: : $(500 + 600) \approx 0{,}09$. Найдем значение статистики критерия:

$$U = \frac{0{,}12 - 0{,}07}{\sqrt{0{,}09 \cdot 0{,}91 \cdot \left(\frac{1}{500} + \frac{1}{600}\right)}} \approx 2{,}85.$$

Найдем критическую точку из соотношения $\Phi_0(u_{\text{кр}}) = 0{,}495$, откуда $u_{\text{кр}} = 2{,}57$. Поскольку $|U| > u_{\text{кр}}$, нулевая гипотеза отвергается. Значит, вероятности изготовления нестандартных деталей на двух станках различны.

Глава 22
КРИТЕРИИ СОГЛАСИЯ

Критерии проверки гипотезы о предполагаемом виде закона распределения случайной величины называются *критериями согласия*.

Следует понимать, что мы проверяем не тот факт, что случайная величина *действительно* имеет определенный закон распределения (например, нормальный), а проверяется лишь, достаточно ли хорошо наблюдаемые данные *согласуются* с некоторым законом распределения, чтобы можно было использовать этот закон для прогнозирования поведения данной случайной величины.

Гипотезы могут быть как простыми, так и сложными. Гипотеза называется *простой*, если проверяется соответствие некоторому закону распределения с заданными параметрами. Гипотеза называется *сложной*, если проверяется соответствие некоторому закону распределения с произвольными параметрами. В этом случае параметры оцениваются по выборке.

Наиболее часто используемые критерии согласия — это критерии Пирсона, Фишера и Колмогорова.

22.1. КРИТЕРИИ СОГЛАСИЯ ПИРСОНА И ФИШЕРА

Критерий согласия Пирсона (хи-квадрат) идеально подходит для проверки гипотез в полиномиальной схеме.

Пусть проводится n независимых испытаний, каждое из которых может иметь r различных исходов $A_1, A_2, \ldots, A_r$. Проверяется гипотеза H_0 о том, что вероятности этих исходов равны $p_1, p_2, \ldots, p_r$, если в последовательности испытаний они встретились $m_1, m_2, \ldots, m_r$ раз. Альтернативная гипотеза H_1 заключается в том, что набор вероятностей исходов отличается от заданного.

Теорема 22.1 (теорема Пирсона). Если основная гипотеза верна, то распределение статистики хи-квадрат

$$\chi^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i}$$

при $n \to \infty$ стремится к распределению хи-квадрат с $r - 1$ степенями свободы. В противном случае эта статистика стремится к бесконечности.

Отсюда получаем критерий (применимый при больших n): если $\chi^2 < \chi^2_{\alpha:r-1}$, то основная гипотеза принимается, иначе — отвергается.

В случае если вероятности $p_1, p_2, ..., p_r$ зависят от неизвестных параметров $\theta_1, \theta_2, ..., \theta_k$, которые можно оценить по $m_1, m_2, ..., m_r$, то их оценивают (как правило, методом максимального правдоподобия), получают соответствующие оценки $p_1, p_2, ..., p_r$ и так же вычисляют статистику хи-квадрат, но в этом случае ее предельное распределение имеет уже $r - k - 1$ степеней свободы.

Тогда, если $\chi^2 < \chi^2_{\alpha:r-k-1}$, гипотеза принимается, иначе — отвергается.

Критерий хи-квадрат можно применять и в более общей схеме для проверки гипотез о распределении случайных величин. В этом случае в качестве исходов $A_1, A_2, ..., A_r$ берут попадания наблюдений в некоторые множества $\Delta_1, \Delta_2, ..., \Delta_r$. Для дискретных величин это могут быть отдельные значения или их объединения. Для непрерывных величин используют обычную группировку, т.е. подсчитывают числа попаданий в некоторые интервалы.

Если распределение не ограничено слева или справа, то крайние интервалы продолжают до бесконечности. Если числа попаданий в какие-то интервалы слишком малы (например, меньше 5), то такие интервалы объединяют с соседними интервалами. Всегда желательно иметь не менее 50 наблюдений в выборке.

В результате есть множества $\Delta_1, \Delta_2, ..., \Delta_r$, находим числа $m_1, m_2, ..., m_r$ попаданий наблюдений в эти множества и теоретические вероятности $p_i = P(\xi \in \Delta_i)$, $i = 1, 2, ..., r$, после чего применяем критерий хи-квадрат.

Заметим, что критерий хи-квадрат не позволяет проверить основную гипотезу против альтернативы, что случайная величина имеет *другое* распределение, но с *таким же* набором вероятностей p_i. Однако такая ситуация на практике крайне маловероятна и поэтому не рассматривается.

Разберем более подробно следующие случаи.

1. Требуется по выборке $x_1, x_2, ..., x_n$ проверить нулевую гипотезу о том, что генеральная совокупность имеет функцию распределения $F_0(x)$, если известны значения параметров закона распределения, т.е. имеет место простая гипотеза.

Для проверки этой гипотезы область значений $x_1, x_2, ..., x_n$ случайной величины ξ разбивают на r непересекающихся промежутков: $[C_0, C_1)$; $[C_1, C_2)$; ...; $[C_{r-1}, C_r)$, где C_0 и C_r — левая и правая границы возможных значений случайной величины. В частности, если речь идет о распределении на всей числовой прямой (например, нормальном), то полагают $C_0 = -\infty$, $C_r = +\infty$; если о распределении на положительной полуоси (например, показательном), то $C_0 = 0$, $C_r = +\infty$; если о распределении на отрезке $[a; b]$ (например, равномерном), то $C_0 = a$, $C_r = b$.

В предположении, что справедлива основная гипотеза, находим теоретические вероятности попадания случайной величины в частичные интервалы:

$$p_i = P(C_{i-1} \le \xi < C_i) = F_0(C_i) - F_0(C_{i-1}), \quad \text{где } p_i > 0, \sum_{i=1}^{r} p_i = 1.$$

Рекомендуется выбирать разбиение так, чтобы выполнялось условие $np_i \ge 10$ для всех $i = 1, 2, ..., r$.

Обозначим через m_i число наблюдений, попавших в $\Delta_i = [C_{i-1}, C_i)$. По теореме Пирсона получаем, что если $x_1, x_2, ..., x_n$ — выборка из генеральной совокупности с функцией распределения $F_0(x)$, то статистика $\chi^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i}$ имеет при $n \to \infty$ распределение хи-квадрат с $r - 1$ степенями свободы, если основная гипотеза верна. В противном случае статистика стремится к бесконечности. Поэтому в качестве критической области выбирается область больших значений. Вероятность попадания статистики критерия в эту область в предположении истинности основной гипотезы должна быть равна принятому уровню значимости α: $P(\chi^2 \ge \chi^2_{\alpha,r-1}) = \alpha$. Таким образом, если $\chi^2 < \chi^2_{\alpha;r-1}$, то основная гипотеза принимается, иначе — отвергается.

2. Если значения параметров гипотетической функции распределения $F_0(x)$ неизвестны, то имеем сложную гипотезу. Основная гипотеза H_0 заключается в том, что функция распределения имеет вид $F_0(x) = F(x, \theta_1, ..., \theta_k)$ при некоторых неизвестных значениях параметров $\theta_1, \theta_2, ..., \theta_k$. В этом случае вероятности $p_1, p_2, ..., p_r$ также зависят от параметров. Статистика критерия принимает вид

$$\chi^2 = \sum_{i=1}^{r} \frac{(m_i - np_i(\theta_1, \theta_2, ..., \theta_k))^2}{np_i(\theta_1, \theta_2, ..., \theta_k)}.$$

При известных значениях параметров имел бы место первый случай. Но так как истинные значения $\theta_1, \theta_2, ..., \theta_k$ неизвестны, то, подставляя их оценки, найденные методом максимального правдоподобия, получаем критерий χ^2 с меньшим числом степеней свободы, а именно $\nu = r - k - 1$, где r — число интервалов, на которые разбит весь диапазон наблюдаемых значений; k — число неизвестных параметров функции распределения. Сравнивая наблюдаемое значение статистики критерия с критической точкой $\chi^2_{\alpha,r-k-1}$, делаем заключение об истинности нулевой гипотезы: гипотеза принимается, если $\chi^2 < \chi^2_{\alpha,r-k-1}$, и отвергается в противном случае.

Критерий хи-квадрат для простой гипотезы (т.е. в случае известных параметров) называют *критерием хи-квадрат Пирсона*, а критерий хи-квадрат для сложной гипотезы (с оцениванием параметров) — *критерием хи-квадрат Фишера*.

Пример 22.1. В следующей таблице представлены данные о числе крупных сделок, заключенных на фондовой бирже за квартал, для 517 инвесторов.

i	0	1	2	3	4	5	6	7
m_i	112	168	130	68	32	5	1	1

В первой строке приведено число сделок, во второй — число инвесторов, заключивших указанное количество сделок за квартал.

Проверить, используя критерий Пирсона, что на уровне значимости $\alpha = 0{,}05$ число крупных сделок, заключенных одним инвестором за квартал, распределено по закону Пуассона с параметром $\lambda = 1{,}5$.

Решение. Поскольку распределение Пуассона дискретно, в качестве различных исходов здесь можно принять сами значения случайной величины. Заметим, что два последних значения (6 и 7) встретились слишком мало раз, поэтому их следует объединить с предыдущим (5). Кроме того, распределение Пуассона не ограничено справа, и следует учесть все значения, превышающие число 7 (которые не встретились ни разу). Таким образом, в качестве множеств Δ_i выберем $\{0\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $[5, +\infty)$. Здесь $r = 6$.

Найдем теоретические вероятности по формуле распределения Пуассона:

$$P(\xi = j) = \frac{\lambda^j}{j!}e^{-\lambda}, \quad j = 0, 1, 2, \dots .$$

При $\lambda = 1{,}5$ получаем:

$p_1 = P(\xi = 0) \approx 0{,}2231$; $\quad p_4 = P(\xi = 3) \approx 0{,}1255$;
$p_2 = P(\xi = 1) \approx 0{,}3347$; $\quad p_5 = P(\xi = 4) \approx 0{,}0471$;
$p_3 = P(\xi = 2) \approx 0{,}2510$; $\quad p_6 = P(\xi \geq 5) \approx 0{,}0186$.

Умножим эти величины на число инвесторов $n = 517$ и составим таблицу:

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2/(np_i)$
0	112	115,34	−3,34	0,10
1	168	173,04	−5,04	0,15

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2/(np_i)$
2	130	129,77	0,23	0,00
3	68	64,88	3,12	0,15
4	32	24,35	7,65	2,40
≥5	7	9,62	−2,62	0,71

Суммируя значения в последнем столбце, получаем значение статистики хи-квадрат: $\chi^2 = 3{,}51$.

По таблице критических точек распределения хи-квадрат по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $r - 1 = 5$ находим критическую точку: $\chi^2_{\text{кр}} = 11{,}1$. Поскольку $\chi^2 < \chi^2_{\text{кр}}$, можно считать, что число сделок, заключенных одним инвестором за квартал, распределено по закону Пуассона с параметром $\lambda = 1{,}5$.

Замечание. Если бы значение параметра $\lambda = 1{,}5$ было оценено по самой выборке, следовало бы задать число степеней свободы $r - 2 = 4$. Тогда получается $\chi^2_{\text{кр}} = 9{,}5$, следовательно, гипотеза также принимается.

Пример 22.2. В таблице приведены сгруппированные данные о коэффициентах соотношения заемных и собственных средств на 100 малых предприятиях региона.

Номер интервала	Интервал	Середина интервала	Наблюдений в интервале
1	5,05–5,15	5,1	5
2	5,15–5,25	5,2	8
3	5,25–5,35	5,3	12
4	5,35–5,45	5,4	20
5	5,45–5,55	5,5	26
6	5,55–5,65	5,6	15
7	5,65–5,75	5,7	10
8	5,75–5,85	5,8	4

На уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что коэффициенты можно описать нормальным распределением.

Решение. В данном случае параметры распределения явно не заданы и их следует оценить по сгруппированным данным. Находим выборочное среднее 5,46 и выборочную дисперсию 0,03.

Теоретические вероятности оцениваем по формулам:

$$p_i = P(C_{i-1} < \xi < C_i) = \Phi\left(\frac{C_i - \bar{x}}{s}\right) - \Phi\left(\frac{C_{i-1} - \bar{x}}{s}\right), \quad i = 1, 2, \ldots, 8.$$

Следует продолжить крайние интервалы и считать $C_0 = -\infty$, $C_8 = +\infty$, поскольку нормальное распределение не ограничено с обеих сторон.

С учетом полученных значений построим таблицу:

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2/(np_i)$
≤5,15	5	3,67	1,32	0,47
5,15–5,25	8	7,59	0,41	0,02
5,25–5,35	12	15,00	–3,00	0,60
5,35–5,45	20	21,43	–1,43	0,09
5,45–5,55	26	22,14	3,86	0,67
5,55–5,65	15	16,53	–1,53	0,14
5,65–5,75	10	8,93	1,07	0,13
≥5,75	4	4,70	–0,70	0,10

Суммируя значения в последнем столбце, получаем значение статистики хи-квадрат: $\chi^2 = 2{,}22$.

По таблице критических точек распределения хи-квадрат по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $r - 1 - k = 8 - 1 - 2 = 5$ находим критическую точку: $\chi^2_{\text{кр}} = 11{,}1$. Поскольку $\chi^2 < \chi^2_{\text{кр}}$, можно считать, что коэффициенты хорошо описываются нормальным распределением.

Замечание. В принципе здесь можно было бы объединить крайние интервалы с соседними. Вычисления показывают, что и в этом случае гипотеза принимается.

3. Экономисты используют также критерий χ^2 как *критерий однородности*.

Пусть имеется $k \geq 2$ независимых выборок, содержащих соответственно $n_1, n_2, \ldots, n_k$ независимых наблюдений $(x_1, x_2, \ldots, x_{n_1})$; $(y_1, y_2, \ldots, y_{n_2})$; ...; $(z_1, z_2, \ldots, z_{n_k})$.

Гипотеза однородности H_0 предполагает, что генеральные совокупности, из которых извлечены выборки, одинаковы (или все выборки произведены из одной генеральной совокупности) и им соответствуют одинаковые функции распределения. Альтернативная гипотеза H_1 состоит в том, что распределения различны.

Наиболее часто в приложениях встречается случай, когда $k = 2$. Пусть имеется два ряда наблюдений некоторого признака и каждый ряд разбит на r групп по значениям этого признака. Сгруппированные ряды имеют вид:

$$n_1:\quad m_1 \quad m_2 \quad \ldots \quad m_i \quad \ldots \quad m_r$$
$$n_2:\quad l_1 \quad l_2 \quad \ldots \quad l_i \quad \ldots \quad l_r$$

Пусть m_i и l_i — количества выборочных значений в i-й группе соответственно для первого и второго рядов наблюдений. Тогда статистика критерия для проверки истинности нулевой гипотезы принимает вид

$$\chi^2 = n_1 n_2 \sum_{i=1}^{r} \frac{\left(\frac{m_i}{n_1} - \frac{l_i}{n_2}\right)^2}{m_i + l_i},$$

и в случае истинности основной гипотезы при $n \to \infty$ она имеет предельное распределение χ^2_{r-1}. Критическими точками, соответствующими уровню значимости α, будут $\chi^2_{\alpha;r-1}$, и проверка гипотезы проводится по общей схеме: если $\chi^2 < \chi^2_{\alpha;r-1}$, то нулевая гипотеза принимается, в противном случае — отвергается.

Если положить $w_i = \dfrac{m_i}{m_i + l_i}$ и $w = \dfrac{n_1}{n_1 + n_2}$, то формулу можно представить в виде, часто более удобном для практических расчетов:

$$\chi^2 = \frac{(n_1 + n_2)^2}{n_1 n_2}\left(\sum_{i=1}^{r} \frac{m_i^2}{m_i + l_i} - \frac{n_1^2}{n_1 + n_2}\right) = \frac{1}{w(1-w)}\left(\sum_{i=1}^{r} m_i w_i - n_1 w\right).$$

Пример 22.3. Распределение доходов среди рабочих и служащих согласно шведской переписи 1930 г. приведено в таблице. С помощью критерия Пирсона проверить гипотезу об одинаковости распределения доходов у двух возрастных групп для заводских мастеров и всех рабочих и служащих [15].

Доходы (тыс. крон)	Все рабочие и служащие			Заводские мастера		
	Возрастная группа			Возрастная группа		
	m_i	l_i	w_i	m_i	l_i	w_i
	40–50	50–60		40–50	50–60	
0–1	7831	7558	0,508869	71	54	0,568000
1–2	26 740	20 685	0,563837	430	324	0,570291
2–3	35 572	24 186	0,595267	1072	894	0,545269
3–4	20 009	12 280	0,629784	1609	1202	0,572394
4–6	11 527	6776	0,629787	1178	903	0,566074
>6	6919	4222	0,621039	158	112	0,585185

Решение. Рассматривается гипотеза однородности: обе выборки (по возрастным группам) извлечены из одной генеральной совокупности. Для заводских мастеров получаем $\chi^2 = 4{,}27$. Отсюда $\chi^2 < \chi^2_{0{,}05;5} = 11{,}1$, и можно считать, что две выборки извлечены из одной генеральной совокупности. Но если сравнить распределение доходов у возрастных групп всех промышленных рабочих и служащих, то получаем $\chi^2 = 840{,}62$, что говорит об очень высокой степени различия между распределениями.

При этом видно, что числа w_i имеют тенденцию возрастать с ростом доходов (т.е. доля более молодых среди имеющих большие доходы оказывается больше).

22.2. КРИТЕРИЙ СОГЛАСИЯ КОЛМОГОРОВА

Критерий согласия Колмогорова применяется для проверки гипотез о законах распределения только непрерывных случайных величин. Проверяется гипотеза H_0: $F(x) = F_0(x)$ против альтернативной H_1: $F(x) \neq F_0(x)$.

Критерий основан на том факте, что распределение супремума разности между теоретической и эмпирической функциями распределения $D_n = \sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)|$ одинаково для любой $F(x)$. Величину D_n называют *статистикой Колмогорова.*

При малых n для статистики Колмогорова существуют таблицы критических точек $D_{\text{кр}}$. Если $D_n < D_{\text{кр}}$, то гипотеза H_0 принимается, иначе — отвергается. При больших n используют *предельное распределение Колмогорова*, описываемое следующей теоремой.

Теорема 22.2 (теорема Колмогорова).

$$P(\sqrt{n}D_n < x) \to Q(x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2}, \quad n \to \infty.$$

Для распределения Колмогорова $Q(x)$, предельного для статистики $\lambda = \sqrt{n}D_n$, также существуют таблицы критических точек $\lambda_{\text{кр}}$. Практически их используют уже при $n > 20$. Если $\lambda < \lambda_{\text{кр}}$, то гипотеза H_0 принимается, иначе — отвергается.

Статистику Колмогорова можно вычислить по эквивалентным формулам:

$$D_n = \max_{1\le i\le n}\left\{\left|F_0(x_{(i)}) - \frac{i-1}{n}\right|, \left|\frac{i}{n} - F_0(x_{(i)})\right|\right\}$$

и

$$D_n = \max_{1\le i\le n}\left|F_0(x_{(i)}) - \frac{2i-1}{2n}\right| + \frac{1}{2n},$$

где $x_{(i)}$ — члены вариационного ряда.

Заметим, что критерий Колмогорова, строго говоря, *нельзя* применять в случаях сгруппированных данных при неизвестных параметрах распределения. Тем не менее он иногда применяется на практике и в подобных ситуациях. Однако при этом статистики критерия получаются *заниженными*, что увеличивает ошибку первого рода.

В таких случаях предпочтительней пользоваться критерием хи-квадрат Пирсона.

Проверка гипотезы однородности с помощью *критерия Колмогорова—Смирнова* состоит в следующем. Пусть $x_1, x_2, ..., x_{n_1}$ и $y_1, y_2, ..., y_{n_2}$ — выборки из двух генеральных совокупностей. Требуется проверить нулевую гипотезу H_0 о совпадении законов распределения генеральных совокупностей, из которых произведены выборки, против гипотезы H_1, что эти распределения различны.

Определим эмпирические функции распределения $F_1^{(n_1)}$ и $F_2^{(n_2)}$ для первой и второй выборки соответственно. Для проверки гипотезы вводится статистика

$$D_{n_1,n_2} = \sup_{-\infty<x<\infty} \left|F_1^{(n_1)}(x) - F_2^{(n_2)}(x)\right|.$$

Пусть $n_1, n_2 \to \infty$, $n_0 = \dfrac{n_1 n_2}{n_1 + n_2}$ и предельные функции $F_1(x)$ и $F_2(x)$ непрерывны. Тогда в условиях истинности нулевой гипотезы статистика $D_{n_1,n_2}\sqrt{n_0}$ имеет в пределе распределение Колмогорова. Критической областью является область больших значений: основная гипотеза отклоняется, если $D_{n_1,n_2}\sqrt{n_0} > \lambda_\alpha$, где λ_α — критическая точка распределения Колмогорова, соответствующая уровню значимости α.

Пример 22.4. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение пяти поездок фиксировал время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 (мин). Проверить гипотезу о том, что время ожидания равномерно распределено на отрезке [0; 10] на уровне значимости $\alpha = 0{,}05$.

Решение. Составим вариационный ряд: 1,2; 3,7; 4,8; 5,1; 9,2. С учетом того, что в данном случае $F_0(x) = x/10$, $0 \le x \le 10$, построим таблицу:

i	$x_{(i)}$	$F_0(x_{(i)})$	$(2i-1)/(2n)$	$\lvert F_0(x_{(i)}) - (2i-1)/(2n)\rvert$
1	1,2	0,12	0,1	0,02
2	3,7	0,37	0,3	0,07
3	4,8	0,48	0,5	0,02
4	5,1	0,51	0,7	0,18
5	9,2	0,92	0,9	0,02

Таким образом, значение статистики Колмогорова составляет $D_5 = 0{,}18 + 0{,}1 = 0{,}28$.

По таблице критических точек при $\alpha = 0{,}05$ и $n = 5$ находим $D_{кр} = 0{,}56$. Поскольку $D < D_{кр}$, то нулевая гипотеза (о равномерности распределения) принимается.

З а м е ч а н и е. На самом деле по таким небольшим выборкам конечно же нельзя делать далеко идущие выводы.

Пример 22.5. Выборка из 10 наблюдений приведена в таблице. Проверить с помощью критерия Колмогорова гипотезу о том, что эта выборка из генеральной совокупности, равномерно распределенной на отрезке [0; 1]. Уровень значимости $\alpha = 0{,}05$.

Решение.

i	$x_{(i)}$	$F_n(x) = \frac{i}{n}$	$F_0(x_{(i)})$	$\frac{i}{n} - F_0(x_{(i)})$	$F_0(x_{(i)}) - \frac{i-1}{n}$
1	0,01	0,1	0,01	0,09	0,01
2	0,09	0,2	0,09	0,11	–0,01
3	0,1	0,3	0,1	0,2	–0,1
4	0,25	0,4	0,25	0,15	–0,05
5	0,33	0,5	0,33	0,17	–0,07
6	0,35	0,6	0,35	0,25	–0,15
7	0,52	0,7	0,52	0,18	–0,08
8	0,73	0,8	0,73	0,07	0,03
9	0,76	0,9	0,76	0,14	–0,04
10	0,86	1	0,86	0,14	–0,04

Из таблицы получаем $D_{10} = 0{,}25$.

По заданному уровню значимости $\alpha = 0{,}05$ находим критическую точку $D_{кр} = 0{,}41$. Поскольку $D_{10} < D_{кр}$, то гипотеза принимается.

ЛИТЕРАТУРА

1. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
2. *Бочаров П.П., Печинкин А.В.* Теория вероятностей. Математическая статистика. — М.: Гардарика, 1998.
3. *Ван дер Варден Б.* Математическая статистика. — М.: Высшая школа, 1975.
4. *Венецкий Г.И., Кильдышев А.Н.* Математическая статистика. — 3-е изд. — М.: Статистика, 1986.
5. *Вентцель Е.С.* Теория вероятностей и математическая статистика. — М.: Мир, 1972.
6. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1998.
7. *Грачева М.В., Фадеева Л.Н., Черемных Ю.Н.* Количественные методы в экономических исследованиях. — М.: ЮНИТИ, 2004.
8. *Кибзун А.И., Горяинова Е.Р., Наумов А.В., Сиротин А.Н.* Теория вероятностей и математическая статистика. — М.: Физматлит, 2002.
9. *Колемаев В.А., Староверов О.В., Турундаевский В.Б.* Теория вероятностей и математическая статистика. — М.: Высшая школа, 1991.
10. *Крамер Гаральд.* Математические методы статистики. — М.: Мир, 1975.
11. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
12. *Сулицкий В.Н.* Методы статистического анализа в управлении. — М.: Дело, 2002.
13. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. — М.: ИНФРА-М, 2003.
14. *Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Ященко И.В.* Теория вероятностей и статистика. — М.: МЦНМО, 2004.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Раздел 6
ЭКОНОМЕТРИЧЕСКИЙ АСПЕКТ

Глава 23
КОЛИЧЕСТВЕННЫЕ ВЕЛИЧИНЫ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

Мы рассматривали абстрактные количественные понятия и методы в математике, лишь сопровождая их экономическими иллюстрациями. В этом разделе степень включения экономики и менеджмента значительно возрастет. Раздел можно рассматривать как решение задач управления и экономических задач математическими методами. Экономическая теория и наука управления в XXI веке предполагают существенно более высокий уровень формализации, чем это было раньше. Познакомимся с понятиями в экономике и управлении, которые можно выразить количественно. Анализ и осмысление целого ряда экономических явлений, а также принятие эффективных управленческих решений невозможны без использования эластичности, построения экономико-математических моделей, изучения соотношений между суммарными, средними и предельными величинами.

23.1. СРЕДНИЕ И ПРЕДЕЛЬНЫЕ ВЕЛИЧИНЫ В УПРАВЛЕНИИ

В количественном анализе и принятии управленческих решений большую роль играют средние и предельные величины. Кроме того, при анализе работы фирмы для выбора наилучшего решения важны как абсолютные, так и относительные показатели.

Средняя величина отражает величину одного экономического показателя в расчете на единицу другого. Предельная величина дает прирост экономического показателя в расчете на единицу прироста другого показателя. Пусть фирма, оценивая свои выручку и издержки, изучает вопрос о расширении производства продукции. Если средняя выручка фирмы превышает средние издержки, то фирма получает прибыль и производить продукцию выгодно. Если при этом предельная выручка превышает предельные издержки, то фирме выгодно расширять производство, увеличивая объем прибыли. Соответственно, если средние издержки превышают среднюю выручку, то фирма терпит убытки, а если предельные издержки превышают предельную выручку, то объем производства нужно сократить.

Введем понятия накопленной величины, суммарной, средней и предельной величины. Изучим свойства и соотношения этих величин.

Накопленная величина — это суммирование некоторого показателя в зависимости от другого показателя. Накопленная величина описывается функцией независимой переменной. В роли суммарной величины может выступать, например, доход как функция времени. Накопленная величина есть неубывающая функция своего аргумента.

Суммарная величина есть значение одного показателя в зависимости от другого. Доход как функция объема выпуска, объем выпуска как функция от количества вложенного труда — примеры суммарной величины $F(x)$.

Средняя величина $AF(x)$ определяется как отношение суммарной величины к независимой переменной: $AF(x) = \dfrac{F(x)}{x}$. Буква A — сокращение от *average* (средняя). Средняя величина может обозначаться также следующим образом $\bar{F} \equiv AF(x)$. Примерами средней величины может служить среднедушевой объем потребления, средний доход и др.

Предельная (маржинальная) величина $MF(x)$ определяется как предел приращения суммарной величины к приращению показателя — аргумента:

$$MF(x) = \lim_{\Delta x \to 0} \frac{\Delta F(x)}{\Delta x}.$$

Как известно из курса математического анализа, предельная величина есть производная $MF(x) = F'(x)$ суммарной величины по независимой переменной x. Если суммарная величина меняется дискретно, то в определении предельной величины предел опускается: $MF(x) = \dfrac{\Delta F}{\Delta x}$. Тогда предельную величину можно интерпретировать как изменение суммарной величины, вызванное увеличением независимой переменной на единицу. Примеры предельных величин: предельный доход, предельные издержки, предельная полезность и т.д.

По одной из этих величин можно найти другие, причем как аналитически, так и графически.

Нахождение средней величины по суммарной

Аналитически задача решается с помощью определения: $AF(x) = \frac{F(x)}{x}$. Например, если $F(x) = ax^2 + bx + c$, то $AF(x) = ax + b + \frac{c}{x}$.

Для графического решения этой задачи необходимо провести отрезок, соединяющий начало координат с точкой графика функции, имеющей координаты $(x, F(x))$. Его тангенс угла наклона в прямоугольном треугольнике $\operatorname{tg}\beta = \frac{F(x)}{x}$ будет по определению численно равен средней величине $AF(x) = \operatorname{tg}\beta(x)$ (рис. 23.1).

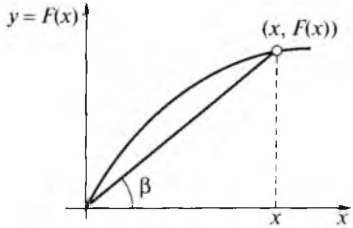

Рис. 23.1

Нахождение суммарной величины по средней (обратная задача)

Аналитически задача решается просто: $F(x) = x \cdot AF(x)$. Если зависимость средней величины $F(x)$ от показателя x задана в виде графика, то суммарную величину при данном значении независимой переменной x можно определить как площадь прямоугольника с вершинами в начале координат и точке графика средней величины (рис. 23.2), имеющей координаты $(x, AF(x))$, и сторонами x и $AF(x)$.

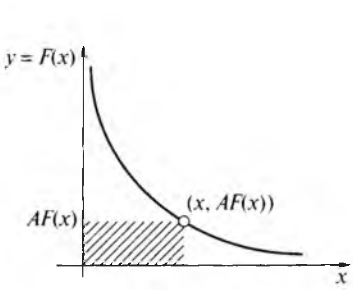

Рис. 23.2

Задавая различны значения аргументу x и вычисляя соответствующие площади, мы можем построить график суммарной величины в зависимости от параметра x.

Нахождение предельной величины по суммарной

Задача решается нахождением производной

$MF(x) = F'(x)$.

Для графического решения этой задачи необходимо через точку графика суммарной величины, имеющую координаты $(x, F(x))$, провести касательную к графику (рис. 23.3).

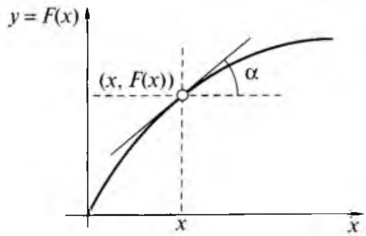

Рис. 23.3

Тангенс угла наклона касательной в произвольной точке x будет численно равен производной суммарной величины, а следовательно, являться предельной величиной:

$MF(x) = F'(x) = \operatorname{tg}\alpha_{кас}(x)$.

При изменении независимой переменной x угол наклона касательной α также изменяется. Увеличение этого угла с увеличением x свидетельствует о возрастании предельной величины, а уменьшение — об убывании.

Нахождение суммарной величины по предельной означает нахождение функции $F(x)$, производная которой $F'(x) = MF(x)$ известна. Для решения этой задачи служит операция интегрирования:

$$F(x) = \int MF(x)dx.$$

Площадь под графиком функции предельной величины в диапазоне изменения независимой переменной от нуля до x будет равна суммарной величине с точностью до некоторой постоянной C или $F(x) = S(x) + C$. Если при $x \to 0$ площадь $S(x) \to 0$, то константу можно найти как значение суммарной величины при $x = 0$. Соотношение между средними и предельными величинами в дискретном случае имеет простую интерпретацию.

23.2. ЭЛАСТИЧНОСТЬ

Пусть величина y зависит от x и эта зависимость описывается функцией $y = f(x)$. Изменение Δx независимой переменной x приводит к большому или малому изменению Δy переменной y. Как велико это изменение? Другими словами, какова чувствительность функции y к изменениям аргумента? Ответ на этот вопрос дает производная

$$y'_x = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

характеризующая скорость изменения функции с изменением аргумента x. Однако этот показатель неудобен тем, что он зависит от выбора единиц измерения.

Поэтому для измерения чувствительности изменения функции к изменению аргумента изучают связь не абсолютных изменений переменных x и y (Δx и Δy), а их относительных или процентных изменений.

О п р е д е л е н и е. Эластичностью функции $E_x(y)$ называется предел отношения относительного приращения функции $y = f(x)$ к относительному приращению аргумента x при $\Delta x \to 0$:

$$E_x(y) = \lim_{\Delta x \to 0} \frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}} = \frac{x}{y} \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{x}{y} y'.$$

Если эластичность функции представить в виде

$$E_x(y) = \lim_{\Delta x \to 0} \frac{\frac{\Delta y}{y} \cdot 100\%}{\frac{\Delta x}{x} \cdot 100\%},$$

то легко увидеть, что эластичность функции показывает приближенно, на сколько процентов изменится функция $y = f(x)$ при изменении независимой переменной x на 1%.

Пользуясь понятием дифференциала, эластичность можно представить иначе:

$$\mathrm{E}_x(y) = \frac{x}{y}\frac{dy}{dx} = \frac{\frac{dy}{y}}{\frac{dx}{x}} = \frac{d(\ln y)}{d(\ln x)}.$$

Геометрическая интерпретация

Эластичность функции $y = f(x)$ можно найти из графика этой функции. По определению эластичности $E_x(y) = \frac{x}{y}y' = \frac{x}{y}\operatorname{tg}\alpha$, где α — угол наклона касательной к функции $y = f(x)$ в точке $C(x_0, y_0)$ (рис. 23.4).

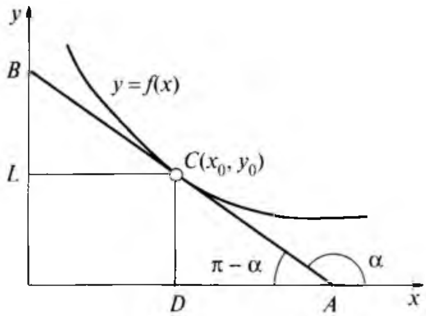

Рис. 23.4

Из треугольника ACD

$$\frac{CD}{AC} = \frac{y_0}{AC} = \sin(\pi - \alpha) = \sin\alpha.$$

Из треугольника BCL

$$\frac{LC}{BC} = \frac{x_0}{BC} = \cos(\pi - \alpha) = -\cos\alpha.$$

Отсюда

$$\frac{BC}{AC} = \frac{\frac{x_0}{-\cos\alpha}}{\frac{y_0}{\sin\alpha}} = -\frac{x_0}{y_0}f'(x_0) = -E_x(y),$$

т.е. эластичность убывающей функции равна отношению расстояний по касательной от точки C с координатами (x_0, y_0) до ее пересечения с осями ординат и абсцисс, взятому со знаком минус. Таким образом, если аккуратно построить график функции $y = f(x)$ и провести касательную к кривой в исследуемой точке $C(x_0, y_0)$, можно приблизительно определить величину эластичности функции в этой точке.

При приближенном определении эластичности по дискретному набору данных определение эластичности уже не столь однозначно.

как в непрерывном случае, поскольку в относительном изменении $\delta x = \frac{\Delta x}{x} = \frac{x_2 - x_1}{x}$ не ясно, что брать в качестве x:

- первоначальное значение ($x = x_1$);
- конечное значение ($x = x_2$);
- среднее значение $x = \frac{x_1 + x_2}{2}$.

В зависимости от этого выбора различают:

- конечную (процентную) эластичность: $E_x(y) = \dfrac{\frac{y_1 - y_2}{y_1}}{\frac{x_2 - x_1}{x_1}}$;
- среднюю (дуговую) эластичность: $E_x(y) = \dfrac{\frac{2(y_1 - y_2)}{y_1 + y_2}}{\frac{2(x_2 - x_1)}{x_1 + x_2}}$;
- логарифмическую эластичность: $E_x(y) = \dfrac{\Delta \ln y}{\Delta \ln x} = \dfrac{\ln y_2 - \ln y_1}{\ln x_2 - \ln x_1} = \dfrac{\ln \frac{y_2}{y_1}}{\ln \frac{x_2}{x_1}}$.

Все эти выражения мало отличаются друг от друга при небольших относительных (процентных) изменениях величин x и y.

Свойства эластичности функции

Пусть функция $y = f(x)$ имеет конечную или бесконечную производную на промежутке. Вспомним, что производная есть отношение дифференциалов: $y' = \frac{dy}{dx}$.

1. Эластичность есть безразмерная величина:

$E_x(y) = E_{ax}(by)$.

Доказательство очевидно:

$$E_{ax}(by) = \frac{ax}{by}\frac{d(by)}{d(ax)} = \frac{x}{y}\frac{dy}{dx} = \frac{x}{y}y'.$$

2. Эластичности обратных функций есть обратные величины:

$$E_x(y) = \frac{1}{E_y(x)}.$$

Для доказательства также перейдем к дифференциалам:

$$E_x(y) = \frac{x}{y}\frac{dy}{dx} = \frac{1}{\frac{y}{x}\frac{dx}{dy}} = \frac{1}{E_y(x)}.$$

3. Эластичность произведения функций $u = u(x)$ и $v = v(x)$ равна сумме их эластичностей:

$E_x(uv) = E_x(u) + E_x(v).$

При доказательстве свойства воспользуемся следующим свойством дифференциала: $d(uv) = v\,du + u\,dv$. Тогда

$$E_x(uv) = \frac{x}{uv}\frac{d(uv)}{dx} = \frac{x}{uv}\frac{vdu + udv}{dx} = \frac{x}{u}\frac{du}{dx} + \frac{x}{v}\frac{dv}{dx} = E_x(u) + E_x(v).$$

4. Эластичность отношения функций $u = u(x)$ и $v = v(x)$ равна разности их эластичностей:

$$E_x\left(\frac{u}{v}\right) = E_x(u) - E_x(v).$$

Доказательство аналогично:

$$E_x\left(\frac{u}{v}\right) = \frac{x}{\frac{u}{v}}\frac{d\left(\frac{u}{v}\right)}{dx} = \frac{x}{\frac{u}{v}}\frac{vdu - udv}{v^2dx} = \frac{x}{u}\frac{du}{dx} - \frac{x}{v}\frac{dv}{dx} = E_x(u) - E_x(v).$$

5. Эластичность суммы функций $u = u(x)$ и $v = v(x)$ равна сумме их эластичностей, взятых с соответствующими весами:

$$E_x(u + v) = \frac{u}{u+v}E_x(u) + \frac{v}{u+v}E_x(v).$$

Доказательство:

$$E_x(u + v) = \frac{x}{u+v}\frac{d(u+v)}{dx} = \frac{x}{u+v}\frac{du}{dx}\frac{u}{u} + \frac{x}{u+v}\frac{dv}{dx}\frac{v}{v} =$$

$$= \frac{u}{u+v}\cdot\frac{x}{u}\frac{du}{dx} + \frac{v}{u+v}\cdot\frac{x}{v}\frac{dv}{dx} = \frac{u}{u+v}E_x(u) + \frac{v}{u+v}E_x(v).$$

Эластичность элементарных функций

Вычислим эластичности некоторых функций.

1. Степенная функция $y = x^\alpha$. Ее эластичность является постоянной величиной:

$$E_x(x^\alpha) = \frac{x}{x^\alpha}\frac{d(x^\alpha)}{dx} = \frac{x}{x^\alpha}\frac{\alpha x^{\alpha-1}dx}{dx} = \alpha.$$

2. Показательная функция $y = a^x$. Эластичность показательной функции пропорциональна величине x:

$$E_x(a^x) = \frac{x}{a^x}\frac{d(a^x)}{dx} = \frac{x}{a^x}\frac{a^x \ln a \cdot dx}{dx} = x\ln a.$$

3. Логарифмическая функция $y = \ln x$.

$$E_x(\ln x) = \frac{x}{\ln x}\frac{d(\ln x)}{dx} = \frac{1}{\ln x}.$$

4. Линейная функция $y = ax + b$.

$$E_x(ax+b) = \frac{x}{ax+b}\frac{d(ax+b)}{dx} = \frac{ax}{ax+b}.$$

Если график линейной функции имеет отрицательный наклон ($a < 0$), то эластичность функции меняется от нуля в точке y_m пересечения графиком оси y до минус бесконечности ($-\infty$) в точке пересечения оси x, проходя через значение (-1) в средней точке. Таким образом, хотя прямая имеет постоянный наклон, ее эластичность зависит не только от наклона, но и от того, в какой точке x мы ее находим (рис. 23.5).

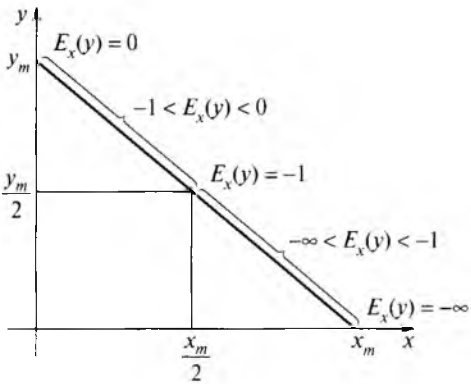

Рис. 23.5

Функция в зависимости от величины своей эластичности может быть:

- совершенно эластичная $|E_x(y)| = +\infty$;
- эластичная $1 < |E_x(y)| < +\infty$;
- неэластичная $0 < |E_x(y)| < 1$;
- совершенно неэластичная $E_x(y) = 0$.

Рассмотрим различные виды эластичностей

- спроса по цене:

$$E_p(q) = \frac{\frac{dq}{q}}{\frac{dp}{p}} = \frac{dq}{dp}\frac{p}{q},$$

характеризует относительное изменение, выраженное в процентах, величины спроса на товар при изменении цены этого товара на 1% и характеризует чувствительность потребителей к изменению цен на продукцию;

- спроса по доходу:

$$E_i(q) = \frac{\frac{dq}{q}}{\frac{dp}{p}} = \frac{dq}{di}\frac{i}{q},$$

показывает относительное изменение в процентах величины спроса на какое-либо благо при изменении дохода потребителей этого блага на 1%. Положительная эластичность спроса по доходу характеризует нормальные (качественные) товары, а отрицательная величина — малоценные (некачественные) товары.

Так, высокий положительный коэффициент спроса по доходу в отрасли указывает, что ее вклад в экономический рост больше, чем доля в структуре экономики, и она имеет шансы на расширение и процветание в будущем. Наоборот, если коэффициент эластичности спроса на продукцию отрасли по доходу имеет небольшое положительное или отрицательное значение, то ее может ожидать застой и перспектива сокращения производства;

- перекрестная эластичность спроса по цене:

$$E_{p_j}(q_i) = \frac{\frac{dq_i}{q_i}}{\frac{dp_j}{p_j}} = \frac{dq_i}{dp_j}\frac{p_j}{q_i},$$

описывает относительное изменение в процентах величины спроса на один товар при изменении цены на другой товар, замещающий или дополняющий его в потреблении, на 1%. Положительный знак перекрестной эластичности спроса по цене свидетельствует о замещаемости товара, а отрицательный — о его дополняемости;

- ценовая эластичность ресурсов:

$$E_{p_i}(R_i) = \frac{\frac{dR_i}{R_i}}{\frac{dp_i}{p_i}} = \frac{dR_i}{dp_i}\frac{p_i}{R_i},$$

указывает относительное изменение в процентах величины спроса на какой-либо ресурс, например труд, при изменении цены этого ресурса, в данном случае заработной платы, на 1%;

- эластичность замещения одного ресурса другим:

$$E_{R_j}(R_i) = \frac{\frac{dR_i}{R_i}}{\frac{dR_j}{R_j}} = \frac{dR_i}{dR_j}\frac{R_j}{R_i},$$

характеризует необходимое изменение в процентах величины одного ресурса, например капитала, при изменении количества другого ресурса, например труда, на 1%, с тем чтобы выпуск при этом не изменился.

На эластичность спроса влияют различные факторы. Это возможность при желании поменять один вид товара на другой, т.е. замещаемость товара, доля в расходах потребителя, фактор времени, делающий разными краткосрочную и долгосрочную эластичности, и др.

1. *Замещаемость товара.* Эластичность спроса по цене на товар увеличивается с ростом замещаемости этого товара. Замещаемость блага зависит от количества замещающих товаров, от степени объединения в одном товаре различных свойств, желательных для потребителя, другими словами, от агрегированности товара. Чем больше у потребителей возможностей заместить потребление данного товара потреблением других товаров, тем выше эластичность спроса на него. Например, товар «хлеб» слабо заменяется другими продуктами в отличие от товара «мясо», которое может быть при необходимости заменено рыбой или овощами. Пословицы, например «хлеб — всему голова», рекомендации отцов церкви, врачей-диетологов способствуют уменьшению эластичности хлеба как товара. Мясо не рекомендуют потреблять при некоторых отклонениях в состоянии здоровья. Верующим указывают на выработанную веками необходимость соблюдать постные дни и недели. Мясо — сравнительно более эластичный товар.

Чем выше степень агрегированности товара, тем меньше у потребителей возможностей заместить потребление данного товара потреблением других и тем ниже эластичность спроса на него. Например, эластичность спроса на мясные изделия ниже, чем на колбасу, а тем более конкретного наименования.

Введение в 1980 г. 6%-го налога на бензин в Вашингтоне (округ Колумбия), эластичность спроса на который, по оценкам аналитиков, составляла 0,2, привело к 33%-му падению спроса (что соответствует эластичности 5,5), и через 2 месяца налог был отменен. Причина этого — «узкое» определение бензина в штате *Washington D.C.*, не включившее в себя бензин из соседних штатов *Meriland* и *Virginia*, которым потребители и стали заменять подорожавший в Вашингтоне бензин.

2. *Доля в расходах.* Эластичность спроса по цене тем выше, чем выше удельный вес расходов на данный товар в расходах потребителя. В Дании и соседних странах при резком повышении цен на бензин в 1980 г. люди пересели на велосипеды. В то же время спрос потребителя на спички практически не изменится, даже если их цена возрастет в несколько раз.

3. *Фактор времени.* Эластичность спроса по цене, как правило, увеличивается со временем. Это обычно обосновывается тем, что за долгосрочный промежуток времени потребители могут изменить свои привычки, увеличить свои доходы и найти больше заменителей данному благу. Многие российские граждане за последние 10 лет пересели с отечественных автомобилей на иномарки, хотя цены на автомобили различаются в несколько раз.

23.3. ПОНЯТИЕ ПРОИЗВОДСТВЕННОЙ ФУНКЦИИ. МАКРОЭКОНОМИЧЕСКАЯ ФУНКЦИЯ КОББА—ДУГЛАСА

Работа экономического субъекта может быть описана с помощью *производственной функции*. Цель такого описания — получать количественные оценки успешности работы, находить пути оптимизации издержек и максимизации прибыли.

Производственной функцией (в дальнейшем — ПФ) называется зависимость между объемами затрачиваемых или используемых ресурсов и объемом выпускаемой продукции. В качестве одного из ресурсов рассматривают объем используемого в течение некоторого промежутка времени капитала: $x_1 = K$. Другим ресурсом является затрачиваемый в течение этого промежутка времени живой труд: $x_2 = D$. Третьим фактором может служить объем используемых за это время природных ресурсов.

Общий вид производственной функции (ПФ)

$$y = f(x_1, x_2, \ldots, x_n),$$

где переменные $x_1, x_2, \ldots, x_n$ есть объемы затрачиваемых ресурсов, а зависимая переменная y имеет смысл величины объема выпуска продукции. В экономике обычно принимается условие

$x_i \geq 0$, где $i = 1, 2, \ldots, n$.

Если переменные $x_1, x_2, \ldots, x_n$ являются независимыми величинами, ПФ называется статической. В случае зависимости переменных $x_1, x_2, \ldots, x_n$ от времени

$x_1 = x_1(t), x_2 = x_2(t), \ldots, x_n = x_n(t)$

ПФ становится динамической и сложной функцией аргумента t.

Требования к производственной функции

Предполагается, что функция f удовлетворяет некоторым условиям (не обязательно всем), вытекающим из общеэкономических соображений.

1. $f(0, 0, \ldots, 0) = 0$.

Условие означает, что выпуск при отсутствии ресурсов невозможен.

2. При $x'_i > x_i$, где $i = 1, 2, \ldots, n$, справедливо неравенство $f(x'_1, x'_2, \ldots, x'_n) > f(x_1, x_2, \ldots, x_n)$.

Функция является возрастающей по всем переменным, т.е. при увеличении затрат всех ресурсов выпуск будет расти.

3. $\dfrac{\partial f}{\partial x_i} \geq 0$ для всех $i = 1, 2, \ldots, n$.

При увеличении затрат любого ресурса и неизменных остальных выпуск не уменьшается.

4. $\dfrac{\partial^2 f}{\partial x_i^2} \leq 0$ для всех $i = 1, 2, \ldots, n$.

Это свойство означает, что с ростом затрат любого ресурса при неизменных остальных прирост выпуска на каждую дополнительную единицу этого ресурса не растет.

5. $\dfrac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ для всех $i = 1, 2, \ldots, n; j = 1, 2, \ldots, n$.

Неотрицательность всех смешанных производных второго порядка означает, что эффективность затрат любого из ресурсов при увеличении затрат какого-либо другого ресурса и неизменных остальных не снижается.

6. $f(t \cdot x_1, t \cdot x_2, \ldots, t \cdot x_n) = t^p f(x_1, x_2, \ldots, x_n)$.

Равенство утверждает однородность функции $y = f(x_1, x_2, \ldots, x_n)$ со степенью $p > 0$. При $p > 1$ увеличение затрат ресурсов в t раз приводит к увеличению объема выпуска в $t^p > t$ раз.

Иногда требуется, чтобы ПФ, помимо указанных свойств, обладала и некоторыми другими.

Конкретный вид ПФ выводят, исходя из гипотез об экономической деятельности и ее характеристиках. В литературе предложено множество конкретных ПФ. Рассмотрим одну из них, построенную на основе функции Кобба—Дугласа двух переменных.

Функция Кобба—Дугласа как макроэкономическая производственная функция

Изучая эмпирические связи между объемом выпуска Y, капитальными затратами K и затратами труда D, экономист Пол Дуглас в 1927 г. обнаружил определенную закономерность. Математик Чарльз Кобб, к которому он обратился за помощью, предложил следующую формулу:

$$Y = aK^{\alpha}D^{1-\alpha},$$

наилучшим образом описывающую эмпирическую зависимость.

В дальнейшем зависимость была обобщена добавлением показателя β, в результате чего зависимость приобрела в научной литературе название производственной функции Кобба—Дугласа:

$$Y = aK^{\alpha}D^{\beta}.$$

Поскольку увеличение затрат производственных факторов должно вызывать рост выпуска, в отношении α и β были сделаны предположения: $\alpha > 0$, $\beta > 0$. Эмпирически многократно подтверждена и другая гипотеза: $\alpha < 1$, $\beta < 1$, ибо увеличение затрат капитала и труда приводит к более медленному, чем прямо пропорциональная зависимость, увеличению выпуска.

РЕЗЮМЕ

Рассмотренный в этом разделе инструментарий количественных методов дает краткое представление о возможности использования функций и графиков, производных и интегралов в управлении. В количественном анализе и принятии эффективных управленческих решений в процессе комплексного исследования экономической ситуации важны как накопленные и суммарные величины, так и средние и предельные показатели. Базирующееся на этом фундаменте важное в количественном анализе понятие эластичности позволяет исследовать относительное изменение исследуемого экономического показателя под действием единичного относительного изменения экономического фактора, от которого он зависит, при неизменных остальных влияющих на него факторах.

Глава 24
ЭЛЕМЕНТЫ ЛИНЕЙНОГО РЕГРЕССИОННОГО, КОРРЕЛЯЦИОННОГО И ДИСПЕРСИОННОГО АНАЛИЗА

24.1. ФУНКЦИОНАЛЬНАЯ, СТАТИСТИЧЕСКАЯ И КОРРЕЛЯЦИОННАЯ ЗАВИСИМОСТИ

Для описания, анализа и прогнозирования явлений и процессов в экономике применяют математические модели в форме уравнений или функций. Модели экономического объекта, отражая основные его свойства и абстрагируясь от второстепенных, позволяют судить о его поведении в определенных условиях.

Одной из основных задач математической статистики является исследование зависимости между двумя или несколькими переменными. Две переменные X и Y могут быть независимыми или связанными функциональной или статистической зависимостью. Строгая функциональная зависимость реализуется редко, так как хотя бы одна из переменных подвержена случайным факторам.

Статистической зависимостью называется такая зависимость, при которой изменение одной из величин влечет за собой изменение распределения другой.

Если при изменении одной из величин изменяется среднее значение другой, то в этом случае статистическая зависимость является корреляционной. *Корреляционной зависимостью* между двумя переменными величинами называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

Основная цель изучения зависимостей между случайными величинами заключается в прогнозировании с данной вероятностью области значений одной случайной величины на основании наблюдаемых значений другой случайной величины.

На практике при исследовании зависимости между случайными величинами X и Y часто ограничиваются исследованием зависимости между X и условным математическим ожиданием Y.

Напомним, что зависимости такого рода называются *регрессионными зависимостями*. Функцию

$$\varphi(x) = \int_{-\infty}^{+\infty} yp(y \mid x)dy = M(Y \mid X = x)$$

называют функцией регрессии первого рода (или *модельной функцией регрессии*) Y по X, а ее график — *линией регрессии* Y по X.

Аналогично функцию

$$\psi(y) = \int_{-\infty}^{+\infty} xp(x \mid y)dx = M(X \mid Y = y)$$

называют функцией регрессии X по Y.

Например, уравнение регрессии Y по X двумерной случайной величины (X, Y), распределенной по нормальному закону, имеет вид

$$\varphi(x) = a_y + r\frac{\sigma_y}{\sigma_x}(x - a_x),$$

т.е. функция является линейной и график ее имеет вид прямой линии, проходящей через центр распределения $C(a_x, a_y)$. Аналогично линией регрессии X по Y будет прямая

$$\psi(y) = a_x + r\frac{\sigma_x}{\sigma_y}(y - a_y).$$

Уравнение регрессии $\hat{y} = \varphi(x)$ делает возможными точечные предсказания значений условных математических ожиданий составляющей Y двумерной случайной величины (X, Y) по значению составляющей $X = x$. Однако для такого прогноза необходимо знать закон распределения двумерной случайной величины (X, Y), который на практике, как правило, неизвестен.

Поэтому исследование зависимости случайной величины от ряда неслучайных и случайных величин приводит к моделям регрессии на базе выборочных данных. В качестве оценок условных математических ожиданий принимаются условные средние, которые находятся по данным выборки (x_i, y_i), $i = 1, 2, ..., n$ объема n.

Условным средним $\bar{y}_x$ называется среднее арифметическое наблюдавшихся значений y, соответствующих значению $X = x$.

Например, если при $x = 2$ величина y принимает значения $y_1 = 5$, $y_2 = 6$, $y_3 = 10$, $y_4 = 8$, тогда $\bar{y}_x = \dfrac{5 + 6 + 10 + 8}{4} = \dfrac{29}{4}$.

Условным средним $\bar{x}_y$ называется среднее арифметическое наблюдавшихся значений X, соответствующих $Y = y$.

Оценку функции регрессии называют эмпирической регрессией или функцией регрессии второго рода. Поскольку функции регрессии обладают свойством минимальности, т.е. среднее квадратическое отклонение случайной величины Y от функции $\varphi(x)$ является наименьшим, то для нахождения эмпирических уравнений регрессии $\hat{y} = \varphi(x, a, ..., c)$ применяется метод наименьших квадратов.

Регрессионный анализ — это анализ функций регрессии первого и второго рода. Перед проведением регрессионного анализа необходимо по статистическим данным выбрать общий вид эмпирической функции регрессии. На практике часто считают, что функция регрессии — это линейная функция: $\hat{y} = a + bx$. Проверка этой гипотезы основывается на результатах оценки коэффициента корреляции.

Если установлено, что зависимость между некоторыми наблюдаемыми величинами существует, то на практике важно знать, какая она — сильная или слабая, положительная или отрицательная. Для выяснения этих обстоятельств используется корреляционный анализ.

Корреляционный анализ — это анализ оценок коэффициента корреляции. Он позволяет ответить на вопрос, существует ли линейная функциональная зависимость между случайными величинами X и Y, и позволяет измерить степень близости статистической зависимости к функциональной.

24.2. РЕГРЕССИОННЫЕ МОДЕЛИ КАК ИНСТРУМЕНТ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

В случае применения регрессионных моделей результат действия экономической системы или объекта в виде одного или нескольких выходных показателей можно представить как функцию влияющих на него факторов. Некоторые из них оказывают существенное влияние на результат, другие — весьма незначительное. Как правило, существенных факторов мало, в то время как несущественных — достаточно много, поэтому пренебрегать полностью ими нельзя.

К числу основных факторов при изучении экономического объекта обычно относятся настоящий труд (или трудовые ресурсы) и прошлый труд (энергия, сырье, материалы, оборудование, здания, сооружения и т.д.). Вместе с тем труд прилагается при определенном состоянии внешней среды, т.е. при определенных экономических и природных условиях. Поэтому соответствующие факторы также должны найти отражение в модели.

В процессе производства и распределения продукции имеют место массовые многократно повторяющиеся события. Несмотря на развитие экономики и ее отдельных частей, на протяжении относительно небольших временных периодов и в пределах отдельных экономических подсистем имеет место стабильность в условиях совершения массовых событий. Так, прежде всего в процессе прогнозирования подразумевается возможность многократного повторения производственной ситуации при наличии других существенных и несущественных факторов, однако при относительно стабильном комплексе внешних условий.

Таким образом, приходим к следующей теоретико-вероятностной схеме. Результирующий показатель y является функцией существенных ($x_1, x_2, ..., x_k$) и несущественных факторов ($\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$): $y = F(x_1, ..., x_k, \varepsilon_1, ..., \varepsilon_n)$.

Несмотря на инертность поведения экономических систем, особенно больших, возможно их резкое изменение в связи с достижениями научно-технического прогресса. В новых условиях, вообще говоря, следует рассматривать другую функцию. Однако исходя из предположения эволюторности поведения экономической системы на относительно небольшом временном интервале и при сравнительно малых изменениях переменных, в принципе, с помощью полиномов можно с высокой степенью приближения описать любую эволюторную функцию. В случае же невозможности линеаризировать функцию (разлагая в ряд или преобразуя ее) при небольшом числе коэффициентов используют методы нелинейного оценивания.

Линейная регрессия двух переменных и ее различные экономические приложения более подробно разобраны ранее.

24.3. ВЫБОРОЧНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

Основными характеристиками, описывающими степень связи между составляющими X и Y двумерной случайной величины (X, Y), являются ковариация $\mathrm{cov}(X, Y) = \mu_{xy} = M(X - a_x)(Y - a_y)$ и коэффициент корреляции $\rho_{xy} = \dfrac{\mu_{xy}}{\sigma_x \sigma_y}$, который является мерой линейной зависимости между X и Y.

Так как в случае статистических данных закон распределения двумерной случайной величины неизвестен, то для оценки тесноты связи применяется эмпирическая (или выборочная) ковариация $\hat{\mu}_{xy} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ и эмпирический (или выборочный) коэффициент корреляции.

Выборочным коэффициентом корреляции для выборки вида (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) называется выборочная характеристика

$$r_{xy} = \frac{\hat{\mu}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Выборочный коэффициент корреляции может принимать значения от -1 до $+1$. Он сходится к теоретическому коэффициенту корреляции соответствующих случайных величин, если тот существует. По абсолютной величине и знаку коэффициента можно судить о степени зависимости (сильной или слабой) и ее характере (положительном или отрицательном).

Выборочный коэффициент корреляции обычно используется в предположении о нормальности данных, так как в данном случае из равенства нулю теоретического коэффициента ρ следует независимость случайных величин (в более общем случае это неверно). В случае нормального распределения можно проверить гипотезу $\rho = 0$. Пусть $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. Если гипотеза $\rho = 0$ верна, то T подчиняется распределению Стьюдента с $n - 2$ степенями свободы. При уровне значимости α выберем критическую точку $t_{кр} = t_{кр}(\alpha; n - 2)$ для двусторонней области. Если $|T| < t_{кр}$, гипотеза $\rho = 0$ принимается, в противном случае — отвергается.

В случае когда нормальность данных нарушается, применение выборочного коэффициента корреляции может вести к ошибкам. Либо мы «не заметим» зависимость между величинами, либо получим *ложную* корреляцию. Существуют коэффициенты и методы, свободные от предположения о нормальности.

Наблюдения всегда можно упорядочить по возрастанию какой-либо переменной (x или y). *Рангом наблюдения* называется его номер в таком ряду. Если какое-то значение переменной встречается несколько раз, ему приписывается *средний* ранг. Обозначим ранги наблюдений по возрастанию x и y через r_i и s_i соответственно. Пусть

$$S = \sum_{i=1}^{n}(r_i - s_i)^2.$$

Выборочным коэффициентом ранговой корреляции Спирмена называется величина

$$r_S = 1 - \frac{6S}{n^3 - n}.$$

Этот коэффициент также может принимать значения от –1 до +1. Аналогичным образом он отражает силу и характер зависимости между величинами. Для проверки гипотезы о независимости случайных величин существуют специальные таблицы критических точек. Однако при больших n можно проверять гипотезу так же, как для обычного выборочного коэффициента корреляции.

Заметим, что с помощью коэффициента Спирмена можно анализировать также ситуации, когда некоторый признак объекта («качество», «привлекательность» и т.п.) нельзя строго выразить численно, но можно упорядочить объекты по его возрастанию или убыванию, т.е. *проранжировать* их.

Пример 24.1. В таблице представлены средние цены на растительное масло и сахар-песок (руб.) в 12 городах Центрального района России на июнь 1996 г. (рис. 24.1).

Город	Цена на масло	Цена на сахар
Брянск	7726	3410
Владимир	7880	3183
Иваново	6182	3209
Калуга	8237	3400
Кострома	8750	3600
Москва	11024	4418
Орел	8456	3634
Рязань	9172	4033
Смоленск	8320	3909
Тверь	7083	3416
Тула	8259	3486
Ярославль	7991	3938

Вычислить выборочный коэффициент корреляции между ценами на растительное масло и сахар. Проверить нулевую гипотезу на уровне значимости $\alpha = 0{,}1$.

Решение. Для проверки нулевой гипотезы вычисляем выборочный коэффициент корреляции, равный $r = 0{,}82$, и наблюдаемое значение $T = \dfrac{0{,}82\sqrt{10}}{\sqrt{1-0{,}82^2}} \approx 4{,}53$. Из таблицы критических точек распределения Стьюдента по заданному уровню значимости $\alpha = 0{,}1$ и числу степеней свободы $n = 10$ находим критическую точку: $t_{кр}(0{,}1;\ 10) = 1{,}81$. Поскольку $|T| > t_{кр}$, нулевая гипотеза отвергается. Связь между ценами на растительное масло и сахар оказывается довольно сильной и положительной.

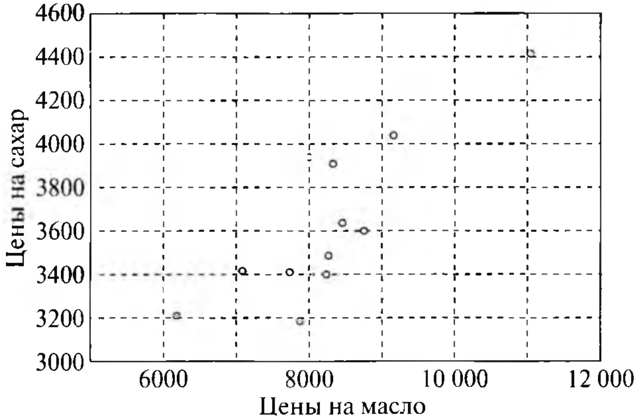

Рис. 24.1. Цены на масло и сахар в городах Центральной России

Пример 24.2. Два эксперта проранжировали 11 фирм в порядке их привлекательности для инвестиций. Получены следующие последовательности рангов фирм:

r_i	1	2	3	4	5	6	7	8	9	10	11
s_i	1	2	3	5	4	9	8	11	6	7	10

С помощью коэффициента Спирмена проверить, насколько согласуются мнения экспертов. Проверить нулевую гипотезу на уровне значимости $\alpha = 0{,}05$.

Решение. Вычисляем $S = 1^2 + 1^2 + 3^2 + 1^2 + 3^2 + 3^2 + 3^2 + 1^2 = 40$ и находим $r_S = 1 - \dfrac{6 \cdot 40}{11^3 - 11} \approx 0{,}82$. Для проверки нулевой гипотезы вычисляем $T = \dfrac{0{,}82\sqrt{9}}{\sqrt{1 - 0{,}82^2}} \approx 4{,}29$ и из таблицы критических точек распределения Стьюдента по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $n = 9$ находим критическую точку: $t_{\text{кр}}(0{,}05;\ 9) = 2{,}26$. Поскольку $|T| > t_{\text{кр}}$, нулевая гипотеза отвергается. Мнения экспертов достаточно хорошо согласуются между собой.

24.4. ЭЛЕМЕНТЫ ДИСПЕРСИОННОГО АНАЛИЗА

Предположим, что имеется k выборок с объемами $n_1, n_2, \dots n_k$, $N = n_1 + n_2 + \dots + n_k$, и наблюдения можно представить в виде: $x_{ij} = a_j + \varepsilon_{ij}$, где i — номер наблюдения в выборке; j — номер выборки; a_j — групповые математические ожидания; ε_{ij} — случайные ошибки с $M\varepsilon_{ij} = 0$, о которых предполагается, что они независимы и одинаково распределены.

Подобная ситуация возникает, когда существует некий *фактор*, принимающий k различных значений (называемых *уровнями*), и каждая группа объектов, чьи признаки мы измеряем, подвергается воздействию определенного уровня этого фактора. Методы математической статистики, изучающие воздействие одного фактора на объекты и их признаки, называют в совокупности *однофакторным анализом*.

Предположим, что ошибки нормально распределены: $\varepsilon_{ij} \in N(0, \sigma^2)$. Тогда можно изучать влияние фактора, вычисляя дисперсии некоторых величин. Совокупность этих методов называется *однофакторным дисперсионным анализом*.

Основной гипотезой, нуждающейся в проверке, является гипотеза о равенстве групповых средних — H_0: $a_1 = a_2 = ... = a_k$. Иными словами, мы проверяем гипотезу о том, что фактор *вообще не влияет* на наблюдения. В случае нормальных ошибок ее можно проверить, вычислив две разные оценки дисперсии σ^2, а именно:

$$\sigma^{2*} = \frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n_i}(x_{ij} - x_{.j})^2; \quad \sigma^{2**} = \frac{1}{k-1}\sum_{j=1}^{k} n_j(x_{.j} - \bar{x})^2,$$

где через $x_{.j}$ обозначены групповые выборочные средние. Первую величину называют *остаточной дисперсией*, а вторую — *факторной*.

При нарушении гипотезы H_0 оценка σ^{2**} имеет тенденцию к возрастанию тем большему, чем больше отклонение от H_0. Проверить гипотезу можно с помощью уже знакомого нам дисперсионного отношения Фишера: $F = \sigma^{2**}/\sigma^{2*}$, которое должно иметь распределение Фишера с $k - 1$ и $N - k$ степенями свободы. При заданном уровне значимости α можно найти критическую точку $F_{кр} = F_{кр}(a; k - 1; N - k)$. Если $F < F_{кр}$, то нулевая гипотеза принимается, иначе — отвергается. Заметим, что если $F < 1$, то следует сразу принять гипотезу H_0, поскольку $F_{кр}$ всегда больше единицы.

Если гипотеза о равенстве средних не подтверждается, имеет смысл оценивать величины a_j по отдельности. Получаем для них следующие доверительные интервалы (с надежностью γ):

$$x_{.j} - \frac{\sigma^*}{\sqrt{n_j}}t_\gamma < a_j < x_{.j} + \frac{\sigma^*}{\sqrt{n_j}}t_\gamma,$$

где t_γ — критическая точка распределения Стьюдента с $N - k$ степенями свободы (для двусторонней области): $t_\gamma = t_{кр}(1 - \gamma; N - k)$.

Существуют и более сложные методы однофакторного анализа, не требующие предположения о нормальности наблюдений. Однако здесь мы будем придерживаться классического дисперсионного анализа.

Пример 24.3. Банк имеет по четыре отделения в трех городах. Текущие объемы денежных вкладов (в условных единицах) представлены в следующей таблице.

Город / Отделение	1	2	3
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34

Можно ли утверждать на уровне значимости $\alpha = 0{,}05$, что в среднем дела идут одинаково хорошо во всех трех городах? Построить доверительные интервалы для средних по городам с надежностью 90%.

Решение. Вычисляем групповые средние: $x_{.1} = 35$, $x_{.2} = 25$, $x_{.3} = 27$ и общее среднее $\bar{x} = 29$. Затем вычисляем оценки дисперсий: $\sigma^{2**} = 112$; $\sigma^{2*} \approx 22{,}67$. Получаем $F \approx 4{,}94$. По таблице находим критическую точку: $F_{\text{кр}}(0{,}05;\ 2;\ 9) = 4{,}26$. Имеем $F > F_{\text{кр}}$, следовательно, групповые средние различаются. Чтобы построить доверительные интервалы, находим $\sigma^* \approx 4{,}76$ и $t_\gamma = t_{\text{кр}}(0{,}1;\ 9) = 1{,}83$. Точность доверительных интервалов во всех группах составляет $\frac{\sigma^* t_\gamma}{\sqrt{n_j}} \approx 4{,}36$, отсюда $30{,}64 < a_1 < 39{,}36$; $20{,}64 < a_2 < 29{,}36$; $22{,}64 < a_3 < 31{,}36$. Таким образом, лучше всего дела идут в первом городе с явным отрывом от двух других, где ситуация примерно одинакова (доверительные интервалы перекрываются).

Пример 24.4. В цехе работает четыре бригады по семь рабочих. Выработка каждого (в условных единицах) представлена в следующей таблице.

Бригада / Рабочий	1	2	3	4
1	51	52	56	54
2	59	58	56	58
3	53	66	58	62
4	59	69	58	64
5	63	70	70	66
6	69	72	74	67
7	72	74	78	69

Можно ли утверждать на уровне значимости $\alpha = 0{,}05$, что все бригады работают в среднем одинаково?

Решение. Вычисляем групповые средние: $x_{.1} = 60{,}9$; $x_{.2} = 65{,}9$; $x_{.3} = 64{,}3$; $x_{.4} = 62{,}9$ и общее среднее $\bar{x} = 63{,}5$. Затем вычисляем оценки дисперсий: $\sigma^{2**} = 31{,}67$; $\sigma^{2*} \approx 60{,}25$. Получаем $F < 1$, так что следует принять гипотезу о равенстве групповых средних. Все бригады работают в среднем одинаково.

Глава 25
ЭЛЕМЕНТЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

25.1. ОСНОВНЫЕ ПОНЯТИЯ В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

Временным рядом называется ряд наблюдений некоторой величины, сделанных в последовательные моменты времени (обычно через равные интервалы).

В экономических приложениях это могут быть, например, объемы производства или продаж каких-либо товаров, цены акций, курсы валют и т.п. Данные могут регистрироваться ежегодно, ежемесячно или ежедневно, а учитывая современные компьютерные технологии — даже ежечасно и ежеминутно.

Основными целями анализа временных рядов является выявление присущих им закономерностей и прогнозирование на будущее (с учетом этих закономерностей).

Существует великое множество методов анализа временных рядов, которым посвящена обширная литература. Эти методы могут применяться по отдельности, в различных комбинациях и в разной последовательности. В зависимости от этого могут получаться различные результаты. Однако для «хороших» временных рядов они будут не очень сильно различаться между собой.

Предполагается, что временной ряд можно разделить на три составляющие:

1) *тренд*, т.е. некая общая закономерность к росту или убыванию;
2) *сезонная* (или *циклическая*) *компонента*;
3) *случайная компонента* (шум).

В качестве тренда обычно берут достаточно простую функцию. Наиболее часто используются линейный и экспоненциальный тренды. Первый означает, что величина в среднем растет или убывает в арифметической прогрессии (на сколько-то в единицу времени), второй — что она меняется в геометрической прогрессии (во сколько-то раз или на сколько-то процентов в единицу времени). Иногда рассматривают и более сложные тренды, например полиномиальные.

Линейный тренд легко построить с помощью линейной регрессии, применяя метод наименьших квадратов к исходным данным. Экспоненциальный тренд можно построить, если перейти к логарифмам наблюдаемых величин, провести линейную регрессию для них, а затем взять экспоненту от полученной линейной функции.

Сезонной компоненты во временном ряду может не быть совсем, их может быть одна или несколько. Например, во временном ряду из ежемесячных данных, скорее всего, будет сезонная компонента с периодом 12 (годичный цикл). В ряду ежедневных наблюдений могут быть компоненты с периодом 7 (неделя), около 30 (месяц) и 365 (год) и т.д. В ежегодных наблюдениях могут присутствовать многолетние циклы (например, экономические циклы Кондратьева).

Сезонная компонента может быть аддитивной (когда ее амплитуда относительно «среднего» значения остается в среднем постоянной) и мультипликативной (когда колебания происходят пропорционально текущему «среднему», и их амплитуда может расти или убывать со временем согласно тренду). Для описания сезонной компоненты часто используются *сезонные индексы*, показывающие, насколько (или во сколько раз) значение в данной точке цикла отличается от некоторого «среднего» значения.

Следует отметить, что существуют временные ряды с *ложными* трендами и *ложными* циклами, когда закономерности, выявленные на одном отрезке наблюдений, не подтверждаются на других. Однако здесь мы такие ряды рассматривать не будем.

Общая схема анализа временного ряда обычно такова, что выявленные в нем закономерности мы исключаем путем некоторых преобразований и далее изучаем получившиеся величины, называемые *остатками*, пытаясь выявить закономерности уже в их поведении, и т.д. Так продолжается до тех пор, пока мы не дойдем до чисто случайных колебаний (независимых случайных величин).

25.2. ПРОСТЫЕ МЕТОДЫ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Рассмотрим некоторые простые методы анализа и прогнозирования временных рядов, не требующие глубокой теории, на примере нескольких практических задач.

Пример 25.1. В табл. 25.1 представлены данные об урожайности зерновых культур в СССР за 1951–1970 гг. (ц/га).

Построить линейный тренд.

Решение. Методом наименьших квадратов находим $\hat{tr}(t) = 6{,}97 + 0{,}34(t - 1950)$, где t — год наблюдения. На рис. 25.1 представлен график урожайности и линейный тренд. Видно, что в поведении остатков (отклонений от тренда) не проявляется какой-либо закономерности, и их можно считать чисто случайными.

Таблица 25.1

Год	Урожайность	Год	Урожайность
1951	7,4	1961	10,7
1952	8,6	1962	10,9
1953	7,8	1963	8,3
1954	7,7	1964	11,4
1955	8,4	1965	9,5
1956	9,9	1966	13,7
1957	8,4	1967	12,1
1958	11,1	1968	14,0
1959	10,4	1969	13,2
1960	10,9	1970	15,6

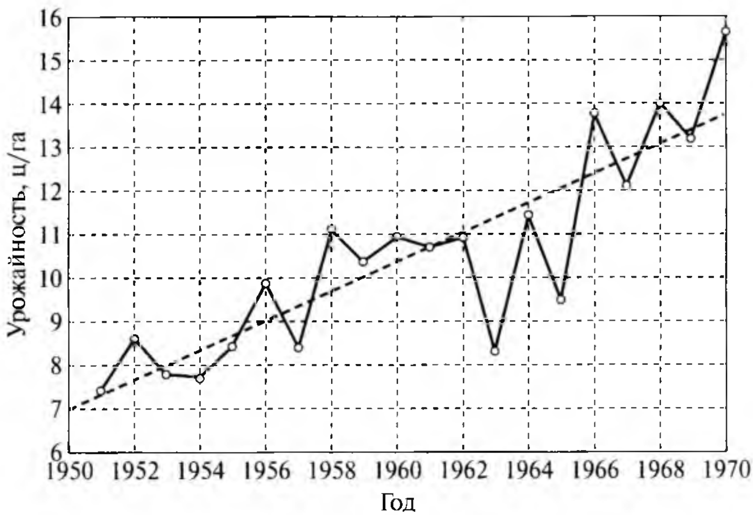

Рис. 25.1. Урожайность зерновых культур в СССР

Пример 25.2. В табл. 25.2 представлены данные о производстве молока (тыс. т) в России с января 1992 г. по октябрь 1996 г. (рис. 25.2).

Построить линейный тренд, вычислить сезонные индексы по данным 1992–1995 гг. Сделать прогноз на 1996 г. и сравнить его с реальными данными.

Решение. Методом наименьших квадратов получаем $\widehat{tr}(t) = 2899{,}9 - 26{,}64t$, где t — номер месяца (от 1 до 24). Производство молока имеет тенденцию к сокращению, обусловленную сокращением поголовья молочного скота, и подвержено сильным сезонным колебаниям с максимумом летом и минимумом зимой. При этом величина сезонных колебаний пропорциональна среднему уровню производства.

Таблица 25.2

Месяц \ Год	1992	1993	1994	1995	1996
Январь	2015	1759	1510	1172	1038
Февраль	2123	1773	1484	1226	1104
Март	2624	2361	1988	1651	1439
Апрель	2891	2649	2211	1859	1521
Май	3335	3203	2559	2392	1827
Июнь	4071	3936	3209	2864	2446
Июль	4040	3861	3204	2714	2369
Август	3392	3321	2687	2420	2081
сентябрь	2467	2438	2031	1925	1577
Октябрь	2092	1760	1506	1338	1081
Ноябрь	1494	1299	1050	984	
Декабрь	1562	1345	1054	1020	

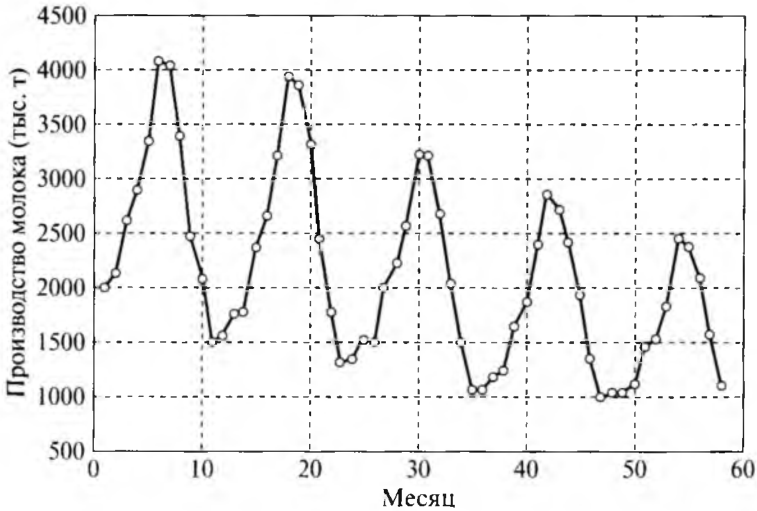

Рис. 25.2. Производство молока в России

Вычислим значения $\frac{x(t)}{\hat{tr}(t)}$, выразим их в процентах, а затем усредним для каждого месяца за 1992–1995 гг. и, таким образом, получим сезонные индексы (табл. 25.3).

Теперь, чтобы построить прогноз для какого-либо месяца следующих лет, надо умножить значение, получаемое из тренда, на соответствующий сезонный индекс. Для 1996 г. получаем следующую таблицу (результаты округлены до целых) (табл. 25.4).

Таблица 25.3.

Год / Месяц	1992	1993	1994	1995	Сезонный индекс (среднее)
Январь	70,13	68,88	67,59	61,22	66,96
Февраль	74,58	70,16	67,23	64,95	69,23
Март	93,05	94,43	91,16	88,71	91,84
Апрель	103,50	107,09	102,64	101,34	103,64
Май	120,54	130,89	120,29	132,32	126,01
Июнь	148,57	162,62	152,75	160,80	156,18
Июль	148,89	161,29	154,47	154,69	154,83
Август	126,25	140,30	131,23	140,06	134,46
Сентябрь	92,74	104,17	100,50	113,16	102,64
Октябрь	79,44	76,06	75,52	79,90	77,73
Ноябрь	57,31	56,79	53,37	59,71	56,80
Декабрь	60,54	59,49	54,30	62,91	59,31

Таблица 25.4

Год / Месяц	Тренд на 1996 г.	Прогноз на 1996 г.	Реальные данные
Январь	1595	1068	1038
Февраль	1568	1085	1104
Март	1541	1415	1439
Апрель	1515	1569	1521
Май	1488	1875	1827
Июнь	1461	2282	2446
Июль	1435	2221	2369
Август	1408	1893	2081
Сентябрь	1381	1418	1577
Октябрь	1355	1053	1081

Приведенные данные свидетельствуют о хорошем согласии прогноза с реальными данными в первые пять месяцев — относительная погрешность не превышает 3%. В последующие четыре месяца прогноз ниже реальных данных на 6–10%. Относительная ошибка за октябрь не превышает 3%. Заметим, что наибольшие различия прогноза с реальными данными наблюдаются в летние месяцы, сезонный индекс которых подвержен наибольшим колебаниям. Эти колебания из года в год имеют порядок 10–20%. С учетом этого можно признать в целом хорошее согласие прогноза с реальными данными.

Пример 25.3. Годовые прибыли фирмы (тыс. долл.) за пять лет представлены в табл. 25.5.

Таблица 25.5

Год	1	2	3	4	5
Прибыль	99	112	120	135	144

Построить экспоненциальный тренд и сделать прогноз на следующий год.

Решение. Построим временной ряд из логарифмов $y(t) = \ln x(t)$ (табл. 25.6).

Таблица 25.6

Год, t	1	2	3	4	5
$y(t)$	4,60	4,72	4,79	4,91	4.97

Проводя линейную регрессию, получаем линейный тренд $\hat{tr}_y(t) = 4{,}52 + 0{,}09t$.

Возвращаясь к исходным величинам, получаем $\hat{tr}_x(t) = \exp(4{,}52 + 0{,}09t)$.

В качестве прогноза на следующий год имеем $\hat{tr}_x(6) = 157{,}6$ (тыс. долл.).

25.3. СТАЦИОНАРНОСТЬ. АВТОКОРРЕЛЯЦИИ. ПЕРИОДОГРАММА

Временной ряд можно рассматривать как последовательность наблюдений некоторого случайного процесса $X(t)$, $t > 0$, в целочисленные моменты времени.

Случайный процесс $X(t)$ называется *стационарным в узком смысле*, если для любого $s > 0$ и любых $0 < t_1 < t_2 < \ldots < t_n$ распределение многомерного вектора $(X(t_1), X(t_2), \ldots, X(t_n))$ совпадает с распределением вектора $(X(t_1 + s), X(t_2 + s), \ldots, X(t_n + s))$.

Таким образом, стационарный в узком смысле случайный процесс ведет себя с любого момента так же (в статистическом смысле), как и сначала. Однако проверить это условие на практике бывает затруднительно.

Определим характеристики случайного процесса: *функцию математического ожидания* $m(t) = MX(t)$ и *ковариационную функцию* $B(t, s) = \text{cov}(X(t), X(s))$, $t, s > 0$.

Отсюда можем выразить дисперсию $DX(t) = B(t, t)$ и среднее квадратическое отклонение $\sigma(t) = \sqrt{DX(t)} = \sqrt{B(t,t)}$.

Корреляционной функцией называется $R(t, s) = B(t, s)/(\sigma(t)\sigma(s))$.

Случайный процесс $X(t)$ называется *стационарным в широком смысле*, если для него $m(t) = \text{const}$ и $B(t_1 + s, t_2 + s) = B(t_1, t_2)$ для любых $t_{1,2}$, $s > 0$.

Это условие уже проще проверить статистическими методами.

Из него также следует, что $R(t_1 + s, t_2 + s) = R(t_1, t_2)$ для любых $t_{1,2}$, $s > 0$.

Автокорреляционной функцией стационарного случайного процесса называется функция $r(t) = R(s, s+t)$, где $s > 0$ — любое.

Пример 25.4. Пусть $X(t) = \xi\cos(wt+\varphi)$, где $w > 0$ — число; $\xi \in N(0, 1)$, а φ равномерно распределено от 0 до 2π, и эти случайные величины независимы. Данный процесс стационарен как в широком, так и в узком смысле, причем $r(t) = \cos(wt)$.

Действительно, поскольку $M\xi = 0$ и $M\xi^2 = 1$, то $m(t) \equiv 0$ и

$$B(t_1, t_2) = MX(t_1)X(t_2) = M\{\xi^2\cos(wt_1+\varphi)\cos(wt_2+\varphi)\} =$$
$$= \int_0^{2\pi}\cos(wt_1+u)\cos(wt_2+u)\frac{du}{2\pi} =$$
$$= \frac{1}{2}\int_0^{2\pi}\{\cos(w(t_1-t_2)) + \cos(w(t_1+t_2)+2u)\}\frac{du}{2\pi} =$$
$$= \frac{1}{2}\cos(w(t_1-t_2)) + \frac{1}{4\pi}\sin(w(t_1+t_2)+2u)\Big|_0^{2\pi} = \frac{1}{2}\cos(w(t_1-t_2)),$$

так что $B(t_1+s, t_2+s) = B(t_1, t_2)$ для любых $t_{1,2}$, $s > 0$. Кроме того, $\sigma(t) = \sqrt{B(t,t)} = 1/\sqrt{2}$ и $R(t_1, t_2) = 2B(t_1, t_2) = \cos(w(t_1-t_2))$, откуда следует $r(t) = \cos(wt)$.

В случае временных рядов имеют дело с автокорреляционной функцией от целых чисел, т.е. $r(k)$. Ее можно оценить статистически следующим образом.

Выборочной автокорреляционной функцией называют

$$\hat{r}(k) = \frac{\sum_{i=1}^{n-k}(x_i-\bar{x})(x_{i+k}-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Данной формулой имеет смысл пользоваться только при достаточно больших объемах ряда n (не менее 50). Отметим, что точность приближения снижается с ростом k, поэтому на практике обычно ограничиваются изучением ограниченного числа первых значений $r(k)$ (как правило, не более $n/4$). Кроме того, полученная оценка оказывается смещенной (со смещением порядка $1/n$).

График выборочной автокорреляционной функции называют *коррелограммой*.

Построение коррелограммы для реальных временных рядов может давать различные результаты.

Если все значения выборочной автокорреляционной функции достаточно велики и испытывают слабую тенденцию к убыванию, это может свидетельствовать о наличии тренда (или общей нестационарности ряда).

Если после удаления тренда и анализа остатков получаем коррелограмму, похожую на график периодической функции, это может свидетельствовать о наличии одной сезонной (циклической) компоненты с соответствующим периодом.

Если после удаления тренда и сезонной компоненты коррелограмма имеет хаотический характер, причем значения выборочной автокорреляционной функции малы и попадают (все или почти все) в 95%-ный доверительный интервал $-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$, это свидетельствует о том, что остатки можно считать за «белый шум» (последовательность независимых одинаково распределенных случайных величин).

Некоторые другие случаи поведения коррелограмм будут рассмотрены в следующем параграфе.

Бывают ситуации, когда временные ряды содержат несколько циклических компонент с разными периодами, которые накладываются друг на друга и создают сложную картину. Для их анализа можно применить следующий метод, предложенный А. Шустером в 1898 г. Введем функции

$$A(\lambda) = \frac{2}{n}\sum_{k=1}^{n} x_k \cos\frac{2\pi k}{\lambda}, \quad B(\lambda) = \frac{2}{n}\sum_{k=1}^{n} x_k \sin\frac{2\pi k}{\lambda},$$

$$S^2(\lambda) = A^2(\lambda) + B^2(\lambda).$$

График функции $S^2(\lambda)$ называют *периодограммой*.

Построенная функция должна принимать большие значения (иметь локальные максимумы — пики) для тех значений λ, которые являются периодами имеющихся у временного ряда циклических составляющих. На практике в силу наличия случайности временного ряда это не всегда так и некоторые максимумы могут не иметь отношения к реальным периодам. Поэтому к анализу периодограммы следует подходить с осторожностью. В настоящее время разработаны также другие, более точные и более сложные методы.

25.4. МОДЕЛИ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО

До сих пор мы предполагали, что случайная компонента временного ряда представляет собой «белый шум» (последовательность независимых одинаково распределенных случайных величин). Однако на практике это может оказаться не так и величины могут быть зависимы. Эта зависимость может возникать и в результате каких-то преобразований ряда, что также необходимо учитывать. Для описания этих явлений, в частности, используются следующие модели.

Процессом скользящего среднего порядка p называется процесс вида

$$X(k) = \xi_k + \theta_1\xi_{k-1} + \ldots + \theta_p\xi_{k-p},$$

где θ_k — некоторые коэффициенты, а ξ_k образуют последовательность независимых одинаково распределенных случайных величин. Определенный таким образом процесс всегда является стационарным.

В коррелограмме процесса скользящего среднего достаточно велики только значения $\hat{r}(k)$ при k от 1 до p (те, для которых коэффициенты θ_k отличны от нуля), а остальные малы и хаотичны.

Процессом авторегрессии порядка q называется процесс вида

$$X(k) = \phi_1 X(k-1) + \ldots + \phi_q X(k-q) + \xi_k,$$

где ϕ_k — некоторые коэффициенты, а ξ_k образуют последовательность независимых одинаково распределенных случайных величин. Такой процесс является стационарным только при дополнительном условии: все корни *характеристического уравнения*

$$\lambda^q = \phi_1\lambda^{q-1} + \ldots + \phi_q$$

должны быть по модулю меньше единицы.

Коррелограмма процесса авторегрессии имеет вид экспоненциального убывания или затухающих колебаний (при $q > 1$).

В простейшем случае, при $q = 1$, автокорреляционная функция представляет собой просто геометрическую прогрессию: $r(k) = \phi_1^k$.

Как правило, рассматриваются процессы скользящего среднего и авторегрессии небольших порядков (первого или второго).

В случае когда мы применяем модель авторегрессии первого порядка, т.е.

$$X(k) = \phi_1 X(k-1) + \xi_k,$$

можно также провести линейную регрессию $X(k)$ по $X(k-1)$.

Пример 25.5. Временной ряд в табл. 25.7 описывает значения выхода 70 последовательных партий продукта в химическом производственном процессе (данные расположены по столбцам).

Выборочное среднее значение выхода продукции составляет 51,13 ед., а выборочное среднее квадратическое отклонение — 11,91 ед.

Тренд и циклические компоненты отсутствуют.

Видно, что в ряду имеется отчетливая тенденция к чередованию «вверх-вниз», однако невозможно предсказать точно выход следующей партии. Проведем линейную регрессию $X(k)$ по $X(k-1)$. Результат представлен на рис. 25.3.

Таблица 25.7

47	71	51	50	48	38	68
64	35	57	71	55	59	38
23	57	50	56	45	55	50
71	40	60	74	57	41	60
38	58	45	50	50	53	39
64	44	57	58	62	49	59
55	80	50	45	44	34	40
41	55	45	54	64	35	57
59	37	25	36	43	54	54
48	74	59	54	52	45	23

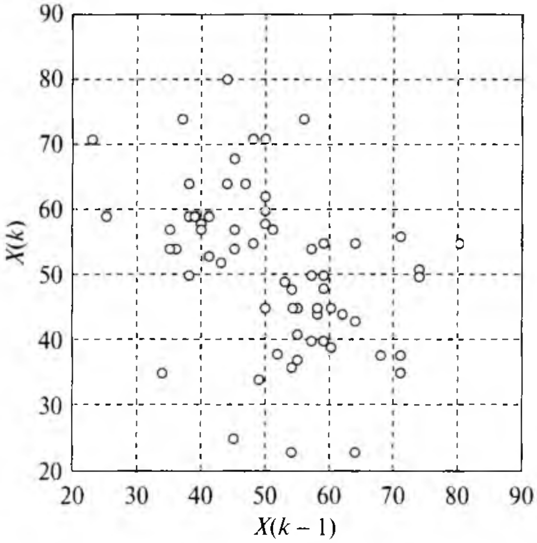

Рис. 25.3

В качестве оценки коэффициента ϕ_1 (как автокорреляции) получаем $\hat{\phi}_1 \approx -0{,}4$. Подобная отрицательная зависимость часто наблюдается в данных о выпуске продукции и является следствием эффекта «переноса». В данном примере высокий выход цикла производства приводил к получению дегтярных остатков, которые не полностью устранялись из емкости и отрицательно влияли на выход следующего цикла.

ЛИТЕРАТУРА

1. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
2. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика в задачах и упражнениях. — М.: ЮНИТИ, 2001.
3. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. — М.: Мир, 1974.
4. *Боровиков В.* STATISTICA. Искусство анализа данных на компьютере. — СПб.: Питер, 2003.
5. *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 2001.
6. *Кузнецов Ю.А.* Оптимальное управление экономическими системами. — Н.-Новгород: Изд-во Нижегородского госуниверситета, 2008.
7. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. — М.: Дело, 1997.
8. *Уотшем Т.Дж., Паррамоу К.* Количественные методы в финансах. — М.: ЮНИТИ, 1999.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Раздел VII
МИКРОЭКОНОМИЧЕСКИЙ АСПЕКТ

Глава 26
КЛАССИЧЕСКИЕ МЕТОДЫ ОПТИМИЗАЦИИ

26.1. ЛОКАЛЬНЫЙ ЭКСТРЕМУМ

Пусть функция $z = z(x, y)$ существует в некоторой области и точка $M(x_0, y_0)$ — внутренняя точка этой области.

Определение. Точка $M(x_0, y_0)$ называется точкой *локального максимума* (*минимума*) функции $z = z(x, y)$, если существует окрестность точки $M(x_0, y_0)$ такая, что для всех точек $M(x, y)$ из этой окрестности $z(x_0, y_0) \geq z(x, y)$ $(z(x_0, y_0) \leq z(x, y))$.

Определение можно интерпретировать в терминах приращений. Если приращение функции $\Delta z = z(x_0 + \Delta x, y_0 + \Delta y) - z(x_0, y_0)$, соответствующее приращению аргументов Δx и Δy, в некоторой окрестности точки $M(x_0, y_0)$ сохраняет знак, то точка $M(x_0, y_0)$ является точкой локального максимума (минимума).

Значение функции в точке $M(x_0, y_0)$ называется *локальным максимумом* или *локальным минимумом* функции. Точки максимума и точки минимума функции называются *точками экстремума* функции, а сами максимумы и минимумы функции — ее *экстремумами*.

Будем различать *строгий экстремум* (его условие $z(x_0, y_0) > z(x, y)$ или $z(x_0, y_0) < z(x, y)$) и *нестрогий экстремум* (условие $z(x_0, y_0) \geq z(x, y)$ или $z(x_0, y_0) \leq z(x, y)$). Например, функция $z = \sqrt{x^2 + y^2}$ имеет в точке (0, 0) строгий локальный минимум, равный нулю (рис. 26.1), а функция $z = |x - y|$ — нестрогий локальный минимум, равный нулю на линии $y = x$ (рис. 26.2).

Необходимые условия локального экстремума

Теорема 26.1 (о необходимом условии экстремума). Если $M(x_0, y_0)$ — точка экстремума функции, то все частные производные в этой точке равны нулю либо не существуют.

Если частные производные обращаются в нуль в точке $M(x_0, y_0)$, она может и не оказаться точкой экстремума. Например, функция

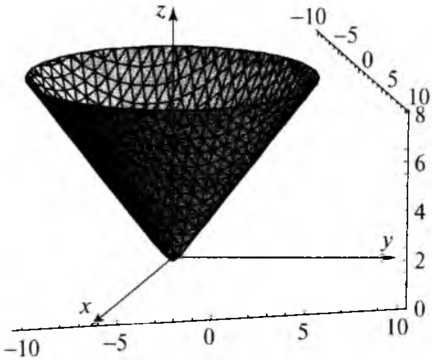

Рис. 26.1

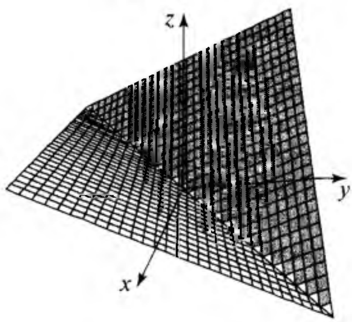

Рис. 26.2

$z = x^2 - y^2$ имеет частные производные $\frac{\partial z}{\partial x} = 2x$ и $\frac{\partial z}{\partial y} = -2y$, которые в точке (0, 0) равны нулю. Из графика (рис. 26.3) видно, что в точке (0, 0), указанной стрелкой, минимума или максимума нет. Таким образом, необходимые условия не оказываются достаточными.

Точки, в которых производные $\frac{\partial z}{\partial x} = 0$, $\frac{\partial z}{\partial y} = 0$ либо не существуют, называются *критическими* точками функции. Те из критических точек, в которых $\frac{\partial z}{\partial x} = 0$, $\frac{\partial z}{\partial y} = 0$, называются *стационарными*. Среди стационарных точек есть точки, в которых экстремум функции не достигается. Они называются *седловыми* точками или точками *минимакса*.

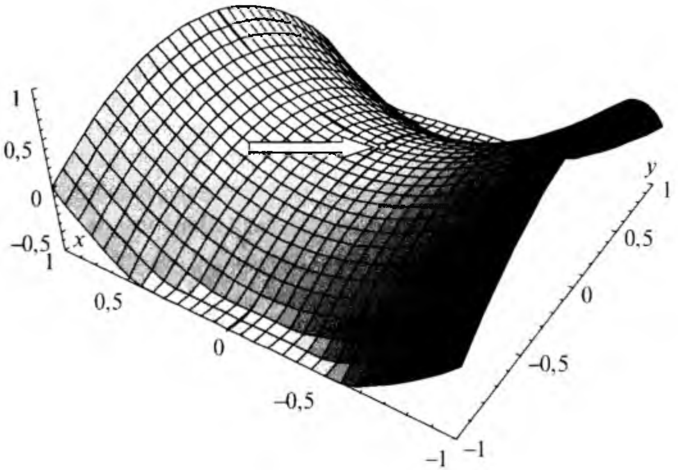

Рис. 26.3

Достаточные условия локального экстремума

Достаточные условия рассмотрим вначале для функции двух переменных, затем перейдем к общему случаю.

Теорема 26.2 (об исследовании на экстремум по угловым минорам). Пусть всюду в окрестности точки (x_0, y_0):

1) определена функция $z = z(x, y)$;
2) z'_x, z'_y непрерывны, причем $z'_x(x_0, y_0) = 0$, $z'_y(x_0, y_0) = 0$;
3) z''_{xx}, z''_{xy}, z''_{yy} непрерывны.

В этом случае комбинации вторых производных

$$M_1 = z''_{xx}(x_0, y_0) \quad \text{и} \quad M_2 = z''_{xx}(x_0, y_0)z''_{yy}(x_0, y_0) - (z''_{xy}(x_0, y_0))^2$$

определяют поведение функции $z = (x, y)$ в точке (x_0, y_0), причем:

1) если $M_1 > 0$, $M_2 > 0$, функция $z = (x, y)$ имеет в точке (x_0, y_0) минимум;

2) если $M_1 < 0$, $M_2 > 0$, функция $z = (x, y)$ имеет в точке (x_0, y_0) максимум;

3) если $M_2 < 0$, функция $z = (x, y)$ не имеет в точке (x_0, y_0) экстремума.

Пример 26.1. Функцию $z = x^3 - y^3 - 3xy$ исследовать на локальный экстремум.

Решение. Найдем критические точки функции:

$$\begin{cases} z'_x = 3x^2 - 3y = 0, \\ z'_y = -3y^2 - 3x = 0. \end{cases}$$

Решениями системы являются координаты двух точек:

$$\begin{cases} x_1 = 0, \\ y_1 = 0 \end{cases} \quad \text{и} \quad \begin{cases} x_2 = -1, \\ y_2 = 1. \end{cases}$$

Найдем вторые частные производные:

$z''_{xx} = 6x, \quad z''_{xy} = -3, \quad z''_{yy} = -6y.$

1. Точка (0, 0). $M_2 = (-36xy - 9)|_{(0,0)} = -9$. Экстремума нет.

2. Точка (–1, 1). $M_2 = (-36xy - 9)|_{(-1,1)} = 27$. Экстремум есть;

$M_1 = 6x|_{(-1,1)} = -6$. Следовательно, в точке (–1, 1) функция имеет локальный максимум $z_{\max} = (x^3 - y^3 - 3xy)|_{(-1,1)} = 1$.

На рис. 26.4 приведен график функции $z = x^3 - y^3 - 3xy$. Левой стрелкой указано положение максимума функции, правой стрелкой — седловая точка (точка минимакса).

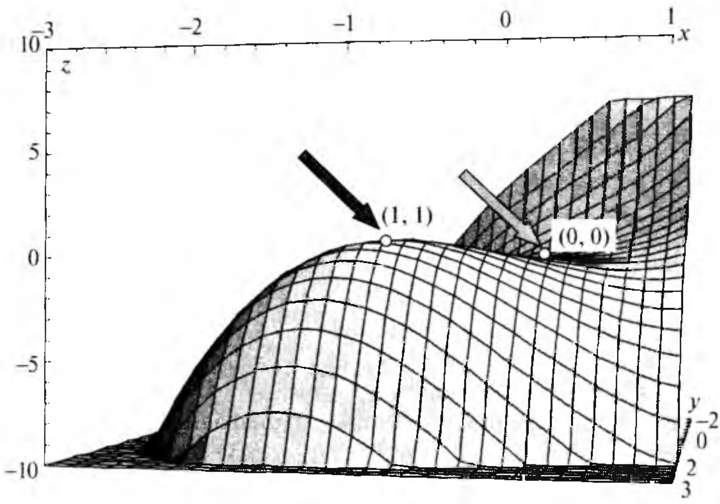

Рис. 26.4

26.2. УСЛОВНЫЙ ЭКСТРЕМУМ

Определение. *Условным экстремумом* функции $z = z(x, y)$ называется максимум или минимум функции, достигнутые при условии, что ее аргументы связаны некоторым уравнением $g(x, y) = 0$:

$$\begin{cases} z = z(x, y), \\ g(x, y) = 0. \end{cases}$$

В дальнейшем экстремумы, не являющиеся условными, будем называть *безусловными*.

При нахождении условных экстремумов функции $z = z(x, y)$ аргументы x и y уже нельзя рассматривать как независимые переменные. Они связаны между собой соотношением $g(x, y) = 0$, которое называется *уравнением связи*.

Для пояснения различия между локальным (безусловным) и условным экстремумом рассмотрим функцию

$$z = z(x, y) = x^2 + y^2.$$

Она описывает так называемый параболоид вращения и имеет безусловный минимум в точке, указанной темной стрелкой (рис. 26.5). Добавим уравнение связи (ограничение в виде равенства)

$$g(x, y) = y - x + 1 = 0,$$

описывающее плоскость. Задача теперь формулируется так: найти экстремум функции $z = x^2 + y^2$, рассматривая среди всех значений (x, y) только те, которые в совокупности образуют плоскость P. Другими словами, экстремум следует искать среди точек, принадлежащих одновременно обеим поверхностям, изображенным на рис. 26.5. Минимальное значение (условный минимум) достигается в точке, указанной белой стрелкой.

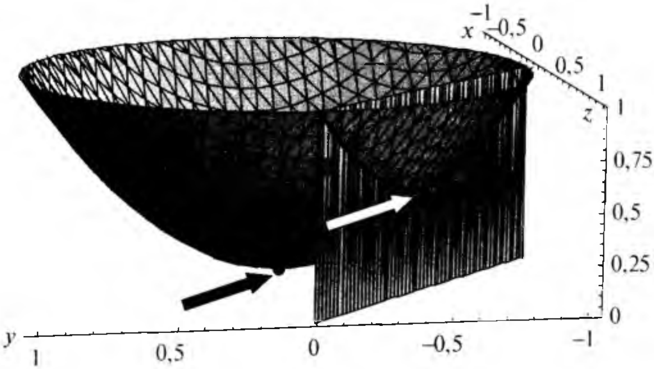

Рис. 26.5

Опишем два метода поиска условного экстремума.

П е р в ы й   м е т о д нахождения условного экстремума. Пусть уравнение связи $g(x, y) = 0$ может быть разрешено относительно зависимой переменной: $y = f(x)$. Подставим функцию $y = f(x)$ в исследуемую на экстремум функцию $z = z(x, y)$. Получим функцию одного аргумента

$$z = z(x, f(x)) = z(x),$$

в которой учтено условие связи. Далее надо исследовать функцию $z = z(x)$ на локальный экстремум, который будет являться для функции $z = z(x, y)$ условным экстремумом.

Пример 28.2. Исследовать на условный экстремум функцию $z = xy$ при условии $x - y + 1 = 0$.

Решение. Из уравнения связи найдем $y = x + 1$ и подставим в исследуемую функцию:

$$z = x(x + 1).$$

Исследуем ее на экстремум:

$$z' = 2x + 1 = 0.$$

Критическая точка $x_0 = -\frac{1}{2}$. Вторая производная $z'' = 2 > 0$. Поэтому в точке $x_0 = -\frac{1}{2}$ существует минимум функции $z = x(x + 1)$, соответственно в точке с координатами $x_0 = -\frac{1}{2}$, $y_0 = \frac{1}{2}$ — условный минимум функции $z = xy$. Его значение $z_0 = -\frac{1}{4}$. На рис. 26.6 представлены описываемые обеими функциями пересекающиеся поверхности. На точку условного минимума указывает стрелка.

Второй метод нахождения условного экстремума (метод Лагранжа). Идея метода состоит в замене функции $z = z(x, y)$, исследуемой на условный экстремум, на функцию, которая может быть исследована на локальный экстремум. Эта функция $L(x, y)$, называемая функцией Лагранжа, составлена из функции $z = z(x, y)$, функции ограничения $g(x, y)$ и некоторого коэффициента λ, их соединяющего (множителя Лагранжа):

$$L(x, y, \lambda) = z(x, y) + \lambda g(x, y).$$

Варьируя величину λ, можно добиться совпадения условного экстремума функции $z = z(x, y)$ со стационарной точкой функции Лагранжа.

Для иллюстрации сказанного рассмотрим задачу

$$\begin{cases} z = 1 - x^2 - y^2, \\ x + y = 1. \end{cases}$$

На рис. 26.7 изображены функции Лагранжа $L(x, y, \lambda)$ при различных значениях коэффициента λ (0; 0,5; 1,0; 1,5; 2,0). Это параболоиды вращения. Кроме того, изображено множество точек $x + y = 1$ в виде плоскости, а также функция $z(x, y)$ (частный случай функции

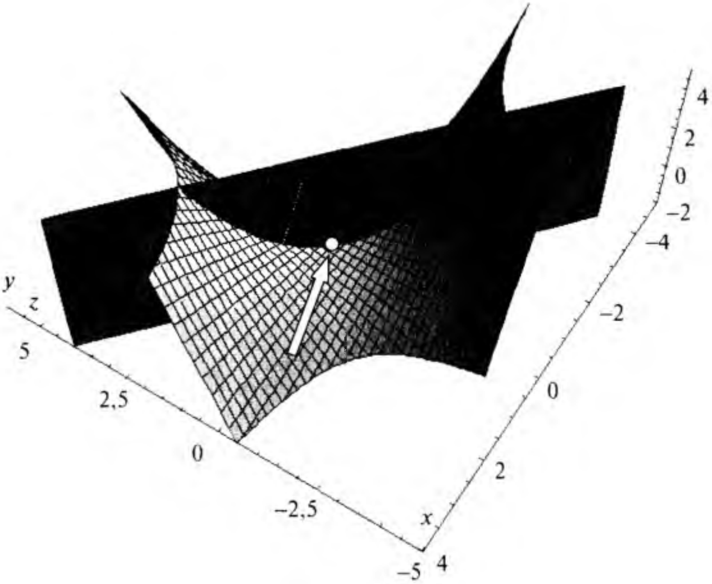

Рис. 26.6

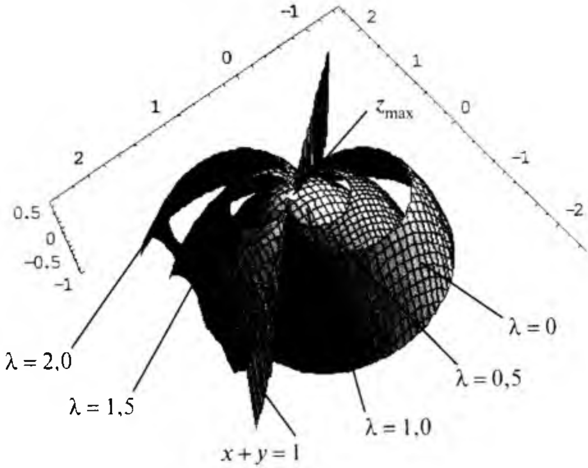

Рис. 26.7

Лагранжа при $\lambda = 0$). Меняя λ, находим ту из функций $L(x, y)$, локальный экстремум которой совпадает с условным экстремумом функции $z(x, y)$.

Как видно из рисунка, это достигается при $\lambda = 1$. Для лучшего обзора картины на рисунке вырезаны части параболоидов.

Геометрическая интерпретация необходимых условий экстремума

Интересна геометрическая интерпретация решений системы. Введем векторы

$$\operatorname{grad} z(x, y) = (z'_x, z'_y) \quad \text{и} \quad \operatorname{grad} g(x, y) = (g'_x, g'_y).$$

Из системы легко получить равенство

$$\lambda = -\frac{z'_y}{g'_y} = -\frac{z'_x}{g'_x},$$

откуда следует коллинеарность векторов $\operatorname{grad} z(x, y)$ и $\operatorname{grad} g(x, y)$. Построим фрагмент карты линий уровня функции $z = z(x, y)$ (рис. 26.8). Пунктиром обозначен график функции $g(x, y) = 0$. Градиенты $\operatorname{grad} z(x_0, y_0)$ и $\operatorname{grad} z(x_1, y_1)$ перпендикулярны линиям уровня в точках M_0 и M_1 соответственно. Векторы $\operatorname{grad} g(x_0, y_0)$ и $\operatorname{grad} g(x_1, y_1)$ перпендикулярны линии $g(x, y) = 0$ также в точках M_0 и M_1. При движении справа налево вдоль кривой $g(x, y) = 0$ пересекаются линии уровня функции $z = z(x, y)$, причем каждое следующее пересечение проис-

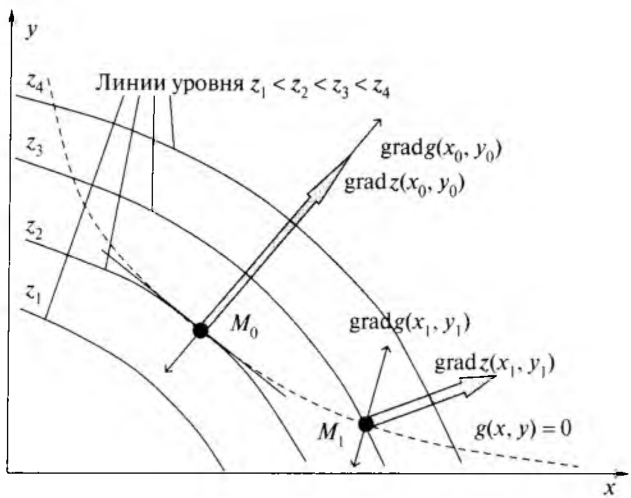

Рис. 26.8

ходит с линией более низкого уровня. Градиенты grad$z$ и grad$g$ в каждой точке направлены под разными углами, как это имеет место, например, в точке M_1. Коллинеарность векторов grad$z$ и grad$g$ возникает в точке M_0. Следовательно, в этой точке выполнены необходимые условия локального экстремума функции Лагранжа. При дальнейшем движении вдоль кривой уравнения связи $g(x, y) = 0$ пересекаются линии более высокого уровня. Можно заключить, что в точке M_0 имеется минимум. Если бы кривая уравнения связи, выйдя из точки M_0, продолжала пересекать линии все более низкого уровня, точка M_0 оказалась бы седловой (или точкой минимакса).

Окаймленный гессиан

Для функции Лагранжа $L(x, y)$ получены необходимые условия локального экстремума. Теорема о достаточных условиях локального экстремума функции $L(x, y)$ подобна ранее сформулированной теореме о достаточных условиях локального экстремума для функции $u = u(x_1, x_2, ..., x_n)$, но между ними существуют отличия.

Замечание. Следует учитывать, что условный минимум функции $z = z(x, y)$ может и не соответствовать безусловному минимуму функции Лагранжа в той же точке. Условному экстремуму с ограничением типа равенства соответствует лишь стационарность функции Лагранжа.

Теорема 26.3 (об исследовании на экстремум по окаймленному гессиану). Пусть всюду в окрестности точки $M_0(x_0, y_0)$:

1) определена функция $L(x, y) = z(x, y) + \lambda g(x, y)$;

2) обе частные производные первого порядка L'_x, L'_y непрерывны, причем $L'_x\big|_{M_0} = 0$, $L'_y\big|_{M_0} = 0$;

3) все частные производные второго порядка L''_{xx}, L''_{xy}, L''_{yy} непрерывны;

4) дифференциалы dx и dy связаны между собой соотношением $g'_x dx + g'_y dy = 0$, причем $g'_y\big|_{M_0} \neq 0$;

5) второй дифференциал d^2L в точке M_0 является знакоопределенной квадратичной формой.

Тогда функция $L(x, y)$ имеет в точке M_0 локальный экстремум, а функция $z = z(x, y)$ — условный экстремум:

а) при $d^2L > 0$ — локальный минимум;

б) при $d^2L < 0$ — локальный максимум.

Доказательство теоремы заключается в обосновании использования в данной теореме выводов ранее рассмотренной теоремы о достаточных условиях для функции $u = u(x_1, x_2, ..., x_n)$ при исследовании на локальный экстремум.

Уравнение связи $g(x, y) = 0$ приводит к существованию зависимости между дифференциалами dx и dy. Как известно, второй дифференциал не обладает инвариантностью формы. Поэтому выражение для d^2L требует уточнения. Найдем дифференциал от дифференциала $dL = L'_x dx + L'_y dy$ в точке M_0, считая y зависимой переменной:

$$d^2L = d(dL) = d(L'_x dx + L'_y dy) = dx(L''_{xx}dx + L''_{xy}dy) +$$
$$+ (L''_{yx}dx + L''_{yy}dy)dy + L'_y \cdot d^2y.$$

Здесь $L''_{xy} = L''_{yx}$ в силу непрерывности вторых производных, $L'_y = 0$ в стационарной точке M_0. Поэтому второй дифференциал функции Лагранжа в точке M_0 будет совпадать со вторым дифференциалом функции Лагранжа двух независимых переменных:

$$d^2L = L''_{xx}dx^2 + 2L''_{xy}dxdy + L''_{yy}dy^2.$$

Следовательно, можно воспользоваться выводами о знакоопределенности d^2L и соответственно знаке ΔL по знакам угловых миноров матрицы Гессе:

1) если $M_1 > 0$, $M_2 > 0$, функция $L(x, y)$ имеет в точке M_0 минимум;

2) если $M_1 < 0$, $M_2 > 0$, функция $L(x, y)$ имеет в точке M_0 максимум.

Если сделать вывод о наличии экстремума невозможно при произвольных dx и dy, найдем dy из равенства $g'_x dx + g'_y dy = 0$ и подставим в d^2L:

$$d^2L = L''_{xx}dx^2 + 2L''_{xy}dx\left(-\frac{g'_x}{g'_y}dx\right) + L''_{yy}\left(-\frac{g'_x}{g'_y}dx\right)^2.$$

Вынесем за скобки множитель $-\dfrac{dx^2}{(g'_y)^2}$:

$$d^2L = -\frac{dx^2}{(g'_y)^2} \cdot (-(g'_y)^2 L''_{xx} + 2g'_x g'_y L''_{xy} - (g'_x)^2 L''_{yy}).$$

Эта формула может быть приведена к удобному для использования виду. Второй сомножитель в произведении вычисляется как определитель матрицы в точке (x_0, y_0):

$$H_3 = \begin{pmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{xy} & L''_{yy} \end{pmatrix}_{(x_0,\, y_0)}$$

Матрица H_3 называется *окаймленной матрицей Гессе*, а определитель $|H_3|$ — *окаймленным гессианом*. Если гессиан $|H_3|_{(x_0,\, y_0)} > 0$, то $d^2L < 0$, что указывает на безусловный максимум функции $L(x, y)$ и условный максимум функции $z = z(x, y)$. Если $|H_3|_{(x_0,\, y_0)} < 0$, то $d^2L > 0$. Это соответствует безусловному минимуму функции $L(x, y)$ и условному минимуму функции $z = z(x, y)$. Обоснование закончено.

Замечание 1. Если $|H_3|_{(x_0,\, y_0)} = 0$, метод не работает. Поскольку точка M_0 является стационарной для функции Лагранжа, в этой точке возможен условный экстремум функции $z = z(x, y)$.

Замечание 2. Метод Лагранжа не указывает условий, при которых условный экстремум отсутствует.

Замечание 3. При составлении функции Лагранжа можно брать неопределенный множитель с любым знаком. Выбор определяется соображениями удобства.

Замечание 4. Вычисляемый неопределенный множитель λ может оказаться любым действительным числом, включая иррациональное число. В частном случае при $\lambda = 0$ условный экстремум исследуемой функции совпадает с ее локальным экстремумом, если таковой существует. На рис. 26.9 представлено графическое решение задачи на условный экстремум функции $z = 1 - x^2 - y^2$ с уравнением связи $g(x, y) = y - x = 0$. Стрелкой указан условный экстремум функции $z = 1 - x^2 - y^2$, совпадающий с ее локальным максимумом. Вычисленное значение множителя $\lambda = 0$.

Последовательность действий при отыскании условных экстремумов функции двух переменных

Пример 26.3. Найти условный экстремум функции $z = z(x, y)$ при наличии уравнения связи $g(x, y) = 0$. Краткая формулировка:

$$\begin{cases} z = z(x, y), \\ g(x, y) = 0. \end{cases} \tag{26.1}$$

Решение.

1. Составляем функцию Лагранжа $L(x, y) = z(x, y) + \lambda g(x, y)$.

2. Находим производные функции $L(x, y)$ и приравниваем их нулю. Присоединяем уравнение связи, получая систему из трех уравнений для определения координат возможных точек экстремума (x_0, y_0) и множителя Лагранжа λ.

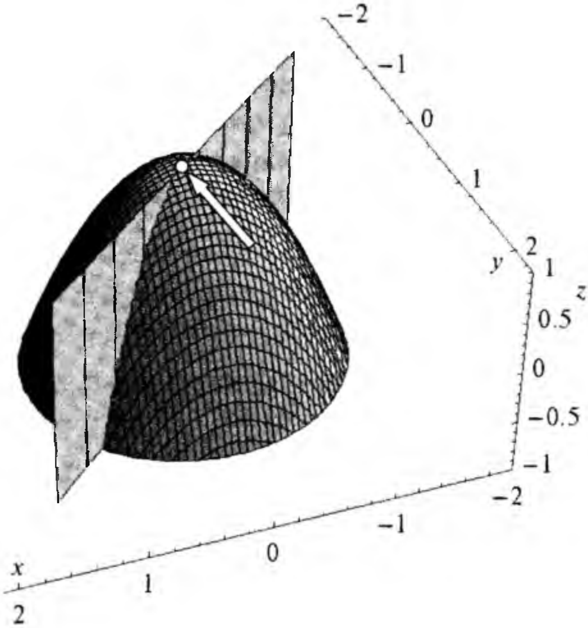

Рис. 26.9

3. Строим окаймленную матрицу Гессе, вычисляем окаймленный гессиан в точке (x_0, y_0) и делаем вывод о наличии условного экстремума. Находим значение функции $z = z(x, y)$ в точке условного экстремума.

Задача 26.4. Провести исследование на условный экстремум

$$\begin{cases} z = x^2 + 12xy + 2y^2, \\ 4x^2 + y^2 = 25. \end{cases}$$

Решение. Составим функцию Лагранжа

$L(x, y) = x^2 + 12xy + 2y^2 - \lambda(4x^2 + y^2 - 25)$

и запишем систему уравнений из первых производных и уравнения связи:

$$\begin{cases} L'_x = 2x + 12y - 8\lambda x = 0, \\ L'_y = 12x + 4y - 2\lambda x = 0, \\ 4x^2 + y^2 = 25. \end{cases} \tag{26.2}$$

Из первых двух уравнений получаем

$$\begin{cases} (1 - 4\lambda)x = -6y, \\ 6x = (\lambda - 2)y. \end{cases} \tag{26.3}$$

Очевидно, что точка с координатами (0, 0) является решением системы (26.3), но не удовлетворяет последнему уравнению системы (26.2). Полагая теперь $x \neq 0$, $y \neq 0$, разделим первое уравнение системы (26.3) на второе

$$\frac{1-4\lambda}{6} = -\frac{6}{\lambda-2}$$

и решим полученное уравнение. Корни уравнения $\lambda_1 = -2, \lambda_2 = \frac{17}{4}$. Подставив $\lambda_1 = -2$ в первое или второе уравнения системы (26.2), найдем связь между переменными x и y: $x = -\frac{2}{3}y$. Подстановка выражения $x = -\frac{2}{3}y$ в третье уравнение системы (26.1) дает $y = \pm 3$. Соответственно $x = \pm 2$. Для $\lambda_2 = \frac{17}{4}$ получим $y = \pm 4$, $x = \pm\frac{3}{2}$. Итак, получены четыре стационарные точки (x_i, y_i, λ_i), $i = 1, \ldots, 4$:

$$A_1(2, -3, -2);\ A_2 = (-2, 3, -2);\ A_3\left(\frac{3}{2}, -4, \frac{17}{4}\right);\ A_4\left(-\frac{3}{2}, -4, \frac{17}{4}\right).$$

Найдем все вторые производные функции Лагранжа:
$L''_{xx} = 2 - 8\lambda$; $L''_{xy} = 12$; $L''_{yy} = 4 - 2\lambda$.
Составим матрицу Гессе из вторых производных:

$$\begin{pmatrix} 2-8\lambda & 12 \\ 12 & 4-2\lambda \end{pmatrix} = 2\begin{pmatrix} 1-4\lambda & 6 \\ 6 & 2-\lambda \end{pmatrix}.$$

Подстановка любого значения множителя из $\lambda_1 = -2$, $\lambda_2 = \frac{17}{4}$ приводит к равенству нулю минора M_2. Поэтому обратимся к окаймленной матрице Гессе:

$$H = \begin{pmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{xy} & L''_{yy} \end{pmatrix} = \begin{pmatrix} 0 & 8x & 2y \\ 8x & 2-8\lambda & 12 \\ 2y & 12 & 4-2\lambda \end{pmatrix} = 2\begin{pmatrix} 0 & 4x & y \\ 4x & 1-4\lambda & 6 \\ y & 6 & 2-\lambda \end{pmatrix}.$$

Рассмотрим последовательно четыре стационарные точки:

1) $A_1(2, -3, -2)$. $|H| = 2\begin{vmatrix} 0 & 8 & -3 \\ 8 & 9 & 6 \\ -3 & 6 & 4 \end{vmatrix} = -1250 < 0.$

В точке $A_1(2, -3, -2)$ достигается условный минимум $z_{\min} = -50$;

2) $A_2(-2, 3, -2)$. $|H| = 2\begin{vmatrix} 0 & -8 & 3 \\ -8 & 9 & 6 \\ 3 & 6 & 4 \end{vmatrix} = -1250 < 0.$

Здесь также условный минимум $z_{\min} = -50$;

3) $A_3\left(\frac{3}{2}, 4, \frac{17}{4}\right)$. $\quad |H| = 2\begin{vmatrix} 0 & 6 & 4 \\ 6 & -16 & 6 \\ 4 & 6 & -\frac{9}{4} \end{vmatrix} = 1250 > 0.$

В точке $A_3\left(\frac{3}{2}, 4, \frac{17}{4}\right)$ достигается условный максимум $z_{\max} = \frac{425}{4}$;

4) $A_4\left(-\frac{3}{2}, -4, \frac{17}{4}\right)$. $\quad |H| = 2\begin{vmatrix} 0 & -6 & -4 \\ -6 & -16 & 6 \\ -4 & 6 & -\frac{9}{4} \end{vmatrix} = 1250 > 0.$

Здесь также условный максимум $z_{\max} = \frac{425}{4}$.

На рис. 26.10 и 26.11 представлено графическое решение задачи на условный экстремум.

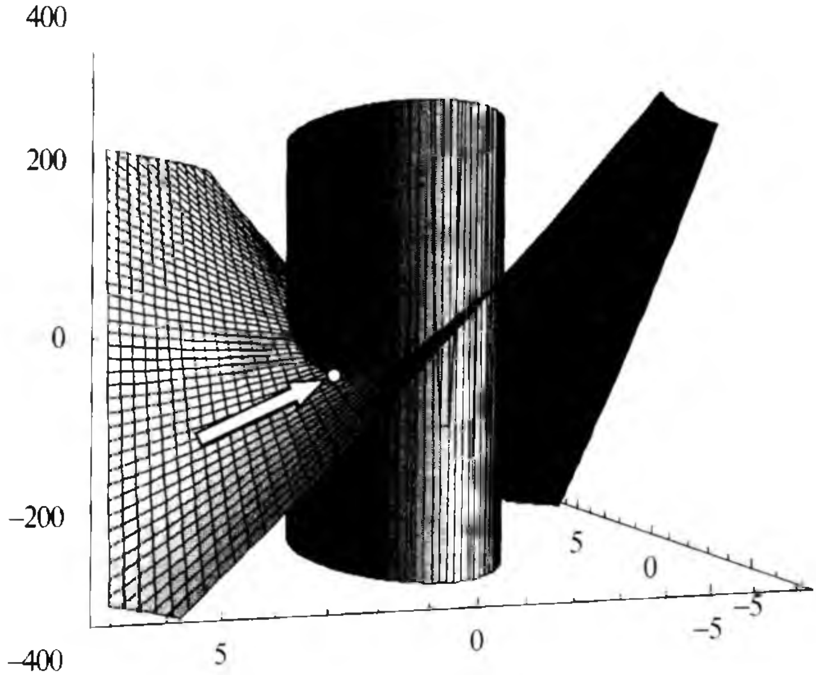

Рис. 26.10

Пересечения поверхностей изображены с двух точек обзора, отличающихся приблизительно на 90°. Стрелками показан один из минимумов (рис. 26.10) и один из максимумов (рис. 26.11).

Для случая *трех переменных* исследование функции на условный экстремум при одном уравнении связи

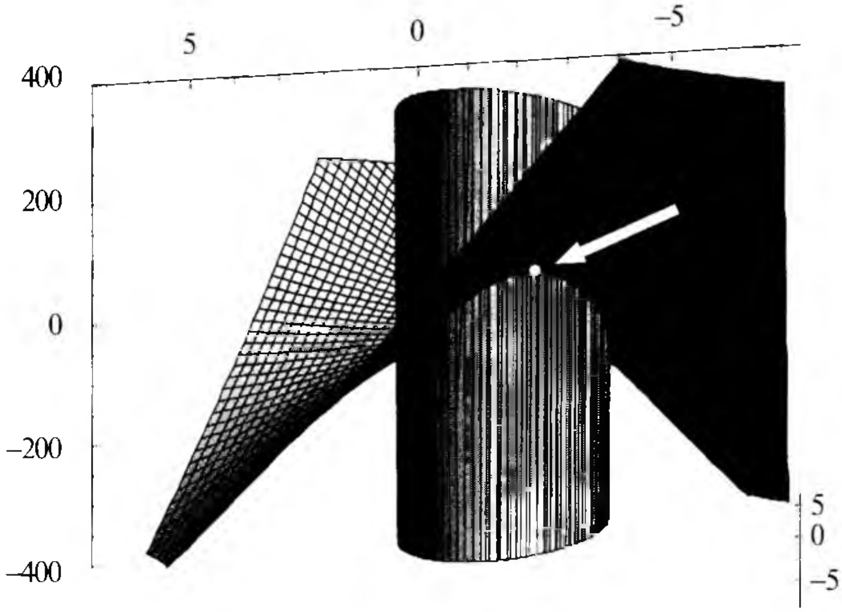

Рис. 26.11

$$\begin{cases} u = u(x, y, z), \quad u \to c.\text{extr}, \\ g(x, y, z) = 0. \end{cases}$$

методом Лагранжа приводит к окаймленной матрице Гессе 4-го порядка

$$\begin{pmatrix} 0 & g'_x & g'_y & g'_z \\ g'_x & L''_{xx} & L''_{xy} & L''_{xz} \\ g'_y & L''_{yx} & L''_{yy} & L''_{yz} \\ g'_z & L''_{zx} & L''_{zy} & L''_{zz} \end{pmatrix}.$$

В стационарной точке $(x_0, y_0, z_0, \lambda_0)$ рассматриваются окаймленные гессианы 3-го $|H_3|$ и 4-го $|H_4|$ порядков, где

$$|H_3| = \begin{vmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{yx} & L''_{yy} \end{vmatrix} \quad \text{и} \quad |H_4| = \begin{vmatrix} 0 & g'_x & g'_y & g'_z \\ g'_x & L''_{xx} & L''_{xy} & L''_{xz} \\ g'_y & L''_{yx} & L''_{yy} & L''_{yz} \\ g'_z & L''_{zx} & L''_{zy} & L''_{zz} \end{vmatrix}.$$

Условия $|H_3| < 0$, $|H_4| < 0$ являются достаточными для достижения в данной точке *условного минимума*.

Условия $|H_3| > 0$, $|H_4| < 0$ являются достаточными для достижения в данной точке *условного максимума*.

Условие $|H_4| > 0$ является достаточным для утверждения об отсутствии в данной точке экстремума.

Условный экстремум может существовать при наличии нескольких уравнений связи. Задача нахождения условного экстремума функции трех переменных с двумя уравнениями связи формулируется следующим образом:

$$\begin{cases} u = u(x, y, z), \quad u \to c.\text{extr}, \\ g(x, y, z) = 0, \\ f(x, y, z) = 0. \end{cases}$$

Если решение системы из уравнений связи с последующей подстановкой результатов в исследуемую функцию затруднено, строится функция Лагранжа в виде

$$L(x, y, z, \lambda_1, \lambda_2) = u(x, y, z) + \lambda_1 g(x, y, z) + \lambda_2 f(x, y, z).$$

Необходимые условия содержат пять уравнений с пятью переменными $(x, y, z, \lambda_1, \lambda_2)$:

$$\begin{cases} L'_x = u'_x + \lambda_1 g'_x + \lambda_2 f'_x = 0, \\ L'_y = u'_y + \lambda_1 g'_y + \lambda_2 f'_y = 0, \\ L'_z = u'_z + \lambda_1 g'_z + \lambda_2 f'_z = 0, \\ g(x, y, z) = 0, \\ f(x, y, z) = 0. \end{cases}$$

Решив систему уравнений и найдя стационарные точки, исследуем окаймленный гессиан

$$|H_5| = \begin{vmatrix} 0 & 0 & g'_x & g'_y & g'_z \\ 0 & 0 & f'_x & f'_y & f'_z \\ g'_x & f'_x & L''_{xx} & L''_{xy} & L''_{xz} \\ g'_y & f'_y & L''_{yx} & L''_{yy} & L''_{yz} \\ g'_z & f'_z & L''_{zx} & L''_{zy} & L''_{zz} \end{vmatrix}.$$

Если $|H_5| > 0$ в рассматриваемой стационарной точке, то достигается минимум, если $|H_5| < 0$, то максимум.

Экстремум в системе функций

Экстремумы математических объектов, задаваемых несколькими функциями, имеют экономические приложения, включая экономико-математические учебные задачи. Одной из них является задача на условный максимум производственной функции $z = f(x_1, x_2)$ при бюджетном ограничении $p_1x_1 + p_2x_2 = I(z)$. Денежные средства I, отпущенные на закупку ресурсов, уже не являются постоянной величиной I_0, а зависят от величины выпуска z и «подпитываются» долей дохода:

$$I(z) = I_0 + kz.$$

Чем лучше работает предприятие, тем большие средства оно может отпустить на закупку сырья, энергообеспечение, аренду и т.д.

Задача на условный экстремум с переменным бюджетным ограничением формулируется так:

$$\begin{cases} z = f(x_1, x_2) \Rightarrow \text{условный экстремум}, \\ p_1x_1 + p_2x_2 = I_0 + kz. \end{cases}$$

Простым преобразованием второго уравнения система сводится к исследованию на экстремум зависимой переменной $z = z(x_1, x_2)$ в системе функций

$$\begin{cases} z = f(x_1, x_2) \Rightarrow \text{условный экстремум}, \\ z = \dfrac{p_1x_1 + p_2x_2 - I_0}{k}. \end{cases}$$

При $k = 0$ получаем классический условный экстремум с постоянным бюджетным ограничением.

Задача на условный экстремум с переменным бюджетным ограничением является частным случаем более общей задачи поиска экстремума функции $z = z(x, y)$, заданной системой функций

$$\begin{cases} z = f(x, y), \\ z = g(x, y). \end{cases}$$

Экстремум функции $z = z(x, y)$ следует назвать условным, что ясно из геометрических соображений. В трехмерном пространстве разыскивается совокупность линий пересечения поверхностей, описываемых заданными функциями. На линиях пересечения находим наибольшее или наименьшее значение z_0.

Поскольку каждое уравнение системы при исследовании на экстремум можно привести к виду $z = f(x, y(x))$ и $z = g(x, y(x))$, то в критической точке должен быть равен нулю дифференциал каждой функции:

$$dz = f'_x dx + f'_y dy = 0,$$
$$dz = g'_x dx + g'_y dy = 0.$$

Умножим первое и второе равенства на λ_1 и λ_2 соответственно, причем λ_1, λ_2 не равны нулю одновременно:

$$(\lambda_1 f'_x + \lambda_2 g'_x)dx + (\lambda_1 f'_y + \lambda_2 g'_y)dy = 0.$$

Слева в равенстве стоит дифференциал функции, которую назовем функцией Лагранжа и обозначим через $L(x, y)$:

$$dL = (\lambda_1 f'_x + \lambda_2 g'_x)dx + (\lambda_1 f'_y + \lambda_2 g'_y)dy = 0.$$

Сама функция имеет вид

$$L(x, y) = \lambda_1 f(x, y) + \lambda_2 g(x, y).$$

Находя стационарные точки для $L(x, y)$, подберем один из множителей λ_1, λ_2 так, чтобы в некоторой точке M_0 было выполнено равенство

$$L'_x\big|_{M_0} = \lambda_1 f'_x + \lambda_2 g'_x = 0.$$

Поскольку дифференциал функции $L(x, y)$ равен нулю, то

$$(\lambda_1 f'_y + \lambda_2 g'_y)dy = 0$$

также, причем $dy \neq 0$ как дифференциал независимой переменной.

Таким образом, система

$$\begin{cases} \lambda_1 f'_x + \lambda_2 g'_x = 0, \\ \lambda_1 f'_y + \lambda_2 g'_y = 0 \end{cases}$$

выражает необходимые условия локального экстремума функции $L(x, y)$. Система содержит две переменные величины x, y и два параметра λ_1 и λ_2. Для нахождения четырех неизвестных добавим уравнение связи $f(x, y) = g(x, y)$ и одно из равенств $L(x, y)\big|_{M_0} = f(x, y)\big|_{M_0}$ или $L(x, y)\big|_{M_0} = g(x, y)\big|_{M_0}$.

Итак, условия первого порядка при нахождении локального экстремума функции Лагранжа и условного экстремума функции $z(x, y)$ сформулированы:

$$\begin{cases} \lambda_1 f'_x + \lambda_2 g'_x = 0, \\ \lambda_1 f'_y + \lambda_2 g'_y = 0, \\ f(x, y) = g(x, y), \\ L(x, y)\big|_{M_0} = f(x, y)\big|_{M_0}. \end{cases}$$

Последнее равенство можно заменить на условие нормировки $\lambda_1 + \lambda_2 = 1$, поскольку

$$L(x, y)\big|_{M_0} = \lambda_1 f(x, y)\big|_{M_0} + \lambda_2 g(x, y)\big|_{M_0} = \lambda_1 z_{\text{extr}} + \lambda_2 z_{\text{extr}} = \\ = z_{\text{extr}}(\lambda_1 + \lambda_2) = z_{\text{extr}}.$$

Из первых двух уравнений системы найдем $\frac{\lambda_1}{\lambda_2}$, считая $f'_x \neq 0$ и $f'_y \neq 0$:

$$\frac{\lambda_1}{\lambda_2} = \frac{g'_x}{f'_x} = \frac{g'_y}{f'_y}.$$

Найденное выражение представляет необходимые условия в задаче на условный экстремум. Решая систему уравнений, находим все стационарные точки функции Лагранжа $M_0(x_0, y_0, \lambda_1, \lambda_2)$.

Замечание. Следует учитывать, что условный минимум функции $z = z(x, y)$ может и не соответствовать безусловному минимуму функции Лагранжа в той же точке. Условному экстремуму с ограничением типа равенства соответствует лишь стационарность функции Лагранжа.

Перейдем к достаточным условиям условного экстремума. Возьмем дифференциал от обеих частей в уравнении связи $f(x, y) = g(x, y)$:

$$f'_x dx + f'_y dy = g'_x dx + g'_y dy,$$

откуда, считая, что $f'_y \neq g'_y$, найдем dy:

$$dy = -\frac{f'_x - g'_x}{f'_y - g'_y} dx.$$

Подставим выражение во второй дифференциал функции Лагранжа:

$$d^2L = L''_{xx} dx^2 + 2L''_{xy} dx\left(-\frac{f'_x - g'_x}{f'_y - g'_y} dx\right) + L''_{yy}\left(-\frac{f'_x - g'_x}{f'_y - g'_y} dx\right)^2.$$

Преобразуя, получим

$$d^2L = -\frac{dx^2}{(f'_y - g'_y)^2} \times$$

$$\times(-L''_{xx}(f'_y - g'_y)^2 + 2(f'_x - g'_x)(f'_y - g'_y)L''_{xy} - (f'_x - g'_x)^2 L''_{yy}).$$

Второй сомножитель в скобках переписываем как окаймленный гессиан:

$$|H| = \begin{vmatrix} 0 & f'_x - g'_x & f'_y - g'_y \\ f'_x - g'_x & L''_{xx} & L''_{xy} \\ f'_y - g'_y & L''_{yx} & L''_{yy} \end{vmatrix}.$$

Его знак в критической точке указывает на локальный минимум или максимум функции L, а следовательно, на условный минимум или максимум функции $z(x, y)$:

- если $|H|\big|_{M_0} > 0$, то $z(x, y)$ имеет условный максимум;
- если $|H|\big|_{M_0} < 0$, то $z(x, y)$ имеет условный минимум.

В качестве примера рассмотрим экстремум зависимой переменной z, определяемой системой уравнений

$$\begin{cases} z = 3 - x^2 - y^2, \\ z = x + 1. \end{cases}$$

Составим функцию Лагранжа:

$L(x, y, \lambda_1, \lambda_2) = \lambda_1(3 - x^2 - y^2) + \lambda_2(x + 1)$.

Необходимые условия имеют вид

$$\begin{cases} L'_x = -2\lambda_1 x + \lambda_2 = 0, \\ L'_y = -2\lambda_1 y = 0, \\ 3 - x^2 - y^2 = x + 1, \\ \lambda_1 + \lambda_2 = 1. \end{cases}$$

Решение системы из четырех уравнений с четырьмя переменными дает две стационарные точки $M(x, y, \lambda_1, \lambda_2)$: $M_1\left(1, 0, \frac{1}{3}, \frac{2}{3}\right)$ и $M_2\left(-2, 0, -\frac{1}{3}, \frac{4}{3}\right)$. В каждой точке вычисляем окаймленный гессиан $|H|$.

В точке M_1 его величина $|H| = \begin{vmatrix} 0 & -3 & 0 \\ -3 & -\frac{2}{3} & 0 \\ 0 & 0 & -\frac{2}{3} \end{vmatrix} = 6$. В соответствии с достаточными условиями следует сделать вывод о наличии максимума: $z_{\max}(1, 0) = 2$.

В точке M_2 величина окаймленного гессиана отрицательна: $|H| = \begin{vmatrix} 0 & 3 & 0 \\ 3 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{vmatrix} = -6$. В этой точке достигается минимум $z_{\min}(-2, 0) = -1$. Геометрическая интерпретация задачи представлена на рис. 26.12.

Таким образом, экстремум в системе функций является более общим по сравнению с классическим условным экстремумом с ограничением типа равенства

$$\begin{cases} z = f(x_1, x_2) \Rightarrow \text{условный экстремум}, \\ g(x, y) = 0. \end{cases}$$

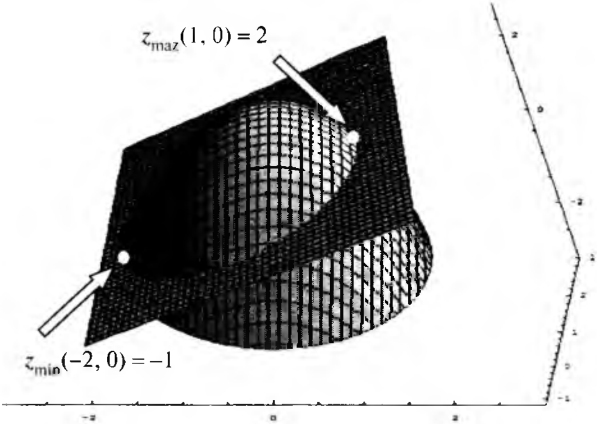

Рис. 26.12

26.3. ГЛОБАЛЬНЫЙ ЭКСТРЕМУМ

Глобальным экстремумом называется наибольшее и наименьшее значения функции в замкнутой ограниченной области D. Пусть функция $z = z(x, y)$ непрерывна в этой области. Тогда найдется точка (x_0, y_0), в которой функция принимает наибольшее (наименьшее) значение. Если точка (x_0, y_0) лежит внутри области D, она является стационарной и в ней может достигаться локальный максимум или минимум. Наибольшее или наименьшее значение функция может принимать также на границе области. Следовательно, задачу исследования функции $z = z(x, y)$ на глобальный (global) экстремум в области $g(x, y) \le 0$ можно сформулировать так:

$$\begin{cases} z = z(x, y) \to \text{gl. extr}, \\ g(x, y) \le 0. \end{cases}$$

Решение задачи разбивается на две части:

1) исследование функции $z = z(x, y)$ на локальный (local) экстремум в области $g(x, y) < 0$:

$$\begin{cases} z = z(x, y) \to \text{loc. extr.}, \\ g(x, y) < 0; \end{cases}$$

2) исследование функции $z = z(x, y)$ на условный (conditional) экстремум на границе области $g(x, y) = 0$:

$$\begin{cases} z = z(x, y) \to \text{cond. extr.}, \\ g(x, y) = 0. \end{cases}$$

Наибольшее (наименьшее) из всех этих чисел и будет искомым наибольшим (наименьшим) значением функции $z = z(x, y)$ в области D.

Пример 26.5. Исследовать на глобальный экстремум функцию $z = 2 - x^2 - y^2$ в области $x^2 + y^2 \geq 1$, $|x| \leq 2$, $|y| \leq 2$.

Решение. Для решения задачи

$$\begin{cases} z = 2 - x^2 - y^2 \to \text{gl. extr.}, \\ x^2 + y^2 \geq 1, \\ |x| \leq 2, |y| \leq 2 \end{cases}$$

надо рассмотреть несколько случаев:

$$1) \begin{cases} z = 2 - x^2 - y^2 \to \text{loc. extr.}, \\ x^2 + y^2 > 1, \\ |x| < 2, |y| < 2. \end{cases}$$

Находим стационарные точки $z'_x = -2x = 0$. Единственная стационарная точка (0, 0) не входит в рассматриваемую область;

$$2) \begin{cases} z = 2 - x^2 - y^2 \to \text{cond. extr.}, \\ x^2 + y^2 = 1, \\ |x| \leq 2, |y| \leq 2. \end{cases}$$

Подставим выражение $x^2 + y^2 = 1$ в исследуемую функцию. Получаем $z = 2 - (x^2 + y^2) = 2 - 1 = 1$. Во всех точках окружности функция принимает одно и то же постоянное значение $z = 1$;

$$3) \begin{cases} z = 2 - x^2 - y^2 \to \text{cond. extr.}, \\ |x| = 2, |y| \leq 2, \\ x^2 + y^2 \geq 1. \end{cases}$$

В силу симметрии задачи из двух случаев достаточно рассмотреть один, например $x = 2$. Подставим $x = 2$ в функцию z: $z = 2 - 4 - y^2 = -2 - y^2$. Это парабола с ветвями, направленными вниз. Критическая точка $y = 0$, в которой функция принимает значение, равное $z = -2$. На концах отрезка $[-2; 2]$ функция равна $z = -6$;

$$4) \begin{cases} z = 2 - x^2 - y^2 \to \text{cond. extr.}, \\ |x| \leq 2, |y| = 2, \\ x^2 + y^2 \geq 1. \end{cases}$$

Система рассматривается аналогично случаю 3. Собираем все полученные значения функции z:

$$z|_{x^2+y^2=1} = 1,$$

$z(2, 0) = z(-2, 0) = z(0, 2) = z(0, -2) = -2,$

$z(2, 2) = z(-2, -2)z(2, -2)z(-2, 2) = -6.$

Выбираем наибольшее и наименьшее значения из всех этих чисел:

$z_{gl.max} = 1$, $z_{gl.min} = -6$.

График функции $z = 2 - x^2 - y^2$ в области $x^2 + y^2 \geq 1$, $|x| \leq 2$, $|y| \leq 2$ приведен на рис. 26.13.

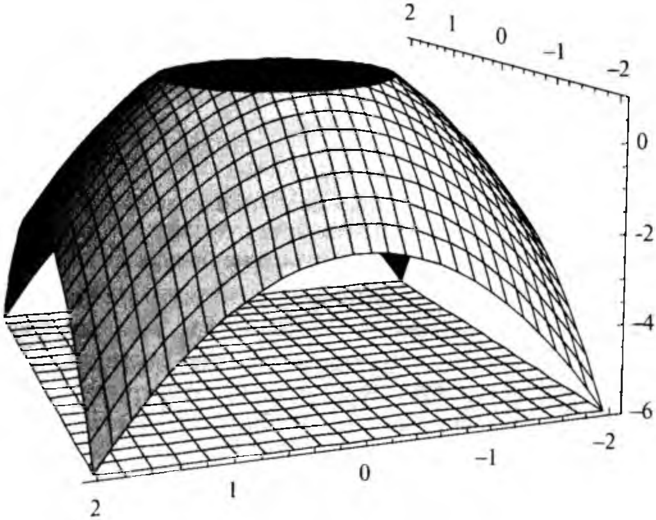

Рис. 26.13

Замечание. При исследовании функции на глобальный экстремум можно ограничиться отысканием значения функции в критических точках. Достаточные условия обычно не используют.

Глава 27
КОЛИЧЕСТВЕННЫЕ МЕТОДЫ В МИКРОЭКОНОМИКЕ

27.1. МАКСИМИЗАЦИЯ ВЫПУСКА ПРИ НАЛИЧИИ ЛИМИТА НА РЕСУРСЫ

В этой главе мы решим важные экономические задачи по оптимизации, а также «покачаем» некоторые экзогенные параметры в оптимальной точке. Посмотрим, как будет меняться решение оптимизационных задач.

Решение в общем виде задачи максимизации выпуска при наличии лимита на ресурсы

Пусть производственная функция предприятия, занимающегося выпуском продукции, равна $y = f(x_1, x_2)$, где x_1, x_2 — закупаемые по цене p_1, p_2 ресурсы; C — денежный лимит на закупки. Потребуем выполнения условий: $y = f(x_1, x_2)$ — дважды непрерывно дифференцируемая функция, $x_1 > 0$, $x_2 > 0$, $p_2 > 0$. Решается задача на условный экстремум:

$$\begin{cases} y = f(x_1, x_2) \to \text{c.extr.}, \\ p_1x_1 + p_2x_2 = C. \end{cases}$$

Составим функцию Лагранжа:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda(C - p_1x_1 - p_2x_2).$$

Необходимые условия существования экстремума:

$$\begin{cases} L'_{x_1}(x_1, x_2, \lambda) = f'_{x_1}(x_1, x_2) - \lambda p_1 = 0, \\ L'_{x_2}(x_1, x_2, \lambda) = f'_{x_2}(x_1, x_2) - \lambda p_1 = 0. \end{cases}$$

Перенеся вторые слагаемые направо и разделив одно равенство на другое, найдем

$$\frac{f'_{x_1}(x_1, x_2)}{f'_{x_2}(x_1, x_2)} = \frac{p_1}{p_2}.$$

Выразим из уравнения переменную $x_2 = x_2(p_1, p_2, x_1)$ и подставим в уравнение связи

$$p_1x_1 + p_2x_2 = x_2(p_1, p_2, x_1) = C,$$

что позволяет найти критическое значение переменной $x_1^0 = x_1^0(p_1, p_2, C)$. Аналогично находим $x_2^0 = x_2^0(p_1, p_2, C)$. В этих точках находится значение максимума выпуска при наличии лимита на ресурсы.

Теоремы о маргинальных значениях (выпуск продукции при вариациях лимита на ресурсы или цены за ресурс)

Величины x_1^0, x_2^0 задают критическую точку, положение которой зависит от параметров задачи p_1, p_2, C. Изменение параметров приводит к вариациям максимума производственной функции $y_{\max} = f(x_1^0, x_2^0)$. Поэтому величины x_1^0, x_2^0 могут рассматриваться как функции переменных p_1, p_2, C:

- $x_1^0 = x_1^0(p_1, p_2, C)$ — функция условного спроса на первый ресурс;
- $x_2^0 = x_2^0(p_1, p_2, C)$ — функция условного спроса на второй ресурс со стороны предприятия на рынке.

Подставим x_1^0, x_2^0 в исследуемую производственную функцию и в уравнение связи, предварительно убрав для удобства верхние индексы:

$$\begin{cases} y_{\max} = f(x_1(p_1, p_2, C), x_2(p_1, p_2, C)), \\ p_1 x_1(p_1, p_2, C) + p_2 x_2(p_1, p_2, C) = C. \end{cases}$$

Второе уравнение связывает две функции x_1 и x_2 и три аргумента p_1, p_2, C. Будем опираться на эту систему при выводе некоторых важных утверждений экономической теории.

Продифференцируем оба равенства системы по C и воспользуемся необходимыми условиями $f'_{x_1} = \lambda p_1$ и $f'_{x_2} = \lambda p_2$:

$$\begin{cases} \dfrac{\partial y_{\max}}{\partial C} = \dfrac{\partial f}{\partial x_1}\dfrac{\partial x_1}{\partial C} + \dfrac{\partial f}{\partial x_2}\dfrac{\partial x_2}{\partial C} = \lambda p_1 \dfrac{\partial x_1}{\partial C} + \lambda p_2 \dfrac{\partial x_2}{\partial C} = \lambda\left(p_1 \dfrac{\partial x_1}{\partial C} + p_2 \dfrac{\partial x_2}{\partial C}\right) = \lambda, \\ p_1 \dfrac{\partial x_1}{\partial C} + p_2 \dfrac{\partial x_2}{\partial C} = 1. \end{cases}$$

Следовательно, *предельное предложение по лимиту* $\dfrac{\partial y_{\max}}{\partial C}$ *равно множителю Лагранжа* λ. Поясним это детальнее.

Поскольку $\dfrac{\partial y_{\max}}{\partial C} \approx \dfrac{y_{\max}}{\Delta C}$, то первое уравнение системы можно записать так:

$$\Delta y_{\max} \approx \lambda \cdot \Delta C.$$

Если предприятие оптимизировало выпуск продукции, то дальнейшее увеличение выпуска пропорционально увеличению затрат на сырье с коэффициентом пропорциональности λ.

С л е д с т в и е. Используя необходимые условия $\lambda = \frac{(y_{\max})'_{x_i}}{p_i}$, где $i = 1, 2$, а также найденное выражение для предложения по лимиту $\lambda = \frac{\partial y_{\max}}{\partial C}$, можем записать

$$\frac{1}{p_i}\frac{\partial y_{\max}}{\partial x_i} = \frac{\partial y_{\max}}{\partial C}.$$

Отношение предельной полезности ресурса к цене за ресурс равно предельному предложению по лимиту и не зависит от вида ресурса.

Продифференцируем оба равенства системы по p_1 и используем необходимые условия:

$$\begin{cases} \frac{\partial y_{\max}}{\partial p_1} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial p_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial p_1} = \lambda p_1\frac{\partial x_1}{\partial p_1} + \lambda p_2\frac{\partial x_2}{\partial p_1} = \\ \qquad = \lambda\left(p_1\frac{\partial x_1}{\partial p_1} + p_2\frac{\partial x_2}{\partial p_1}\right) = -\lambda x_1, \\ x_1 + p_1\frac{\partial x_1}{\partial p_1} + p_2\frac{\partial x_2}{\partial p_1} = 0. \end{cases}$$

Итак, $\frac{\partial y_{\max}}{\partial p_1} = -\lambda x_1$. Отсюда следует, что *предельное предложение по цене за первый ресурс пропорционально объему ресурса*. Дадим пояснение. Заменим $\frac{\partial y_{\max}}{\partial p_1}$ на $\frac{\Delta y_{\max}}{\Delta p_1}$. Тогда

$$y_{\max} = -\lambda x_1 \cdot \Delta p_1.$$

Предприятие, работающее в оптимальном режиме, уменьшит выпуск продукции пропорционально как увеличению цены на ресурс, так и объему закупаемого ресурса.

Вариации выпуска продукции при изменении цены на второй ресурс находятся аналогично.

С л е д с т в и е. Найдем λ из необходимых условий и формулы предельного предложения по цене и приравняем выражения. Из равенства легко получить

$$\frac{\dfrac{\partial y_{\max}}{\partial x_i}}{\dfrac{\partial y_{\max}}{\partial p_i}} = -\frac{p_i}{x_i}.$$

Отношение предельной полезности ресурса к предельной полезности денег за ресурс равно цене единицы этого ресурса, взятой со знаком минус, и не зависит от вида ресурса.

Обоснованные в параграфе зависимости вариаций максимального выпуска продукции от параметров задачи получили название *теорем о маргинальных значениях.*

27.2. МИНИМИЗАЦИЯ ИЗДЕРЖЕК ПРИ ФИКСИРОВАННОМ ОБЪЕМЕ ВЫПУСКА

Работа предприятия с фиксированным объемом выпуска продукции y_0 описывается производственной функцией $y_0 = f(x_1, x_2)$. Закупая ресурсы x_1, x_2 по иенам p_1, p_2, предприятие несет издержки на сумму C д.е. Задача минимизации издержек при фиксированном объеме сводится к решению задачи на условный экстремум при выполнении требований: $y = f(x_1, x_2)$ — дважды непрерывно дифференцируемая функция, $x_1 \geq 0$, $x_2 \geq 0$, $p_1 > 0$, $p_2 > 0$:

$$\begin{cases} p_1x_1 + p_2x_2 = C \Rightarrow c.\min, \\ y_0 = f(x_1, x_2). \end{cases}$$

Воспользуемся методом Лагранжа. Для функции Лагранжа

$$L(x_1, x_2, \lambda) = p_1x_1 + p_2x_2 + \lambda(y_0 - f(x_1, x_2))$$

укажем необходимые условия экстремума:

$$\begin{cases} L'_{x_1}(x_1, x_2, \lambda) = p_1 - \lambda f'_{x_1}(x_1, x_2) = 0, \\ L'_{x_2}(x_1, x_2, \lambda) = p_2 - \lambda f'_{x_2}(x_1, x_2) = 0. \end{cases}$$

Перенеся вторые слагаемые направо и разделив одно равенство на другое, найдем

$$\frac{f'_{x_1}(x_1, x_2)}{f'_{x_2}(x_1, x_2)} = \frac{p_1}{p_2}.$$

Выразим из уравнения переменную $x_2 = x_2(p_1, p_2, x_1)$ и подставим в уравнение связи:

$$y_0 = f(x_1, x_2(p_1, p_2, x_1)).$$

Найдем в критической точке значение переменной $x_1^0 = x_1^0(p_1, p_2, y_0)$.

Аналогично получим $x_2^0 = x_2^0(p_1, p_2, y_0)$.

Перейдем к достаточным условиям экстремума:

$$\begin{cases} L''_{x_1x_1} = -\lambda f''_{x_1x_1}, \\ L''_{x_1x_2} = -\lambda f''_{x_1x_2}, \\ L''_{x_2x_2} = -\lambda f''_{x_2x_2}. \end{cases}$$

Составим окаймленный гессиан:

$$|H| = \begin{vmatrix} 0 & p_1 & p_2 \\ p_1 & L''_{x_1x_1} & L''_{x_1x_2} \\ p_2 & L''_{x_1x_2} & L''_{x_2x_2} \end{vmatrix} = \begin{vmatrix} 0 & f'_{x_1} & f'_{x_2} \\ f'_{x_1} & -\lambda f''_{x_1x_1} & -\lambda f''_{x_1x_2} \\ f'_{x_2} & -\lambda f''_{x_1x_2} & -\lambda f''_{x_2x_2} \end{vmatrix}.$$

Условие $|H| < 0$ есть условие минимума функции Лагранжа. Раскрыв определитель, получим неравенство

$$(f'_{x_1})^2 \cdot \lambda f''_{x_2x_2} + (f'_{x_2})^2 \cdot \lambda f''_{x_1x_1} - 2f'_{x_1}f'_{x_2} \cdot \lambda f''_{x_1x_2} < 0.$$

Рассматривая левую часть неравенства как квадратичную форму относительно переменных f'_{x_1}, f'_{x_2}, выделим матрицу квадратичной формы:

$$\hat{H} = (f'_{x_1}, f'_{x_2})\begin{pmatrix} \lambda f''_{x_2x_2} & -\lambda f''_{x_1x_2} \\ -\lambda f''_{x_1x_2} & -\lambda f''_{x_1x_1} \end{pmatrix}\begin{pmatrix} f'_{x_1} \\ f'_{x_2} \end{pmatrix} < 0.$$

Квадратичная форма является отрицательно определенной, если угловые миноры ее матрицы чередуют знаки, начиная с отрицательного:

$$M_1 = \lambda f''_{x_2x_2} < 0, \quad M_2 = \begin{vmatrix} \lambda f''_{x_2x_2} & -\lambda f''_{x_1x_2} \\ -\lambda f''_{x_1x_2} & \lambda f''_{x_1x_1} \end{vmatrix} > 0.$$

Эти неравенства с учетом следствия из них $f''_{x_1x_1} < 0$, записанные в виде системы

$$\begin{cases} \lambda f''_{x_1x_1} < 0, \\ \lambda f''_{x_2x_2} < 0, \\ \lambda^2 f''_{x_1x_1} f''_{x_2x_2} - \lambda^2 (f''_{x_1x_2})^2 > 0, \end{cases}$$

задают достаточные условия существования локального минимума функции Лагранжа и условного минимума издержек.

Теоремы о маргинальных значениях (издержки при вариации объема выпуска или цены за ресурс). Подставим x_1^0, x_2^0 в функцию издержек и в уравнение связи:

$$\begin{cases} C_{\min} = p_1 \cdot x_1^0(p_1, p_2, y_0) + p_2 \cdot x_2^0(p_1, p_2, y_0), \\ y_0 = f(x_1^0(p_1, p_2, y_0), x_2^0(p_1, p_2, y_0)). \end{cases}$$

Второе уравнение связывает две функции x_1^0, x_2^0 и три аргумента p_1, p_2, y_0. Из этой системы получим некоторые важные утверждения, предварительно убрав для удобства верхние индексы.

Продифференцируем оба равенства системы по y_0 и воспользуемся необходимыми условиями $\lambda f'_{x_1} = p_1$ и $\lambda f'_{x_2} = p_2$:

$$\begin{cases} \dfrac{\partial C_{\min}}{\partial y_0} = p_1 \dfrac{\partial x_1}{\partial y_0} + p_2 \dfrac{\partial x_2}{\partial y_0} = \lambda \dfrac{\partial f}{\partial x_1} \dfrac{\partial x_1}{\partial y_0} + \lambda \dfrac{\partial f}{\partial x_2} \dfrac{\partial x_2}{\partial y_0} = \\ \qquad = \lambda \left(\dfrac{\partial f}{\partial x_1} \dfrac{\partial x_1}{\partial y_0} + \dfrac{\partial f}{\partial x_2} \dfrac{\partial x_2}{\partial y_0} \right) = \lambda, \\ 1 = \dfrac{\partial f}{\partial x_1} \dfrac{\partial x_1}{\partial y_0} + \dfrac{\partial f}{\partial x_2} \dfrac{\partial x_2}{\partial y_0}. \end{cases}$$

Получено равенство $\dfrac{\partial C_{\min}}{\partial y_0} = \lambda$, означающее, что *предельные издержки по объему* $\dfrac{\partial C_{\min}}{\partial y_0}$ *равны множителю Лагранжа* λ.

Перейдем в первом уравнении к приращениям:

$\Delta C_{\min} \approx \lambda \cdot \Delta y_0$.

В случае оптимизации издержек их рост пропорционален росту выпуска продукции с коэффициентом пропорциональности λ.

Следствие. Найдем коэффициент λ из необходимых условий и приравняем полученному значению предельных издержек по объему. После преобразований получим

$$\frac{\partial y_0}{\partial x_i} \cdot \frac{\partial C_{\min}}{\partial y_0} = p_i, \quad i = 1, 2.$$

Произведение предельной полезности ресурса $\dfrac{\partial y_0}{\partial x_i}$ на предельные издержки по объему $\dfrac{\partial C_{\min}}{\partial y_0}$ равно цене за ресурс и не зависит от вида ресурса.

Вернувшись к системе, продифференцируем оба равенства по p_1:

$$\begin{cases} \dfrac{\partial C_{\min}}{\partial p_1} = x_1 + p_1\dfrac{\partial x_1}{\partial p_1} + p_2\dfrac{\partial x_2}{\partial p_1} = x_1 + \lambda\dfrac{\partial f}{\partial x_1}\dfrac{\partial x_1}{\partial p_1} + \lambda\dfrac{\partial f}{\partial x_2}\dfrac{\partial x_2}{\partial p_1} = \\ \qquad = x_1 + \lambda\left(\dfrac{\partial f}{\partial x_1}\dfrac{\partial x_1}{\partial p_1} + \dfrac{\partial f}{\partial x_2}\dfrac{\partial x_2}{\partial p_1}\right) = x_1, \\ \dfrac{\partial y_0}{\partial p_1} = 0 = \dfrac{\partial f}{\partial x_1}\dfrac{\partial x_1}{\partial p_1} + \dfrac{\partial f}{\partial x_2}\dfrac{\partial x_2}{\partial p_1}. \end{cases}$$

Вывод из полученных равенств: $\dfrac{\partial C_{\min}}{\partial p_1} = x_1$. Значит, *предельные издержки по цене за ресурс равны величине условного спроса на этот ресурс*. Зависимость от p_2 аналогична: $\dfrac{\partial C_{\min}}{\partial p_2} = x_2$.

В приращениях: $C_{\min} \approx x_1 \cdot \Delta p_1$. Увеличение издержек пропорционально увеличению цены на ресурс и зависит от величины потребляемого ресурса.

27.3. ОПТИМИЗАЦИЯ ПОТРЕБИТЕЛЬСКОГО ПОВЕДЕНИЯ

Спрос на товары при максимизации целевой функции

Рассмотрим задачу зависимости спроса на товары от дохода и цен при максимизации объема личного потребления. Пусть $U(x)$ — целевая функция потребления, характеризующая предпочтения потребителя, $x = (x_1, x_2, ..., x_n)$ — вектор объемов потребления благ x_i, n — количество рассматриваемых благ. Покупая для личного потребления товары x_i по ценам p_i, покупатель ограничен размером своего дохода I. Ставится задача оптимизации потребительского поведения с бюджетным ограничением

$$\begin{cases} U(\boldsymbol{x}) \to \max, \\ \sum\limits_{i=1}^{n} p_i x_i = I. \end{cases}$$

Решением задачи потребительского выбора является функция спроса — вектор $\boldsymbol{x}^0 = \boldsymbol{x}^0(\boldsymbol{p}, I) = (x_1^0(\boldsymbol{p}, I), x_2^0(\boldsymbol{p}, I), ..., x_n^0(\boldsymbol{p}, I))$, каждая координата которого зависит от вектора цен $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ и дохода I. Получение функции спроса требует знания структуры функции потребления, поэтому задача может быть решена только для частных случаев, один из которых мы рассмотрим ниже.

Решается задача на условный экстремум целевой функции Кобба—Дугласа двух переменных

$$\begin{cases} U(\boldsymbol{x}) = ax_1^{\alpha}x_2^{\beta} \to \max, \\ p_1x_1 + p_2x_2 = I. \end{cases}$$

Будем предполагать, что целевая функция дважды непрерывно дифференцируема, $p_1 > 0$, $p_2 > 0$, $x_1 > 0$, $x_2 > 0$, $a > 0$, $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$.

Составим функцию Лагранжа:

$$L(x_1, x_2) = ax_1^{\alpha}x_2^{\beta} + \lambda(I - p_1x_1 + p_2x_2)$$

и запишем систему уравнений из первых производных и уравнения связи:

$$\begin{cases} L'_{x_1} = \alpha a x_1^{\alpha-1}x_2^{\beta} - \lambda p_1 = 0, \\ L'_{x_2} = \beta a x_1^{\alpha}x_2^{\beta-1} - \lambda p_2 = 0, \\ p_1x_1 + p_2x_2 = I. \end{cases}$$

Перенеся вторые слагаемые направо и разделив первое уравнение на второе, придем к системе

$$\begin{cases} \dfrac{\alpha x_2}{\beta x_1} = \dfrac{p_1}{p_2}, \\ p_1x_1 + p_2x_2 = I, \end{cases}$$

решением которой являются

$$x_1^0 = \frac{\alpha I}{p_1}, \quad x_2^0 = \frac{\beta I}{p_2}.$$

В стационарной точке (x_1^0, x_2^0) функция Кобба—Дугласа $U(\boldsymbol{x})$, как доказывалось ранее, максимальна. Векторная функция спроса имеет вид

$$\boldsymbol{x}^0 = (x_1^0, x_2^0) = \left(\frac{\alpha I}{p_1}, \frac{\beta I}{p_2}\right).$$

Следовательно, *количество приобретаемого товара прямо пропорционально доходу и обратно пропорционально цене товара.*

27.4. ИЗМЕНЕНИЕ СПРОСА НА ТОВАРЫ ПРИ ВАРИАЦИИ ДОХОДА ПОТРЕБИТЕЛЯ

Пусть в задаче оптимизации потребительского поведения с бюджетным ограничением

$$\begin{cases} y = U(x_1, x_2), \\ p_1x_1 + p_2x_2 = I \end{cases}$$

выполнены условия: $y = U(x_1, x_2)$ — дважды непрерывно дифференцируемая функция, $x_1 > 0$, $x_2 > 0$, $p_1 > 0$, $p_2 > 0$.

Для функции Лагранжа

$$L = U(x_1, x_2) + \lambda(I - p_1x_1 - p_2x_2)$$

напишем необходимые условия:

$$\begin{cases} L'_{x_1} = U'_{x_1} - \lambda p_1 = 0, \\ L'_{x_2} = U'_{x_2} - \lambda p_2 = 0. \end{cases}$$

из которых, а также бюджетного ограничения следует система

$$\begin{cases} p_2U'_{x_1} = p_1U'_{x_2}, \\ p_1x_1 + p_2x_2 = I. \end{cases}$$

Полагая, что аргументом в системе является переменная I, а p_1, p_2 — параметры системы, перейдем к дифференциалам:

$$\begin{cases} p_2(U''_{x_1x_1}dx_1 + U''_{x_1x_2}dx_2) = p_1(U''_{x_2x_1}dx_1 + U''_{x_2x_2}dx_2), \\ p_1dx_1 + p_2dx_2 = dI. \end{cases}$$

Раскрыв скобки и перераспределив слагаемые, запишем равенства как линейную систему с двумя переменными dx_1, dx_2:

$$\begin{cases} (p_2U''_{x_1x_1} - p_1U''_{x_1x_2})dx_1 + (p_2U''_{x_1x_2} - p_1U''_{x_2x_2})dx_2 = 0, \\ p_1dx_1 + p_2dx_2 = dI. \end{cases}$$

По теореме Крамера

$$dx_1 = \frac{\begin{vmatrix} 0 & p_2U''_{x_1x_2} - p_1U''_{x_2x_2} \\ dI & p_2 \end{vmatrix}}{\begin{vmatrix} p_2U''_{x_1x_1} - p_1U''_{x_1x_2} & p_2U''_{x_1x_2} - p_1U''_{x_2x_2} \\ p_1 & p_2 \end{vmatrix}} =$$

$$= \frac{p_1U''_{x_2x_2} - p_2U''_{x_1x_2}}{p_1^2U''_{x_2x_2} - 2p_1p_2U''_{x_1x_2} + p_2^2U''_{x_1x_1}}dI.$$

Окончательно получим

$$dx_1 = \frac{\begin{vmatrix} p_1 & p_2 \\ U''_{x_1x_2} & U''_{x_2x_2} \end{vmatrix}}{|H|}dI.$$

Аналогичные преобразования для dx_2 приведут к формуле

$$dx_2 = \frac{\begin{vmatrix} p_1 & p_2 \\ U''_{x_1x_1} & U''_{x_1x_2} \end{vmatrix}}{|H|} dI.$$

Интерпретация изменений спроса на товары dx_1, dx_2 при бюджетной надбавке dI требует конкретизации целевой функции. Вернемся к функции Кобба—Дугласа с соответствующими достаточно общими ограничениями.

Тогда

$$dx_1 = \frac{p_1U''_{x_2x_2} - p_2U''_{x_1x_2}}{p_1^2U''_{x_2x_2} - 2p_1p_2U''_{x_1x_2} + p_2^2U''_{x_1x_1}} dI =$$

$$= \frac{p_1\beta a(\beta-1)x_1^{\alpha}x_2^{\beta-2} - p_2\alpha\beta a x_1^{\alpha-1}x_2^{\beta-1}}{p_1^2\beta(\beta-1)ax_1^{\alpha}x_2^{\beta-2} - 2p_1p_2\alpha\beta a x_1^{\alpha-1}x_2^{\beta-1} + p_2^2\alpha(\alpha-1)ax_1^{\alpha-2}x_2^{\beta}} dI.$$

Умножим числитель и знаменатель на $x_1^{2-\alpha}x_2^{2-\beta}$ и воспользуемся равенством $\alpha + \beta = 1$:

$$dx_1 = \frac{-p_1x_1^2 - p_2x_1x_2}{-p_1^2x_1^2 - 2p_1p_2x_1x_2 - p_2^2x_2^2} dI = \frac{x_1(p_1x_1 + p_2x_2)}{(p_1x_1 + p_2x_2)^2} dI = \frac{x_1}{I} dI.$$

Соответственно, $dx_2 = \dfrac{x_2}{I} dI.$

Увеличение спроса на товары пропорционально увеличению дохода потребителя и зависит от количества товара потребления.

Результат согласуется с решением задачи потребительского выбора по нахождению функции спроса. Действительно, из двух последних соотношений следует $\dfrac{dx_i}{x_i} = \dfrac{dI}{I}$, где $i = 1, 2$. После интегрирования $\int \dfrac{dx_i}{x_i} = \int \dfrac{dI}{I}$ получим $\ln x_i = \ln I + \ln c$, откуда $x_i = cI$. Коэффициент c, как было найдено, равен $\dfrac{\alpha}{p_1}$ или $\dfrac{\beta}{p_2}$.

С л е д с т в и е. Ранее мы получили выражение для предельного предложения по лимиту, которое в новых переменных имеет вид $\dfrac{\partial U}{\partial I} = \lambda$. Из необходимых условий в задаче оптимизации потребительского поведения с бюджетным ограничением найдем λ: $\lambda = \dfrac{U'_{x_i}}{p_i}$.

Приравняв выражения, получим

$$\frac{1}{p_i}\frac{\partial U}{\partial x_i} = \frac{\partial U}{\partial I}.$$

Отношение предельной полезности ресурса к цене за ресурс равно предельной полезности денег и не зависит от вида ресурса.

27.5. МАКСИМИЗАЦИЯ ПРИБЫЛИ В ПРОЕКТНОМ АНАЛИЗЕ

В процессе принятия инвестиционных решений приходится решать задачу максимизации планируемой прибыли. Одна из функций, с помощью которой в проектном анализе принимается инвестиционное решение (критерий), называется *чистым дисконтированным доходом NPV* (*Net Present Value*). Экономический смысл этой функции состоит в максимизации разности между всеми проектными доходами и затратами, т.е. прибыли. Однако, следуя теории проектного анализа, в этом расчете необходимо учесть время произведения затрат и получения доходов, а также процентную ставку. Все это находит отражение в критерии NPV через применение теории изменения ценности денег во времени.

Кладя деньги в банк, мы рассчитываем на денежный прирост. Он определяется по формуле сложных процентов:

$$X(t) = X(0)\left(1 + \frac{r}{100}\right)^t,$$

где $X(0)$ — первоначальный вклад, положенный в банк на t лет под r годовых процентов. Поставим теперь задачу так. Какую сумму надо положить в банк под r % в год, чтобы через t лет получить $X(t)$ денежных единиц? Другими словами, найти стоимость суммы денег, полученной через t лет, но приведенной к настоящему времени. Очевидно, она равна

$$X(0) = \frac{X(t)}{\left(1 + \frac{r}{100}\right)^t}.$$

Величину r можно назвать *ставкой дисконта*[1], отражающей изменение стоимости денег во времени.

Предположим, рассчитывается проект, который принесет через год доход в размере $b(1)$ денежных единиц, через два года — $b(2)$ де-

---

[1] Дисконт — от англ. *discount* — скидка.

нежных единиц и т.д. Тогда, пересчитанный к настоящему времени, он составит в течение t лет величину, равную

$$B_{\text{enefits}} = \frac{b(1)}{1+\frac{r}{100}} + \frac{b(2)}{\left(1+\frac{r}{100}\right)^2} + \ldots + \frac{b(t)}{\left(1+\frac{r}{100}\right)^t} = \sum_{n=1}^{t} \frac{b(n)}{\left(1+\frac{r}{100}\right)^n}.$$

Реализация проекта потребует затрат в каждом году жизненного цикла проекта $e(1)$, $e(2)$, ..., $e(t)$. Пересчитанные к настоящему времени, ежегодные затраты составят величину

$$E_{\text{xpenses}} = \frac{e(1)}{1+\frac{r}{100}} + \frac{e(2)}{\left(1+\frac{r}{100}\right)^2} + \ldots + \frac{e(t)}{\left(1+\frac{r}{100}\right)^t} = \sum_{n=1}^{t} \frac{e(n)}{\left(1+\frac{r}{100}\right)^n}.$$

Рассчитаем чистый дисконтированный доход *NPV* (*Net Present Value*), составляющий, как уже указывалось, разность между приведенными к настоящему времени будущими потоками доходов и затрат с учетом инвестиции $I_{\text{nvestment}}$ в начальный момент. Получим

$$NPV = -I_{\text{nvesment}} + B_{\text{enefits}} - E_{\text{xpenses}} = -I + \sum_{n=1}^{t} \frac{b(n)-e(n)}{\left(1+\frac{r}{100}\right)^n}.$$

Чистый дисконтированный доход *NPV*, как уже говорилось, характеризует эффективность проекта и является одним из основных критериев при выборе проекта. Рассмотрим *NPV* подробнее. Он представляет собой функцию двух переменных: ставки дисконта r и времени жизни проекта (или горизонта его планирования) t. Время жизни проекта оказывается конечной величиной, обычно равной 5–10 годам, после чего требуется радикальная реорганизация, новое инвестирование или ликвидация проекта. В противном случае проект начнет приносить убытки.

Функция чистого дохода (прибыли) $P_{\text{rofit}} = b(n) - e(n)$ (доходы минус затраты) для каждого проекта индивидуальна. Однако некоторые свойства функции являются общими для реальных проектов: первоначальный рост, достижение максимума и последующий спад.

Конкретизируем немного проект. Пусть проект рассчитан на 10 лет. Рассмотрим два различных случая, отличающиеся различным поведением на заключительной стадии проекта.

1. В начале стадии убывания чистого дохода менеджеры поддерживают проект различными мерами, добиваясь превышения доходов над затратами. Пусть функция прибыли имеет вид (рис. 27.1)

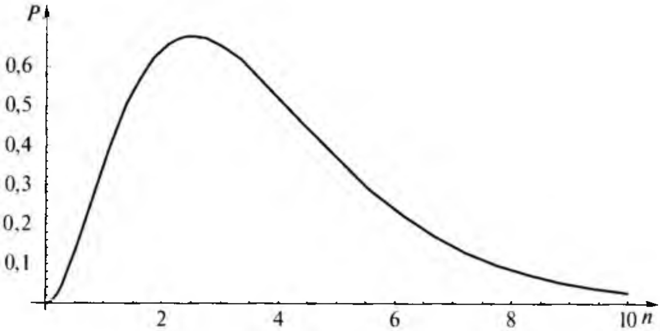

Рис. 27.1

$P = 0{,}8n^2 \cdot e^{-0{,}8n}$.

В начале проекта прибыль растет медленно, в течение трех лет выходит на максимум. Затем в течение последующих семи лет она падает по экспоненте до 10% от своей максимальной величины.

Зависимость функции NPV от ставки дисконта r в пределах 5–55% и времени развития проекта n по годам приведена на рис. 27.2. Плоскость на рисунке проведена на уровне $NPV = 0$ и отсекает участки с $NPV < 0$. Для каждого года проекта можно определить величину r (норму дисконта IRR), при которой затраты на проект окупятся. Для $r = 44\%$ это произойдет только в конце проекта (NPV станет больше нуля). При $r > 44\%$ проект станет убыточным. На рис. 27.2 темная линия отсекает ставку процента $r = 44\%$. Если ставка процента равна, например, 27%, проект окупится к четвертому году (дисконтированный период окупаемости — критерий $DPBP = 3$). Светлой стрелкой указана точка на поверхности, соответствующая $IRR = 27\%$ и $DPBP = 3$.

2. Проект не поддерживается новыми идеями, инвестициями, живет за счет внутренних ресурсов. Вся прибыль извлекается из проекта. Или в проекте обнаружились проблемы, которые привели во второй половине жизненного цикла проекта к превышению расходов над доходами. Пусть в этом случае функция прибыли имеет вид (рис. 27.3)

$P = 0{,}2n^2(1 - 0{,}17n)$.

В начале проекта прибыль растет медленно, в течение четырех лет она выходит на максимум. Затем в течение последующих двух лет прибыль падает до нуля. Расходы продолжают расти.

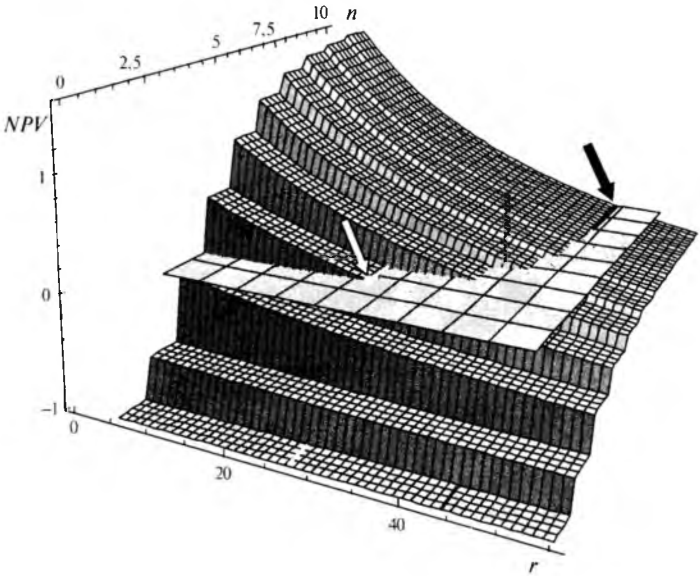

Рис. 27.2

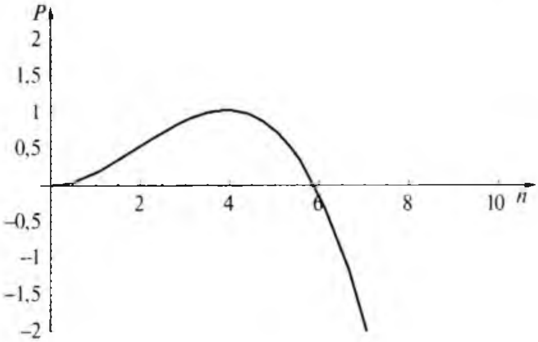

Рис. 27.3

Зависимость в этом случае функции *NPV* от ставки дисконта *r* в пределах 5–55% и времени развития проекта *n* по годам приведена на рис. 27.4. По-прежнему плоскость на рисунке проходит через ординату со значением *NPV* = 0. В первые годы проект развивается благополучно. При невысокой ставке процента прибыль появляется на

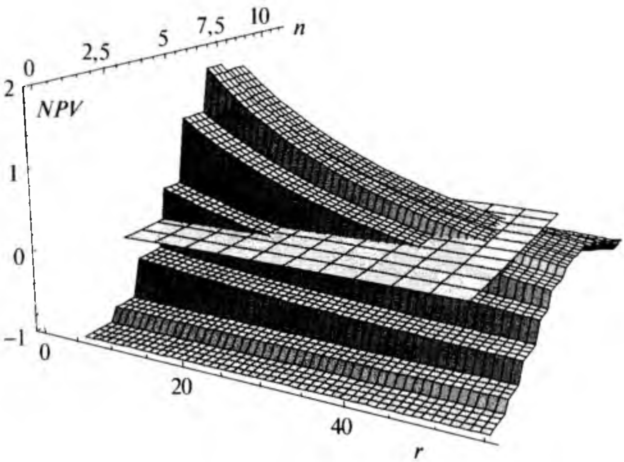

Рис. 27.4

четвертом году реализации и достигает на 5–6-м годах максимальных значений. Однако в последние три года могут развиться катастрофические события. Что на первый взгляд удивительно, катастрофа возникает при малых ставках дисконта, в то время как при больших ставках проект затухает медленно. Это видно на рис. 27.5. Причина этого явления в том, что при низкой процентной ставке деньги с течением времени остаются «дорогими». Потери денежных средств

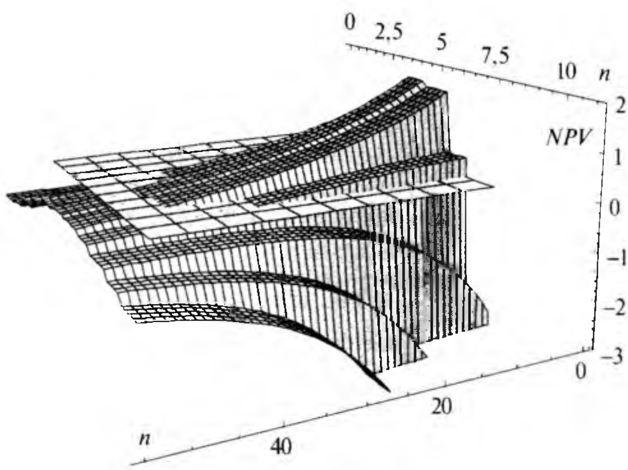

Рис. 27.5

на 7—10-м годах проекта в настоящем времени весьма чувствительны. При высокой ставке процента деньги на 7—10-м годах проекта становятся «дешевыми». Их потеря не столь чувствительна.

Действительно, доход, полученный от проекта на n-м году жизненного цикла, пересчитанный к началу проекта, равен

$$\frac{b(n)-e(n)}{\left(1+\frac{r}{100}\right)^n}.$$

Отношение прибылей, а в данном случае убытков при низкой ставке r_1 и высокой ставке r_2 составит величину

$$\frac{\text{Убытки}_{r_1}}{\text{Убытки}_{r_2}}=\frac{\dfrac{b(n)-e(n)}{\left(1+\frac{r_1}{100}\right)^n}}{\dfrac{b(n)-e(n)}{\left(1+\frac{r_2}{100}\right)^n}}=\frac{\left(1+\frac{r_2}{100}\right)^n}{\left(1+\frac{r_1}{100}\right)^n}=\left(\frac{100+r_2}{100+r_1}\right)^n.$$

Для $r_1 = 5\%$ и $r_2 = 50\%$ на 7-м году жизни проекта будем иметь

$$\left(\frac{100+r_2}{100+r_1}\right)^n=\left(\frac{150\%}{105\%}\right)^7=12{,}14.$$

Это означает, что потери составят при низкой процентной ставке в 12 раз большую величину, чем при высокой. Отсюда вывод: при низкой ставке процента следует более зорко следить за прибылями-убытками. При достижении максимума прибыли и начале спада следует принимать решительные меры, ибо «промедление смерти подобно». Проект при низкой процентной ставке и неблагоприятном стечении обстоятельств может быть потерян для инвестора с большими убытками.

Глава 28
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

Моделью называется представление объекта или процесса в некоторой форме, которая достаточно хорошо отражает изучаемые свойства объекта или процесса. Например, можно построить модель самолета настолько точно, что она будет неотличима от прототипа. Зачем ее строить? Дело в том, что, во-первых, модель стоит гораздо дешевле. Например, нужно создать копию, повторив внешний вид реального самолета. Копия будет сделана из дешевых материалов и не будет иметь внутренней начинки. Или, сделав точную копию конструируемого самолета в пропорции 1:30, можно исследовать его некоторые характеристики в аэродинамической трубе.

Мы будем строить математические модели, т.е. представлять изучаемые свойства объекта или процессы, используя математические символы и соотношения. Свойство объекта, переведенное на язык формул, превращается в математическую модель и становится более доступным для изучения. Математическая модель подчиняется четким и ясным законам. К ней нельзя применить в виде исключения другой закон только потому, что это кому-то выгодно.

Трудность при построении математической модели в другом. Нужно суметь передать в ней суть изучаемого явления и указать ограничения, при которых ею можно пользоваться. Поэтому при создании математической модели должен применяться системный подход.

28.1. СИСТЕМНЫЙ ПОДХОД

Математическое моделирование широко применяется в естественно-научных дисциплинах и в конструировании, в экономике и управлении. Моделирование в управлении позволяет:

- выделить и аналитически описать наиболее важные, существенные связи экономических переменных в объекте;
- применить для изучения объекта управления высокую степень формализации;
- путем математических вычислений получить новые знания об объекте;
- оценить форму и параметры зависимостей переменных объекта, в наибольшей степени соответствующие имеющимся наблюдениям;

- точно и кратко изложить положения экономико-математической теории, сформулировать ее понятия и выводы.

Системный подход — это всестороннее исследование объектов и явлений как единого целого с учетом всех их взаимосвязей и свойств.

Системный анализ является самостоятельной научной дисциплиной, которая выделилась в результате развития системного подхода в управлении исследуемыми объектами. В частности, системный анализ применяется в процессе принятия решений. Системный анализ различных процессов позволяет представить их как динамические системы. Изучение системы зависит от цели исследования. Рассматривая объект с определенной целью, мы выделяем в нем некоторые элементы и взаимосвязи между ними, с тем чтобы полученного описания было достаточно для объяснения функционирования системы с точки зрения избранной цели. Следовательно, структура понятия системы такова:

Объект — Цель — Элементы — Взаимосвязи.

Системный подход появился в связи с необходимостью исследования вопросов, находящихся на «стыке» наук. Создание коммерческих фирм, народнохозяйственных объектов, анализ экономических, управленческих, экологических проблем часто требуют проведения исследований общего характера, или системных исследований.

Потребность не просто изучать объекты, но и влиять на их поведение определяет основные задачи теории систем:

Анализ, синтез, оптимизация.

Задача *анализа* сводится к исследованию объекта, к сбору и обобщению имеющейся информации об исследуемом объекте как о системе. Задача *синтеза* сводится к созданию новой системы или теории, позволяющей увязать различные факты, увидеть перспективы развития систем на основе имеющейся информации. Задача *оптимизации* позволяет сделать выбор наилучшего варианта из числа возможных.

Каждая система (объект исследования) не существует сама по себе. Она взаимодействует с окружающим ее миром (*внешней средой*). Воздействие внешней среды на систему часто называют *входом*, а воздействие системы на внешнюю среду — *выходом*. При принятии управленческого решения учитывается информация о цели системы, внешней среде и входе и выходе системы.

Наличие *обратной связи* является характерной чертой всякого процесса управления. Обратная связь дает возможность достигать цели практически с любой точностью, позволяет системе функционировать в изменяющихся условиях и развиваться.

В экономических системах наиболее приемлемой формой количественного исследования является изучение функциональных зависимостей накопленных, суммарных, средних и предельных величин. Исследование по возможности проводится на аналитическом уровне и содержит графическую составляющую.

Системное применение в этом случае математических методов количественной оценки и анализа не только уточняет представления об изучаемых объектах и процессах, но и предсказывает ход их дальнейшего развития. Огромная степень абстракции, строгая внутренняя логика математики дают возможность моделировать реальную экономику и тем самым проводить экономические лабораторные исследования и эксперименты, проверку гипотез.

Системный анализ логической структуры экономической проблемы позволяет выделить следующие основные элементы:

- цель, достижение которой означает, что проблема решена;
- действия для достижения цели;
- затраты ресурсов при выполнении каждого способа действий;
- модели, в которых находят выражение связи между целями, альтернативами и затратами;
- критерий (показатель эффективности), по которому сравниваются цели и затраты и находится лучшее управленческое решение.

Слово «модель» ведет свое происхождение от латинского слова *modulus*, что значит мера, мерило, образец, норма. Под моделированием понимается конструирование модели и работа с ней, состоящие из ряда последовательных и взаимосвязанных стадий или этапов: постановка задачи, построение модели, ее исследование, проверка и оценка полученного на основе модели решения, реализация результатов решения. Исследование модели позволяет получить наилучшее из всех возможных вариантов (оптимальное) решение поставленной задачи. Методы поиска оптимального решения зависят от степени изученности реального объекта и вида построенной модели.

Экономико-математические методы применяются для решения различных задач. Методы делятся на несколько направлений, представленных следующей таблицей (табл. 28.1).

Таблица 28.1

Экономико-математические методы

Направление	Области применения
Экономико-статистические методы: методы математической статистики (включая корреляционный анализ и теорию планирования эксперимента)	Используются при большом количестве статистических данных об объекте. Изучение последних позволяет установить тенденцию изменения параметров (например, цен) и сделать краткосрочный прогноз развития экономического процесса. Широко применяются для анализа конъюнктуры товаров, прогнозирования изменения цен на мировом рынке
Методы кибернетики	Тесно связаны с моделями функционирования экономических систем, поскольку основываются на системном исследовании объекта. Из известных научных методов здесь используются методы распознавания образов и теория классификации, теория автоматического регулирования (особенно для объектов с быстро меняющимися параметрами), теория автоматизированных систем управления
Методы исследования операций и нахождения оптимальных решений	Применяются для решения проблемы нахождения наилучшего решения из числа возможных для достижения определенной цели. При этом сами возникающие экономические ситуации и объекты исследования могут быть весьма разнообразны. К исследованию операций относятся методы математического программирования (линейное, нелинейное, динамическое и т.д.), теория управления запасами, теория расписания и массового обслуживания, сетевые методы планирования и управления, теория игр, различные эвристические методы нахождения оптимальных решений

28.2. КРАТКИЙ ИСТОРИЧЕСКИЙ ОБЗОР

Математические модели в экономике использовались с иллюстративными и исследовательскими целями А. Смитом (классическая макроэкономическая модель), Д. Рикардо (модель международной торговли). Начиная со второй половины XIX в., когда возникла так называемая математическая школа, в экономических исследованиях стали придавать решающее значение математике как методу изучения экономических явлений. Видными представителями этой школы являются Л. Вальрас, В. Парето, Ф. Эджуорт и другие. Особенность теоретических построений математической школы — ориентирование на маржинализм. Активное использование предельных категорий (предельная полезность, предельная эффективность, предельная производительность) позволили им получить из сформулированных на математическом языке предпосылок выводы, которые иным путем получить было нельзя.

В соответствии со своей моделью равновесия народного хозяйства Вальрас ориентировался на показатели, характеризующие поведение отдельных производителей и потребителей. Каждый производитель характеризуется функцией предложения, каждый потребитель — функцией спроса. В модели с помощью равновесных цен обеспечивается равенство спроса и предложения по каждому товару. Был осуществлен анализ условий равновесия рыночной экономики. Полученные результаты оказали большое влияние на экономистов середины XX в., когда были созданы первые межотраслевые модели народного хозяйства.

Пионерские попытки разработки научного подхода к организации труда и производства, к учету человеческого фактора в промышленности, предпринятые Ч. Бэббиджем, Ф. Тейлором и другими, позволили получить эффективные решения целого ряда конкретных задач. Например, введение в Великобритании в 1840 г. почтовой оплаты в 1 пенни существенно упростило процедуру обработки корреспонденции. Бэббидж, изучая операции в почтовом ведомстве, нашел, что большая часть стоимости письма приходится на его обработку при сортировке, а вовсе не на дальность путешествия от отправителя к получателю, как это считалось ранее.

Начало XX в. отмечено попытками создать теорию управления инвестициями (Ф. Харрис), теорию массового обслуживания (теория очередей (А. Эрланг)). Существенное продвижение было сделано при попытках разрешения целого ряда проблем управления, возникших непосредственно перед и в ходе Второй мировой войны. Наиболее известным примером могут служить результаты работы британской группы экспертов, состоявшей из 11 человек, оказавшие заметное влияние на исход битвы за Англию и сражений в Северной Атлантике (группа «*Blackett's Circus*», в которую входили физиологи, математики, физики, геодезист, астрофизик и военный). В первой половине XX в. начали разрабатывать элементы научного подхода к поиску решений задач управления, а схемы, хорошо показавшие себя при проведении естественно-научных и инженерно-технических изысканий, стали пытаться приспосабливать к решению управленческих задач.

Достижения в теории игр и в программировании, а также создание новых средств вычислений обеспечили существенный прорыв в переводе задач управления на язык математики и помогли разработать методы решения и анализа. В середине XX в. оформилась новая экономико-управленческая дисциплина — исследование операций. Математические методы моделирования стали применяться очень широко, с их использованием связаны практически все работы, удостоенные Нобелевской премии по экономике (Д. Хикс, Р. Солоу, В. Леонтьев, П. Самуэльсон и др.).

В России в 1960–80-е годы экономико-математическое направление (В.С. Немчинов, В.В. Новожилов, Л.В. Канторович) было связано с попытками математически описать «систему оптимального функционирования социалистической экономики». Строились многоуровневые системы моделей народнохозяйственного планирования, оптимизационные модели отраслей и предприятий. Сейчас важной задачей является моделирование процессов переходного периода.

Любое экономическое исследование всегда предполагает объединение теории (экономической модели) и практики (статистических данных). Мы используем теоретические модели для описания и объяснения наблюдаемых процессов и собираем статистические данные с целью эмпирического построения и обоснования моделей.

28.3. МОДЕЛИРОВАНИЕ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

Конструктивно каждая модель представляет собой совокупность взаимосвязанных математических зависимостей (уравнений или неравенств), отображающих определенные группы реальных экономических зависимостей. Параметры описываемых экономических объектов выступают в модели в качестве либо известных, либо неизвестных величин. Известные величины рассчитываются вне модели и вводятся в нее в готовом виде, поэтому их часто называют *экзогенными* (от греческих корней *exo* — снаружи, вне и *genos* — происхождение). Значения неизвестных величин определяются только в результате решения экономической задачи в рамках модели, поэтому их часто называют *эндогенными* (от греч. *endon* — внутри и *genos* — происхождение).

Как правило, экономико-математические модели конструируются таким образом, что значения эндогенных переменных определяются в них неоднозначно. Здесь открывается возможность управлять экономическими решениями и выбирать среди допустимых значений неизвестных такие, которые соответствовали бы формализованным представлениям об их наилучших вариантах. Формализованные представления о качестве наилучших решений принимают логическую форму критериев эффективности управления и адекватную им математическую форму целевых функций в задачах оптимизации управления экономическими системами.

Модели экономических процессов можно классифицировать по самым различным признакам.

По целевому назначению они могут быть либо теоретико-аналитическими, предназначенными для научного исследования механизма протекания соответствующих процессов, либо прикладными, предназначенными для решения задач конкретно-экономического анализа и планирования народного хозяйства.

По характеру зависимостей в связи с временным фактором экономические модели могут быть либо *статическими*, все зависимости которых относятся к одному периоду (году), либо *динамическими*, описывающими процесс изменения объекта во времени.

По глубине временного горизонта динамические модели подразделяются на долгосрочные (15–20 лет), среднесрочные (5 лет) и краткосрочные (1 год).

По способу непосредственного отображения временного фактора и соответственно по характеру применяемого математического аппарата модели подразделяются на модели с *непрерывным изменением* переменных (аппарат дифференциального и интегрального исчисления) и с *дискретным изменением* переменных (аппарат разностных уравнений).

По характеру отображения причинно-следственных связей в управлении экономической системы модели можно разделить на *детерминированные* (параметры заданы достоверно) и *стохастические* (изменение их параметров связано с воздействием случайных величин).

По характеру взаимосвязи между переменными модели могут быть либо *линейными*, либо *нелинейными*.

По степени структуризации своего реального прототипа можно выделить *малоразмерные* (в том числе так называемые *однопродуктовые*) и *многоразмерные*.

Наконец, *по характеру требований к результатам решения задач управления* экономическим процессом модели могут быть либо *балансовыми*, либо *оптимизационными*.

Так как нас интересует в первую очередь математические модели экономических систем, то рассмотрим классификацию моделей по средствам моделирования. По этому признаку модели и методы моделирования подразделяются на материальные и идеальные.

Материальным называется моделирование, в котором исследование ведется на основе модели, воспроизводящей основные геометрические, физические, динамические и функциональные характеристики изучаемого объекта. Непосредственное исследование такой модели проводится в виде натурального эксперимента с моделью.

От материального (предметного) моделирования принципиально отличается *идеальное* моделирование, основывающееся на аналогии идеальной, мыслимой.

В идеальном моделировании различают интуитивное и знаковое моделирование. *Интуитивное* моделирование основано на личном опыте и знаниях исследователя. *Знаковое* моделирование — это формализованное моделирование, где моделями служат различные знаковые образования: схемы, графики, чертежи, формулы и т.д., причем знаковые образования и их элементы всегда задаются вместе с законами, по которым можно оперировать с ними. Важнейшим видом знакового моделирования является математическое моделирование, осуществляемое средствами языка математики и логики. Исследования на основе идеальных моделей носят теоретический характер, и их роль особенно велика в исследованиях по управлению экономикой, поскольку возможности проведения натурального эксперимента и эксперимента с натуральными моделями в них ограничены. Долгое время единственным методом анализа процессов управления было интуитивное моделирование, когда человек, принимающий управленческое решение, руководствовался той или иной неформализованной моделью рассматриваемой ситуации. Поскольку разные люди обладают различной степенью подготовленности, знаний, умений, реакции, то понимать интуитивную цель они могут по-разному, а следовательно, и давать на ее основе различные ответы на один и тот же вопрос. Проникновение математики в исследования по управлению экономикой создало основу для точного и строгого описания процессов и объяснения выводов, получаемых на основе экономико-математических моделей.

Для формализованного отображения экономико-математических систем и процесса их развития используются различные типы функциональных зависимостей. Тип зависимости между элементами системы может быть установлен на основе:

- аналитического выражения физических законов или общепринятых правил учета хозяйственной деятельности, например:

  Запасы (t) = *Запасы* ($t - I$) + *Производство* (t) – *Сбыт* (t);

- эмпирических данных за прошлые годы;

• нормативных данных, которые устанавливают, как переменные должны быть связаны между собой, а не то, как они были связаны в прошлом.

Часто нормативные соотношения устанавливаются на основании теоретического описания системы. При моделировании экономико-математических систем в большинстве случаев используется линейная функциональная связь элементов как самая простая с точки зрения выявления и описания. Одновременно предполагается, что система стационарна, т.е. ее реакция на всякое воздействие не зависит ни от времени воздействия, ни от предыстории системы.

Анализ реальных экономико-математических систем показывает, что они крайне редко бывают линейными. Стационарность таких систем — обычно также абстракция. Ввиду крайней сложности обращения с нелинейными и нестационарными системами при исследовании систем стараются выделить в них линейные и стационарные составляющие, с тем чтобы впоследствии можно было выразить через них другие характеристики систем. Так же поступают и при анализе систем в других отраслях знаний. Так, если наблюдения показывают заведомую неприемлемость гипотезы о линейной связи путь—время (при движении механического объекта), то испытывают гипотезу о линейности связи скорость—время, и если она подтверждается, информации о времени движения и об общих, абстрактных закономерностях движения оказывается достаточно, для того чтобы знать местоположение системы в любой момент времени. Или, если металлоемкость прироста выпуска продукции оказывается переменной, возможно, что окажется стационарной сама тенденция этой переменности. Выявление и логическое обоснование тех или иных линейных и линеаризуемых, стационарных или сводящихся к стационарным причинно-следственных связей составляют основу искусства моделирования.

В процессах управления экономикой причинно-следственные связи проявляются иногда не синхронно, а *с запаздыванием*, причем закономерности запаздывания могут быть самыми разными. Так, предложение, следующее за спросом, отстает от него. Для рациональной организации производства необходимо учесть это запаздывание, чтобы компенсировать его специально рассчитанным «упреждением». Принятие решения запаздывает по отношению к информации, обосновывающей это решение. Результаты принятого решения проявляются уже в новой обстановке, что необходимо учитывать. Обратные связи, приводящие к корректировке решений, также требуют времени на учет и анализ реального состояния системы. Капитальные вложения вызывают ввод основных фондов, но не всегда сразу. Сам по себе ввод вызывает прирост продукции, но не мгновенно, а через процесс освоения.

Все эти примеры — а здесь приведена лишь малая часть их бесконечного списка — содержат признаки одного и того же явления, определяемого обычно как запаздывание (*лаг*) выхода системы по отношению к входу.

Оценка параметров лага представляет собой очень сложную задачу, поскольку априори неизвестны ни форма, ни размах лага. Кроме того, на народнохозяйственном или отраслевом уровне прямое технико-экономическое измерение лага обычно невозможно. Следовательно, исчисление лага приходится производить, опираясь только на информацию о входном и выходном процессах, а моделирование лагов (как, впрочем, и других процессов) следует осуществлять с учетом условий функционирования моделируемой системы в целом. В этом состоит одно из важнейших требований системного подхода к описанию экономических объектов и управления ими.

Степень формализации управленческой задачи во многом определяет и методику поиска ее решения. Различают следующие типы задач: хорошо структурированные, слабоструктурированные и неструктурированные. Резкой грани между ними провести нельзя. К тому же нередко оказывается, что (сначала) слабоструктурированная проблема становится (потом) хорошо структурированной и даже стандартной.

По математическому подходу модели в задачах управления могут использовать статистический анализ, методы математического программирования, теорию очередей, сетевое планирование, теорию игр. Термин «программирование» в данном случае означает, что в задаче разыскивается некоторая программа действий.

Приведем данные по активности применения различных математических подходов при моделировании в задачах управления 125 корпорациями США (табл. 28.2).

Типовые экономические и управленческие модели

Для изучения различных экономических явлений и управления ими специалисты по управлению используют их упрощенные формальные описания, называемые экономическими моделями и моделями управления.

Примерами таких моделей являются:
- модели потребительского выбора;
- модели фирмы;

Таблица 28.2

	Частота использования, % корпораций		
	Редко	Умеренно	Постоянно
Статистический анализ	2	38	60
Имитационное моделирование	13	53	34
Сетевое планирование	26	53	21
Линейное программирование	26	60	14
Теория очередей	40	50	10
Нелинейное программирование	53	39	8
Динамическое программирование	61	34	5
Теория игр	69	27	4

- модели экономического роста;
- модели равновесия на товарных, факторных и финансовых рынках и многие другие.

Строя модели, менеджеры выявляют существенные факторы, определяющие исследуемое явление, и отбрасывают детали, несущественные для решения поставленной проблемы. Формализация основных особенностей функционирования экономических объектов позволяет оценить возможные последствия воздействия на них и использовать такие оценки в управлении.

Процесс построения экономической модели и управления ею включает следующие этапы:

- формулировка предмета и цели исследования;
- выделение в рассматриваемой экономической системе структурных или функциональных элементов, соответствующих данной цели;
- выявление наиболее важных качественных характеристик этих элементов;
- словесное, качественное описание взаимосвязи между элементами модели;
- введение символических обозначений для учитываемых характеристик объекта управления;
- формулирование математической модели через формализацию взаимосвязи между характеристиками объекта управления;
- расчеты по математической модели;
- анализ полученного решения.

После описания общих принципов построения математических моделей и указания особенностей экономико-математических моделей рассмотрим конкретную задачу управления и построим соответствующую ей экономико-математическую модель.

Фирма выпускает четыре вида различных продуктов. В производственном цикле участвуют три вида ресурсов: оборудование, рабочая сила и сырье. Количества их известны и ограничены. Задан расход каждого из ресурсов на производство единицы продукции каждого вида. Определены цены продуктов. Нужно найти объем производства каждого продукта с целью максимизации стоимости произведенной продукции.

Определим экзогенные и эндогенные переменные и параметры. Экзогенные переменные — имеющееся количество оборудования K, рабочей силы L и сырья R. Пусть заданы параметры — коэффициенты их расхода на единицу i-й продукции и k_i, l_i и r_i соответственно, где $i = 1, 2, 3, 4$. Цены продуктов p_i также известны. Введем обозначения для эндогенных переменных, которые определяются в ходе расчетов модели. Это неизвестные объемы производства продукции каждого i-го вида; обозначим их через x_i.

Поскольку количество оборудования K, рабочей силы L и сырья R ограничены, объемы производства каждого вида продукции должны удовлетворять системе неравенств

$$k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4 \le K,$$
$$l_1x_1 + l_2x_2 + l_3x_3 + l_4x_4 \le L,$$
$$r_1x_1 + r_2x_2 + r_3x_3 + r_4x_4 \le R.$$

Добавим требование неотрицательности переменных $x_i \ge 0$. Если бы мы ограничились только равенством в одном из соотношений, т.е. какой-то ресурс должен быть израсходован полностью, это сузило бы допустимое множество решений. Возможно, стоимость произведенной продукции уменьшилась бы. В оптимизационной модели определим целевую функцию, которая по условию задачи максимизируется:

$$f(x_1, x_2, x_3, x_4) = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 \Rightarrow \max.$$

Таким образом, экономико-математическая модель, описывающая конкретное производство, построена:

$$\begin{cases} f(x_1, x_2, x_3, x_4) = p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 \Rightarrow \max, \\ k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4 \le K, \\ l_1x_1 + l_2x_2 + l_3x_3 + l_4x_4 \le L, \\ r_1x_1 + r_2x_2 + r_3x_3 + r_4x_4 \le R, \\ x_i \ge 0, \quad i = 1, 2, 3, 4. \end{cases}$$

Решение данной задачи позволит из всех возможных вариантов использования рабочей силы, сырья и оборудования выбрать тот, при котором будет достигнута максимальная стоимость произведенной продукции. Реальная задача производства, конечно, более сложна. Возможно, дополнительно следует учесть, что:

- ресурсы могут быть взаимозаменяемы;
- затраты ресурсов не строго пропорциональны выпуску, есть постоянные затраты;
- объемы ресурсов могут увеличиваться в процессе производства за счет покупки или аренды;
- цена продукта может зависеть от объема его реализации;
- кроме максимизации стоимости, следует учитывать и другие экономические показатели;
- на ситуацию могут воздействовать случайные факторы, которые необходимо принять во внимание.

Более сложные экономико-математические модели учитывают эти аспекты, но принцип построения моделей остается прежним.

Математические модели, используемые в экономическом управлении, можно подразделять на *классы* по ряду признаков, относящихся к особенностям моделируемого объекта, цели моделирования и используемого инструментария, что отражено в табл. 28.3.

Следует сказать, что микроэкономическое моделирование в научном и учебном плане занимает основную часть экономико-математической теории. Упор делается на исследование стратегического поведения фирм. Преобладает подход в моделировании, основанный на оптимизации.

Экономико-математические модели рассматриваются в трех научных направлениях, сформировавшихся на стыке математики, экономики и управления. Это математическая экономика, эконометрика и, наконец, исследование операций.

Математическая экономика — научное направление, занимающееся анализом наиболее общих свойств и решений математических моделей экономических процессов и управления ими. Это наиболее математизированная часть экономико-математической науки, требующая от исследователя самого высокого математического уровня. Место математической экономики — на стыке математической теории, экономической теории и науки управления. Математическая экономика отделяется обычно от эконометрики, занимающейся статистической оценкой и анализом экономических зависимостей и моделей на основе изучения эмпирических данных. В математической экономике исследуются теоретические модели, основанные на определенных формальных предпосылках: линей-

Таблица 28.3

Класс моделей	Сущность класса моделей
Макроэкономические модели	Описание экономики как единого целого, связывающего между собой укрупненные материальные и финансовые показатели: • ВНП (валовой национальный продукт); • потребление; • инвестиции; • занятость; • процентную ставку; • количество денег и др.
Микроэкономические модели	Описание взаимодействия структурных и функциональных составляющих экономики либо поведения отдельной такой составляющей в рыночной среде
Теоретические модели	Позволяют изучать общие свойства экономики и ее характерных элементов с помощью дедукции выводов из формальных предпосылок
Прикладные модели	Дают возможность оценить параметры функционирования конкретного объекта и сформулировать рекомендации для принятия практических решений. К прикладным относятся прежде всего модели работы конкретного предприятия, оперирующие числовыми значениями экономических переменных и позволяющие статистически значимо оценивать их на основе имеющихся наблюдений
Равновесные модели	Описание таких состояний в управлении экономикой, когда результирующая всех сил, стремящихся вывести ее из данного состояния, равна нулю. В нерыночной экономике неравновесие по одним параметрам (например, дефицит) компенсируется другими факторами (черный рынок, очереди и т.п.). Равновесные модели обычно описательны
Оптимизационные модели	Нахождение наилучшего решения из числа возможных для достижения определенной цели
Статические модели	Описание состояния объекта управления в конкретный момент или период времени
Динамические модели	Учет взаимосвязи переменных во времени
Детерминированные модели	Использование жестких функциональных связей между переменными модели
Стохастические модели	Наличие случайных воздействий на исследуемые показатели и использование инструментария теории вероятностей и математической статистики для их описания

ность, выпуклость, монотонность, конкретность формул взаимосвязи величин.

Математическая экономика, вообще говоря, не занимается изучением степени обоснованности того, что данная зависимость имеет тот или иной вид (например, что величина потребления является линейной возрастающей функцией дохода), — это оставляется для эконометрики. Задачей математической экономики является изучение вопроса о существовании решения модели, условиях его неотрицательности, стационарности, наличия других свойств. Это обычно осуществляется, как и в математике, путем дедуктивного получения следствий (теорем) из априорно сделанных предпосылок (аксиом). Математическую экономику изучают и затем успешно в этой области работают выпускники экономических отделений математических факультетов университетов.

Эконометрика — наука, элементы которой рассматривались в разделе VI. Она, как вы могли прочитать, исследует количественные закономерности и взаимозависимости в экономике при помощи методов математической статистики с целью эффективного управления ими. Основа этих методов — корреляционно-регрессионный анализ. В 1897 г. появилась работа одного из основателей математической школы в экономической теории В. Парето, посвященная статистическому изучению доходов населения в разных странах. Была предложена кривая Парето $y = A(x - a)^{-\alpha}$, где x — величина дохода; y — численность лиц, имеющих доход, больший x; a — минимальный доход; A и a — параметры зависимости, получаемые статистическими методами. Другим примером эконометрического подхода является исследование производственных функций, начатое классической работой Кобба и Дугласа (1928). Эконометрические модели и методы сейчас — это не только мощный инструментарий для получения новых знаний в экономике и управлении, но и широко применяемый аппарат для принятия практических решений в прогнозировании, банковском деле, бизнесе. Эконометрика изучается на экономических отделениях экономических факультетов.

Исследование операций — наука об обосновании разумных решений в управлении экономическими объектами, использующая математический аппарат. Она представляет собой применение научного метода к проблемам, возникающим в управлении большими системами людей, машин, материалов и денег в промышленности, деловой сфере, государственном управлении, обороне и т.д.

Основная задача исследования операций состоит в том, чтобы помочь менеджеру, принимающему решение, научно определить свою политику и действия среди возможных путей достижения поставленных целей. Коротко исследование операций можно назвать научным подходом к проблеме принятия решений. Под проблемой будем понимать разрыв между желаемым и фактически наблюдаемым состояниями (прежде всего целями) той или иной системы. Решением проблемы будем считать средство преодоления такого рода разрыва, выбор одного из многих объективно существующих курсов действий, который позволил бы перейти от наблюдаемого состояния к желаемому. Исследование операций — важная дисциплина для будущих менеджеров, позволяющая на приемлемом математическом уровне самостоятельно построить экономико-математическую модель или разобраться в моделях, предложенных математиками и экономистами, для осознанного принятия решений по управлению экономическим объектом.

Метод формализованного описания объектов заключается в конструировании соответствующих экономико-математических моделей. *Экономико-математическая модель* — это выраженная в формально-математических терминах экономическая абстракция, логическая структура которой определяется как объективными свойствами предмета описания, так и субъективным целевым фактором исследования, для которого это описание предпринимается.

Основоположник советской школы математической экономики академик В.С. Немчинов писал: «Отображая объективную действительность, модель ее упрощает, отбрасывая все второстепенное и побочное. Однако это упрощение не может быть произвольным и грубым. Адекватность реальной действительности — главное требование, предъявляемое к модели. Исследователь может прибегнуть к методу построения модели, идя от наблюдения, от практики к теории, а также идя обратно — от абстрактных теоретических соображений к конкретной реальной действительности. В процессе научного исследования модель может работать также в двух направлениях: от наблюдения реального мира к теории и обратно. Таким образом, построение модели, с одной стороны,— важная ступень к созданию теории, а с другой — одно из средств экспериментального исследования»[1].

В данном разделе книге мы сделаем упор на изучении дисциплины «исследование операций», рассмотрев наиболее важные ее разделы и ограничившись объемом материала, который мог быть прочитан в течение одного семестра.

---

[1] См.: *Немчинов В.С.* Избранные произведения. М.: Наука, 1969. С. 160.

Глава 29
ТЕОРИЯ АНАЛИЗА ЛИНЕЙНЫХ МОДЕЛЕЙ (ИССЛЕДОВАНИЕ ОПЕРАЦИЙ)

Определение политики и действия среди возможных путей достижения поставленных целей есть оптимизационная задача. Методы решения оптимизационных задач называют методами математического программирования. Термин «программирование» указывает на некоторую программу действий, которая вырабатывается в процессе создания экономико-математической модели и последующего решения задачи. В математическом программировании выделяют линейное программирование, целочисленное программирование, выпуклое программирование. Оптимизационная задача, в которой целевая функция является линейной функцией, а в модели присутствуют только линейные уравнения и линейные неравенства, называется задачей *линейного программирования*. Как уже указывалось, линейное программирование возникло в связи с необходимостью нахождения наилучшего варианта решения планово- производственных задач. Основным математическим аппаратом является линейная алгебра. Из огромного многообразия практических задач выделяются наиболее распространенные типы: оптимизация плана производства, оптимальное смешение, оптимальный раскрой, транспортная задача.

29.1. ОБЩАЯ ФОРМУЛИРОВКА ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Математически каждая задача сводится к поиску значений конечного числа неизвестных при условии, что значения этих неизвестных неотрицательны и удовлетворяют соответствующей системе линейных неравенств (уравнений). При этих значениях неизвестных целевая функция задачи (линейная функция) достигает своего максимума или минимума. Общая математическая формулировка задачи линейного программирования может быть представлена в виде

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le (\ge) b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le (\ge) b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le (\ge) b_k, \\ x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0. \end{cases} \tag{29.1}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max/\min, \qquad (29.2)$$

где (29.1) — система ограничений задачи; (29.2) — целевая функция, характеризующая качество решения задачи.

Если система ограничений (29.1) совместна, то существует по крайней мере одна точка, координаты которой удовлетворяют всем ограничениям. Совокупность точек, удовлетворяющих системе неравенств задачи с двумя переменными, есть выпуклый многоугольник, а в случае большего числа переменных — выпуклый многогранник.

Множество точек $X(x_1, x_2, ..., x_n)$, удовлетворяющих системе (29.1), называется *областью допустимых решений* задачи линейного программирования.

Допустимым решением называется любая точка $X(x_1, x_2, ..., x_n)$, удовлетворяющая системе ограничений, т.е. входящая в область допустимых решений.

Оптимальным решением задачи называется такое допустимое решение (не обязательно единственное), при котором целевая функция (29.2) достигает максимума (минимума).

Любая математическая модель задачи линейного программирования может быть сведена к каноническому виду, т.е. система ограничений представлена в виде системы линейных уравнений. Переход от общего вида (29.1), (29.2) к каноническому очевиден.

Пусть имеется система неравенств

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + ... + a_{kn}x_n \le b_k, \\ x_1 \ge 0, x_2 \ge 0, ..., x_n \ge 0. \end{cases}$$

В каждую из строк введем дополнительную переменную $x_{n+1} \ge 0$, ..., $x_{n+k} \ge 0$. Задача принимает канонический вид:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n + x_{n+2} = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + ... + a_{kn}x_n + x_{n+k} = b_k, \\ x_1 \ge 0, x_2 \ge 0, ..., x_{n+k} \ge 0, \end{cases}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max(\min).$$

Заметим, что $F(\bar{X})\max = -F(\bar{X})\min$, что означает возможность свести любую задачу к задаче на максимизацию целевой функции или наоборот.

З а м е ч а н и е. Вводимые дополнительные переменные имеют экономический смысл. Например, при решении задачи производственного планирования введенные дополнительные переменные можно толковать как остатки соответствующих ресурсов, не использованные в процессе производства.

29.2. ГРАФИЧЕСКИЙ МЕТОД РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Простейшие задачи линейного программирования можно проиллюстрировать графически. Рассмотрим задачу с двумя переменными:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \le b_1, \\ a_{21}x_1 + a_{22}x_2 \le b_2, \\ \dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 \le b_k, \\ x_1 \ge 0, x_2 \ge 0, \end{cases} \tag{29.3}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 \to \max. \tag{29.4}$$

Каждое из неравенств системы (29.3) определяет полуплоскость. Общая часть всех полуплоскостей расположена в координатной плоскости Ox_1x_2 и является областью допустимых решений задачи. Она представляет собой выпуклый многоугольник (рис. 29.1). Це-

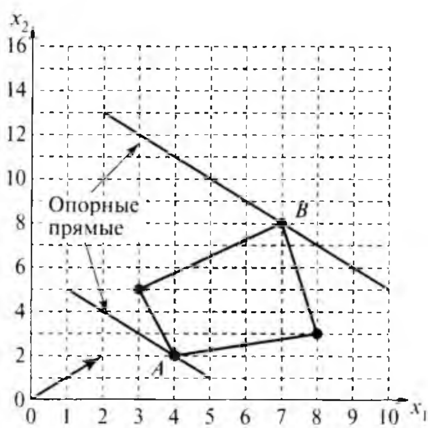

Рис. 29.1

левая функция (29.4) может быть представлена в виде линий уровня $c_1x_1 + c_2x_2 = \text{const}$. Вектор-градиент целевой функции $\bar{n} = (c_1, c_2)$ показывает направление возрастания значения целевой функции.

Среди всех параллельных прямых $F(\bar{X}) = c_1x_1 + c_2x_2$ существуют две таких прямые, которым принадлежат некоторые вершины многоугольника области допустимых решений системы (29.3), но не принадлежат внутренние точки области. Эти прямые называются опорными. Они расположены на разных расстояниях от начала координат. Если каждая опорная прямая проходит лишь через одну вершину области решений, то координаты этой вершины являются решением задачи. При этом для опорной прямой, расположенной ближе к началу координат, имеем минимальное значение целевой функции (точка A на рис. 29.1), а для опорной прямой, расположенной дальше от начала координат — максимум целевой функции (точка B на рис. 29.1).

Если опорная прямая содержит две вершины многоугольника области решений, то она содержит и все точки отрезка, соединяющего эти вершины. В этом случае задача имеет бесконечное множество решений (рис. 29.2).

Рис. 29.2

Если область решений не ограничена и линия уровней (целевая функция) стремится в бесконечность — задача не имеет решений (рис. 29.3).

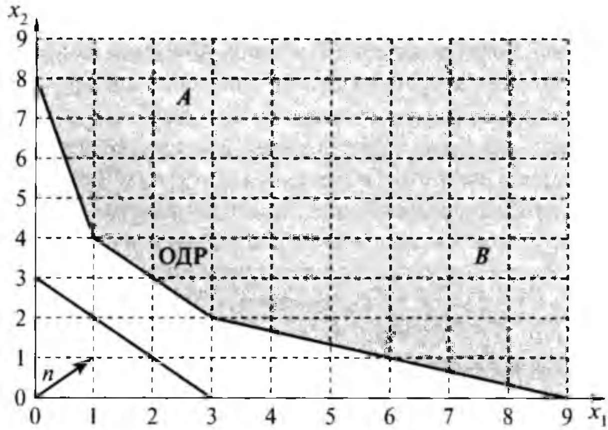

Рис. 29.3

Графически могут решаться задачи:

- заданные в общем виде, содержащие не более двух переменных;
- заданные в канонической форме, содержащие n переменных. При этом если ранг данной системы ограничений равен r, то должно выполняться условие $n - r \leq 2$;
- заданные в общем виде, которые после приведения к каноническому виду будут содержать не более двух свободных переменных ($n - r \leq 2$).

Основные этапы графического метода решения задач линейного программирования

1. Построить область допустимых решений.

2. Построить одну из линий уровня и вектор-градиент целевой функции.

3. Переместить линию уровня до положения одной из опорных прямых. В случае нахождения максимума целевой функции — перемещать линию уровня в направлении, указанном вектором-градиентом. Для нахождения минимального значения целевой функции — в направлении, противоположном указанному вектором-градиентом.

Пример 29.1. Решить задачу производственного планирования.

Допустим, что на некотором малом предприятии планируется выпуск двух видов изделий. На производство одного изделия первого вида затрачивается 2 у.е. производственных мощностей и 6 у.е. сырья. На производство одного изделия второго вида затрачивается 1 у.е. производственных мощностей и 7 у.е. сырья. Расход производственных мощностей не может превысить 20 у.е., а запасы сырья составляют 84 у.е. Известно, что при реализации изделий за одно изделие первого вида предприятие получает 6 у.д.е., а за одно изделие второго вида — 5 у.д.е.

Следует составить такой план выпуска продукции, который бы обеспечил предприятию наибольшую прибыль.

Решение. Обозначив искомые объемы выпуска продукции через x_1, x_2 соответственно, составим систему ограничений задачи, характеризующую данное производство, и целевую функцию:

$$\begin{cases} 2x_1 + x_2 \le 20, \\ 6x_1 + 7x_2 \le 84, \\ x_1 \ge 0; \quad x_2 \ge 0, \end{cases} \tag{29.5}$$

$$F(x) = 6x_1 + 5x_2 \to \max. \tag{29.6}$$

Построим область допустимых решений (29.4), линию уровней (рис. 29.4) (например, $f = 6x_1 + 5x_2 = 30$), вектор-градиент $\bar{n} = (6, 5)$.

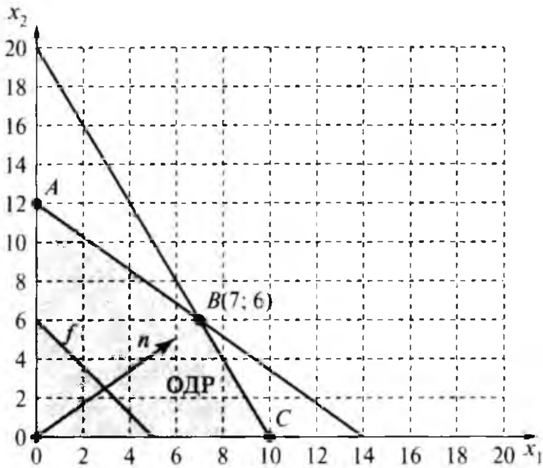

Рис. 29.4

Область допустимых решений — многоугольник $OABC$. Сдвигая линию уровней f параллельно самой себе в направлении вектора-градиента $\bar{n}$, достигаем положения опорной прямой, проходящей через вершину $B(7, 6)$ многоугольника (рис. 29.5). Следовательно, на области допустимых решений целевая функция достигает максимального значения в вершине $B(7, 6)$ и при этом $F_{\max}(\bar{X}) = 72$.

Получено решение задачи: при плане выпуска изделий первого вида в количестве 7 шт. и изделий второго вида в количестве 6 шт. предприятие получает максимальную прибыль в размере 72 у.д.е.

Ответ: $F_{\max}(\bar{X}) = 72$; $\bar{X}_{\text{opt}} = (7, 6)$.

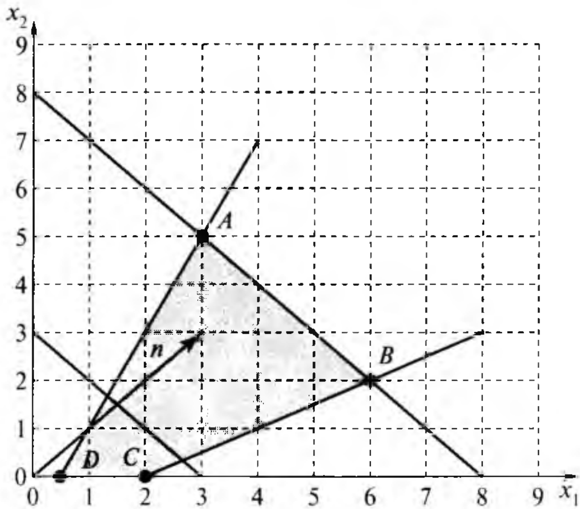

Рис. 29.5

Пример 29.2. Решить графически следующую задачу линейного программирования:

$$\begin{cases} x_1 + x_2 \le 8, \\ 2x_1 - x_2 \ge 1, \\ x_1 - 2x_2 \le 2, \\ x_1 \ge 0, x_2 \ge 0, \end{cases}$$

$$F(\bar{X}) = 3x_1 + 3x_2 \to \max.$$

Решение. Строим область допустимых решений, линию уровней (например, $f = 3x_1 + 3x_2 = 9$), вектор-градиент $\bar{n} = (3, 3)$ (рис. 29.5). Перемещая линию уровня в направлении вектора-градиента, убеждаемся в том, что она совпадает с прямой AB: $x_1 + x_2 = 8$. Решением задачи будут все точки отрезка AB — бесконечное множество решений: $\bar{X} = \alpha A + (1 - \alpha)B$, где точки A и B заданы своими координатами, а $\alpha \in [0; 1]$.

Ответ: $F(\bar{X})\max = 24$; $\bar{X} = (6 - 3\alpha; 2 + 3\alpha)$ при $\alpha \in [0; 1]$.

Пример 29.3. Решить графически следующую задачу линейного программирования:

$$\begin{cases} 3x_1 + 2x_2 \ge 11, \\ -2x_1 + x_2 \le 2, \\ x_1 - 3x_2 \le 0, \\ x_1 \ge 0, x_2 \ge 0, \end{cases}$$

$$F(\bar{X}) = 2x_1 + 4x_2 \to \max.$$

Строим область допустимых решений, линию уровней (например, $f = 2x_1 + 4x_2 = 4$), вектор-градиент $\bar{n} = (2, 4)$. При параллельном перемещении линии уровня в направлении вектора-градиента убеждаемся в том, что такое перемещение не ограничено, так как область допустимых решений открыта (рис. 29.6). Задача не имеет максимума целевой функции.

Ответ: $F(\bar{X})\max \to \infty$.

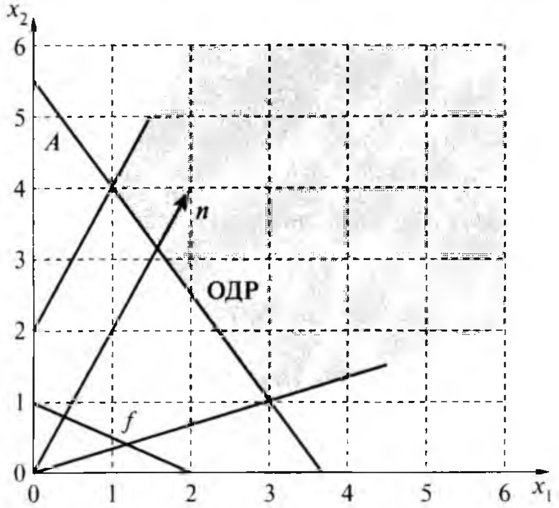

Рис. 29.6

Заметим, однако, что если при заданных выше условиях решать задачу на нахождение минимального значения функции $F(\bar{X}) = 2x_1 + 4x_2 \to \min$, то, перемещая линию уровней в направлении, противоположном вектору-градиенту, получаем единственное решение в точке $D(3, 1)$. Решение задачи: $F(\bar{X})\min = 10$, оптимальный план $\bar{X}_{\text{opt}} = (3, 1)$.

Пример 29.4. Решить графически задачу линейного программирования, заданную в каноническом виде:

$$\begin{cases} 4x_1 - 3x_2 - x_3 + x_4 + x_5 = 6, \\ x_1 + 4x_2 + x_3 + x_5 = 15, \\ 2x_1 - 4x_2 - x_3 + x_4 = 6, \\ x_i \geq 0, \quad i = (1, \ldots, 5), \end{cases} \tag{29.7}$$

$$F(x) = 2x_1 - x_2 \to \max. \tag{29.8}$$

Решение. Решая систему уравнений (29.7) методом Гаусса—Жордана, приведем ее к виду

$$\begin{cases} 2x_1 + x_2 + x_5 = 9, \\ -x_1 + 3x_2 + x_3 = 6, \\ x_1 - x_2 + x_4 = 3, \\ x_i \geq 0, \quad i = (1, \ldots, 5). \end{cases} \tag{29.9}$$

Очевидно, что ранг системы (29.9) равен трем, и следовательно, выполняется условие $n - r = 2$. Выразим базисные переменные через свободные:

$$\begin{cases} x_5 = 9 - 2x_1 - x_2, \\ x_3 = 6 + x_1 - 3x_2, \\ x_4 = 3 - x_1 + x_2. \end{cases} \tag{29.10}$$

Так как значения всех переменных из условия (29.7) не меньше нуля, получаем систему

$$\begin{cases} 9 - 2x_1 - x_2 \geq 0, \\ 6 + x_1 - 3x_2 \geq 0, \\ 3 - x_1 + x_2 \geq 0, \\ x_i \geq 0, \quad i = (1, 2). \end{cases} \tag{29.11}$$

Получаем следующий вид задачи:

$$\begin{cases} 2x_1 + x_2 \leq 9, \\ -x_1 + 3x_2 \leq 6, \\ x_1 - x_2 \leq 3, \\ x_i \geq 0, \quad i = (1, 2), \end{cases} \tag{29.12}$$

$$F(x) = 2x_1 - x_2 \to \max.$$

Изобразим на координатной плоскости Ox_1x_2 область допустимых решений, вектор-градиент $\overline{n} = (2, -1)$ и линию уровней, например f $3x_1 + 3x_2 = 0$ (рис. 29.7).

Область допустимых решений — многоугольник $OABCD$. Передвигая линию уровней в направлении вектора градиента $\overline{n} = (2, -1)$, достигаем максимума целевой функции в вершине C (4, 1). Получим значения переменных:

$x_1 = 4$; $x_2 = 1$; $x_3 = 7$; $x_4 = 0$; $x_5 = 0$.

Ответ: $F(\overline{X})\max = 7$; $\overline{x_{\text{opt}}} = (4, 1, 7, 0, 0)$.

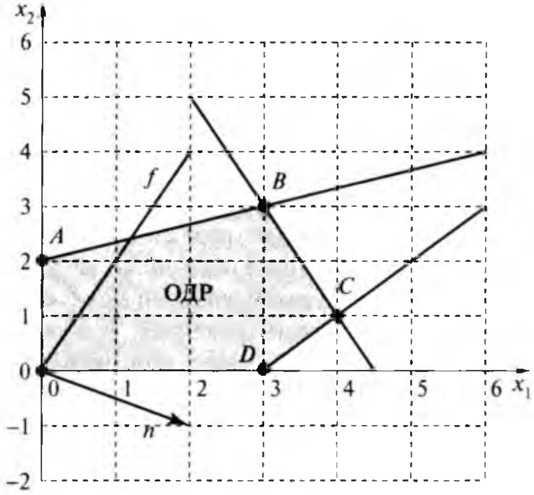

Рис. 29.7

29.3. СИМПЛЕКСНЫЙ МЕТОД РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Симплексный метод был предложен американским математиком Дж. Данцигом в 1951 г. и является общим методом решения задач линейного программирования.

Математическая модель задачи записывается в каноническом виде:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n = b_k, \\ x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0, \end{cases} \tag{29.13}$$

$$F(\overline{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\min). \tag{29.14}$$

С геометрической тоски зрения симплексный метод представляет собой направленный перебор вершин выпуклого многогранника, переходя по ребру от одной вершины к соседней. За конечное число шагов достигается вершина, в которой целевая функция (29.14) получает экстремальное значение. Если такой вершины не найдено, то либо задача не имеет решений (несовместна), т.е. множество допустимых решений пусто, либо задача имеет решения, но $F(\overline{X}) \to \infty$, что означает — область допустимых решений не ограничена. Следует заметить, что в случае, когда область допустимых решений является не плоским многогранным множеством, из каждой вершины выходит более двух ребер и перебор всех вершин — сложная комбинаторная задача. Поэтому переход к новой вершине осуществляется только в том случае, когда целевая функция меняет свое значение в нужном направлении, т.е. если задача решается на максимум целевой функции, то переход осуществляется только к той из соседних вершин, в которой целевая функция увеличивает свое значение.

Введем некоторые необходимые понятия. Решив систему уравнений (29.13) методом Гаусса—Жордана, получим допустимое решение, записанное в общем виде, где базисные переменные выражены через свободные.

Базисным решением задачи называется частное решение, полученное путем задания свободным переменным нулевого значения. Базисное решение представляет собой координаты одной из вершин многогранника области допустимых решений. Для задач линейного программирования прикладного характера базисное решение называют *опорным планом задачи*.

Критерий оптимальности решения задачи линейного программирования при отыскании максимума (минимума)

Если на выбранном допустимом решении задачи в выражении целевой функции отсутствуют отрицательные (положительные) коэффициенты при свободных переменных, то решение оптимально, т.е. $F = F_{\max(\min)}$. Если критерий оптимальности не выполнен, то следует перейти к новому базисному решению.

Процедуру симплексного метода представим в виде последовательности следующих шагов.

Шаг 1. Задача приводится к каноническому виду.

Шаг 2. Находится первый опорный план.

Шаг 3. Оценивается целевая функция. Если согласно критерию оптимальности достигнут оптимум, то решение задачи прекращается и полученный опорный план является оптимальным. Если оптимум не достигнут, то необходим переход к новому опорному плану.

Шаг 4. Переход к новому опорному плану осуществляется введением в базис одной из свободных переменных ранее найденного допустимого решения. Целевую функцию выражаем через новый набор свободных переменных и проверяем выполнение критерия оптимальности (переход к шагу 3). Процесс повторения шагов 3 и 4 заканчивается отысканием оптимального решения задачи, либо устанавливается, что задача неограниченна (F), либо задача не имеет решений.

Рассмотрим реализацию идей симплексного метода на примере модели (29.3).

Исходная математическая модель распределения ресурсов имела вид:

$$\begin{cases} 2x_1 + x_2 \le 20, \\ 6x_1 + 7x_2 \le 84, \\ x_1 \ge 0; \quad x_2 \ge 0, \end{cases}$$

$$F(x) = 6x_1 + 5x_2 \to \max.$$

1. Приведем задачу к каноническому виду, введя дополнительные переменные $x_3 \ge 0$; $x_4 \ge 0$:

$$\begin{cases} 2x_1 + x_2 + x_3 = 20, \\ 6x_1 + 7x_2 + x_4 = 84, \\ x_1 \ge 0; \quad x_2 \ge 0; \quad x_3 \ge 0; \quad x_4 \ge 0, \end{cases}$$

$$F(x) = 6x_1 + 5x_2 \to \max.$$

Запишем первое базисное решение данной системы уравнений:

$$x_3 = 20 - 2x_1 - x_2,$$

$$x_4 = 84 - 6x_1 - 7x_2.$$

Давая свободным переменным нулевые значения, получим первое базисное решение или первый опорный план задачи $\overline{X}_1 = (0,\ 0,\ 20,\ 84)$. При этом решении значение целевой функции $F_1 = 0$. (На рис. 29.5 этому решению соответствует вершина O многоугольника $OABC$.) Полученное решение означает, что объемы выпуска продукции x_1 и x_2 равны нулю, а значения дополнительных переменных x_3 и x_4 соответствуют объемам запасов сырья, и так как производство ничего не выпускает, то прибыль $F = 0$.

2. Оценим вид целевой функции, используя критерий оптимальности задачи линейного программирования: $F_1(x) = 6x_1 + 5x_2 + 0x_3 + 0x_4 = 0$. Очевидно, что если ввести в базис переменные x_1 или x_2, то значение целевой функции возрастет.

3. Введем в базис, например, переменную x_1. С точки зрения роста значения целевой функции чем больше значение переменной x_1, тем больше значение целевой функции. Однако следует учитывать ограничения на объем выпуска первого продукта, накладываемые имеющимися объемами запасов сырья. Запасы первого

вида сырья (20 ед.) позволяют выпустить не более 10 ед. изделий первого вида ($2x_1 + x_2 + x_3 = 20$, следовательно, $x_{1\max} = \frac{20}{2} = 10$).

Запасы второго вида сырья (84 ед.) позволяют выпустить не более 14 единиц изделий первого вида ($6x_1 + 7x_2 + x_4 = 84$, следовательно $x_{1\max} = \frac{84}{6} = 14$). Следовательно, реально можно выпустить максимально 10 ед. изделий первого вида. Это означает, что переменная x_1 вводится в базис из первого уравнения исходной системы и заменяет собой в базисе переменную x_3. Теперь имеем выражение новых базисных переменных:

$$\begin{cases} x_1 = 10 - \frac{1}{2}x_2 - \frac{1}{2}x_3, \\ x_4 = 24 - 4x_2 + 3x_3. \end{cases}$$

Новый опорный план $\bar{X}_2 = (10, 0, 0, 24)$, который соответствует вершине C многоугольника $OABC$, приведен на рис. 29.5.

Теперь система уравнений с учетом новых базисных переменных имеет вид

$$\begin{cases} x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 = 10, \\ 4x_2 - 3x_3 + x_4 = 24, \\ x_1 \ge 0; \quad x_2 = 0; \quad x_3 = 0; \quad x_4 \ge 0. \end{cases}$$

Целевая функция при новых свободных переменных принимает вид $F_2(x) = 60 + 2x_2 - 3x_3$, $F_2 = 60$.

4. Оценим вновь вид целевой функции. Очевидно, что если значение переменной x_2 будет отлично от нуля, то значение целевой функции возрастет.

5. Введем в базис переменную x_2, оценив, как в п. 3, какую из прежних базисных переменных она может заменить.

Из уравнения $x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 = 10$ ясно, что $x_{2\max} = 20$, из второго уравнения $4x_2 - 3x_3 + x_4 = 24$ ясно, что $x_{2\max} = 6$. Следовательно, реально с использованием двух видов сырья можно выпустить продукции второго вида не более 6 ед. Выражаем переменную x_2 из второго уравнения, где ограничение на объем выпуска более жесткое, и получаем следующие выражения для новых базисных переменных:

$$x_1 = 7 + \frac{7}{8}x_3 - \frac{1}{8}x_4;$$

$$x_2 = 6 + \frac{3}{4}x_3 - \frac{1}{4}x_4.$$

Теперь система уравнений и целевая функция имеют вид

$$\begin{cases} x_1 + \frac{7}{8}x_3 - \frac{1}{8}x_4 = 7, \\ x_2 - \frac{3}{4}x_3 + \frac{1}{4}x_4 = 6, \\ x_1 \ge 0; \quad x_2 \ge 0; \quad x_3 \ge 0; \quad x_4 \ge 0, \end{cases}$$

$$F(x) = 72 - \frac{3}{2}x_3 - \frac{1}{2}x_4 \to \max.$$

6. Применив критерий оптимальности, убеждаемся, что целевая функция достигла своего оптимального значения $F_2 = F_{\max} = 72$ и оптимальный план выпуска $\bar{X}_{opt}(7, 6, 0, 0)$, что соответствует вершине B многоугольника $OABC$ на рис. 29.4.

29.4. ИСПОЛЬЗОВАНИЕ СИМПЛЕКСНЫХ ТАБЛИЦ

Практически все расчеты при решении задач линейного программирования удобно проводить в виде так называемых симплексных таблиц. В этом случае задача линейного программирования должна быть представлена в каноническом виде, целевая функция записывается в виде уравнения $F(x) - c_1x_1 - c_2x_2 - \ldots - c_nx_n = 0$, найдено первое базисное решение.

Пусть дана задача в виде

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\ x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0, \end{cases} \tag{29.15}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max. \tag{29.16}$$

1. Запишем задачу в каноническом виде, введя дополнительные переменные, и запишем целевую функцию в виде «уравнения»:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \ldots + x_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + \ldots + x_{n+k} = b_k, \\ x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0, x_{n+1} \ge 0, \ldots, x_{n+k} \ge 0, \end{cases} \tag{29.17}$$

$$F(\bar{X}) - c_1x_1 - c_2x_2 - \ldots - c_nx_n = 0(\max). \qquad (29.18)$$

2. Составим первую симплексную таблицу (табл. 29.1). (Заметим, что в разной литературе приводятся различные типы таблиц.)

Таблица 29.1

Базисные переменные	X_1	X_2	X_3	...	X_n	X_{n+1}	X_{n+2}	...	X_{n+k}	b_i	Оценки b_i/a_{ij}
X_{n+1}	a_{11}	a_{12}	a_{13}	...	a_{1n}	1	0	...	0	b_1	b_1/a_{12}
X_{n+2}	a_{21}	a_{22}	a_{23}	...	a_{2n}	0	1	...	0	b_2	b_2/a_{22}
...	...	...	...	...	...	...	...	...	...	...	
X_{n+k}	a_{k1}	a_{k2}	a_{k3}	...	a_{kn}	...	...	...	...	b_k	b_k/a_{k2}
Целевая функция	$-c_1$	$-c_2$	$-c_3$	...	$-c_n$	0	0	...	0	0	0

Очевидно, что $x_{n+1}, \ldots, x_{n+k}$ — базисные переменные. Первый опорный план $\bar{X}_1 = (0, 0, \ldots, 0, b_1, b_2, \ldots, b_k)$, а значение целевой функции на этом плане равно нулю (данные предпоследнего столбца таблицы).

3. Оценим строку целевой функции (последняя строка табл. 29.1).

Если при решении задачи на максимум в последней строке нет отрицательных чисел, то полученный опорный план оптимален. Для задач, решаемых на минимум, в последней строке при достижении минимума не должно быть положительных чисел. (Далее будем говорить о задаче, решаемой на максимум.) Если в последней строке таблицы есть несколько отрицательных чисел, то можно выбрать любое из них. Столбец таблицы, в котором стоит выбранное число, называется *ключевым столбцом*. Пусть для примера $-c_2 = 0$. Ключевым столбцом в табл. 29.1 является столбец, содержащий коэффициенты при переменной x_2, т.е. третий столбец. Вводим в базис переменную x_2, а следовательно, переменная x_2 заменит в базисе одну из прежних базисных переменных.

4. Строка табл. 29.1, из которой x_2 будет введена в базис, называется *ключевой строкой*. Для выбора ключевой строки заполняем последний столбец таблицы (оценка), учитывая что для примера мы выбрали в качестве ключевого столбца столбец, соответствующий переменной x_2. Ключевая строка определяется как строка табл. 29.1, содержащая наименьшее положительное частное $\frac{b_i}{a_{i2}}$.

Допустим, что отношение $\frac{b_2}{a_{22}}$ является наименьшим положительным числом в оценочном (последнем) столбце табл. 29.1. Это означает, что ключевой строкой становится вторая значимая строка таблицы, а переменная x_2 заменяет в базисе переменную x_{n+2}. Переходим ко табл. 29.2.

5. Для заполнения второй таблицы делаются следующие преобразования данных табл. 29.1:

а) все элементы ключевой строки делятся на величину a_{22}. Таким образом, на месте значения a_{22} в табл. 29.1 получаем $a'_{22} = 1$;

б) выбираем $a'_{22} = 1$ за ведущий элемент и, преобразовывая строки табл. 29.1 методом Гаусса—Жордана, получаем новую симплексную таблицу (табл. 29.2).

Таблица 29.2

Базисные переменные	X_1	X_2	X_3	...	X_n	X_{n+1}	X_{n+2}	...	X_{n+k}	b_i	Оценки b_i/a_{ij}
X_{n+1}	a'_{11}	0	a'_{13}	...	a'_{1n}	1	$a'_{1,n+2}$	...	0	b'_1	–
X_2	a'_{21}	1	a'_{23}	...	a'_{2n}	0	$a'_{2,n+2}$	...	0	b'_2	–
...	...	...	...	...	...	...	...	...	...	...	–
X_{n+k}	a'_{k1}	0	a'_{k3}	...	a'_{kn}	...	$a'_{k,n+2}$	...	1	b'_k	–
Целевая функция	c'_1	0	c'_3	...	c'_n	0	c'_{n+2}	...	0	F_2	–

Второй опорный план $\bar{X}_2 = (0, b'_2, 0, 0, ..., b'_1, 0, b'_3, ..., b'_k)$, а значение целевой функции на этом плане равно некоторой величине F_2 (данные предпоследнего столбца табл. 29.2).

Анализируем данные строки целевой функции. Если все числа c'_j неотрицательны (неположительны), то найденный опорный план оптимален и $F_2 = F_{\max}(F_{\min})$. Если это не так — продолжаем решать задачу, составляя очередную симплексную таблицу.

Формальное правило составления симплексных таблиц (кроме первой)

1. В последней строке очередной симплексной таблицы выбираем наибольшее по модулю отрицательное число, если задача решается на максимум целевой функции, или наибольшее положительное число, если задача решается на мининум.

2. Выбранный элемент определяет ключевой столбец таблицы (т.е. определяет переменную, вводимую в новый базис).

3. Разделим элементы столбца свободных членов (b_i) на соответствующие положительные элементы ключевого столбца.

4. Из полученных частных выберем наименьшее положительное число ($\min(b_i/a_{ij}) = 0$).

5. Строка таблицы, содержащая выбранное наименьшее частное, является ключевой строкой, а число, стоящее на пересечении ключевого столбца и ключевой строки, является главным элементом.

6. Разделим каждый элемент ключевой строки на главный элемент и, преобразовывая таблицу методом Гаусса—Жордана, введем главный элемент в базис. Получаем новую симплексную таблицу и возвращаемся к п. 1.

29. Повторяем операции 1–6 до тех пор, пока в последней строке очередной таблицы не будет выполнен критерий оптимальности.

Пример 29.5. Рассмотрим уже известную нам задачу распределения ресурсов, математическая модель которой имела следующий вид:

$$\begin{cases} 2x_1 + x_2 \le 20, \\ 6x_1 + 7x_2 \le 84, \\ x_1 \ge 0; x_2 \ge 0, \end{cases}$$

$$F(x) = 6x_1 + 5x_2 \to \max.$$

Приведем задачу к каноническому виду, а целевую функцию запишем согласно требованию для симплексных таблиц:

$$\begin{cases} 2x_1 + x_2 + x_3 = 20, \\ 6x_1 + 7x_2 + x_4 = 84, \\ x_1 \ge 0; \quad x_2 \ge 0; \quad x_3 \ge 0; \quad x_4 \ge 0, \end{cases}$$

$$F(x) - 6x_1 - 5x_2 = 0 \to \max.$$

Составим первую симплексную табл. 29.3.

Таблица 29.3

	Базисные переменные	X_1	X_2	X_3	X_4	b_i	Оценки b_i/a_{ij}
⇐	X_3	2	1	1	0	20	10
	X_4	6	7	0	1	84	14
	Целевая функция	–6	–5	0	0	0	

Первый опорный план задачи $\overline{X}_1 = (0, 0, 20, 84)$, и значение целевой функции $F_1 = 0$. Анализ строки целевой функции показывает, что в качестве ключевого столбца можно выбрать либо столбец, соответствующий переменной x_1, либо столбец, соответствующий переменной x_2. Выберем столбец, соответствующий переменной x_1.

Заполним оценочный (последний) столбец таблицы. Так как наименьшая положительная оценка принадлежит первой содержательной строке таблицы, то эта строка становится ключевой и переменная x_1 вводится в базис вместо переменной x_3. Преобразуем методом Гаусса—Жордана табл. 29.3, вводя в базис новую переменную. Получим новую табл. 29.4.

Таблица 29.4

	Базисные переменные	X_1	X_2	X_3	X_4	b_i	Оценки b_i/a_{ij}
	X_1	1	1/2	1/2	0	10	20
⇐	X_4	0	4	–3	1	24	6
	Целевая функция	0	–2	3	0	60	

Новый опорный план $\overline{X}_2 = (10, 0, 0, 24)$, и $F_2 = 60$. Так как в строке целевой функции есть еще отрицательное число, то согласно уже изложенному правилу выбираем соответствующий ключевой столбец и ключевую строку (в таблице они обозначены стрелками) и, преобразуя таблицу методом Гаусса—Жордана, получаем табл. 29.5.

Таблица 29.5

Базисные переменные	X_1	X_2	X_3	X_4	b_i	Оценки b_i/a_{ij}
X_1	1	0	7/8	–1/8	7	
X_2	0	1	–3/4	1/4	6	
Целевая функция	0	0	3/2	1/2	72	

Оценивая строку целевой функции, видим, что все коэффициенты при свободных переменных положительны. Это означает, что достигнут максимум целевой функции.

Ответ: $F_{\max} = 72$, и оптимальный план $\overline{X}_{\text{opt}} = (7, 6, 0, 0)$.

29.5. МЕТОД ИСКУССТВЕННОГО БАЗИСА

В рассмотренных выше случаях задача линейного программирования приводилась к каноническому виду путем введения неотрицательных дополнительных переменных, которые и играли роль базисных переменных при составлении первого опорного плана.

Однако система ограничений задач может быть изначально задана в виде уравнений или дополнительные переменные вводятся с отрицательным знаком. В этих случаях для нахождения первого опорного плана используется метод искусственного базиса.

Пусть задача линейного программирования задана в виде

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n = b_k, \\ x_j \geq 0,\ j = (1, 2, \ldots, n);\quad b_i \geq 0, i = (1, 2, \ldots, k), \end{cases} \tag{29.19}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\min). \tag{29.20}$$

Введем в систему ограничений (29.19) искусственные переменные $\gamma_1 \geq 0, \ldots, \gamma_k \geq 0$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \gamma_1 = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \gamma_2 = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + \gamma_k = b_k, \\ x_j \geq 0,\ j = (1, 2, \ldots, n);\quad b_i \geq 0, \gamma_i \geq 0, i = (1, 2, \ldots, k). \end{cases} \tag{29.21}$$

Очевидно, что системы уравнений (29.19) и (29.21) эквивалентны только при условии равенства нулю искусственных переменных: $\gamma_1 = 0, \ldots, \gamma_k = 0$. Для задачи с ограничениями (29.21) базисными переменными являются искусственные переменные.

Именно для этого они и были введены. Но как перейти от искусственного базиса системы уравнений (29.21) к первому базисному решению системы (29.19)? Воспользуемся симплексным методом и решим задачу с ограничениями (29.21), минимизируя новую целевую функцию $T = \gamma_1 + \gamma_2 + \ldots + \gamma_k \to \min$. Новая задача имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \gamma_1 = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \gamma_2 = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + \gamma_k = b_k, \\ x_j \geq 0,\quad j = (1, 2, \ldots, n);\quad b_i \geq 0,\quad \gamma_i \geq 0,\quad i = (1, 2, \ldots, k), \\ F(\bar{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n, \end{cases} \tag{29.22}$$

$$T = \gamma_1 + \gamma_2 + \ldots + \gamma_k \to \min. \tag{29.23}$$

После конечного числа шагов симплексного метода будет получен минимум функции $T(\gamma)$. Возможны два случая.

1. $T_{\min}(\gamma) \succ 0$. Это означает, что система (29.22) не имеет неотрицательных решений, при которых значения искусственных пере-

менных были бы равны нулю. Следовательно, и исходная задача (29.19), (29.20) не имеет неотрицательных решений.

2. $T_{min}(\gamma) = 0$. Так как $T = \gamma_1 + \gamma_2 + ... + \gamma_k$ при $\gamma_i \geq 0$, $i = (1, 2, ..., k)$, то искусственные переменные стали свободными, а их место в базисе заняли переменные системы (29.19). Это означает, что найдено первое базисное решение исходной задачи.

Рассмотрим реализацию метода искусственного базиса на примере:

$$\begin{cases} 3x_1 + 2x_2 + x_3 + x_4 = 7, \\ 4x_1 + 3x_2 + 2x_3 + x_4 = 10, \\ x_j \geq 0, \quad j = (1, ..., 4), \end{cases} \tag{29.24}$$

$$F(x) = x_1 - x_2 - 2x_3 + 3x_4 \to \min. \tag{29.25}$$

Введем в ограничения (29.24) искусственные переменные $\gamma_1 \geq 0$, $\gamma_2 \geq 0$ и составим вспомогательную целевую функцию $T = \gamma_1 + \gamma_2 \to \min$. Задача примет вид

$$\begin{cases} 3x_1 + 2x_2 + x_3 + x_4 + \gamma_1 = 7, \\ 4x_1 + 3x_2 + 2x_3 + x_4 + \gamma_2 = 10, \\ x_j \geq 0, \quad j = (1, ..., 4), \end{cases}$$

$$F(x) - x_1 + x_2 + 2x_3 - 3x_4 = 0, \tag{29.26}$$

$$T = \gamma_1 + \gamma_2 \to \min.$$

Так как γ_1 и γ_2 — базисные переменные, то выразим их через свободные переменные:

$$\gamma_1 = 7 - (3x_1 + 2x_2 + x_3 + x_4),$$

$$\gamma_2 = 10 - (4x_1 + 3x_2 + 2x_3 + x_4)$$

и представим соответственно новую целевую функцию

$$T = \gamma_1 + \gamma_2 = 17 - (7x_1 + 5x_2 + 3x_3 + 2x_4) \to \min.$$

Составим первую симплексную табл. 29.6.

Так как целевая функция минимизируется, то выбираем в последней строке наибольшее положительное число, определив таким образом ключевой столбец и, действуя далее согласно правилам симплексного метода, получаем вторую симплексную таблицу 29.7.

Анализируя содержимое последней строки таблицы 29.7 согласно алгоритму симплексного метода, выбираем новую базисную переменную, например x_3, определив тем самым ключевой столбец. Согласно полученным оценкам b_i/a_{ij} ключевой строкой становится вторая строка. Вводим x_3 в базис и получаем табл. 29.8.

Таблица 29.6

	Базисные переменные	⇑ X_1	X_2	X_3	X_4	γ_1	γ_2	b_i	Оценки b_i/a_{ij}
⇐	γ_1	3	2	1	1	1	0	7	7/3
	γ_2	4	3	2	1	0	1	10	5/2
	F	–1	1	2	–3	0	0	0	–
	Целевая функция T	7	5	3	2	1	1	17	–

Таблица 29.7

	Базисные переменные	X_1	X_2	⇑ X_3	X_4	γ_1	γ_2	b_i	Оценки b_i/a_{ij}
	X_1	1	2/3	1/3	1/3	1/3	0	7/3	7
⇐	γ_2	0	1/3	2/3	–1/3	–4/3	1	2/3	1
	F	0	5/3	7/3	–8/3	1/3	0	7/3	–
	Целевая функция T	0	1/3	2/3	–1/3	–4/3	1	2/3	–

Таблица 29.8

Базисные переменные	X_1	X_2	X_3	X_4	γ_1	γ_2	b_i	Оценки b_i/a_{ij}
X_1	1	1/2	0	1/2	1	–1/2	2	–
X_3	0	1/2	1	–1/2	–2	3/2	1	–
F	0	1/2	0	–3/2	5	–7/2	0	–
Целевая функция T	0	0	0	0	0	0	0	–

В табл. 29.8 значение вспомогательной целевой функции T равно нулю. Следовательно, полученный опорный план $\overline{X} = (2, 0, 1, 0)$ является первым опорным планом исходной задачи (29.24), (29.25). Значение целевой функции $F = 0$. С этого момента продолжаем решение задачи симплексным методом, ориентируясь уже на строку целевой функции F.

29.6. ДВОЙСТВЕННОСТЬ

Теоремы двойственности встречаются в разных разделах математики. Рассмотрим взаимно двойственные задачи в линейном программировании. Приведем упрощенную интерпретацию управления экономическим объектом при построении двойственной пары задач. В качестве примера возьмем задачу использования сырья. Некоторое производство, выпускающее n видов продукции, располагает сырьем: m видов сырья в количестве $b_1, b_2, \ldots, b_m$ соответственно. Технология производства требует затрат i-го вида сырья ($i = 1, 2, \ldots, m$) на единицу продукции j-го вида ($j = 1, 2, \ldots, n$) в количестве a_{ij}. Прибыль от реализации единицы продукции j-го вида равна c_j. План выпуска продукции — вектор $X = (x_1, x_2, \ldots, x_n)$ должен обеспечить максимальную прибыль.

Математическая модель такой задачи имеет вид

$$\text{Задача 1}\quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m, \\ x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0, \end{cases} \tag{29.27}$$

$$F(\bar{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max. \tag{29.28}$$

Предположим, что производитель решил перепродать имеющееся у него сырье другому производителю. Каковы же условия продажи сырья? Производитель, продающий сырье, оценивает единицу сырья каждого вида. Пусть $y_1, y_2, \ldots, y_m$ — оценки единицы сырья.

Ему выгодно продать свое сырье, если суммарная стоимость всех видов сырья, расходуемых на производство единицы каждого вида продукции, не меньше прибыли c_j, получаемой при реализации каждого изделия. В свою очередь, производителю, покупающему сырье, выгодно минимизировать свои затраты. В результате получаем новую математическую модель

$$\text{Задача 2}\quad \begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \ge c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \ge c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \ge c_n, \\ y_j \ge 0, \quad j = (1, 2, \ldots, m), \end{cases} \tag{29.29}$$

$$Z(\bar{Y}) = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \min. \tag{29.30}$$

З а м е ч а н и е. В рассматриваемых моделях знаки неравенств в системах ограничений направлены в одну сторону. Если в исходной формулировке это не так, то всегда можно добиться однонаправленности знаков неравенств, умножив несоответствующее неравенство на -1.

Задача 1 и задача 2 составляют двойственную пару задач линейного программирования.

Сравним математические модели задачи 1 и задачи 2.

Матрица коэффициентов задачи 2 представляет собой транспонированную матрицу коэффициентов задачи 1.

Если знаки неравенств в ограничениях исходной задачи «≤», то в ограничениях двойственной задачи знаки неравенств «≥». При этом в первой задаче функция максимизируется, а во второй задаче — минимизируется.

Свободные члены первой задачи становятся коэффициентами целевой функции второй задачи, а коэффициенты при неизвестных в целевой функции первой задачи — свободными членами ограничений второй задачи.

Приведем возможные пары двойственных задач.

Исходная задача в стандартной форме	Двойственная задача
$\sum_{j=1}^{n} a_{ij}x_j \le b_i,\ i = 1, 2, ..., m,\ x_j \ge 0,$ $j = 1, 2, ..., n,\ F(\bar{X}) = \sum_{j=1}^{n} c_j x_j \to (\max)$	$\sum_{i=1}^{m} a_{ij}y_i \ge c_j,\ j = 1, 2, ..., n,\ y_i \ge 0,$ $i = 1, 2, ..., m,\ Z(\bar{Y}) = \sum_{i=1}^{m} b_i y_i \to (\min)$
$\sum_{j=1}^{n} a_{ij}x_j \ge b_i,\ i = 1, 2, ..., m,\ x_j \ge 0,$ $j = 1, 2, ..., n,\ F(\bar{X}) = \sum_{j=1}^{n} c_j x_j \to (\min)$	$\sum_{i=1}^{m} a_{ij}y_i \le c_j,\ j = 1, 2, ..., n,\ y_i \ge 0,$ $i = 1, 2, ..., m,\ Z(\bar{Y}) = \sum_{i=1}^{m} b_i y_i \to (\max)$
Исходная задача в каноническом виде	**Двойственная задача**
$\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = 1, 2, ..., m,\ x_j \ge 0,$ $j = 1, 2, ..., n,\ F(\bar{X}) = \sum_{j=1}^{n} c_j x_j \to (\max)$	$\sum_{i=1}^{m} a_{ij}y_i \ge c_j,\ j = 1, 2, ..., n,$ y_i — произвольные, $i = 1, 2, ..., m,$ $Z(\bar{Y}) = \sum_{i=1}^{m} b_i y_i \to (\min)$
$\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = 1, 2, ..., m,\ x_j \ge 0.$ $j = 1, 2, ..., n,\ F(\bar{X}) = \sum_{j=1}^{n} c_j x_j \to (\min)$	$\sum_{i=1}^{m} a_{ij}y_i \le c_j,\ j = 1, 2, ..., n,$ y_i — произвольные, $i = 1, 2, ..., m,$ $Z(\bar{Y}) = \sum_{i=1}^{m} b_i y_i \to (\max)$

Рассмотрим несколько примеров составления двойственной задачи к данной.

Пример 29.6. Составить двойственную задачу к данной задаче:

$$\begin{cases} 2x_1 - x_2 + 2x_3 - 3x_4 \le 5, \\ x_1 + 2x_2 - x_3 + x_4 \le 3, \\ x_j \ge 0, \quad j = (1, \ldots, 4), \end{cases}$$

$$F(\bar{X}) = x_1 - 2x_2 + 3x_3 - x_4 \to \max.$$

Решение. Введем переменные двойственной задачи y_1 и y_2. Действуя в соответствии с общими правилами составления двойственной задачи, получаем:

$$\begin{cases} 2y_1 + y_2 \ge 1, \\ -y_1 + 2y_2 \ge -2, \\ 2y_1 - y_2 \ge 3, \\ -3y_1 + y_2 \ge -1, \\ y_i \ge 0 \quad i = 1,2, \end{cases}$$

$$Z(\bar{Y}) = 5y_1 + 3y_2 \to \min.$$

Пример 29.7. Составить двойственную задачу к данной задаче:

$$\begin{cases} x_1 - 2x_2 + x_4 = 8, \\ x_2 + x_3 - 3x_4 = 6, \\ x_j \ge 0, \quad j = (1, \ldots, 4), \end{cases}$$

$$F(\bar{X}) = 3x_2 - x_4 \to \max.$$

Решение. Следуя правилу составления двойственной задачи для случая, когда исходная задана в каноническом виде, получаем:

$$\begin{cases} y_1 \ge 0, \\ -2y_1 + y_2 \ge 3, \\ y_2 \ge 0, \\ y_1 - 3y_2 \ge -1, \end{cases}$$

$$Z(\bar{Y}) = 8y_1 + 6y_2 \to \min.$$

Пример 29.8. Составить двойственную задачу к данной задаче:

$$\begin{cases} 2x_1 - x_2 + x_3 - 3x_4 - x_5 = 10, \\ x_1 + 2x_2 - x_3 + 2x_4 + x_5 \ge 8, \\ 2x_1 - x_2 + 3x_3 - x_4 + 2x_5 \le 4, \\ x_1 \ge 0, x_3 \ge 0, x_4 \ge 0, \end{cases}$$

$$F(\bar{X}) = 2x_1 - x_2 + x_3 - 3x_4 + x_5 \to \min.$$

Прежде всего упорядочим запись задачи, умножив третье ограничение на -1.

Получим:

$$\begin{cases} 2x_1 - x_2 + x_3 - 3x_4 - x_5 = 10, \\ x_1 + 2x_2 - x_3 + 2x_4 + x_5 \ge 8, \\ -2x_1 + x_2 - 3x_3 + x_4 - 2x_5 \ge -4, \\ x_1 \ge 0, x_3 \ge 0, x_4 \ge 0, \end{cases}$$

$$F(\bar{X}) = 2x_1 - x_2 + x_3 - 3x_4 + x_5 \to \min.$$

В соответствие каждой строке ограничений исходной задачи вводим переменные y_1, y_2, y_3. Двойственная задача имеет вид

$$\begin{cases} 2y_1 + y_2 - 3y_3 \le 2, \\ -y_1 + 2y_2 + y_3 = -1, \\ y_1 - y_2 - 3y_3 \le 1, \\ -3y_1 + 2y_2 + y_3 \le -3, \\ -y_1 + y_2 - 2y_3 = 1, \\ y_2 \ge 0, y_3 \ge 0, \end{cases}$$

$$Z(\bar{Y}) = 10y_1 + 8y_2 - 4y_3 \to \max.$$

Теорема 29.1 (первая теорема двойственности). Если исходная задача имеет решение, то и двойственная к ней имеет решение. При этом максимальное значение целевой функции $F(X)$ первой задачи равно минимальному значению целевой функции $Z(Y)$ двойственной задачи: $\max F(\bar{X}) = \min Z(\bar{Y})$. Если область решений исходной задачи не ограничена и $F(\bar{X}) \to \infty$, то двойственная к ней задача несовместна (т.е. ее область решений пуста).

Теорема 29.2. (вторая теорема двойственности). Пусть дана симметричная пара двойственных задач:

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i, \quad i = 1, 2, \ldots, m, \qquad \sum_{i=1}^{m} a_{ij}y_i \ge c_j, \quad j = 1, 2, \ldots, n,$$

$$x_j \ge 0, \quad j = 1, 2, \ldots, n, \qquad y_i \ge 0, \quad i = 1, 2, \ldots, m,$$

$$F(\bar{X}) = \sum_{j=1}^{n} c_j x_j \to (\max), \qquad Z(\bar{Y}) = \sum_{i=1}^{m} b_i y_i \to (\min).$$

Вектор $\bar{X}^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ является допустимым решением исходной задачи, а вектор $\bar{Y}^0 = (y_1^0, y_2^0, \ldots, y_m^0)$ — допустимым решением двойственной задачи.

Для того чтобы $\bar{X}^0$ и $\bar{Y}^0$ являлись оптимальными решениями двойственной пары задач, необходимо и достаточно, чтобы выполнялись следующие равенства:

$$x_j^0\left(\sum_{i=1}^{m}a_{ij}y_i^0-c_j\right)=0,\quad j=1,2,\ldots,n;$$

$$y_i^0\left(\sum_{i=1}^{m}a_{ij}x_j^0-b_i\right)=0,\quad i=1,2,\ldots,m.$$

Эти равенства позволяют найти оптимальное решение одной из задач, если известно оптимальное решение двойственной к ней задачи.

Теорема 29.3 (третья теорема двойственности). Значения переменных оптимального решения двойственной задачи называют объективно обусловленными оценками. Термин «объективно обусловленные оценки» введен одним из основоположников математического программирования академиком Л.В. Канторовичем.

Теорема 29.4. Компоненты оптимального плана двойственной задачи равны значениям частных производных целевой функции:

$$Z(\bar{Y}^0)=\sum_{i=1}^{m}b_iy_i^0\ :\ \frac{\partial Z_{\max}}{\partial b_i}=y_i^0,\quad i=1,2,\ldots,m.$$

Объективно обусловленные оценки определяют, например, степень дефицитности ресурсов исходной задачи. Если в оптимальном плане исходной задачи некоторые ресурсы используются полностью (дефицитны), то объективно обусловленные оценки этих ресурсов ненулевые; и наоборот, если какие-то ресурсы используются не полностью, то соответствующие им объективно обусловленные оценки равны нулю. Кроме того, объективно обусловленные оценки показывают, на сколько денежных единиц изменится максимальная прибыль производителя от реализации продукции при изменении запаса соответствующего ресурса на одну единицу.

Пример 29.9. Дана следующая задача:

$$\begin{cases}-x_1+2x_2\le 6,\\ x_1+x_2\le 9,\\ 3x_1-x_2\le 15,\\ x_1\ge 0, x_2\ge 0,\end{cases}$$

$$F(\bar{X})=4x_1+2x_2\to\max.$$

Найти ее оптимальный план и максимальное значение целевой функции, если известны: оптимальный план двойственной задачи $\bar{Y}^0=\left(0,\frac{5}{2},\frac{1}{2}\right)$ и оптимальное значение целевой функции двойственной задачи $Z(\bar{Y}^0)_{\min}=30$.

Решение.

1. Применяя первую теорему двойственности, утверждаем, что $F(\bar{X})_{\max} = 30$.

2. Запишем двойственную задачу, поставив в соответствие каждому ограничению исходной задачи переменные y_1, y_2, y_3:

$$\begin{cases} -y_1 + y_2 + 3y_3 \ge 4, \\ 2y_1 + y_2 - y_3 \ge 2, \\ y_1 \ge 0, y_2 \ge 0, y_3 \ge 0, \end{cases}$$

$$Z(\bar{Y}) = 6y_1 + 9y_2 + 15y_3 \to \min.$$

3. Подставив данный оптимальный план $\bar{Y}^0 = \left(0, \frac{5}{2}, \frac{1}{2}\right)$ в ограничения двойственной задачи, убеждаемся, что ограничения превращаются в точные равенства.

4. Используя вторую теорему двойственности, приходим к выводу, что, так как $y_1^0 = 0$, $y_2^0 > 0$, $y_3^0 > 0$, исходная задача на оптимальном плане имеет вид

$$\begin{cases} -x_1^0 + 2x_2^0 \le 6, & (29.31) \\ x_1^0 + x_2^0 = 9, & (29.32) \\ 3x_1^0 - x_2^0 = 15, & (29.33) \\ x_1 \ge 0, x_2 \ge 0, \end{cases}$$

$$F(\bar{X})_{\max} = 4x_1^0 + 2x_2^0 = 30.$$

Решая совместно уравнения (29.32) и (29.33), получаем $x_1^0 = 6$, $x_2^0 = 3$ и для найденных значений неизвестных убеждаемся в справедливости неравенства (29.31). Найдено оптимальное решение исходной задачи.

Ответ: $\bar{X}^0 = (6,3)$; $F(\bar{X})_{\max} = 30$.

Замечание. Учитывая значения переменных в оптимальном плане двойственной задачи, приходим к выводу: так как $y_1^0 = 0$, то первый ресурс не дефицитен и незначительное изменение его запасов не приведет к изменению прибыли производителя. Напротив, $y_2^0 = \frac{5}{2}$, $y_3^0 = \frac{1}{2}$, следовательно, второй и третий ресурсы используются полностью и при увеличении (уменьшении) запасов каждого из них на одну единицу максимальная прибыль возрастет (уменьшится) на $\frac{5}{2}$ и $\frac{1}{2}$ денежных единиц.

Методы исследования операций служат надежным инструментом, облегчающим процесс принятия решений на разных уровнях в системе экономического управления: на уровне отдельных производственных участков и бригад, предприятий, отраслей, на уровне экономики в целом.

Глава 30
МАТЕМАТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА МОДЕЛЕЙ ЭКОНОМИКИ И УПРАВЛЕНИЯ

Рассмотрим типовые прикладные задачи экономики и управления, различающиеся по математическому подходу.

30.1. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Оптимизация плана производства

Необходимо определить план производства одного или нескольких видов продукции, который по какому-либо критерию является оптимальным, например обеспечивает максимум прибыли или минимум затрат и т.д. Подобные задачи возникают на разных уровнях управления. Это может быть планирование своей работы отдельным мастером, а может быть планирование производства целой отрасли народного хозяйства.

Экономико-математическая модель оптимизации производства имеет вид

$$\begin{cases} f(x_1, \dots, x_n) = \sum\limits_j c_j x_j \to \max, \\ \sum\limits_j a_{ij} x_j \le b_i, \quad i = 1, 2, \dots, m, \\ x_j \ge 0, \quad j = 1, 2, \dots, n, \end{cases}$$

где $f(x_1, \dots, x_n)$ — целевая функция; n — количество видов выпускаемой продукции; m — количество видов производственных ресурсов; a_{ij} — объем i-го ресурса при выпуске единицы j-й продукции; c_j — прибыль от выпуска единицы j-й продукции; b_i — количество имеющегося ресурса i-го вида; x_j — объем выпуска j-й продукции.

Пример 30.1. Фирма «Хлопушка» занимается изготовлением ворот. Она имеет ресурсы сырья, рабочей силы и станок, необходимый для производства любого из четырех видов производимых ворот. Затраты ресурсов на изготовление единицы данного вида продукции, прибыль, получаемая фирмой, а также запасы ресурсов указаны в табл. 30.1.

На фирму поступил заказ на изготовление 10–12 разных ворот, причем надо изготовить хотя бы одни ворота каждого вида. Необходимо определить, сколько ворот каждого вида надо выпустить, чтобы прибыль была максимальной.

Таблица 30.1

Вид ресурсов	Виды продукции				Запас ресурсов
	1	2	3	4	
Сырье, у.е.	2	2	1	1	20
Рабочая сила, чел.	2	1	2	1	20
Работа станка, чел.	0	1	1	3	10
Прибыль на единицу продукции, тыс. руб.	1	1	2	3	–

Решение. Модель ЛП:

$$\begin{cases} f(x_1, x_2, x_3, x_4) = x_1 + x_2 + 2x_3 + 3x_4 \to \max, \\ 2x_1 + 2x_2 + x_3 + x_4 \le 20, \\ 2x_1 + x_2 + 2x_3 + x_4 \le 20, \\ x_2 + x_3 + 3x_4 \le 10. \end{cases}$$

Оптимальное решение задачи: $f_{\max} = 19$ тыс. руб. при изготовлении $x_1 = 3$ ворот первого вида, $x_2 = 1$ — второго вида, $x_3 = 6$ — третьего вида, $x_4 = 1$ — четвертого вида.

30.2. ОПТИМАЛЬНОЕ СМЕШЕНИЕ

Задачи о смесях возникают при выборе наиболее рационального способа смешения исходных ингредиентов для получения смеси с заданными свойствами. Смесь должна содержать определенное количество компонентов, входящих в состав исходных ингредиентов. Известны стоимостные характеристики ингредиентов. Требуется получить искомую смесь с наименьшими затратами. Задачи такого типа встречаются во многих отраслях промышленности, таких как металлургия, парфюмерия, пищевая промышленность, фармакология, сельское хозяйство. Примерами задач о смесях могут служить: определение кормового рациона скота на животноводческих фермах; составление рецептуры шихты на металлургическом производстве.

За неизвестные в моделях оптимального смешения принимаются доли или количества ингредиентов, идущие на приготовление смеси. Различаются типы моделей: с одной или с большим числом смесей; с ограниченным или неограниченным количеством ингредиентов; по критерию производства смеси.

Однопродуктовая модель оптимального смешения имеет вид

$$\begin{cases} f(x_1, \ldots, x_n) = \sum_j c_j x_j \to \min, \\ \sum_j a_{ij} x_j \ge b_i, \quad i = 1, 2, \ldots, m, \\ x_{ij} \ge 0, \quad j = 1, 2, \ldots, n. \end{cases}$$

Введенные обозначения:

$f(x_1, ..., x_n)$ — целевая функция (минимум затрат на получение смеси); n — количество исходных ингредиентов; m — количество компонентов в смеси; a_{ij} — удельный вес i-го компонента в j-м ингредиенте; c_j — стоимость единицы j-го ингредиента; b_i — минимально допустимое количество i-го компонента в смеси; x_j — количество j-го ингредиента, входящего в смесь.

Пример 30.2. Фермер разводит крупный рогатый скот. В рацион кормления каждого животного в зимний период входят сено, силос и дорогие концентраты. Содержание питательных веществ и минимально необходимые нормы их потребления приведены в табл. 30.2.

Таблица 30.2

Таблица исходных данных

	Белок, г/кг	Кальций, г/кг	Витамины, у.е./кг	Цена, руб./кг
Сено	40	5	2	2
Силос	20	4	1	4
Концентраты	160	4	2	50
Норма потребления	2000	120	40	

Определить рацион, стоимость которого была бы минимальной, если предельные нормы суточной выдачи сена не более 31 кг, силоса не более 26 кг, концентратов не более 5 кг.

Решение. Экономико-математическая модель имеет вид

$$\begin{cases} f = 2x_1 + 4x_2 + 50x_3 \to \min, \\ 40x_1 + 20x_2 + 160x_3 \geq 120, \\ 5x_1 + 4x_2 + 4x_3 \geq 40, \\ x_1 \leq 31, x_2 \leq 26, x_3 \leq 5. \end{cases}$$

Оптимальное решение задачи: $x_1 = 31$, $x_2 = 26$, $x_3 = 1{,}5$; оптимальное значение целевой функции: $f = 241$. Следовательно, суточный рацион животного: 31 кг сена, 26 кг силоса, 1,5 кг концентратов. Минимальная стоимость рациона 241 руб.

Оптимальный раскрой

Задача оптимального раскроя материала возникает перед технологами предприятий, перед менеджерами производящих продукцию фирм. Она обусловлена необходимостью раскроя материала, закупаемого и поставляемого в виде полуфабрикатов стандартных размеров, на заготовки определенной формы или размера для подачи в производственные цеха или для изготовления фирменного продукта в мастерской фирмы. Оптимальный раскрой позволяет сэкономить значительные средства на закупках сырья.

Существует несколько математических моделей оптимального раскроя в зависимости от целевой функции. Рассмотрим задачу раскроя материала с минимальным отходом. Имеются полуфабрикаты (листы металла, или куски кожи, или рулоны тканей). Требуется разделить их на комплекты заготовок для производства металлических изделий, или кожаных перчаток, или простынь и наволочек.

Пусть $f(x_1, ..., x_n)$ — целевая функция (минимум расхода материала); i — количество разных видов полуфабрикатов, где $i = 1, 2, ..., n$; j — количество заготовок разного вида, полученных из полуфабрикатов i-го вида. Здесь $j = 1, 2, ..., m$. Пусть a_{ij} — количество заготовок j-го вида, полученных при раскрое полуфабриката i-го вида; x_i — необходимое количество полуфабриката i-го вида; c_i — величина отходов при раскрое i-го вида полуфабриката; b_j — число заготовок j-го вида в комплекте.

Экономико-математическая модель минимизации отходов материалов имеет вид

$$\begin{cases} \sum\limits_{i=1}^{n} c_i x_i \to \min, \\ \sum\limits_{j=1}^{m} a_{ij} x_i \ge b_j, \\ x_i \ge 0. \end{cases}$$

Пример 30.3. Фирма занимается изготовлением встроенных шкафов. Она может закупать древесно-стружечные плиты трех видов. По договору со строительным предприятием фирма должна оборудовать встроенными шкафами все этажи строящегося дома. Для изготовления одного шкафа требуются заготовки двух видов. При раскрое одного листа первого вида получается пять заготовок первого вида и три заготовки второго вида. Отходы составляют 2 дм^2. Соответствующие данные представлены в табл. 30.3 в первой строке. В других строках таблицы даны варианты раскроя и получающиеся при этом отходы для листов второго и третьего видов. Какие листы и в каком количестве должна использовать фирма для оборудования встроенными шкафами каждого этажа строящегося дома, если заготовки первого вида требуются в количестве не менее 33 штук, заготовки второго вида — в количестве не менее 25 штук?

Таблица 30.3

Таблица исходных данных

Древесно-стружечные плиты	Первый вид заготовки	Второй вид заготовки	Отходы (дм^2)
1	5	3	2
2	4	5	2
3	3	7	3

Решение. Модель ЛП:

$$\begin{cases} f(x_1, x_2, x_3) = 2x_1 + 2x_2 + 3x_3 \to \min, \\ 5x_1 + 4x_2 + 3x_3 \geq 33, \\ 3x_1 + 5x_2 + 5x_3 \geq 25. \end{cases}$$

Оптимальное решение задачи: при изготовлении шкафов на каждый этаж требуется 5 плит первого вида и 2 плиты второго вида. Отходы будут минимальны и составят 14 дм^2.

Транспортная задача

Транспортная задача представляет собой создание и решение экономико-математической модели транспортировки груза от нескольких поставщиков нескольким потребителям. Пусть однородный груз от m-го количества поставщиков доставляется n потребителям. Запасы груза у поставщиков обозначим через a_i, где $i = 1, 2, \ldots, n$, а потребности в данном грузе каждого потребителя — через b_j, где $j = 1, 2, \ldots, n$. Стоимость перевозки единицы груза от i-го поставщика j-му потребителю обозначим через c_{ij}. Объем перевозимого груза от i-го поставщика j-му потребителю обозначим через x_{ij}. Требуется составить такой план грузоперевозок, который обеспечил бы, с одной стороны, вывоз всех запасов груза у поставщиков, с другой стороны — удовлетворение запросов потребителей. При этом затраты на перевозки должны быть минимальными.

Математическая формулировка имеет вид

$$\begin{cases} f = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min \text{ — целевая функция (минимизация} \\ \qquad\qquad\qquad\qquad \text{полных затрат на перевозки груза),} \\ \sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, 2, \ldots, m \text{ — вывоз груза,} \\ \sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1, 2, \ldots, n \text{ — завоз груза потребителю,} \\ x_{ij} \geq 0. \end{cases}$$

Мы сформулировали простейшую сбалансированную транспортную задачу, в которой общий объем поставок равен объему потребления. Для несбалансированных задач характерен избыток продукта — $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$ или его дефицит — $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$. При решении этих задач вводится либо фиктивный поставщик с объемом запасов груза $a = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$, если суммарная потребность в грузе превышает суммарные запасы поставщиков, либо фиктивный потребитель с объемом потребности $b = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$, если суммарная потребность в грузе меньше суммарных запасов у поставщиков. Цена фиктивной перевозки единицы груза принимается равной нулю.

Другие типы транспортной задачи — задачи с фиксированными перевозками или задачи с ограничениями на пропускную способность магистрали. Например, если объем перевозок из пункта i в пункт j задан величиной v_{ij}, то вводится дополнительное ограничение: $x_{ij} \le v_{ij}$.

Пример 30.4. Три поставщика груза должны обеспечить груз трем потребителям. Запасы груза у поставщиков равны: $a_i = 100, 150, 80$ единиц. Потребности в грузе для каждого потребителя соответственно равны: $b_j = 90, 130, 110$ ед. Затраты на перевозку единицы груза от каждого поставщика каждому потребителю представлены матрицей $C = \begin{pmatrix} 4 & 3 & 5 \\ 10 & 1 & 2 \\ 3 & 8 & 6 \end{pmatrix}$. Какова минимальная стоимость перевозки грузов от поставщиков к потребителям?

Решение. Удобно составить таблицу транспортной задачи.

a_i \ b_j	90	130	110
100	4	3	5
150	10	1	2
80	3	8	6

Экономико-математическая модель в численном виде представлена системой

$$\begin{cases} f = 4x_{11} + 3x_{12} + 5x_{13} + 10x_{21} + x_{22} + 2x_{23} + 3x_{31} + 8x_{32} + 6x_{33} \to \min, \\ x_{11} + x_{12} + x_{13} = 100, \\ x_{21} + x_{22} + x_{23} = 150, \\ x_{31} + x_{32} + x_{33} = 80, \\ x_{11} + x_{21} + x_{31} = 90, \\ x_{12} + x_{22} + x_{32} = 130, \\ x_{31} + x_{32} + x_{33} = 110. \end{cases}$$

Оптимальное распределение перевозок: $x_{11} = 10$, $x_{12} = 90$, $x_{13} = 0$, $x_{21} = 0$, $x_{22} = 40$, $x_{23} = 110$, $x_{31} = 80$, $x_{32} = 0$, $x_{33} = 0$. Минимальные затраты на перевозку составят $f = 810$ ед.

Задача с запретами может резко увеличить минимальные издержки. Предположим, что груз перевозится по железной дороге, пропускная способность которой ограничена по всем направлениям величиной $x_{ij} \le 60$, $i = 1, 2, 3$.

Минимальные затраты на перевозку составят 1180 ед. при следующих объемах перевозок от поставщиков к потребителям $x_{11} = 0$, $x_{12} = 60$, $x_{13} = 40$, $x_{21} = 30$, $x_{22} = 60$, $x_{23} = 60$, $x_{31} = 60$, $x_{32} = 10$, $x_{33} = 10$.

30.3. ЦЕЛОЧИСЛЕННОЕ ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Задачи целочисленного ЛП — это оптимизационные задачи, в которых некоторые переменные или все переменные могут принимать только целые значения. Это, как правило, единственное отличие от общей задачи ЛП. Однако решение этих задач может потребовать специальных методов. Среди них методы отсечений, метод ветвей и границ, метод Беллмана, с которыми можно познакомиться в справочнике для экономистов под ред. В.И. Ермакова. Простые задачи без труда решаются методом перебора. Для каждой совокупности переменных находится целевая функция. Комбинация переменных, при которой целевая функция достигает наибольшего или наименьшего значения, является решением.

Загрузка объема (грузового контейнера, товарного вагона, корабля)

Имеются предметы n видов. Предмет i-го вида имеет массу a_i и ценность c_i, $i = 1, 2, \ldots, n$. Требуется загрузить контейнер грузоподъемностью b так, чтобы ценность груза была максимальной. Пусть x_i — количество предметов i-го вида, которые должны быть помещены в контейнер. Задача целочисленного ЛП принимает вид

$$\begin{cases} \sum_{i=1}^{n} c_i x_i \to \max, \\ \sum_{i=1}^{n} a_i x_i \le b, \\ x_i \ge 0, \\ x_i \text{ – целое}, \quad i = 1, 2, \ldots, n. \end{cases}$$

Пример 30.5. Общественная организация медицинской помощи пострадавшим от стихийных бедствий комплектует грузовой автомобиль с кузовом объемом 15,5 м^3 медицинским оборудованием, лекарствами и предметами первой необходимости для отправки в пострадавший район. Из-за ограниченного объема приходится выбирать такие предметы, которые представляют наибольшую значимость для пострадавших при наименьшем объеме. Перечень части имеющегося на складе представлен в табл. 30.4.

Таблица 30.4

Предметы первой необходимости	Уровень значимости	Объем упаковки, м^3
Перевязочные средства	3	1
Противоэпидемические лекарства	2	2
Палатки	1	5

Создадим экономико-математическую модель для нахождения наибольшей значимости отправляемого оборудования, введя x_1, x_2, x_3 — количество упаковок перевязочных средств, противоэпидемических лекарств и палаток соответственно:

$$\begin{cases} f = 3x_1 + 2x_2 + x_3 \to \max, \\ 4x_1 + 3x_2 + 4x_3 \le 15{,}5, \\ x_i \ge 0, \\ x_i \text{ – целое}, \quad i = 1, 2, 3. \end{cases}$$

Решение. $f_{\max} = 11$ при $x_1 = 3, x_2 = 1, x_3 = 0$. В соответствии с решением следует отправить три упаковки с перевязочными средствами и одну упаковку с противоэпидемическими лекарствами. В этом случае будет достигнута максимальная значимость, а соответственно, и ценность материалов для пострадавших.

Распределение инвестиций с ожидаемым доходом

Банк вкладывает имеющиеся у него средства в несколько проектов. Каждый из проектов в случае его реализации принесет доход.

Требуется определить, какие проекты следует выбрать, чтобы суммарный доход был максимальным. Введем неизвестные:

$$x_i = \begin{cases} 1, & \text{если проект } p_i \text{ реализуется,} \\ 0, & \text{если нет.} \end{cases}$$

Для каждого проекта известны ожидаемый доход при его осуществлении c_i и необходимая величина инвестиций a_i. Экономико-математическая модель:

$$\begin{cases} f = \sum_{i=1}^{n} c_i x_i \to \max, \\ \sum_{i=1}^{n} a_i x_i \le b, \\ x_i \ge 0, \\ x_i = \{0; 1\}, \quad i = 1, 2, \dots, n. \end{cases}$$

Пример 30.6. Малый региональный банк занимается инвестициями в строительстве и располагает для этих целей денежными средствами в сумме 87 млн руб. В портфеле банка находятся три проекта. Каждый из них предполагает строительство нескольких одинаковых объектов, например до четырех магазинов, или до четырех автозаправочных станций, или до пяти одинаковых индивидуальных жилых домов. Ожидаемый доход и величина инвестиций при строительстве одного объекта даны в табл. 30.5

Таблица 30.5

№ п/п	Проект строительства	Количество	Инвестиции в расчете на один объект строительства	Доход в расчете на один объект строительства
1	Магазины	4	10	12
2	Автозаправочные станции	4	16	19
3	Жилые дома	5	7	8

Пусть x_1, x_2, x_3 — количество объектов каждого проекта. Экономико-математическая модель в численном виде:

$$\begin{cases} f = 12x_1 + 19x_2 + 8x_3 \to \max, \\ 10x_1 + 16x_2 + 7x_3 \le 87, \\ x_1 \in \{0, 1, \dots, 4\}, \\ x_2 \in \{0, 1, \dots, 4\}, \\ x_3 \in \{0, 1, \dots, 5\}. \end{cases}$$

Решение. $f_{max} = 102$, {4; 2; 2} или {2; 2; 5}. Максимальный доход от реализации проектов составит 102 млн руб. при условии, что первый проект будет проинвестирован полностью, остальные два — частично. Другой вариант инвестирования при том же максимальном доходе — последний проект инвестируется полностью, а первые два — наполовину.

30.4. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Нелинейное программирование (НП) — раздел дисциплины «Исследование операций», который посвящен анализу математических моделей, в которых имеются функции с нелинейными членами. Упрощенный учет входящих факторов или условия вариации переменных в небольшом диапазоне позволяют строить как экономико-математические, так и инженерно-математические *линейные* модели. Если анализ построенной модели показывает, что она неудовлетворительно отражает реальный объект, добавляются нелинейные слагаемые.

Экономико-математическая модель с нелинейной программой действий содержит целевую непрерывную линейную или нелинейную функцию векторного аргумента $f(\boldsymbol{x}) = f(x_1, x_2, ..., x_n)$, линейные или нелинейные ограничения в виде равенств $h_i(\boldsymbol{x}) = 0$, где $i = 1, 2, ..., m$, и линейные или нелинейные ограничения в виде неравенств $g_j(\boldsymbol{x}) \le 0$ или $g_j(\boldsymbol{x}) \ge 0$, где $j = m + 1, m + 2, ..., p$. Любое уравнение $h_i(\boldsymbol{x}) = 0$ равносильно системе двух неравенств

$$\begin{cases} h_i(\boldsymbol{x}) \le 0, \\ -h_i(\boldsymbol{x}) \le 0. \end{cases}$$

Можно считать, что все ограничения задачи НП задаются только неравенствами. Число неравенств, таким образом, равно $2m + p - m = m + p$.

Тогда задача НП в общем виде может быть сформулирована следующим образом: найти координаты вектора $\boldsymbol{x}$, при которых целевая функция достигает минимума (максимума) при линейных и (или) нелинейных ограничениях в виде неравенств с новым обозначением $g_i(\boldsymbol{x}) \le 0$, где $i = 1, 2, ..., m + p$.

$$\begin{cases} f = f(x_1, x_2, \ldots, x_n) \to \text{extr}, \\ q_i(\boldsymbol{x}) \le 0, \quad i = 1, 2, \ldots, m + p, \\ x_j \ge 0, \quad j = 1, 2, \ldots, n. \end{cases}$$

Интенсивное развитие в последние годы НП позволило систематизировать этот раздел, выделив относительно самостоятельные части: выпуклое программирование; целочисленное программирование, линейную составляющую которого мы уже рассматривали; стохастическое программирование; динамическое программирование.

В задачах *выпуклого программирования* рассматриваются выпуклые или вогнутые целевые функции, заданные на выпуклом замкнутом множестве. Среди задач выпуклого программирования более подробно изучены задачи квадратичного программирования. В этих задачах целевая функция является квадратичной, а ограничения — линейны.

В задачах *стохастического программирования* используются достижения теории вероятностей, поскольку целевая функция или функции ограничений содержат случайные величины.

Динамическое программирование рассматривает в качестве одной из переменных время, в результате чего получаются решения, являющиеся функциями времени.

Решение задач в НП значительно сложнее решения задач ЛП. В ЛП целевая функция задана всегда на выпуклом множестве с конечным числом крайних точек. Разработанный для этих задач симплекс-метод позволяет, перебрав эти точки, найти оптимальное решение. В НП необязательны выпуклость области допустимых значений и конечность числа ее крайних точек. Нет также универсальных методов, подобных симплексу. Применяемые методы ограничены узким классом задач. Это и служит причиной основной трудности решения задач НП.

Рассмотрим метод исследования целевой функции на экстремум, основанный на теореме Куна—Таккера.

Для дифференцируемой функции двух переменных $f(x_1, x_2)$ с двумя ограничениями математическая модель может, например, иметь вид

$$\begin{cases} f = f(x_1, x_2) \to \text{extr}, \\ q(x_1, x_2) \le 0, \\ x_1 \ge 0. \end{cases}$$

Необходимые условия оптимальности для задачи НП используют функцию Лагранжа

$$L(x_1, x_2, \lambda_1, \lambda_2) = f(x_1, x_2) - \lambda_1 q(x_1, x_2) - \lambda_2(-x_1).$$

Возникают отличия от стандартного метода решения задачи на условный экстремум с ограничениями в виде равенств. Знак перед λ_1 и λ_2 отрицателен, причем все λ считаются неотрицательными. Не все ограничения в виде неравенств должны действовать. Теорема Куна—Таккера справляется с этим посредством приведения соответствующих множителей Лагранжа к нулю. Случай равенства нулю множителя Лагранжа для действующего ограничения мы исключим.

Пусть точка $(x_1^0, x_2^0, \lambda_1^0, \lambda_2^0)$ является точкой экстремума функции $f(x_1, x_2)$. Тогда функция Лагранжа удовлетворяет в этой точке следующим условиям:

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 0, \\ \dfrac{\partial L}{\partial x_2} = 0, \\ q(x_1, x_2) \le 0, \\ x_2 \ge 0, \\ \lambda_1 \cdot q(x_1, x_2) = 0, \\ \lambda_2 \cdot x_1 = 0, \\ \lambda_1 \text{ и } \lambda_2 \ge 0. \end{cases}$$

Два последних выражения в системе называются условиями комплиментарности и говорят о том, что недействующие ограничения имеют нулевой множитель Лагранжа. Эта точка, а также все другие, для которых справедлива система, называются точками Куна—Таккера. Оптимальное решение задачи находится среди этих точек.

Пример 30.7. Исследовать на максимум функцию $f = x_1 - x_2$ при условиях $2x_1^2 + 4x_2^2 \le 3$ и $x_1 \ge 0$.

Решение. Для задачи

$$\begin{cases} f = x_1 - x_2 \to \max, \\ 2x_1^2 + 4x_2^2 - 3 \le 0, \\ x_1 \ge 0 \end{cases}$$

построим функцию Лагранжа:

$$L(x_1, x_2, \lambda) = x_1 - x_2 - \lambda_1(2x_1^2 + 4x_2^2 - 3) - \lambda_2(-x_1).$$

Условия Куна—Таккера имеют вид

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 1 - 4x_1\lambda_1 + \lambda_2 = 0, \\ \dfrac{\partial L}{\partial x_2} = -1 - 8x_2\lambda_1 = 0, \\ 2x_1^2 + 4x_2^2 - 3 \le 0, \\ x_1 \ge 0, \\ \lambda_1(2x_1^2 + 4x_2^2 - 3) = 0, \\ \lambda_2 x_1 = 0, \\ \lambda_1 \text{ и } \lambda_2 \ge 0. \end{cases}$$

Решение системы уравнений и неравенств с четырьмя переменными дает точку $M\left(-\frac{1}{2}, 1\right)$. Максимальное значение функции $f_{\max}\left(-\frac{1}{2}, 1\right) = \frac{3}{2}$.

Оптимизация портфеля ценных бумаг

Задача оптимизации инвестиционного портфеля состоит в том, чтобы определить доли покупаемых акций различных компаний с целью оптимизации величины ожидаемого дохода и уровня риска. Известно, что доходность тех акций больше, приобретение которых увеличивает риск убытка. Предположим, что мы желаем скомбинировать два актива в портфель с пропорциями x_1 и x_2, т.е. $x_1 + x_2 = 1$. Пусть ожидаемая процентная ставка по ним 12 и 15%. Ожидаемый доход $Z = 1{,}12x_1 + 1{,}15x_2$. Рискованность одного актива измеряется дисперсией или средним квадратическим отклонением прибыли по этому активу, а риск портфеля — дисперсией или средним квадратическим отклонением прибыли портфеля. Для измерения риска портфеля необходимо знание вариации доходов отдельных ценных бумаг, а также степень, с которой прибыли пар ценных бумаг колеблются вместе, т.е. знание ковариации. Дисперсии и ковариации прибыли по всем возможным парам активов отобразим в ковариационной матрице:

$$\sigma_\rho^2 = (x_1 \quad x_2)\begin{pmatrix} \sigma_{x_1}^2 & \operatorname{cov}(x_1, x_2) \\ \operatorname{cov}(x_2, x_1) & \sigma_{x_2}^2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Вспомним, что ковариация $\operatorname{cov}(x_1, x_2)$ выражается через коэффициент корреляции $\rho(x_1, x_2)$ так:

$$\operatorname{cov}(x_1, x_2) = \rho(x_1, x_2) \cdot \sigma_{x_1}^2 \sigma_{x_2}^2.$$

Определяя при статистическом анализе информации величины $\sigma_{x_1} = 20\%$, $\sigma_{x_2} = 25\%$ и $\rho(x_1, x_2) = -0{,}1$, найдем дисперсию прибыльности портфеля:

$$\sigma_\rho^2 = (0{,}2)^2 2x_1^2 + (0{,}25)^2 x_2^2 - 2 \cdot 0{,}1 \cdot 0{,}2 \cdot 0{,}25 x_1 x_2.$$

Менеджер, управляющий портфелем, устанавливает приоритеты, в соответствии с которыми строится портфель ценных бумаг. Например, при заранее определенной доле риска требуется сформировать портфель так, чтобы его доходность была максимальной. Или целевой функцией может быть минимизация риска, если доход должен быть не меньше некоторой величины.

Пусть менеджер работает с рискованными бумагами, имея прибыль 15%. Можно ли, незначительно снизив прибыль покупкой более надежного актива, существенно уменьшить риск? Решим задачу минимизации риска при условии, что суммарная прибыль пакета активов не должна быть меньше, например, 14%. Соответствующая экономико-математическая модель включает четыре неравенства:

$$\begin{cases} \sigma_\rho^2 = 0{,}04x_1^2 + 0{,}0625x_2^2 - 0{,}01x_1x_2 \to \min, \\ 1 - x_1 - x_2 \geq 0, \\ 1{,}10x_1 + 1{,}15x_2 - 1{,}14 \geq 0, \\ x_1 \geq 0, \\ x_2 \geq 0. \end{cases}$$

Решение задачи указывает доли $x_1 = 0{,}2$ и $x_2 = 0{,}8$ денежных средств на приобретение наиболее эффективного пакета акций с минимальным риском убытка $\sigma_\rho^2 = 0{,}04$. При этом все средства инвестированы. Среднее квадратическое отклонение прибыльности портфеля составит $\sigma_\rho^2 = \sqrt{0{,}04} = 0{,}2$, или 20%. Таким образом, решая задачу оптимизации, менеджер может покупать в основном более рискованные активы с большей доходностью. Уровень риска сбалансированного портфеля, в котором рискованных активов в 4 раза больше, не превысит риска более надежного актива. Но бесполезно оптимизировать портфель, если все акции быстро падают.

30.5. МНОГОКРИТЕРИАЛЬНОЕ ПРОГРАММИРОВАНИЕ

До сих пор, рассматривая экономико-математические модели, мы исследовали на экстремум одну целевую функцию, варьируя переменные. Иногда ставится задача одновременного исследования на экстремум нескольких функций от одних и тех же переменных, причем своего экстремального значения они достигают при разных значениях переменных. Возникает вопрос: что считать оптимальным решением? На каких значениях переменных остановиться, если мы заинтересованы в экстремальном значении каждой функции? На этот вопрос мы и постараемся ответить.

Пусть на плоскости x и y дано множество точек D, определяемое некоторыми условиями (рис. 30.1). Оно является областью определения двух функций: $u = u(x, y)$ и $v = v(x, y)$. Требуется найти на этом множестве точку $U(x^0, y^0)$, в которой одновременно выполняются два условия: $u(x^0, y^0) \to \max$ и $v(x^0, y^0) \to \max$, т.е. на плоскости (u, v) найти точку $N(u(x^0, y^0), v(x^0, y^0))$. Очевидно, что в общем случае задача может не иметь решения. Поставим задачу иначе. Найти значения $u(x^1, y^1)$ и $v(x^1, y^1)$, наиболее близко расположенные к точке N в совокупности (рис. 30.2). Такая задача может быть решена. Но следует определить, что значит «наиболее близко в совокупности». Например, на рис. 30.2 образ N искомой точки U расположен вне области E, являющейся образом множества D. Точка N недостижима и поэтому названа *точкой утопии*.

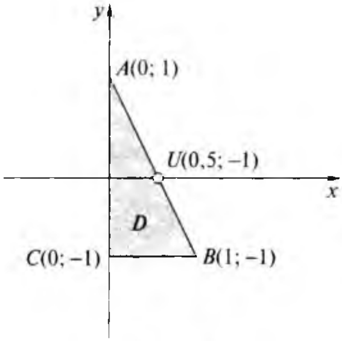

Рис. 30.1

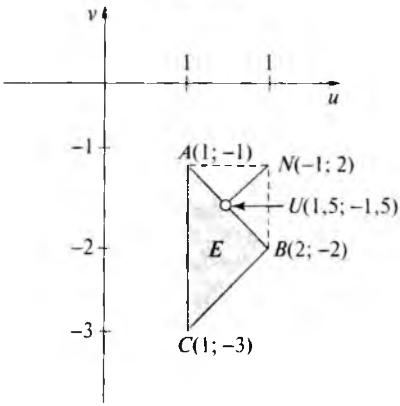

Рис. 30.2

Метод идеальной точки

Перемещаясь по множеству D, будем стремиться к точке N. Достигнув границы множества E, обнаружим, что когда приближаемся к точке N, увеличивая одну координату, другая будет уменьшаться. Назовем *множеством Парето* совокупность таких точек на границе множества E, перемещение по которым приводит к увеличению одной из координат (u или v) при одновременном уменьшении другой координаты. На рис. 30.2 эти точки лежат на отрезке $A'B'$. Метод идеальной точки состоит в отыскании на множестве Парето точки, ближайшей к точке N.

Пример 30.8. На множестве D

$$\begin{cases} x \geq 0, \\ y \geq 0, \\ 2x + y \leq 1 \end{cases}$$

заданы две функции: $u(x, y) = x + 1$ и $v(x, y) = x + y - 2$. Найти точку $U(x^0, y^0)$ на множестве D, в которой эти функции достигнут наибольшего значения.

Решение. На рис. 30.1 изображено множество D, представляющее собой треугольник ABC. Обозначим множество пар допустимых значений u и v через E. Построим множество E в координатах u и v (см. рис. 30.2). В точке A' функция v достигнет своего максимального значения $v_{\max}(0, 1) = -1$. В точке B' функция u будет иметь максимум $u_{\max}(1, -1) = 2$. Из точек A' и B' проведем прямые, параллельные осям u и v, до взаимного пересечения. Получим точку утопии $N(2, -1)$. Теперь на множестве Парето требуется найти точку, ближайшую к точке N. Проведя перпендикуляр из точки N к отрезку $A'B'$, найдем точку $U'(u^0, v^0)$ на множестве Парето, ближайшую к точке утопии. Эта точка получила название *идеальной*. Ее координаты $u^0 = 1{,}5$ и $v^0 = -1{,}5$. Зная u^0 и v^0, определим $x^0 = 0{,}5$ и $y^0 = 0$. Итак, в точке $U(0{,}5; 0)$ функции $u(x, y) = x + 1$ и $v(x, y) = x + y - 2$ достигнут таких значений, что расстояние от идеальной точки до точки утопии окажется минимальным. Назовем их максимальными по совокупности или максимальными по Парето.

Пусть построена многокритериальная модель, в которой одна целевая функция $u = u(x, y)$ исследуется на максимум, а другая — $v = v(x, y)$ на минимум. Сделаем замену $q = -v(x, y)$. Тогда

$$u(x, y) \to \max \quad \text{и} \quad q(x, y) \to \max.$$

Задача сводится к предыдущей — исследованию на максимум двух функций от одних и тех же переменных.

При анализе модели с требованием минимума для обеих функций: $u(x, y) \to \min$ и $v(x, y) \to \min$ — делаются две замены:

$$q = -u(x, y) \to \max \quad \text{и} \quad q = -v(x, y) \to \max.$$

Методы нахождения кратчайшего расстояния от точки утопии до идеальной точки зависят от структуры множества Парето.

Метод сглаживания противоречий

К многокритериальным можно отнести модели, использование которых позволяет найти решение в, казалось бы, безвыходной ситуации. Например, экономико-математическая модель содержит противоречивые условия. В безвыходной ситуации оказывается управляющий предприятием при планировании производства в условиях завышенных требований к объему выпуска товаров и недостаточной обеспеченности производства сырьем. Рассмотрим эту модель подробнее.

Пусть x_j, где $j = 1, 2, ..., m$, — выпускаемый ассортимент продукции; b_i, где $i = 1, 2, ..., n$, — объемы потребления n различных ресурсов; a_{ij} — удельная ресурсоемкость производства, т.е. количество затрачиваемого i-го ресурса в расчете на j-й вид продукта; c_j — план по выпуску j-го вида продукта; d_j — эффективность (например, доходность) выпуска j-го вида продукта.

Исследуется на максимум функция эффективности (или доходности) работы предприятия $\sum_{j=1}^{m} d_j x_j \to \max$ при ограничениях сверху по использованию ресурсов и ограничениях снизу по выпуску каждого вида продукта:

$$\begin{cases} \sum\limits_{i=1}^{n} a_{ij} x_j \le b_i, \quad i = 1, 2, \ldots, n, \\ x_j \ge c_j, \quad j = 1, 2, \ldots, m, \\ \sum\limits_{j=1}^{m} d_j x_j \to \max. \end{cases}$$

Как поступить, если система ограничений противоречива? Возможность приписок при представлении отчетности или обмана потребителя не является математическим решением задачи.

В случае несовместности ограничений следует скорректировать план по выпуску. Пусть z_j — снижение выпуска относительно плана, т.е. $z_j = c_j - x_j$, причем $z_j \ge 0$. Это снижение следует провести с минимальными потерями, после чего из всех вариантов плана выбрать тот, который максимизирует функцию эффективности работы предприятия. Минимизация потерь означает исследование на минимум функции $\sum_{j=1}^{m} e_j z_j$, где e_j — коэффициент снижения эффективности работы предприятия из-за невыполнения плана.

Пример 30.9. Предприятие выпускает два вида продукции x и y с эффективностью выпуска каждой продукции $d_1 = 3$ и $d_2 = 2$, с удельной ресурсоемкостью по каждому продукту $a_1 = 1$ и $a_2 = 2$ соответственно. Объем потребления ресурса не превышает $b = 3$. На предприятие спущен план по выпуску $(c_1, c_2) = (2, 1)$, т.е. $x \geq 2$, $y \geq 1$. В случае невыполнения плана на предприятие налагаются штрафные санкции пропорционально снижению выпуска по каждому продукту с коэффициентами $e_1 = 1$ и $e_2 = 2$. При этом предприятие не имеет права в ущерб одному виду продукции наращивать выпуск другого вида. Найти оптимальный план в этих условиях.

Решение. Экономико-математическая модель работы предприятия следующая:

$$\begin{cases} f = 3x + 2y \to \max, \\ x + 2y \leq 3, \\ x \geq 2, \\ y \geq 1. \end{cases}$$

На рис. 30.3 по осям отложены количества производимых предприятием продуктов. Ограничения на ресурс позволяют предприятию работать только в заштрихованной области. Требования по выпуску превышают реальные возможности предприятия (соответствующая область находится правее и выше точки $B(2; 1{,}5)$). Предприятие корректирует выпуск каждого продукта на величины

$$z_1 = 2 - x, \quad z_2 = 1 - y,$$

за что обязано заплатить штраф

$$\varphi = 1 \cdot (2 - x) + 2 \cdot (1 - y).$$

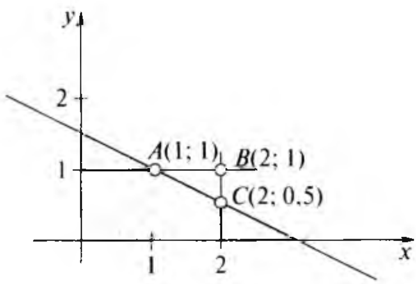

Рис. 30.3

Найдем минимальное значение величины штрафа. При полном использовании предоставленного ресурса второе соотношение системы превращается в равенство, откуда $x = 3 - 2y$. Подстановка величины x в функцию φ дает значение

$$\varphi = 1 \cdot (2 - 3 + 2y) + 2 \cdot (1 - y) = +1.$$

Заплатив штраф, предприятие максимизирует функцию f:

$f = 3x + 2y = 3(3 - 2y) + 2y = 9 - 4y.$

В точке $C(2; 0,5)$ будет достигнута максимальная эффективность

$$f = 9 - 4y = 9 - 4 \cdot \frac{1}{2} = 7.$$

Если штрафные коэффициенты иные: $e_1 = 2$ и $e_2 = 5$, функция штрафа приобретает вид

$\varphi = 2 \cdot (2 - 3 + 2y) + 5 \cdot (1 - y) = 3 - y.$

Получим $\varphi_{min} = 2$ при $x = 1$ и $y = 1$ (точка A на графике). В этих условиях функция f будет равна

$f = 3x + 2y = 3 \cdot 1 + 2 \cdot 1 = 5.$

Таким образом, предприятие, заплатив штрафные налоги, далее «живет спокойно».

30.6. ИМИТАЦИОННОЕ ПРОГРАММИРОВАНИЕ

Моделирование ситуации с учетом случайной составляющей называют *имитацией*. Пояснить, что такое имитационное моделирование, можно на примере определения площади криволинейной фигуры (рис. 30.4). Для нахождения площади фигуры S набросаем множество точек (чем больше, тем лучше) случайным образом в фигуру P, внутри которой расположена фигура S. Подсчитаем число точек N_P, попавшее в фигуру P, а также их часть N_S, попавшую в фигуру S. Площадь фигуры S находим из соотношения $\frac{N_S}{N_P}ab$, где ab — площадь фигуры P.

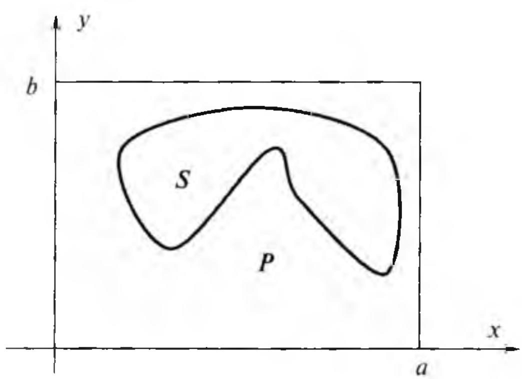

Рис. 30.4

Расстановка приоритетов

Количественные методы дают возможность менеджеру использовать математику в тех областях, где, казалось бы, числам нет места. Одна из задач, часто возникающих у сотрудника, принимающего решения, — расставить приоритеты, т.е. ранжировать совокупность элементов по значимости. Например, если первый проект важен по сравнению со вторым и очень важен по сравнению с третьим, то проекты по важности следует расставить так: первый; второй; третий. Но может оказаться, что надо иметь в виду, что третий проект очень важен по сравнению со вторым. Учет этого обстоятельства проекты 2 и 3 может поменять местами: первый; третий; второй. Под такие рассуждения можно подвести математическую базу.

Пусть у нас есть n гирек разного веса. Расставим их по весам, начав с самой тяжелой, $a_1, a_2, \ldots, a_n$ и построим матрицу A из отношений $\frac{a_i}{a_j}$ следующего вида:

$$A = \begin{pmatrix} \frac{a_1}{a_1} & \frac{a_1}{a_2} & \ldots & \frac{a_1}{a_n} \\ \frac{a_2}{a_1} & \frac{a_2}{a_2} & \ldots & \frac{a_2}{a_n} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{a_n}{a_1} & \frac{a_n}{a_2} & \ldots & \frac{a_n}{a_n} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}.$$

Столбец $\begin{pmatrix} a_1 \\ \ldots \\ a_n \end{pmatrix}$ является собственным вектором матрицы с собственным значением $\lambda = n$.

Действительно,

$$\begin{pmatrix} \frac{a_1}{a_1} & \frac{a_1}{a_2} & \ldots & \frac{a_1}{a_n} \\ \frac{a_2}{a_1} & \frac{a_2}{a_2} & \ldots & \frac{a_2}{a_n} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{a_n}{a_1} & \frac{a_n}{a_2} & \ldots & \frac{a_n}{a_n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix} = \begin{pmatrix} \frac{a_1}{a_1}a_1 & \frac{a_1}{a_2}a_2 & \ldots & \frac{a_1}{a_n}a_n \\ \frac{a_2}{a_1}a_1 & \frac{a_2}{a_2}a_2 & \ldots & \frac{a_2}{a_n}a_n \\ \ldots & \ldots & \ldots & \ldots \\ \frac{a_n}{a_1}a_1 & \frac{a_n}{a_2}a_2 & \ldots & \frac{a_n}{a_n}a_n \end{pmatrix} = n\begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix}.$$

Назовем матрицу *A идеальной*. Если элементы матрицы незначительно изменятся, то и собственный вектор с собственным числом изменятся незначительно.

Пусть теперь мы знаем веса гирек с некоторой погрешностью. Например, грубо измеренный вес гирьки a_2 оказался на 10% больше ее точного веса. Тогда элемент матрицы a_{21} стал равен $1{,}1\frac{a_2}{a_1}$. Учитывая погрешность измерения весов гирек, поставим в ячейки матрицы вместо чисел $\frac{a_i}{a_j}$ мало отличающиеся от них числа $k_{ij}\frac{a_i}{a_j}$, где коэффициент k_{ij} учитывает погрешность измерения. Он близок или равен единице. Тогда произведение матрицы на собственный вектор будет иметь вид

$$\begin{pmatrix} k_{11}\frac{a_1}{a_1} & k_{12}\frac{a_1}{a_2} & \dots & k_{1n}\frac{a_1}{a_n} \\ k_{21}\frac{a_2}{a_1} & k_{22}\frac{a_2}{a_2} & \dots & k_{2n}\frac{a_2}{a_n} \\ \dots & \dots & \dots & \dots \\ k_{n1}\frac{a_n}{a_1} & k_{n2}\frac{a_n}{a_2} & \dots & k_{nn}\frac{a_n}{a_n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} (k_{11}+k_{12}+\dots+k_{1n})a_1 \\ (k_{21}+k_{22}+\dots+k_{2n})a_2 \\ \dots\dots\dots\dots\dots\dots \\ (k_{n1}+k_{n2}+\dots+k_{nn})a_n \end{pmatrix} =$$

$$= n\begin{pmatrix} \frac{\sum_{j=1}^{n} k_{1j}}{n}a_1 \\ \frac{\sum_{j=1}^{n} k_{2j}}{n}a_2 \\ \dots \\ \frac{\sum_{j=1}^{n} k_{nj}}{n}a_n \end{pmatrix} = n\begin{pmatrix} a'_1 \\ a'_2 \\ \dots \\ a'_n \end{pmatrix}.$$

Числа $\frac{\sum_{j=1}^{n} k_{ij}}{n}$, где $i = 1, 2, \dots, n$, мало отличаются от единицы. Если вес одной гирьки мы учли с превышением, то вес другой могли учесть с недостатком. Коэффициенты k_{ij} вносят элемент рассогласования в идеальную матрицу, вследствие чего немного меняются элементы собственного вектора и собственное число. Но расстановка по весам в собственном векторе сохраняется. Одни элементы приобретают незначительно больший вес, другие — незначительно меньший. Это означает, что если вы не знаете точных весов, ваши оценки приблизительны, вы получите, тем не менее, хороший результат.

Следует иметь в виду, что рассогласование с идеальной матрицей не должно быть значительным. Критерий хорошего согласования: $\frac{\lambda - n}{n - 1} < 0{,}1$, где n — порядок квадратной матрицы; λ — собственное число. Если собственных чисел несколько, берется наиболее близкое к n. Оно является максимальным среди других λ.

Для элементов, сравниваемых между собой и не измеряемых числом, надо определить, как их сравнивать. Выберем простую и естественную школьную пятибалльную шкалу сравнений. Пусть один элемент a_1 имеет такую же степень значимости, что и другой a_2. Тогда будем считать, что отношение предпочтения $\frac{a_1}{a_2} = \frac{1}{1}$. Если a_1 более значим, чем a_2, то $\frac{a_1}{a_2} = \frac{2}{1}$. Если элемент a_1 весьма значим по сравнению с a_2, то $\frac{a_1}{a_2} = \frac{3}{1}$. Точно так же при меньшей значимости элемента a_1 по сравнению с элементом a_2 будем писать отношение $\frac{a_1}{a_2} = \frac{1}{2}$. При малой значимости a_1 по сравнению с a_2 поставим $\frac{a_1}{a_2} = \frac{1}{3}$. Естественно считать, что если $\frac{a_1}{a_2} = \frac{3}{1}$, то $\frac{a_2}{a_1} = \frac{1}{3}$. Возможны также и промежуточные оценки, например $\frac{2}{3}$ или $\frac{3}{2}$ (табл. 30.6).

Таблица 30.6

Таблица значимости

Элемент матрицы	Отношение предпочтения				
	Одинаковая	Бóльшая	Много большая	Меньшая	Много меньшая
$\frac{a_i}{a_j}$	$\frac{1}{1}$	$\frac{2}{1}$	$\frac{3}{1}$	$\frac{1}{2}$	$\frac{1}{3}$

Далее составим матрицу сравнений и найдем собственный вектор и собственное число. По ним будем судить о приоритетах и степени доверия к их сравнению.

Замечание 1. Пусть три элемента a_1, a_2, a_3 сравниваются между собой. Было определено, что a_1 значительнее a_2, а a_2 значительнее a_3. Указывать затем, что a_1 менее значимо, чем a_3, было бы нелогично. Поэтому, если критерий хорошего согласования неудовлетворителен, это означает, что при сравнении элементов рассуждения не были вполне логичными. В них велика доля субъективности или произвола. Матрицу сравнений следует пересмотреть.

Замечание 2. Если неясно, в каком отношении элементы находятся друг к другу, следует ограничиться нейтральным отношением $\frac{1}{1}$.

Замечание 3. Подобные задачи желательно решать на компьютере. Во-первых, преобразование матриц с дробными числами вызывает трудности. Во-вторых, если элементов всего два или три, расставить их в приоритетном порядке легко вручную. Интерес представляют задачи управления, где число элементов велико. И тут без компьютера не обойтись. В математических компьютерных программах в разделе линейной алгебры есть опция поиска собственных чисел и собственных векторов. Если научиться пользоваться этой опцией, то мучительная проблема «что предпочесть» сводится к простому сравнению элементов между собой по значимости.

Пример 30.10. В отдел пришли четыре инвестиционных проекта a_1, a_2, a_3, a_4 для решения вопроса о предпочтении в инвестировании. Анализ показал, что проект a_1 более значим по сравнению с проектами a_2 и a_4 и гораздо предпочтительнее проекта a_3. Проект a_2 имеет такую же степень значимости, как и a_3, и предпочтителен при сравнении с a_4. Проект a_4 менее интересен, чем проект a_3. Сформировать приоритетный портфель инвестирования.

Решение. Составим матрицу приоритетов:

$$\begin{pmatrix} \frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{2}{1} \\ \frac{1}{2} & \frac{1}{1} & \frac{1}{1} & \frac{2}{1} \\ \frac{1}{3} & \frac{1}{1} & \frac{1}{1} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{1} & \frac{1}{1} \end{pmatrix}.$$

Собственный вектор и собственное число равны соответственно $\begin{pmatrix} a_1' \\ a_2' \\ \dots \\ a_n' \end{pmatrix} = \begin{pmatrix} 0,77 \\ 0,44 \\ 0,28 \\ 0,37 \end{pmatrix}$ и $\lambda = 4,17$. Проект a_1 имеет высший приоритет. Это сразу видно по отношениям предпочтения в первой строке матрицы сравнений. Второй проект предпочтителен по сравнению с четвертым. На последнем месте по значимости стоит третий проект. Критерий хорошего согласования $\frac{\lambda - n}{n - 1} = \frac{4,17 - 4}{3} < 0,06$ выполнен. Полученным результатам можно доверять.

30.7. МОДЕЛЬ ЛЕОНТЬЕВА МНОГООТРАСЛЕВОЙ ЭКОНОМИКИ

Многоотраслевое хозяйство требует баланса между отдельными отраслями. Каждая отрасль, с одной стороны, является производителем одного определенного набора видов продукции, а с другой — потребителем другого набора видов продукции. Возникает сложная задача: согласовать объемы производства каждой из отраслей, чтобы удовлетворить все потребности в продукте каждой отрасли. Эта задача может быть сформулирована в виде экономико-математической модели межотраслевого баланса (модели Леонтьева), требующей привлечения аппарата матричной алгебры.

Рассмотрим для определенности производственную сферу из n отраслей, каждая из которых производит один (свой) продукт. Выделим определенный период времени, например год, в течение которого все коэффициенты остаются постоянными.

Пусть x_i — общий (валовой) выпуск i-й отрасли ($x_i \geq 0$), $i = 1, 2, ..., n$; x_{ij} — объем продукции i-й отрасли, поставляемой для j-й отрасли в процессе производства; y_i — конечный спрос на продукцию i-й отрасли ($y_j \geq 0$). Сюда относятся личное потребление граждан, содержание государственных и общественных институтов, чистый экспорт, производственное накопление и т.д. Тогда балансовые соотношения примут вид

$$\begin{cases} x_1 = x_{11} + x_{12} + ... + x_{1n} + y_1, \\ x_2 = x_{21} + x_{22} + ... + x_{2n} + y_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_n = x_{n1} + x_{n2} + ... + x_{nn} + y_n. \end{cases} \qquad (30.1)$$

Введем коэффициенты прямых затрат $a_{ij} = \frac{x_{ij}}{x_j} \geq 0,\ i, j = 1, 2, ..., n,$

показывающие затраты i-й отрасли на выпуск одной единицы продукции для j-й отрасли. Заменяя в (30.1) $x_{ij} = a_{ij}x_j$, получим систему n линейных уравнений с n переменными

$$\begin{cases} x_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + y_1, \\ x_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + y_n \end{cases}$$

или в матричной форме

$$X = A \cdot X + Y, \qquad (30.2)$$

где X — матрица (вектор) выпусков отраслей; A — матрица прямых затрат; Y — матрица (вектор) конечного спроса.

Упрощенная экономико-математическая модель межотраслевого баланса составлена. Это матричное уравнение может быть решено. Из (30.2) следует, что

$$(E - A)X = Y. \qquad (30.3)$$

Матрица $(E - A)$ не вырождена, что можно доказать исходя из экономических соображений. Тогда

$$X = (E - A)^{-1}Y.$$

Матрица $S = (E - A)^{-1}$ называется *матрицей полных затрат*. Ее коэффициенты имеют четкий экономический смысл. Зададим конечный спрос, например, в виде вектора $Y_1 = \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix}$, т.е. требуется обеспечить конечный спрос на одну единицу продукции первой отрасли. Матрица (вектор) выпусков отраслей будет иметь вид

$$\begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} = \begin{pmatrix} s_{11} \\ s_{21} \\ \ldots \\ s_{n1} \end{pmatrix}.$$

Следовательно, каждый элемент s_{i1}, $i = 1, 2, \ldots, n$ матрицы полных затрат S есть выпуск продукции каждой из отраслей для обеспечения единицы конечного спроса на продукцию первой отрасли.

Рассмотрим применение модели на примере.

Пример 30.11. В табл. 30.7 приведены данные по балансу между двумя отраслями за некоторый период. Найти необходимый объем валового выпуска каждой отрасли, если конечное потребление первой отрасли увеличится вдвое.

Таблица 30.7

Отрасль		Потребление		Конечный продукт	Валовой продукт
		1	2		
Производство	1	$x_1 = 1$	$x_{12} = 12$	$y_1 = 77$	$x_1 = 100$
	2	$x_{21} = 21$	$x_{22} = 22$	$y_2 = 157$	$x_2 = 200$

Решение. Находим матрицу прямых затрат:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{x_{11}}{x_1} & \dfrac{x_{12}}{x_2} \\ \dfrac{x_{21}}{x_1} & \dfrac{x_{22}}{x_2} \end{pmatrix} = \begin{pmatrix} 0{,}11 & 0{,}06 \\ 0{,}21 & 0{,}11 \end{pmatrix}.$$

Матрица полных затрат

$$S = (E - A)^{-1} = \begin{pmatrix} 0{,}89 & -0{,}06 \\ -0{,}21 & 0{,}89 \end{pmatrix}^{-1} = \begin{pmatrix} 1{,}14 & 0{,}08 \\ 0{,}27 & 1{,}14 \end{pmatrix}.$$

По условию вектор конечного продукта должен быть равен $Y = \begin{pmatrix} 154 \\ 157 \end{pmatrix}$.

Найдем вектор валового выпуска:

$$X = (E - A)^{-1} Y = \begin{pmatrix} 1{,}14 & 0{,}08 \\ 0{,}27 & 1{,}14 \end{pmatrix} \cdot \begin{pmatrix} 154 \\ 157 \end{pmatrix} = \begin{bmatrix} 188{,}12 \\ 220{,}56 \end{bmatrix}.$$

Следовательно, валовой выпуск в первой отрасли необходимо увеличить на 88,12 у.е., а во второй отрасли — на 20,56 у.е.

Данные по балансу между двумя отраслями за следующий период примут вид (табл. 30.8).

Таблица 30.8

Отрасль		Потребление		Конечный продукт	Валовой продукт
		1	2		
Производство	1	$x_{11} = 20{,}69$	$x_{12} = 13{,}23$	$y_1 = 154$	$x_1 = 188{,}12$
	2	$x_{21} = 39.5$	$x_{22} = 24{,}26$	$y_2 = 157$	$x_2 = 220{,}56$

Расчеты проводились с точностью до второй значащей цифры после запятой, поэтому возникла погрешность в десятых долях.

Глава 31
ЭКОНОФИЗИЧЕСКИЙ ПОДХОД К ПРОБЛЕМАМ ЭКОНОМИКИ И УПРАВЛЕНИЯ

31.1. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИСТОРИЧЕСКОГО РАЗВИТИЯ ФИЗИКИ И ЭКОНОМИКИ

Обладая универсальностью своих законов, физика оказала воздействие на развитие всех естественных наук. Образовался ряд смежных дисциплин: физическая химия, биофизика, геофизика, астрофизика и др. Из гуманитарных наук физика активно влияла на философию и сама обогащалась ее идеями. Возникла связь физики и искусствоведения. В последние годы физика стала проникать в экономику.

Если рассматривать физику и экономику в их сравнительном историческом развитии, можно отметить ряд особенностей. Физика определяется как наука, изучающая простейшие и, вместе с тем, наиболее общие закономерности явления природы[1]. Понятие «экономика» имеет несколько значений, одно из которых — отрасль науки, «занимающаяся изучением объективных закономерностей экономического строя общества[2]. Заметим попутно, что в определениях используются различные *стандарты строгости*. Физика как наука начала формироваться в XVII в. с трудов Г. Галилея по механике, хотя он известен больше своими астрономическими наблюдениями. Становление экономики как науки происходило в XVIII в. и связано с А. Смитом, благодаря исследованиям которого политическая экономия превратилась в сравнительно разработанную систему экономических знаний.

Период развития физики в XVII–XIX вв. (300 лет) привел к созданию основ классической физики. Явления физики сводились к механике, основные положения которой таковы.

1. Всякое свободное от действия сил тело находится в покое или в состоянии равномерного и прямолинейного движения.

2. Сила F, действуя на тело массой m, вызывает его ускорение величиной a:

$F = ma.$

Сила действия первого тела на второе F_{12} равна силе действия второго тела на первое F_{21}, причем направления сил противоположны:

---

[1] См.: Большой Советский энциклопедический словарь.
[2] Там же.

$$\sum F_{ij} = 0.$$

Основные положения классической физики известны как законы механики Ньютона.

Экономика как научное направление интенсивно развивалась также в течение приблизительно 300 лет, в XVIII–XX вв. Ее достижения оформлены как классическая экономика, или ортодоксальная экономика, или классическая экономическая теория.

Основные положения классической экономики сводятся к следующему.

1. Производители и потребители преследуют каждый свои цели. Цель производителей — максимум прибыли, цель потребителей — максимальное удовлетворение потребностей. Пусть $F(X)$ есть производственная функция, компоненты вектора X — затраты труда, капитала, учет научного прогресса и т.д. или $F(X)$ — функция потребления, аргументы которой выступают как потребляемые блага. Тогда

$$F(X) \to \max.$$

Экономические отношения стремятся к рыночному равновесию, в результате которого достигается баланс спроса и предложения труда, денег, товаров. Например, суммы величины спроса F_{12} со стороны потребителей и предложения F_{21} со стороны производителей сбалансированы (закон Вальраса):

$$\sum F_{ij} = 0.$$

К концу XIX в. физика представлялась современникам почти завершенной. Все физические явления можно было отнести к механическому взаимодействию, сопровождаемому движением. Физические явления прекрасно описывались математически. В частности, используя закон всемирного тяготения, с высокой точностью удалось рассчитать движение Луны, планет и комет Солнечной системы. Существовало всего несколько фактов, которые классическая физика не могла объяснить. Но они не тормозили технический прогресс, общество было удовлетворено современным состоянием физической науки. Дальнейшее развитие физики обусловливалось в основном любознательностью образованных представителей общества.

Иная ситуация сложилась в экономической науке. К концу XX в. классическая экономика оказалась хорошо оснащена математически. На протяжении длительного периода сформулированные экономические законы объясняли явления экономической жизни общества. Разработанные методики доказали свою работоспособность, что проверялось по краткосрочным прогнозам, когда основные макропараметры рынка оставались постоянными. Можно было подвести некоторые итоги, заявить об адекватном отображении экономической наукой современного экономического состояния общества, объяснить динамику экономического развития и спрогнозировать поведение на ближайшие годы.

События конца XX в. повергли общество в разочарование. Выяснилось, что возникающие фундаментальные экономические проблемы экономическая наука не в состоянии решить. В своей работе «Кризис экономической теории», описывая последствия для России, академик В. Полтерович указывает: «В России прогноз инфляции был занижен в тысячи раз; совершенно неожиданными оказались кризисы неплатежей, глубочайший спад производства и криминализация общества; практически во всех теоретических работах по приватизации предполагалось, что она ведет к быстрому увеличению эффективности, что оказалось неверным». Классическая экономическая теория не только не предсказала катастрофические события в экономике России в 90-е гг. прошлого века, но и ничего не может ответить на вопрос, а что же нам теперь делать.

Двигаемая любопытством и подталкиваемая требованиями технического прогресса, физика в XX в. продолжала развиваться. Стали накапливаться необъяснимые с позиций классической физики факты: выяснилось, что атомы не элементарны, что электромагнитные процессы невозможно свести к механическим, что электромагнитная энергия излучается отдельными порциями — квантами и т.д. Пытаясь объяснить эти и множества других фактов, открытых впоследствии, физика в XX столетии совершила грандиозный скачок. Были созданы теория относительности (физическая теория пространства, времени и тяготения), квантовая механика с соотношением неопределенностей Гейзенберга, согласно которому координата и импульс частицы не могут одновременно иметь точных значений, ядерная физика, приведшая к созданию принципиально новых мощных источников энергии для человечества. В начале XXI в. наши современники воодушевлены идеей создания большого адронного коллайдера в целях поиска бозона Хиггса — «частицы Бога», которая, как предполагают, поможет раскрыть секреты рождения Вселенной. Хотя на современном этапе многие теоретические и экспериментальные исследования призваны удовлетворить лишь любознательность ученых, развитые страны затрачивают многомиллиардные суммы на эти физические проекты.

В настоящее время экономическая наука находится в том состоянии, которое переживала физика конца XIX столетия. Период классического развития заканчивается, нужны новые идеи, другие подходы, свежий взгляд. Различие же состоит в том, что если дальнейшее развитие физики двигалось любознательностью современников, то дальнейшее развитие экономической науки вызвано *крайней необходимостью*. Почему так случилось? Физика изучает законы неживой природы, заметная эволюция которой совершается за миллионы лет. Экономика изучает некоторые закономерности развития человеческого общества, ощутимые эволюционные движения в котором происходят за сотни и даже десятки лет. Если в XVIII–XIX вв. экономическая наука, возможно, опережала развитие общества, то в XX в. она стала отставать.

Исследователи называют некоторые причины сложившейся в экономике ситуации.

1. В классической экономике не нашли отражения значительные *качественные преобразования*, произошедшие в реальной экономике: изменение структуры многих отраслей; быстрое изменение структуры товаров, технологий; увеличение доли неравновесных процессов; информатизация экономики; тенденции глобализации.

2. Большие *количественные изменения* и *длительные временные интервалы* приводят к изменению макропараметров, что выходит за пределы ограничений классической экономической теории.

3. *Самоизоляция* препятствует развитию экономической науки, не давая возможности черпать идеи из других наук. В период интеграции наук и развития смежных дисциплин это становится заметным.

4. На экономику активное воздействие оказывают *политика* и СМИ. Как следствие, в этой среде часто фигурируют не строгие утверждения, а догмы и мифы. К их числу можно отнести следующие: 1) рыночное равновесие единственно; 2) эмиссия денег всегда ведет к инфляции; 3) государство не должно вмешиваться в экономику, рыночные отношения способны сами все отрегулировать и т.д.

На другой канал проникновения искажений в экономические знания указывает В. Полтерович: «Учебники же по экономике и весь процесс обучения построены так, что создают у студентов впечатление, будто они изучают дисциплину, принципиально ничем не отличающуюся от естественных наук. Этому способствует достаточно сложный математический аппарат, обилие формальных доказательств... Подобная точка зрения распространяется в обществе и создает завышенные ожидания, которые экономика не может удовлетворить».

5. Экономика в своем движении под действием политических, природных и других факторов иногда подходит к состоянию, близкому *к потере устойчивости*, она как бы приближается к точкам ветвления на несколько каналов эволюционного развития, к точкам бифуркации. Выбор траектории дальнейшего развития в точке бифуркации будет зависеть от случайных воздействий на экономическое состояние в этот момент, которые могут быть и очень малыми. Достаточно вспомнить инициированные одним-двумя правительственными чиновниками августовские события 1991 г. в России, приведшие к запрету КПСС, исчезновению одних и появлению других экономических механизмов. Другой пример — обвал акций компании «МЕТЧЕЛ» более чем на 40% в августе 2008 г., произошедший на мировых рынках после шутки премьер-министра РФ о приглашении доктора к заболевшему президенту компании. Но и в точках бифуркации действуют законы природы, которые изучают естественные науки.

Подводя итоги сравнительного исторического развития физики и экономики за последние столетия, сведем результаты в табл. 31.1.

Таблица 31.1

	Физика	Экономика
Определение	Корректное по БСЭ	Некорректное по БСЭ Употреблены слова «объективные закономерности», экономика определяется через экономический строй
Начало становления науки	XVII в., связано с работами Г. Галилея	XVIII в., связано с работами А. Смита
Период развития классической науки	XVII–XIX вв. (300 лет)	XVIII–XX в. (300 лет)
Основные положения	Сформулированы (законы Ньютона)	Сформулированы (закон Вальраса и др.)
Достижения классической науки	Физические явления прекрасно описываются открытыми законами и сводятся к механическому взаимодействию, сопровождаемому движением; математически точно рассчитываются	Экономические явления хорошо описываются статистически, краткосрочные прогнозы с высокой вероятностью сбываются. Имеется развитый математический аппарат

	Физика	Экономика
Время завершения периода	Конец XIX в.	Конец XX в.
Причины завершения	Считалось, что физика как наука свою миссию выполнила. Остались второстепенные задачи и некоторые необъяснимые факты, не имеющие принципиального значения	Неумение решить фундаментальные проблемы современности: неучет качественных преобразований в экономических отраслях; серьезные расхождения в длительном прогнозе с реальностью; самоизоляция; негативное воздействие политики и СМИ
Стимулы дальнейшего развития	Любознательность, иногда без стимулирующей роли экспериментов и требования технического прогресса	Крайняя необходимость в понимании экономических явлений и создания прогноза различных возможных путей экономического развития для принятия правительственных решений
Начало развития следующего этапа	Начало XX в.	Начало XXI в.
Достижения	Создание теории относительности, релятивистской и квантовой механики, физики атомного ядра и элементарных частиц	–

Пессимистические слова В. Полтеровича «в экономической науке не происходит накопления фундаментальных эмпирических закономерностей. Скорее наоборот: ранее обнаруженные и, казалось бы, фундаментальные связи между параметрами впоследствии не подтверждаются» заставляют задуматься над тем, что, может быть, у экономической науки свой самобытный путь развития, отличный от естественных наук. Что экономическая теория должна быть построена на других научных принципах.

Физические законы *глобальны*. Всюду, где бы мы ни встретились с двумя материальными телами, в макро- или микромире, на Земле или в космосе, они будут притягиваться друг к другу в соответствии с законом всемирного тяготения.

Может быть, экономику отличает *локальность* законов? В каждой экономической формации, на каждом этапе экономического развития действуют свои коротко живущие, математически описываемые связи. Иными словами, экономические законы эволюционируют с развитием человеческого общества. Меняются значения коэффициентов, добавляются или исчезают переменные, меняется даже характер взаимосвязей. Было бы интересно провести работу по сопоставлению каждого этапа экономического развития общества и тех устойчивых экономических взаимосвязей, которые наилучшим образом в каждый период работали.

Физический закон всемирного тяготения не подвергается сомнению. Однако физики-теоретики в XX в. обратили внимание на то, что одни физические фундаментальные постоянные *более постоянны, чем другие*. Если допустить наличие эволюции гравитационной постоянной вместе с эволюцией Вселенной, то согласие некоторых экспериментальных данных с теорией улучшается. Например, английский физик-теоретик Поль Дирак (1902–1984), осмысливая достижения космологии, выдвинул предположение, что гравитационная постоянная уменьшается с течением космологического времени. Отсюда высокий интерес к экспериментам по прецизионному измерению гравитационной постоянной.

Вопрос о путях развития экономической науки в XXI в. остается открытым. Академик РАН РФ, член Европейской академии, лауреат премий им. Н.Д. Кондратьева и им. Л.В. Канторовича российский экономист-математик В.М. Полтерович заявляет: «*...я не вижу ясных путей выхода из кризиса*». Известный западный экономист Фрэнк Найт формулирует минимальную задачу экономики: «*выяснить, чего делать заведомо не следует*». И далее ему приписываются слова: «Самое вредное — это вовсе не невежество, а знание чертовой уймы вещей, которые на самом деле не верны». Дж. фон Нейман и О. Моргенштерн в дискуссии о самобытном пути развития экономики, о приложении к экономике совершенно новых, пока неизвестных науке принципов указывают, что пока «было бы неразумным рассматривать что-либо иное, чем трактовку задач тем способом, который уже привел к созданию физической науки».

Для понимания причин происходящего в современной экономике разумно естественное желание исследователей обратиться к наукам, которые интенсивно развивались в последние десятилетия, в которых достигнуты значительные результаты, где расчет, эксперимент и реальность находятся в хорошем согласии. Уже в середине XX в. экономисты прибегали к иллюстрациям из физики для освещения экономических концепций. Включение физики в решение экономических проблем вылилось в XXI в. в новое научное направление — физическую экономику или *эконофизику*. Название «физическая экономика» предложил Линден Ларуш, крупный американский экономист, создатель рейгономики. Под словом «физическая» Л. Ларуш понимает экономику, построенную по образу и подобию естественных наук. Первые, пусть маленькие, шаги в этом направлении уже сделаны. Сложившаяся в эконофизике ситуация напоминает стартовую позицию в развитии квантовой физики, но без драматизма экономики начала XXI в. Не претендуя на полноту обзора, приведем некоторые факты проникновения форм физического знания в экономическую науку.

31.2. ИСПОЛЬЗОВАНИЕ ЗАКОНОВ ФИЗИКИ И ФИЗИЧЕСКИХ ИДЕЙ

Флуктуации

Ключевое положение экономической теории — движение к равновесию — оставляет в стороне вопрос об экономическом развитии. Нужен добавочный механизм. Это явления *флуктуаций* (с точки зрения физической науки) или *мутации* (взгляд со стороны биологии), т.е. случайное воспроизводство альтернативных возможностей. Руководствуясь идеями и понятиями квантовой статистической термодинамики, описывающей рождение флуктуаций, авторы применяют его к «общественным самовоспроизводящимся системам». Они указывают, что случайное воспроизводство альтернативных возможностей осуществляется людьми, занимающимися предпринимательством в условиях конкуренции и отбора.

Отдельная мутация, как правило, означает скорее ухудшение, чем улучшение системы в смысле эволюции. Но если возникает благоприятная возможность и она усиливается, то эволюция делает шаг вперед. Прогресс экономики происходит именно тогда, когда новаторы вторгаются в экономическое пространство, когда своими технологиями и продуктами они возбуждают новые потребности, ломают сложившуюся равновесную структуру спроса и предложения. Благодаря случайным процессам эволюция обретает скачкообразный характер.

Каждая мутация означает возмущение в установившемся на некоторое время равновесии. Мутант-агент возбуждает систему, проверяя ее на устойчивость относительно появившейся инновации. Если мутация не дает никаких преимуществ по сравнению с существовавшими ранее агентами, то новый агент-мутант исчезает в процессе отбора. Система оказывается устойчивой относительно возмущения.

Если оказывается, что мутант обладает определенным преимуществом по сравнению с первоначально имевшимися агентами, то отбор ведет к росту нового агента-мутанта. Система оказывается неустойчивой относительно возмущения. Усиливая эти возмущения (мутации), система поднимается на более высокий уровень эволюции. Ориентируясь на добавочную прибыль, новаторы расширяют свое производство, привлекают дополнительную рабочую силу. В это время консерваторы экономят сырье и материалы, сокращают число занятых на своих предприятиях. В дальнейшем подстраиваются к движению новаторов, в противном случае исчезают. Система переходит в новое селекционное равновесие. Затем цикл повторяется.

В заключение отметим, что физические идеи либо находят экспериментальное подтверждение посредством целенаправленных усилий ученых-физиков, что воплощается затем в законах, либо отвергаются. Физические понятия позволяют ясно формулировать идеи и законы. В экономике роль экспериментаторов играют чаще всего правительства. Групповые интересы, противоречивые законы и постановления, непоследовательные действия делают эксперимент «грязным» с точки зрения науки, поэтому подтверждение или опровержение теоретических разработок становится проблематичным.

Термодинамическое распределение частиц в силовом поле (закон Фоккера—Планка)

Один из разделов физики — молекулярная физика изучает свойства вещества, исходя из молекулярно-кинетической теории. В ее основе среди других лежат такие постулаты: атомы и молекулы находятся в непрерывном и хаотическом движении, они упруго сталкиваются между собой. Молекулярно-кинетическая теория изучает макроскопические явления, не вдаваясь в подробности движения отдельных частиц. Суммарным результатом действия огромного количества частиц являются наблюдаемые свойства среды — давление, температура, плотность, распределение по скоростям и т.д., а также их изменения. Система находится в состоянии термодинамического равновесия, когда частицы системы взаимодействуют между собой, движутся по сложным траекториям, но макроскопические параметры остаются неизменными. Всякий термодинамический процесс перехода из одного равновесного состояния в другое может быть совершен с такой скоростью, что состояние системы в каждый момент времени остается равновесным.

В работе [7] и ссылках на предыдущие публикации получено уравнение баланса доходов и расходов семей с учетом случайных факторов. После преобразования и некоторых допущений оно оказалось эквивалентным известному в молекулярной физике интегро-дифференциальному уравнению Фоккера—Планка, описывающему статистическое распределение частиц по скоростям в потенциальном поле. Это уравнение было в свое время решено для нужд физики. Известные решения этого уравнения были применены к изучению экономической структуры общества. Развивая свои идеи и используя российские статистические данные за последние десятилетия, авторы применяют найденные распределения по накоплениям и доходам в налоговой системе, в вопросе ценообразования, в адресной эмиссии денег. Не вдаваясь в подробности, приведем наиболее важные результаты, полученные на основе рассмотренной физико-экономико-математической модели.

1. Рыночное равновесие в общем случае не единственно. Даже при одних и тех же макроэкономических параметрах страна может перейти либо в высокопродуктивное, либо в низкопродуктивное состояние в зависимости от начальных условий. При этом оба состояния являются стационарными и в них имеет место рыночное равновесие цен и накоплений.

2. Модель позволяет получить ответ на вопрос, какие именно факторы могут вызвать кризис (переход в низкопродуктивное состояние) или привести к экономическому чуду, т.е переходу в высокопродуктивное состояние.

3. Современная Россия находится в низкопродуктивном состоянии и без усилий со стороны государства не сможет из него выйти. Этот переход не может произойти за счет рыночной самоорганизации.

Притяжение материальных тел (закон всемирного тяготения Ньютона)

Нашел отражение в экономике и социологии ньютоновский закон всемирного тяготения, который формулируется следующим образом: любые два материальных тела притягивают друг друга с силой F, прямо пропорциональной их массам m_1 и m_2 и обратно пропорциональной квадрату расстояния r между ними и направленной вдоль прямой, соединяющей центры масс:

$$F = G\frac{m_1 m_2}{r^2},$$

где G — гравитационная постоянная.

Американский исследователь Стюарт [8] ввел аналогичное силе тяготения понятие демографической силы, взяв в качестве масс количество населения в городах p_i и p_j и расстояние между городами d_{ij}. Тогда величина демографической силы определяется как

$$F = \alpha\frac{p_i p_j}{d_{ij}^2},$$

где α — некоторый коэффициент.

Каждое тело с массой m_1 создает вокруг себя поле тяготения (гравитационное поле) с гравитационным потенциалом V, связанным с силой тяготения зависимостью grad $V = F$. Соответствующий гравитационному демографический потенциал равен

$$V_{ij} = \beta\frac{p_j}{d_{ij}},$$

где β — некоторая константа. Следовательно, демографический потенциал, создаваемый, например, в центре Филадельфии массой (населением) Нью-Йорка, как указывает Стюарт, равен константе, умноженной на количество населения Нью-Йорка и деленной на 90 миль.

Силы тяготения обладают свойством аддитивности. Продолжая аналогию, получим, что при наличии нескольких масс суммарный потенциал в точке i равен

$$V_i = \beta\sum_{j=1}^{n}\frac{p_j}{d_{ij}}.$$

В связи с этим можно вычислить суммарный потенциал для любого населенного пункта, составить карту потенциалов с эквипотенциальными линиями. Такая работа была сделана для территории США. Интерпретируя понятие демографического потенциала, Стюарт и его сотрудники провели ряд эмпирических исследований. Они отметили высокую степень корреляции демографического потенциала с плотностью сельского населения, с густотой железнодорожной сети на квадратную милю, с ценой на сельскохозяйственные участки и даже с коэффициентом смертности сельского населения. Дальнейшие работы уточняли и развивали эти идеи [9].

Обратная связь

Как известно, в физике уже в XVIII в. в расчетах и конструкциях активно использовалось понятие обратной связи. Экономика XIX в. ввела отрицательную обратную связь. Ее применение приводило к установлению равновесия между ценами, установленными продавцами, и контролируемыми ими долями рынка. А. Маршалл в труде «Принципы экономики», изданном в 1890 г., впервые сформулировал действие также и положительной обратной связи. Однако на протяжении почти 100 лет ей не уделяли должного внимания.

К числу физических идей, которые обещают оказать плодотворное влияние на экономическую науку, можно отнести обратную связь. Обратная связь (ОС) — процесс передачи выходного сигнала обратно на вход, при котором частично погашается или увеличивается входной сигнал, родилась в недрах физики и хорошо изучена теоретически. ОС может быть амплитудно-зависимой, когда коэффициент ОС зависит от величины выходного сигнала. При частотно-зависимой ОС коэффициент усиления будет зависеть от частоты. Знак ОС определяется целью ее применения. Положительная ОС используется для создания генераторов, отрицательная ОС уменьшает коэффициент усиления, но при этом улучшает другие параметры, например устраняет искажения и нелинейность, сглаживает частотную характеристику, делает поведение схемы предсказуемым. Чем глубже отрицательная ОС, тем меньше внешние параметры влияют на работу схемы. Все эти утверждения математически рассчитаны и экспериментально многократно подтверждены.

Пример работы механизма ОС — слежение антенны радиолокационной станции за движущимся над поверхностью Земли объектом. Угол между диаграммой направленности антенны и направлением на движущийся объект, отслеживаемый посредством датчиков, является управляющим сигналом. Поступая на поворотное устройство антенны, этот сигнал заставляет ее поворачиваться вслед за движением объекта.

Примеры ОС были обнаружены в окружающем мире, философски осмыслены и обобщены. Сделан вывод о том, что замкнутая система может успешно функционировать только при наличии в ней механизма приспособления к изменению условий этого функционирования — механизма ОС. Наличие ОС — объективное условие существования живой природы, любого саморегулирующегося образования, каким является человеческое общество. Экономический рынок регулируется механизмом ОС. Движение к балансу между спросом и предложением, материальное стимулирование, изменение налогов и тарифов — это работа ОС в экономической практике.

Рассмотрим некоторые черты ОС, а также обстоятельства, которые следует учитывать при ее изучении и использовании.

1. *ОС как элемент управления.* ОС является субъектом управления или элементом управления, или регулятором, задача которого, по словам Норберта Винера, «состоит в управлении механической тенденцией к дезорганизации». Работа ОС представлена на рис. 31.1. Работа объекта управления (предприятия, учебного заведения, железной дороги) имеет своим результатом материально или интеллектуально значимый выходной продукт (продукцию предприятия, новые знания, перемещения как часть производственного процесса). Вырабатывается управляющий сигнал, поступающий на исполнительный орган, регулирующий как входное воздействие (предположим, закупку сырья), так и порядок работы объекта управления. Результатом является переход объекта управления в новое состояние, что отражается на структуре выходного продукта.

Рис. 31.1

2. *Замкнутая цепь — необходимое условие существования ОС.* Объект и средства ОС должны быть замкнутой системой. Управление существует только там, где есть замкнутый на объект управления канал связи, по которому передается сигнал управления, т.е. существует замкнутая цепь. Разрыв кольца зависимости приводит к параличу управления.

3. *Средства связи.* ОС классифицируется в соответствии с природой субъекта управления, посредством которого она осуществля-

ется. На механический или электрический элемент управления поступает механический, электрический или оптический сигнал. В аналитический отдел отдела управления приходит информация в виде материалов по итогам хозяйственной деятельности, статистической и бухгалтерской отчетности, материалов учета и контроля, докладов, сообщений, отчетов работников об их деятельности. В результате анализа вырабатывается управляющий сигнал.

4. *Разнообразие каналов. Ручное и автоматическое управление.* Управляющий сигнал ОС может быть электрическим сигналом, силовым воздействием, звуковым или письменным сообщением. Управление ОС может передаваться вручную после принятия соответствующего решения или автоматически — механической, оптической или электрической системой управления.

5. *Направленность.* Обратное воздействие может происходить в двух противоположных направлениях. Связь, в которой протекает нарастающий процесс обратного воздействия, называют положительной. При включении положительной ОС система переходит в новое состояние, которое может быть устойчивым или приводить к возбуждению, взрыву, разрушению системы. Вливание денежных средств в экономику есть положительная ОС, осуществляемая регулятором-государством.

Действие ОС непосредственно связано и с конкуренцией. Если оказалось, что фирма получила преимущества по сравнению с другими, имея патент, купив современное оборудование, выпустив на рынок новый, хорошо разрекламированный товар, естественный отбор приведет к увеличению прибыли фирмы. Ориентируясь на добавочную прибыль, фирма расширит свое производство, привлечет дополнительную рабочую силу, в то время как фирмы-консерваторы вынуждены экономить сырье и материалы, сокращать число занятых на своих предприятиях. В дальнейшем они подстраиваются к движению фирмы-новатора, в противном случае исчезают. Рынок переходит в новое устойчивое равновесие. Когда это происходит в целой экономической системе, она поднимается на более высокий уровень эволюции. Затем цикл повторяется.

ОС, приводящая к депрессии процесса, называется отрицательной ОС. Экономика использует отрицательную ОС, которая приводит к балансу между спросом и предложением, к равновесию между ценами, назначаемыми продавцами, и контролируемыми ими долями рынка. Происходит стабилизация экономических отношений. Любые значительные изменения компенсируются.

Изъятие денежных средств в виде налогов, их повышения, перевод средств за границу — также явления отрицательной ОС. Введенный в 1991 г. налог на добавленную стоимость в размере 28% привел к замедлению реализации продукции на рынке из-за удорожания товаров и услуг. Трудности в реализации привели к уменьшению выручки и паузам в закупке сырья и материалов для производства товаров. Результат — увеличение налоговых неплатежей. Не было соблюдено условие осторожного экспериментирования при поиске оптимальности в управлении. В результате субъекту управления, т.е. органам государственного регулирования, пришлось отказаться от первоначальной цифры, снизив налог до 20%.

6. *Внешняя и внутренняя ОС.* Если ОС осуществляется подключением внешних элементов, цепей, органов, она называется внешней ОС. При функционировании ОС в рамках отдельного устройства, системы, предприятия ОС остается внутренней.

7. *Функциональность.* В одних системах роль ОС является определяющей. Без нее замкнутая система приходит к разрушению, перестает функционировать. В других — возникает непреднамеренно и носит паразитный характер. Например, гул в телефонной трубке или радиоприемнике — пример действия возникшей паразитной ОС.

8. *Многозначность.* ОС в замкнутой системе часто вырабатывается многими элементами управления, включенными последовательно или параллельно. Сигнал управления, отправленный начальным звеном цепи управления, может проходить последовательно через корректирующие цепи, например надзорные органы, перед поступлением на исполнительный орган. Возможно вырабатывание сигналов несколькими элементами управления с поступлением их на исполнительный элемент параллельно. Не исключена в таком случае нулевая реакция на управление. Наиболее распространенным явлением становится противоречивость указаний разных звеньев управления.

9. *Линейность.* Если элемент управления содержит только пропорциональные звенья, ее выходное воздействие пропорционально входному и является линейным. При наличии, например, сотрудника-управленца в составе элемента управления ОС, вероятно, станет нелинейной или даже дискретной, если требуется время для сбора и осмысления информации.

10. *Статичность и динамичность.* Если управляющий сигнал вырабатывается в зависимости от величины выходного сигнала, ОС можно назвать статической. В связях без статизма не участвует величина выходного сигнала, важны лишь его изменения. И чем они

больше, тем выше значение сигнала управления. На языке эконофизики это означает, что фигурируют производные выходного сигнала, стремящиеся к нулю с окончанием переходных процессов. Например, текущее управление процессом производства или учебы или движением поездов на железной дороге осуществляется имеющимися звеньями управления, начальниками цехов, прорабами, учебной частью, диспетчерами. Но если процесс неожиданно остановился, может понадобиться вмешательство стратегического звена управления — руководителя предприятия — для возобновления работы.

11. *Управление по отклонениям.* Управляющие воздействия часто ставят своей задачей обеспечить следование объекта управления определенной программе действий. По каналу ОС субъект управления получает информацию о возможных отклонениях от установленной программы и вырабатывает управляющие воздействия, подавляющие, ликвидирующие такие отклонения.

12. *Запаздывание.* Регулятор ОС может мгновенно передавать сигнал управления на исполнительный орган, но возможно и запаздывание в функционировании ОС. Для вырабатывания сигнала управления иногда требуется накопление информации, ее анализ, наконец, прохождение очередного полного отчетного периода. Запаздывание характерно для системы налогов. После введения определенного налога должно пройти время, чтобы это отразилось на хозяйственной деятельности предприятия и его финансовом состоянии.

13. *Оптимальность управления.* Управление является оптимальным, если оно обеспечивает перевод управляемой системы из исходного состояния в желаемое за минимально возможное время, с наименьшими затратами при одновременном соблюдении ограничительных условий, т.е. признанных законов, запретов, общепринятых морально-этических правил и норм.

14. *Выбор устойчивого состояния посредством ОС.* Регулируя ОС посредством налогов, административного ресурса, других средств ОС, можно перевести экономическую систему, отрасль экономики, отдельное предприятие в новое устойчивое состояние более высокой продуктивности. Желательно правильно рассчитать время для перехода из одного стабильного состояния в другое. Экономическая теория может способствовать определению таких состояний и моментов, помогая в выборе оптимальных действий.

15. *Колебания при воздействии ОС.* Наличие ОС может стать не только причиной усиления, ослабления, нового устойчивого со-

стояния или взрыва системы, но и причиной возникновения в ней колебаний. Неоптимальная величина налога на имущество предприятий, по которому налогооблагаемая база складывается из остатков активов по состоянию на дату предоставления отчетности, приводит к тому, что на определенный период эти остатки занижаются. Имея недостаток сырья, готовой продукции, предприятие вынуждено снижать загрузку производства, уменьшать временно хозяйственный оборот. Периодические сокращения или колебания объемов реализации приводят к неравномерному начислению других налогов: НДС, налога на прибыль.

16. *Принцип детерминированности*. Рассчитывая ОС как элемент управления, следует проводить максимально точные расчеты петли ОС. В противном случае накапливающиеся со временем погрешности могут разорвать петлю. Управляя экономическими процессами, надо прогнозировать последствия принимаемых решений в той степени, в которой это представляется возможным.

17. *Осторожное экспериментирование*. Очень важно во избежание негативных последствий для экономической деятельности государства осуществлять управление ОС так, чтобы их влияние на субъект управления не приводило к его деформации. При невозможности точного расчета системы с ОС в гуманитарных областях, например в финансовой системе, использование сплошного и непрерывного экспериментирования чрезвычайно опасно. Велика разрушительная сила неудачного эксперимента. Только локальный, кратковременный, с возможностью возврата в исходное состояние эксперимент путем проб и ошибок может помочь в нахождении такого элемента управления ОС, который выполнит поставленную задачу. В этой связи, перед тем как осуществлять управление ОС, особое внимание следует обратить на сигналы, посылаемые управляемой системой, особенно когда обратные зависимости существенно опаздывают, а качества, накопленные в них, могут проявиться не сразу, а по истечении некоторого времени.

Механизм ОС имеет непосредственное отношение к производству на предприятии. Прибыль вкладывается в производство. Это обеспечивает условия дальнейшего роста прибыли. Ниже предлагается экономико-математическая модель работы предприятия при наличии ОС. При превышении выпуска над запланированным часть прибыли предприятия вкладывается в расширение производства, в дополнительную закупку ресурсов. Формируется ОС между выпуском и инвестированием в ресурсы, которая характеризуется величиной коэффициента ОС. Выбирая коэффициент в разумных пределах, можно добиться перехода предприятия в новое экономически устойчивое состояние с получением более высокой прибыли. Работа предприятия всегда производится в оптимальном режиме. Экономико-математическая модель предполагает известной производственную функцию предприятия. Известны также стоимости ресурсов и нормативные средства, отпущенные на их закупку.

31.3. УСЛОВНЫЙ ЭКСТРЕМУМ ПРОИЗВОДСТВЕННОЙ ФУНКЦИИ С ПЕРЕМЕННЫМ БЮДЖЕТНЫМ ОГРАНИЧЕНИЕМ

Перейдем к рассмотрению экономико-математической модели, описывающей условный экстремум производственной функции с переменным бюджетным ограничением. Исследование функции на максимум при условии, что переменные связаны между собой некоторым соотношением, проводится при постоянном бюджетном ограничении. Обычно задача формулируется следующим образом: найти экстремум функции $z = f(x_1, x_2)$, если переменные x_1 и x_2 удовлетворяют ограничению типа уравнения $g(x_1, x_2) = 0$:

$$\begin{cases} z = f(x_1, x_2) \to \text{extr}, \\ g(x_1, x_2) = 0. \end{cases}$$

В экономико-математическом моделировании предполагается, что $z = f(x_1, x_2)$ — это, например, производственная функция предприятия, отвечающая за выпуск продукта, тогда переменные x_1, x_2 — сырьевые ресурсы, закупаемые по ценам p_1 и p_2, на сумму I_0. Следовательно,

$$p_1x_1 + p_2x_2 = I_0.$$

Эта задача, можно сказать, из советского прошлого. Предприятие настраивается, казалось бы, на максимальный выпуск продукции (находится экстремум) при ограничении по затратам, но не может управлять инвестированием средств на закупку ресурсов. Бюджетное ограничение раз и навсегда задано и равно I_0.

Более реалистичным представляется вариант, при котором в бюджетном ограничении кроме постоянной составляющей имеется еще переменная вида $k \cdot Z$, позволяющая закупать больше ресурсов при росте выпуска и продаж продукта ($k > 0$) и уменьшать выпуск при перепроизводстве ($k < 0$). Бюджетное ограничение становится управляемым. Представим его в виде линейной функции по Z:

$$I = I_0 + k \cdot Z.$$

Тогда задача на условный экстремум с переменными затратами на закупку ресурсов I формулируется так:

$$\begin{cases} Z = f(x_1, x_2) - I_0 \to \text{extr}, \\ p_1x_1 + p_2x_2 = I_0 + k \cdot Z. \end{cases}$$

Функция $Z = f(x_1, x_2) - I_0$ есть доход от продажи произведенного продукта $z = f(x_1, x_2)$ за вычетом установленных нормами средств на закупку ресурсов I_0. Из средств Z производятся отчисления на амортизацию, аренду, налоги, коммунальные услуги и т.д. При наличии резервов увеличения производства продукта часть этих средств $k \cdot Z$, где $k < 1$, направляется на дополнительную закупку ресурсов. Если $k > 0$, возникает положительная ОС между выпуском и инвестированием в ресурсы. При $k < 0$ предприятие уменьшает величину закупок ресурсов, тем самым уменьшая выпуск. Возникает отрицательная обратная зависимость.

Задача на условный экстремум с переменным инвестированием в ресурсы сводится к исследованию на максимум переменной Z в системе функций

$$\begin{cases} Z = f(x_1, x_2) - I_0 \to \text{extr}, \\ Z = \dfrac{p_1x_1 + p_2x_2 - I_0}{k}. \end{cases}$$

С целью решения задачи составляется функция Лагранжа:

$$L(x_1, x_2, \lambda_1, \lambda_2) = \lambda_1(f(x_1, x_2) - I_0) + \lambda_2 \frac{p_1x_1 + p_2x_2 - I_0}{k}.$$

Находятся стационарные точки $M_0(x_1^0, x_2^0, \lambda_1^0, \lambda_2^0)$ из системы (необходимые условия экстремума):

$$\begin{cases} L'_{x_1} = \lambda_1 f_{x_1} + \lambda_2 \dfrac{p_1}{k} = 0, \\ L'_{x_2} = \lambda_1 f_{x_2} + \lambda_2 \dfrac{p_2}{k} = 0, \\ f(x_1, x_2) - I_0 = \dfrac{p_1x_1 + p_2x_2 - I_0}{k}, \\ \lambda_1 + \lambda_2 = 1. \end{cases} \tag{31.1}$$

Берутся вторые производные функции Лагранжа и строится окаймленный гессиан в точке M_0:

$$|H| = \begin{vmatrix} 0 & f'_{x_1} - \frac{p_1}{k} & f'_{x_2} - \frac{p_2}{k} \\ f'_{x_1} - \frac{p_1}{k} & \lambda_1 f''_{x_1x_1} & \lambda_1 f''_{x_1x_2} \\ f'_{x_2} - \frac{p_2}{k} & \lambda_1 f''_{x_2x_1} & \lambda_1 f''_{x_2x_2} \end{vmatrix}. \quad (31.2)$$

Определение его знака является достаточным условием обнаружения экстремума. Если $|H| > 0$, в точке M_0 достигается максимум, при $|H| < 0$ имеем минимум.

Выпуск при увеличении закупок ресурсов. Рассмотрим стандартную количественную задачу с производственной функцией Кобба—Дугласа (функцией *KD*) $Z = f(x_1, x_2) = 3x_1^{\frac{1}{2}}x_2^{\frac{1}{2}}$ и переменным инвестированием в ресурсы в виде линейной функции. Существует единственный продукт — выпуск. Доход и выпуск совпадают. Выпуск производится при помощи двух факторов производства x_1 и x_2. После отчислений установленных нормами средств на закупку ресурсов (примем $I_0 = 1$) часть выпуска потребляется $\left(\frac{1}{2}Z\right)$, остаток $\frac{1}{2}Z$ сберегается и инвестируется в ресурсы. Затраты на ресурсы имеют тем самым постоянную составляющую и переменную, зависящую от размеров дохода: $I = 1 + \frac{1}{2}Z$. Система уравнений в задаче на условный экстремум с переменными затратами на ресурсы приобретает вид

$$\begin{cases} Z = 3x_1^{\frac{1}{2}}x_2^{\frac{1}{2}} - 1 \to \text{extr}, \\ \frac{1}{2}x_1 + 2x_2 = 1 + \frac{1}{2}Z. \end{cases}$$

Переходя к решению задачи, составим функцию Лагранжа:

$$L(x_1, x_2, \lambda_1, \lambda_2) = \lambda_1\left(3x_1^{\frac{1}{2}}x_2^{\frac{1}{2}} - 1\right) + \lambda_2(x_1 + 4x_2 - 2).$$

Рассмотрим систему уравнений из четырех неизвестных x_1, x_2, λ_1, λ_2, содержащую условия первого порядка:

$$\begin{cases} L'_{x_1} = \dfrac{3}{2}\lambda_1 x_1^{-\frac{1}{2}} x_2^{\frac{1}{2}} + \lambda_2 = 0, \\ L'_{x_2} = \dfrac{3}{2}\lambda_1 x_1^{\frac{1}{2}} x_2^{-\frac{1}{2}} + 4\lambda_2 = 0, \\ 3x_1^{\frac{1}{2}} x_2^{\frac{1}{2}} - 1 = x_1 + 4x_2 - 2, \\ \lambda_1 + \lambda_2 = 1. \end{cases}$$

Ее решение есть стационарная точка

$$M_0(x_1^0, x_2^0, \lambda_1^0, \lambda_2^0) = M_0\left(2, \frac{1}{2}, 4, -3\right).$$

Найдя вторые производные функции Лагранжа, построим окаймленный гессиан и вычислим его в точке M_0:

$$|H| = \begin{vmatrix} 0 & \dfrac{3}{2}\lambda_1 x_1^{-\frac{1}{2}} x_2^{\frac{1}{2}} - 1 & \dfrac{3}{2}\lambda_1 x_1^{\frac{1}{2}} x_2^{-\frac{1}{2}} - 4 \\ \dfrac{3}{2}\lambda_1 x_1^{-\frac{1}{2}} x_2^{\frac{1}{2}} - 1 & -\dfrac{3}{4}\lambda_1 x_1^{-\frac{3}{2}} x_2^{\frac{1}{2}} & \dfrac{3}{4}\lambda_1 x_1^{-\frac{1}{2}} x_2^{-\frac{1}{2}} \\ \dfrac{3}{2}\lambda_1 x_1^{\frac{1}{2}} x_2^{-\frac{1}{2}} - 4 & \dfrac{3}{4}\lambda_1 x_1^{-\frac{1}{2}} x_2^{-\frac{1}{2}} & -\dfrac{3}{4}\lambda_1 x_1^{\frac{1}{2}} x_2^{-\frac{3}{2}} \end{vmatrix} =$$

$$= \begin{vmatrix} 0 & 2 & 8 \\ 2 & -\dfrac{3}{4} & 3 \\ 8 & 3 & -12 \end{vmatrix} = 192 > 0.$$

Следовательно, при наличии коэффициента положительной ОС $k = +\dfrac{1}{2}$ в точке $M_0\left(2, \dfrac{1}{2}, 4, -3\right)$ достигается максимум

$$Z_{\max}^{+\text{обр.св}}\left(2, \frac{1}{2}\right) = 2.$$

Для сравнения приведем решение стандартной задачи

$$\begin{cases} Z = 3x_1^{\frac{1}{2}} x_2^{\frac{1}{2}} - 1 \to \text{extr}, \\ \dfrac{1}{2}x_1 + 2x_2 = 1 \end{cases}$$

на условный экстремум без ОС.

Функция Лагранжа имеет вид

$$L(x_1, x_2, \lambda) = \left(3x_1^{\frac{1}{2}}x_2^{\frac{1}{2}} - 1\right) + \lambda(x_1 + 4x_2 - 2).$$

Условия первого порядка

$$\begin{cases} L'_{x_1} = \dfrac{3}{2}x_1^{-\frac{1}{2}}x_2^{\frac{1}{2}} + \lambda = 0, \\ L'_{x_2} = \dfrac{3}{2}x_1^{\frac{1}{2}}x_2^{-\frac{1}{2}} + 4\lambda = 0, \\ x_1 + 4x_2 = 2 \end{cases}$$

дают стационарную точку $M_1(x_1^0, x_2^0, \lambda^0) = M_1\left(1, \frac{1}{4}, -\frac{3}{4}\right)$, в которой достигается стандартный условный максимум

$$Z_{\max}^{\text{станд}}\left(1, \frac{1}{4}\right) = \frac{1}{2}.$$

На рис. 31.2 приведен график функции KD, сдвинутый по оси Z на единицу вниз, а также графики функций, описывающих постоянные и изменяющиеся с доходом затраты на ресурсы. Белыми стрелками указаны условные максимумы при постоянном, а также

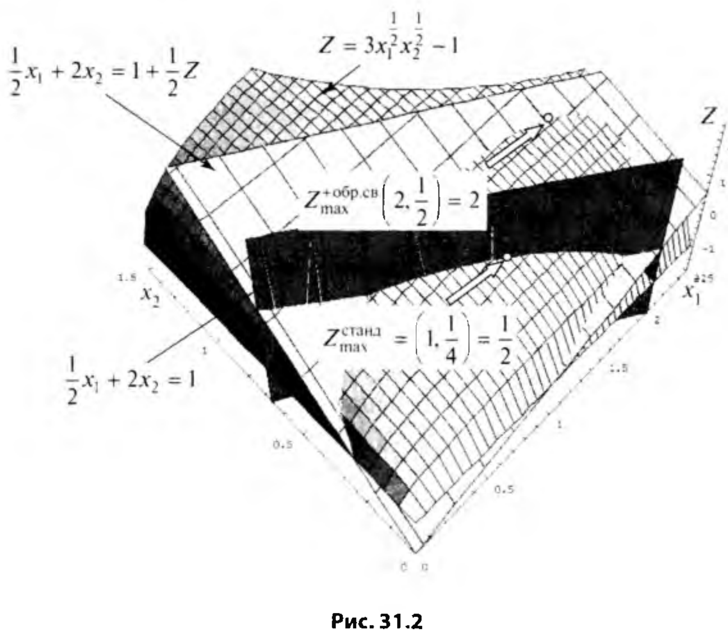

Рис. 31.2

переменном, зависящем от выпуска ограничении на закупку ресурсов.

Выпуск при уменьшении закупок ресурсов. Аналогично решается задача на условный экстремум с отрицательной ОС $k = -\frac{1}{2}$ при тех же начальных условиях:

$$\begin{cases} Z = 3x_1^{\frac{1}{2}}x_2^{\frac{1}{2}} - 1 \to \text{extr}, \\ \frac{1}{2}x_1 + 2x_2 = 1 - \frac{1}{2}Z. \end{cases}$$

Точно так же используем функцию Лагранжа с двумя множителями, что приводит к условиям первого порядка в виде системы из четырех уравнений с четырьмя переменными. Ее решение — стационарная точка $M_2(x_1^0, x_2^0, \lambda_1^0, \lambda_2^0) = M_2\left(\frac{6}{7}, \frac{3}{14}, \frac{4}{7}, \frac{3}{7}\right)$, в которой имеется условный максимум $Z_{\max}^{-\text{обр.св}}\left(\frac{6}{7}, \frac{3}{14}\right) = \frac{2}{7}$.

Метод решения экономико-математических задач с переменными ограничениями на закупку ресурсов или, говоря языком эконофизики, с положительной и отрицательной ОС позволяет расширить круг рассчитываемых экономических моделей, учитывая взаимное влияние параметров задачи, что на практике часто имеет место. В экономической теории рассматриваются бесконечно малые вариации параметров, что вызвано использованием дифференциалов. ОС позволяет преодолеть это затруднение. Регулировкой коэффициента ОС можно добиться любых изменений параметра, оставаясь в оптимальной точке. В частности, получить решение задачи при значительном инвестировании предприятием средств в ресурсы или исследовать целевую функцию потребителя при увеличении его дохода, связанном, например, с повышением по службе, со сменой им места работы.

Аналитическое решение третьего уравнения из системы (31.1) возможно не при любом виде производственной функции, в том числе и функции *KD* с $\alpha > 0$, $\beta > 0$. Графическое решение уравнения, где в качестве производственной функции предприятия использована функция *KD*

$$Z = ax_1^{\alpha}\left(\frac{I_0 + kZ - p_1x_1}{p_2}\right)^{\beta} - I_0,$$

позволяет преодолеть это затруднение и найти зависимость $Z = Z(x_1)$ при условии $Z \geq 0$. На рис. 31.3 приведено семейство кривых зависимости выпуска Z от количества используемого ресурса x_1 при различных положительных и отрицательных значениях коэффициента ОС. Для расчета использованы данные, приведенные на рис. 31.3. Зависимость $Z = Z(x_1)$ при $k = 0$ получена из решения задачи на стандартный условный экстремум без ОС. Зная потребность в продукции предприятия Z_0, обладая ограниченными средствами I_0, можно по графику определить оптимальный вариант использования ресурсов x_1^0, x_2^0, при котором будет достигнут максимум выпуска $Z_0 = Z_{max}$ требуемой продукции. Например, потребность в продукте предприятия в данном регионе составляет $Z_0 = 0{,}90$ у.е. Имея ограничения на закупку ресурсов в размере $I_0 = 1$ у.д.е., по графику находим, что $k = 0{,}3$, $x_1^0 = 1{,}25$ у.е. Из уравнения связи с $p_1 = \frac{1}{2}$, $p_2 = 2$ вычисляем $x_2^0 = 1{,}54$ у.е. Таким образом, предприятие будет работать в оптимальном режиме, инвестируя в ресурсы $I_0 = 1$ у.д.е., приобретая их в объеме $x_1^0 = 1{,}25$ у.е., $x_2^0 = 1{,}54$ у.е. и имея нераспределенную прибыль $(Z + I_0) - (I_0 + kZ) = (1 - k)Z = 0{,}7 \cdot 0{,}9 = 0{,}63$ у.д.е.

Семейство кривых удобно использовать и при практическом решении другой задачи: оптимизировать выпуск продукции при ограничении на использование ресурса x_1. Аналогично строится зависимость $Z = Z(x_2)$.

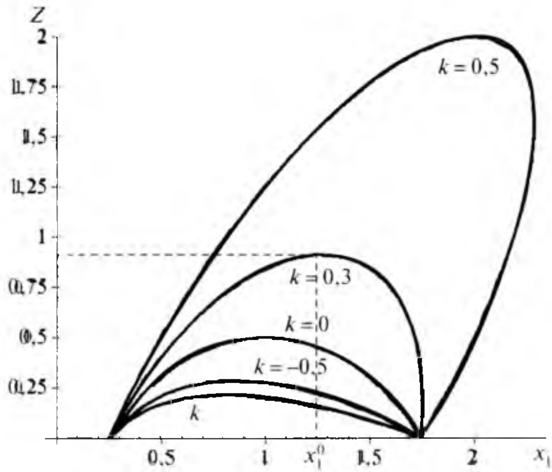

Рис. 31.3

Затраты на дополнительную закупку ресурсов не приводят к потере прибыли. Найдем величину

$$\frac{\text{Доход} - \text{Расход на ресурсы}}{\text{Расход на ресурсы}} \cdot 100\% = \frac{(Z + I_0) - (I_0 + kZ)}{I_0 + kZ} \cdot 100\% =$$

$$= \frac{(1-k)Z}{I_0 + kZ} \cdot 100\%,$$

пренебрегая другими расходами. Данные возьмем из рассмотренной выше модели. Величина прибыли при разных значениях коэффициента ОС представлена в табл. 31.2.

Таблица 31.2

Коэффициент ОС k	$-0,5$	0	$0,5$
$\frac{\text{Доход} - \text{Расход на ресурсы}}{\text{Расход на ресурсы}} \cdot 100\%$	$\frac{\left(1+\frac{1}{2}\right)\cdot\frac{2}{7}}{1-\frac{1}{2}\cdot\frac{2}{7}} \times 100\% = 50\%$	$\frac{\frac{1}{2}}{1} \cdot 100\% = 50\%$	$\frac{\left(1-\frac{1}{2}\right)\cdot 2}{1+\frac{1}{2}\cdot 2} \times 100\% = 50\%$
Прибыль в абс. единицах	$\frac{3}{7}$	$\frac{1}{2}$	1

Превышение доходов над расходами на приобретение ресурсов оказалось одно и то же и равно 50%. Можно показать в общем виде, что при выборе оптимального режима работы предприятия эта величина при любом допустимом значении k будет константой. Абсолютная прибыль с ростом выпуска при оптимизации производства будет расти.

Рассмотренная экономико-математическая модель оптимизации производства позволяет построить *алгоритм* оптимального запуска производственных мощностей. На рис. 31.4 построены кривые зависимости выпуска от количества использованного сырьевого ресурса x_1 при различных значениях параметров I_0 и k. Запуск производства и технологическая отладка проводятся при небольших значениях I_0 ($I_0 = 1$), достаточных для получения первоначальной прибыли (кривая 1). Работа предприятия оптимизируется в низкопродуктивном состоянии $Z^1_{1\text{опт}}$ при оптимальном значении $x^1_{1\text{опт}}$ и, соответственно, $x^1_{2\text{опт}}$. Когда предприятие ритмично заработало в оптимальном режиме (точка A на кривой 1) и выпускает кондиционную готовую продукцию, имеющую спрос на рынке, появилась прибыль, включается обратная связь. Дискретно повышая величину коэффи-

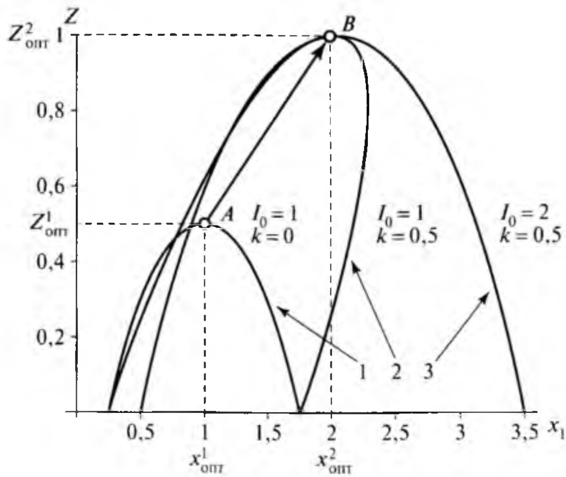

Рис. 31.4

циента обратной связи $k = 0, ..., 0{,}1, ..., 0{,}5, ... < 1$, направляя часть прибыли на дополнительное к I_0 инвестирование в ресурсы, управляющее звено переводит работу предприятия в новое высокопродуктивное состояние $Z^2_{опт}$, характеризующееся полной загрузкой мощности (точка B на кривой 2). Дискретность обусловлена цикличностью работы предприятия: сырье — товар — деньги. Сырьевой ресурс x_1 начинает потребляться в количестве $x^2_{1опт}$, ресурс x_2 — в количестве $x^2_{2опт}$ (в нашей модели это соответствует величине $k = 0{,}5$). Движение от низкопродуктивного режима работы предприятия к высокопродуктивному указано вектором $\overrightarrow{AB}$. Альтернатива этому алгоритму — выделение на закупку сырья сразу значительных денежных средств в размере $I_0 = 2$ для закупки сырья в количестве $x^2_{1опт}$, $x^2_{2опт}$ с последующим его консервированием до выхода предприятия на проектную мощность (кривая 3).

31.4. ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ИСПОЛЬЗУЮЩИЕ УСЛОВНЫЙ ЭКСТРЕМУМ ФУНКЦИИ *KD*

Наличие ОС позволяет моделировать процессы, в которых функция $KD\, f(x_1, x_2) = ax_1^{\alpha}x_2^{\beta}$ исследуется на *условный минимум*. Анализ показывает, что необходимым условием минимума является возрастающий эффект от масштаба, т.е. $\alpha + \beta > 1$. Эмпирические оценки степени отдачи от масштаба производства продемонстрировали реальность этого условия.

Дополнительное очевидное в экономико-математических задачах условие $I_0 > 0$ в уравнении бюджетного ограничения приводит к появлению и условного минимума, и условного максимума у функции *KD*. Варьируя параметры задачи, можно сблизить условный минимум и максимум вплоть до их слияния с образованием седловой точки. Этот вариант имеет место при выполнении условий

$$\begin{cases} \alpha + \beta > 1, \\ I_0 > 0, \\ \dfrac{(ax_1^{\alpha}x_2^{\beta})_{x_1}}{p_1/k} = \dfrac{(ax_1^{\alpha}x_2^{\beta})_{x_2}}{p_2/k} = 1. \end{cases}$$

Уравнения множества точек, принадлежащих обеим функциям, примут простой вид:

$$z = \frac{p_1}{\alpha k}x_1, \qquad z = \frac{p_2}{\beta k}x_2.$$

Точка пересечения линий и будет седловой точкой.

В качестве иллюстрации решена количественная задача с нахождением обоих экстремумов:

$$\begin{cases} z = 5x_1 \cdot x_2 \to \text{extr}, \\ 2x_1 + 3x_2 = 2 + \dfrac{1}{3}z. \end{cases}$$

Решение выявило две стационарные точки $M(x_1^0, x_2^0, \lambda_1^0, \lambda_2^0)$, а именно: $M_1\left(\frac{3}{5}, \frac{2}{5}, 2, -1\right)$ и $M_2\left(3, 2, -\frac{2}{3}, \frac{5}{3}\right)$. В одной из них достигается условный максимум функции *KD* $Z_{\max}\left(\frac{3}{5}, \frac{2}{5}\right) = \frac{6}{5}$, в другой — функция принимает условный минимум $Z_{\min}(3, 2) = 30$. На рис. 31.5 в координатах (x_1, x_2, z) представлены трехмерные поверхности, описываемые функцией *KD* и функцией переменных затрат на закупку ресурсов.

Модели, использующие условный экстремум с ОС, могут быть применены в области экологии, медицины, социологии. Например, добыча и использование ресурсов требуют такой организации производства, при которой вредное воздействие на окружающую среду, описываемое функцией модели, минимально.

В заключение отметим, что физические идеи либо находят экспериментальное подтверждение посредством целенаправленных

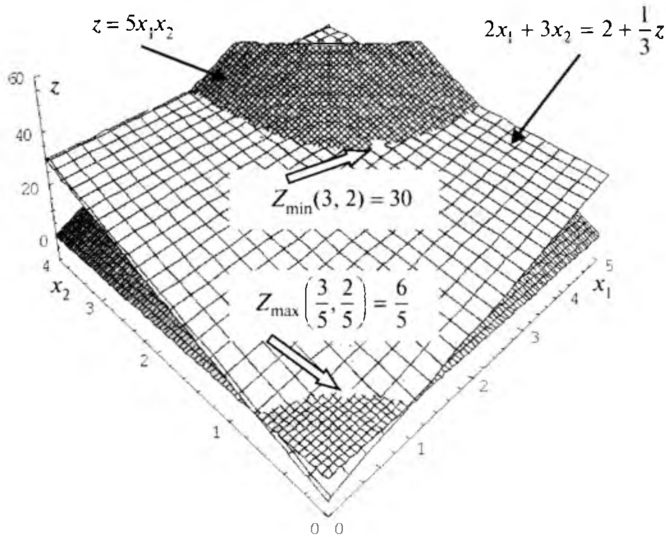

Рис. 31.5

усилий ученых-физиков, экспериментаторов, что воплощается затем в законах, либо отвергаются. Физические понятия позволяют легко формулировать идеи и законы. В экономике роль экспериментаторов играют чаще всего правительства. Групповые интересы, противоречивые законы и постановления, непоследовательные действия делают эксперимент «грязным» с точки зрения науки, поэтому подтверждение или опровержение теоретических разработок становится проблематичным.

ЛИТЕРАТУРА

1. *Акулич И.Л.* Математическое программирование в примерах и задачах. — М.: Высшая школа, 1986.
2. *Аронович А.Б., Афанасьев М.Ю., Суворов Б.П.* Сборник задач по исследованию операций. — М.: Изд-во Московского ун-та, 1997.
3. *Бабайцев В.А., Браилов А.В., Солодовников А.С.* Математика в экономике. — М.: Финансы и статистика, 1998.
4. *Гасс С.* Линейное программирование. — М.: Физматгиз, 1961.
5. Исследование операций в экономике / Под ред. Н.Ш. Кремера. — М.: ЮНИТИ-ДАНА, 2006.
6. *Калихман И.Л.* Сборник задач по математическому программированию. — М.: Высшая школа, 1975.
7. *Красс М.С., Чупрынов Б.П.* Математика в экономике. Математические модели и методы. — М.: Финансы и статистика, 2007.
8. *Федосеев В.В., Эриашвили Н.Д.* Экономико-математические модели и методы в маркетинге. — М.: Финстатпром, 1996.
9. *Шикин Е.В., Чхартишвили А.Г.* Математические методы и модели в управлении. — М.: Дело, 2000.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

ПРИЛОЖЕНИЕ

1. ТАБЛИЦА ЗНАЧЕНИЙ ФУНКЦИИ

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014

X	Десятые доли x				
	0	2	4	6	8
4	0,0001338	0000589	0000249	0000101	0000040
5	0,0000015				

2. ТАБЛИЦА ЗНАЧЕНИЙ ФУНКЦИИ ЛАПЛАСА

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$
0,00	0,0000	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,0040	0,33	0,1293	0,65	0,2422	0,97	0,3340
0,02	0,0080	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,0120	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,0160	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,1480	0,70	0,2580	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,10	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,1700	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643

x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$
0,15	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,80	0,2881	1,12	0,3686
0,17	0.0675	0,49	0,1879	0,81	0,2910	1,13	0,3708
0,18	0,0714	0.50	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,1950	0,83	0,2967	1,15	0,3749
0,20	0,0793	0.52	0,1985	0,84	0,2995	1,16	0,3770
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,3790
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,3810
0,23	0,0910	0,55	0,2088	0,87	0,3078	1,19	0,3830
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,2190	0,90	0,3159	1,22	0,3883
0,27	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,28	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,30	0,1179	0,62	0,2324	0,94	0,3264		
0,31	0,1217	0,63	0,2357	0,95	0,3289		
1,26	0,3962	1,59	0,4441	1,92	0,4726	2,50	0,4938
1,27	0,3980	1,60	0,4452	1,93	0,4732	2,52	0,4941
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,54	0,4945
1,29	04015	1,62	0,4474	1,95	0,4744	2,56	0,4948
1,30	0,4032	1,63	0,4484	1,96	0,4750	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,60	0,4953
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,62	0,4956
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,64	0,4959
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,66	0,4961
1.35	0.4115	1,68	0,4535	2,02	0,4783	2,68	0,4963
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,70	0,4965
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,72	0,4967
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,74	0,4969
1,39	0,4177	1,72	0.4573	2,10	0,4821	2,76	0,4971
1,40	0,4192	1,73	0,4582	2,12	0,4830	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,80	0,4974
1,42	0,4222	1,75	0,4599	2,16	0,4846	2,82	0,4976
1,43	0.4236	1,76	0,4608	2,18	0,4854	2,84	0,4977
1,44	0,4251	1,77	0,4616	2,20	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,4980

x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,90	0,4981
1,47	0,4292	1,80	0,4641	2,26	0,4881	2,92	0,4982
1,48	0,4306	1,81	0,4649	2,28	0,4887	2,94	0,4984
1,49	0,4319	1,82	0,4656	2,30	0,4893	2,96	0,4985
1,50	0,4332	1,83	0,4664	2,32	0,4898	2,98	0,4986
1,51	0,4345	1,84	0,4671	2,34	0,4904	3,00	0,49865
1,52	0,4357	1,85	0,4678	2,36	0,4909	3,20	0,49931
1,53	0,4370	1,86	0,4686	2,38	0,4913	3,40	0,49966
1,54	0,4382	1,87	0,4693	2,40	0,4918	3,60	0,499841
1,55	0,4394	1,88	0,4699	2,42	0,4922	3,80	0,499928
1,56	0,4406	1,89	0,4706	2,44	0,4927	4,00	0,499968
1,57	0,4418	1,90	0,4713	2,46	0,4931	4,50	0,499997
1,58	0,4429	1,91	0,4719	2,48	0,4934	5,00	0,499997

3. КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ ХИ-КВАДРАТ ПИРСОНА

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	10,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

4. КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)						
	0,2	0,1	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,7	318,3	637,0
2	1,89	2,92	4,30	6,97	9,92	22,33	31,6
3	1,64	2,35	3,18	4,54	5,84	10,22	12,9
4	1,53	2,13	2,78	3,75	4,60	7,17	8,61
5	1,48	2,01	2,57	3,37	4,03	5,89	6,86
6	1,44	1,94	2,45	3,14	3,71	5,21	5,96
7	1,42	1,89	2,36	3,00	3,50	4,79	5,40
8	1,40	1,86	2,31	2,90	3,36	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	4,03	4,44
12	1,36	1,78	2,18	2,68	3,05	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,69	4,01
17	1,33	1,74	2,11	2,57	2,90	3,65	3,95
18	1,33	1,73	2,10	2,55	2,88	3,61	3,92

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)						
	0,2	0,1	0,05	0,02	0,01	0,002	0,001
19	1,33	1,73	2,09	2,54	2,86	3,58	3,88
20	1,33	1,73	2,09	2,53	2,85	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,51	3,79
23	1,32	1,71	2,07	2,50	2.81	3,49	3,77
24	1,32	1,71	2,06	2,49	2,80	3,47	3,74
25	1,32	1,71	2,06	2,49	2,79	3,45	3,72
26	1,32	1,71	2,06	2,48	2,78	3,44	3,71
27	1,31	1,71	2,05	2,47	2,77	3,42	3,69
28	1,31	1,70	2,05	2,46	2,76	3,40	3,66
29	1,31	1,70	2,05	2,46	2,76	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,39	3,65
40	1,30	1,68	2,02	2.42	2,70	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,28	1,65	1,96	2,33	2,58	3,09	3,29
Число степеней свободы k	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)						

5. КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ ФИШЕРА—СНЕДЕКОРА

(k_1 — число степеней свободы в числителе, k_2 — число степеней свободы в знаменателе)

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
Уровень значимости $\alpha = 0,01$												
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99.33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34.12	30.81	29.46	28,71	28.24	27,91	27,67	27,49	27,34	27,23	27.13	27,05
4	21,20	18.00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11.39	10,97	10,67	10,45	10,27	10.15	10,05	9.96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8.10	7,98	7,87	7,79	7,72

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6, 84	6, 71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45
Уровень значимости $\alpha = 0,05$												
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4.95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

6. КРИТИЧЕСКИЕ ТОЧКИ ДЛЯ СТАТИСТИКИ КОЛМОГОРОВА D_n

Объем выборки n	Уровень значимости α			
	0,10	0,05	0,02	0,01
1	0,95	0,98	0,99	0,995
2	0,78	0,84	0,90	0,93
3	0,64	0,71	0,78	0,83
4	0,57	0,62	0,69	0,73
5	0,51	0,56	0,62	0,67
6	0,47	0,52	0,58	0,62
7	0,44	0,48	0,54	0,58
8	0,41	0,45	0,51	0,54
9	0,39	0,43	0,48	0,51
10	0,37	0,41	0,46	0,49
11	0,35	0,39	0,44	0,47
12	0,34	0,38	0,42	0,45
13	0,33	0,36	0,40	0,43
14	0,31	0,35	0,39	0,42
15	0,30	0,34	0,38	0,40
16	0,29	0,33	0,37	0,39
17	0,29	0,32	0,36	0,38
18	0,28	0,31	0,34	0,37
19	0,27	0,30	0,34	0,36
20	0,26	0,29	0,33	0,35

7. КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ КОЛМОГОРОВА

$$Q(\lambda) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2\lambda^2}$$

α	0,10	0,05	0,02	0,01
$\lambda_{кр}$	1,23	1,36	1,52	1,63

8. ЗНАЧЕНИЯ ФУНКЦИИ ПУАССОНА

$$P(m, \lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$$

m	λ									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,90484	0,81873	0,74082	0,67032	0,60653	0,54881	49659	0,44033	0,40657	0,36788
1	0,09048	0,16375	0,22225	0,26813	0,30327	0,32929	0,34761	0,35946	0,36591	0,36788
2	0,00452	0,01637	0,03334	0,05363	0,07582	0,09879	0,12166	0,14379	0,16466	0,18394
3	0,00015	0,00109	0,00333	0,00715	0,01264	0,01976	0,02839	0,03834	0,04940	0,06131
4		0,00005	0,000250	0,00072	0,00158	0,00296	0,00497	0,00767	0,01111	0,01533
5			0,00002	0,00006	0,00016	0,00036	0,00070	0,00123	0,00200	0,00307
6					0,00001	0,00004	0,00008	0,00016	0,00030	0,00051
7							0,00001	0,00002	0,00004	0,00007
8										0,00001

m	λ									
	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0
0	0,22313	0,13534	0,08208	0,04979	0,03020	0,01832	0,01111	O.00674	0,00409	00248
1	0,33740	0,27067	0,20521	0,14936	0,10569	0,07326	0,04999	0,03369	0,02248	0,01487
2	0,25102	0,27067	0,25652	0,22404	0,18496	0,14653	0,11248	0,08422	0,06181	0,04462
3	0,12551	0,18045	0,21376	0,16803	0,21579	0,19537	0,16872	0,14037	0,11332	0,08924
4	0,04707	0,09022	0,13360	0,10082	0,18881	0,19537	0,18981	0,17547	0,15582	0,13385
5	0,01412	0,03609	0,06680	0,05041	0,13217	0,15629	0,17083	0,17547	0,17140	0,16062
6	0,00353	0,01203	0,02783	0,02160	0,07710	0,10420	0,12812	0,14622	0,15712	0,16062
7	0,00076	0,00344	0,00994	0,00810	0,03855	0,05954	0,08236	0,10444	0,123245	0,13768
8	0,00014	0,00086	0,00311	0,00270	0,01687	0,02977	0,04633	0,06528	0,08487	0,10326
9	0,00002	0,00019	0,00086	0,00081	0,00656	0,01323	0,02316	0,03627	0,05187	0,06884
10		0,00004	0,00022	0,00022	0,00230	0,00529	0,01042	0,01813	0,02853	0,04130
11		0,00001	0,00005	0,00006	0,00073	0,00192	0,00426	0,00824	0,01426	0,02253
12			0,00001	0,00001	0,00021	0,00064	0,00160	0, 00343	0,00654	0,1126
13					0,00006	0,00020	0,00055	0,00132	0,00277	0,00520
14					0,00001	0,00006	0,00018	0,00047	0,00109	0,00223
15						0,00002	0,00005	0,00016	0,00040	0,00089
16							0,00002	0,00005	0,00014	0,00033
17								0,00001	0,00004	0,00012
18									0,00001	0,00004
19										0,00001

СОДЕРЖАНИЕ

По вопросам приобретения книг обращайтесь:
Отдел продаж «ИНФРА-М» (оптовая продажа):
127282, Москва, ул. Полярная, д. 31в, стр. 1
Тел. (495) 380-4260; факс (495) 363-9212
E-mail: books@infra-m.ru

•

Отдел «Книга–почтой»:
тел. (495) 363-4260 (доб. 232, 246)

Учебное издание

Малугин Виталий Александрович
Фадеева Людмила Николаевна

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

Учебник

Оригинал-макет подготовлен в НИЦ ИНФРА-М

Подписано в печать 25.07.2013.
Формат 60×90/16. Бумага офсетная. Гарнитура Newton.
Усл. печ. л. 39,0. Уч.-изд. л. 41,23 + 11,46 ЭБС.
Тираж 700 экз. Заказ № 5838.
ТК 207500-9804-250713

ООО «Научно-издательский центр ИНФРА-М»
127282, Москва, ул. Полярная, д. 31В, стр. 1
Тел.: (495) 380-05-40, 380-05-43. Факс: (495) 363-92-12.
E-mail: books@infra-m.ru http://www.infra-m.ru

Отпечатано с готовых файлов заказчика
в ОАО «Первая Образцовая типография»,
филиал «УЛЬЯНОВСКИЙ ДОМ ПЕЧАТИ»
432980, г. Ульяновск, ул. Гончарова, 14