

М.В. Воронов
Г.П. Мещерякова

МАТЕМАТИКА

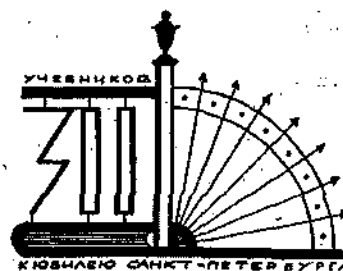
ДЛЯ СТУДЕНТОВ
ГУМАНИТАРНЫХ
ФАКУЛЬТЕТОВ

ВЫСШЕЕ
ОБРАЗОВАНИЕ



Серия «Учебники, учебные пособия»

Санкт-Петербургский
государственный университет
технологии и дизайна



М. В. Воронов, Г. П. Мещерякова

МАТЕМАТИКА для студентов гуманитарных факультетов

Рекомендовано

*Министерством образования Российской Федерации
в качестве учебника для направлений и специальностей
вузов: 030000, 050000, 520000, 630000*

Ростов-на-Дону
«Феникс»
2002

РЕЦЕНЗЕНТЫ

Кафедра высшей математики Российского государственного гидрометеорологического университета

В.Г. Дегтярев — профессор кафедры высшей математики Санкт-Петербургского государственного университета путей сообщений

Воронов М.В., Мещерякова Г.П.
В 75 Математика для студентов гуманитарных факультетов / Серия «Учебники, учебные пособия» — Ростов н/Д: Феникс, 2002. — 384 с.

В учебнике дается представление о математике как об одном из основных инструментов познания реальной действительности, в историческом аспекте прослеживается эволюция математических наук, излагаются некоторые методологические понятия математического моделирования, изложены основы высшей математики в объеме необходимом каждому специалисту гуманитарного профиля, при этом основное внимание уделено базовым понятиям математики, даны основы дискретной математики и теории некоторых пространств, изложены необходимые сведения из математического анализа. Важное место отводится прикладным аспектам математики: математике случайного мира и математическому моделированию.

Учебник предназначен для студентов вузов, обучающихся по гуманитарным специальностям.

ISBN 5-222-02212-9

ББК 22.1я73

©Воронов М.В., Мещерякова Г.П., 2002

© Феникс, оформление, 2002

ПРЕДИСЛОВИЕ

*Тот, кто не знает, и не знает, что он не знает —
 глупец, избегай его.*

*Тот, кто не знает, и знает, что он не знает —
 может научиться, научи его.*

*Тот, кто знает и не знает, что он знает —
 спит, разбуди его.*

*Тот, кто знает и знает, что он знает —
 пророк, учись у него.*

(древнее персидское стихотворение)

Начало третьего тысячелетия будет связано с переходом от индустриального к информационному обществу. Информационное общество — это общество, структуры, техническая база и человеческий потенциал которого приспособлены для оптимального превращения знаний в информационный ресурс и переработки последнего с целью перевода пассивных форм информации (книги, патенты и т. п.) в активные (модели, алгоритмы, программы). В этом обществе большая часть населения будет занята в сфере работы с информацией, а роль информационных систем в значительной мере станет определяющей. Именно поэтому в базовой подготовке каждого человека с высшим образованием математика и тесно связанная с ней информатика занимают одно из ключевых мест.

Для успешной деятельности каждый член информационного общества должен уметь максимально продуктивно использовать имеющуюся у него информацию о рассматриваемом объекте или явлении. Кроме всегда имеющих непреходящую ценность знаний, интуиции и опыта, специалист с высшим образованием в любой области человеческой деятельности должен хорошо ориентироваться в математическом аппарате, математическом моделировании, компьютерных технологиях.

Для специалиста гуманитарного профиля важно понимать роль и место математики в жизни современного общества. Для этого студент должен усвоить сущность математической науки, познакомиться с ее языком и основными методами. Это поможет ему самостоятельно читать ту литературу по специ-

альности, в которой используются математические методы и модели, заниматься повышением своей математической подготовки в ряде конкретных разделов этой науки. После окончания вуза специалист самостоятельно сможет анализировать математические результаты исследований в выбранной им области деятельности.

Учебник предназначен для студентов, обучающихся по специальностям гуманитарного профиля. В результате изучения дисциплины «Математика» студент должен получить ясное представление о математической науке, ее роли и месте в деятельности современного общества, перспективах развития и возможности использования в своей практической деятельности.

Умение логически мыслить и оперировать абстрактными понятиями, понимать место точных формулировок и уметь, где необходимо, обходиться описательными определениями, отличать тривиальные и частные модели от глубоких и общих — вот основные цели, преследуемые при изучении дисциплины «Математика».

В процессе изучения математики студент должен:

- научиться использовать математику как метод мышления, как язык, как средство формулирования и организации понятий;
- уметь формулировать, формализовать и решать с помощью компьютера основные математические задачи;
- уметь строить простейшие математические модели и ориентироваться в возможностях их реализации на вычислительной технике.

Для успешного усвоения изложенного материала достаточен объем знаний, полученный в средней общеобразовательной школе.

На изучение всего материала, представленного в учебнике, предполагается отвести не менее 214 часов аудиторных занятий (119 на лекции и 95 на практические занятия). Однако, в зависимости от специальности и учебного плана, курс может содержать лишь часть разделов и глав, представленных в учебнике. С другой стороны, по ключевым для ряда специальностей темам целесообразно дополнительно предусмотреть выделение часов для проведения семинарских занятий. В качестве неотъемлемой составляющей учебного процесса выступает самостоятельная подготовка студентов, на обеспечение которой в значительной мере направлено данное издание.

Материал учебника изложен во введении и двух частях.

Во введении представлены некоторые сведения о предмете изучения: дается представление о математике как об одном из основных инструментов познания реальной действительности, в историческом аспекте прослеживается эволюция математических наук, излагаются некоторые методологические понятия математического моделирования.

В четырех разделах первой части изложены элементы математики в объеме необходимом каждому специалисту гуманитарного профиля. При этом основное внимание уделено базовым понятиям математики, даны основы дискретной математики и некоторых пространств, изложены наиболее необходимые сведения из математического анализа.

Вторая часть посвящена прикладным аспектам математики: математике случайного мира и математическому моделированию.

В настоящее время имеется очень большое количество отличных учебников и пособий по математическим наукам, предназначенных для подготовки математиков и физиков, инженеров широкого профиля и специалистов по отдельным областям науки и техники. Имеются первоклассные монографии по отдельным разделам математики. В то же время для подготовки специалистов в гуманитарных областях данное издание является одним из первых.

В учебнике не так много доказательств теорем, но большое внимание уделено введению математических понятий и обозначений, много примеров. По мнению авторов, это позволит студентам-гуманитариям получить широкие и общие знания по основным разделам математики, научиться понимать общенаучную и специальную литературу, поможет в дальнейшем повышать свою квалификацию. Для углубленного ознакомления с отдельными разделами рекомендуется обращаться к специальной литературе, список которой приведен в конце книги.

Нумерация формул, рисунков и таблиц произведена отдельно по частям учебника: первая цифра означает номер раздела, вторая — порядковый номер главы, третья — параграфа. Вводимые понятия и определения сведены в Предметный указатель, расположенный в конце книги.

Авторы выражают благодарность своим коллегам — профессорам и преподавателям Санкт-Петербургского государственного университета технологии и дизайна за постоянное внимание к работе над книгой и высказанные ими ценные замечания

и предложения по ее содержанию, Бондарюку А.Л. и Колесовой В.В., вложившим много труда в оформление данного издания, а также рецензентам профессору Дегтяреву В.Г. и доценту Веретенникову В.Н., взявшим на себя труд отрецензировать рукопись книги.

ВВЕДЕНИЕ

Вечные истины значимы совершенно независимо от какого-то ни было фактического состояния действительности, какова бы она ни была.

(Лейбниц)

Что такое математика

Математика — наука о количественных отношениях и пространственных формах действительного мира. При этом в круг изучаемых математикой отношений входят отношения между элементами произвольной природы и все разнообразные формы пространств.

Количественные отношения, в отличие от качественных, характеризуются своим безразличием к конкретной природе тех объектов, которые их связывают. Именно поэтому они могут быть совершенно отделены от их содержания как от чего-то безразличного для них. Например, число 5 может определять и оценку на экзамене, и количество пальцев на руке; отношение $y = ax + b$ определяет линейную зависимость между величинами x и y , независимо от конкретных значений a или b и от того, что они обозначают. Таким образом, количественные отношения есть чистые отношения, сохраняющие от конкретной действительности, от которой они отвлечены, только то, что предусмотрено по их определению.

Математика представляет собой стройную и глубокую совокупность знаний о математических структурах со своими проблемами, с собственными путями развития [1]. Она дает удобные и плодотворные способы описания самых разнообразных явлений реального мира и тем самым выполняет функцию языка. Так еще Г. Галилей сказал: «Философия написана в грандиозной книге — Вселенной, которая открыта нашему пристальному взгляду. Но понять эту книгу может лишь тот, кто научился понимать ее языки и знаки, которыми она изложена. Написана же она на языке математики...»

Владение математикой дает людям мощные методы изучения и познания окружающего их мира, методы исследования как теоретических, так и практических проблем. Широко извест-

тно высказывание Леонардо да Винчи, который по этому поводу писал: «Ни одно человеческое исследование не может называться истинной наукой, если оно не прошло через математические доказательства».

Естественные науки изучают окружающий нас мир, гуманитарные — человеческое общество. Математика же, в отличие от всех других наук, изучает самое себя. Именно это является самым существенным отличием математики от всех других наук, причем оно более существенно, чем отличие естественных наук от гуманитарных. Если в физике, других естественных и гуманитарных науках, критерием истинности является совпадение теоретических и экспериментальных результатов, то критерием истинности математического рассуждения является лишь его логическая безукоризненность. Дело в том, что математика строится чисто умозрительно и ни в какой лаборатории, кроме человеческой головы, не нуждается.

Вместе с тем все мы являемся свидетелями как бы сращения естественных и гуманитарных наук; одним из основных стимулов которого является математизация гуманитарных наук. Именно использование математического моделирования, дедуктивных методов и специального математического аппарата сближает гуманитарные и естественные науки. Наблюдается и обратное явление. В математику проникают подходы и методы гуманитарных наук. Примером тому может явиться теория нечетких множеств [2].

Прежде в гуманитарных науках редко использовался дедуктивный вывод (один из признаков научности), считавшийся у математиков единственно приводящим к «истинному» доказательству. Сегодня в рамках гуманитарных наук развивается математическая лингвистика, математическая психология, литературоведение, правоведение и другие науки [см., например, 3, 4, 5, 6].

Математика никогда не претендовала и не претендует на решение проблем других наук. Математика является служанкой естественных и гуманитарных наук, помогает им, доставляя адекватный аппарат для описания всевозможных фактов и явлений. С другой стороны, без математики науку и за науку признать невозможно, ибо «уровень научности той или иной дисциплины можно измерить объемом применяемых в ней математических рассуждений, глубиной и содержательностью характерных для этой дисциплины дедуктивных выводов [7]».

Математика дает чрезвычайно много изучающим окружающую нас действительность дисциплинам, вооружая их мощным аппаратом исследования всевозможных явлений. В то же время развитие нематематических наук доставляет математике существенные стимулы для расширения поля своей деятельности. Обычно математик, построивший внутренне совершенную и естественную теорию, всегда убежден, что такая красота не может быть случайной, что эта теория отражает какие-то существенные внешние закономерности. Можно привести множество примеров, когда открытия в различных областях науки делались, как говорится, «на кончике пера».

Сила математики заключается в том, что возникшие в ее рамках числовые системы и формальные схемы предоставляют нам некоторый универсальный ключ, годный для отпирания всех на свете замков: они приложимы к физике и социологии, технике и лингвистике. Универсальная приложимость математических конструкций к изучению реально существующих объектов связана, прежде всего, с ограниченностью числа математических схем, возникающих в качестве математических моделей самых разнообразных явлений.

Математика строит модели целых классов. При построении моделей идут в ход понятия (независимо от того, почерпнуты они из наблюдений природы или являются плодами человеческого разума), аксиомы, которые могут быть подсказаны физическими фактами или придуманы людьми, процессы идеализации, обобщения и абстракции, а также интуиция. Доказательство цементирует элементы модели воедино.

В современный мир математика входит не только как метод, позволяющий компенсировать несовершенство наших органов чувств, но и как метод расширения того знания, которое человек способен обрести об окружающем мире. Опираясь на человеческий разум и способность человека к рассуждениям, математика порождает знания о реальном мире. Одна из важнейших задач математики сегодняшнего дня — усиление интеллектуальных возможностей человека; неотъемлемым инструментом решения этой задачи является компьютерная техника.

Историко-философские аспекты развития математики

Математики много раз меняли представление о своей науке и делали это каждый раз под давлением определенных факто-

ров, которые заставляли их отказываться от устоявшихся привычных воззрений. В процессе своего становления и развития она прошла долгий и сложный путь.

Зарождение математики. Потребность в счете предметов привела к простейшим понятиям арифметики натуральных чисел и системе устного счисления. Позднее возникает письменное счисление и вырабатываются приемы выполнения четырех действий арифметики. При этом выполнение четвертого действия — деления — долгое время представляло большие трудности. Под давлением потребностей человеческой практики возникли понятия дробей и приемы выполнения арифметических действий над ними. Таким образом, накапливался материал, складывающийся постепенно в древнейшую математическую науку — арифметику. Измерение площадей и объемов, потребности строительной техники, а затем астрономии, вызывают развитие зачатков другой математической науки — геометрии [8].

Первые дошедшие до нас источники, содержащие арифметические и геометрические задачи датируются 20—18 вв. до н. э. Их относят к древнему Египту и Вавилону, поскольку только здесь использовались носители (пергамент и обожженные глиняные таблички), обеспечившие сохранность нанесенной на них информации на века.

Египетская математика представляла собой исключительно систему правил, подчиненную практическим задачам, скомбинированных по назначению. Она имела целью облегчить календарные расчеты, распределение урожая, организацию общественных работ и т. п. Отличительной чертой науки того периода была статичность. Научные сведения сохранялись в неизменном виде века и тысячелетия. Обособленность же отдельных районов и целых стран приводила еще и к тому, что сделанные где-либо открытия могли долгое время оставаться неизвестными в других местностях.

Важным этапом в развитии древней математики явился вавилонский период (с 19 века до н. э.). Здесь задачи усложняются в своей математической основе, они комбинируются по математическому типу, появляются задачи, не связанные непосредственно с практикой. Вавилоняне обладали высокоразвитой арифметикой как искусством счета, была тщательно разработана система записи чисел и алгоритмы для производства действий над ними. Задачи имели в основном прикладной характер, однако встречались и учебные задачи. Вместе с тем доказательной

системы выработано еще не было, встречались только предписания типа «делай то-то, делай так-то» [9].

Период элементарной математики. Накопление большого конкретного материала постепенно привело к возникновению математики как самостоятельной науки с ясным пониманием ее метода и необходимости систематического развития. Формирование основ математической науки относится к временам древней Греции. Именно здесь в 7—6 вв. до н. э. наблюдается революционный темп развития математики. Набор практических правил превращается в науку, в систему внутренних связанных идей и методов.

Математика появилась как наука, когда сформировалась методика дедуктивного рассуждения, т. е. последовательный вывод одних формул и правил на основе других, — один из основных методов математического доказательства. Важнейшая самобытная особенность древних греков состояла именно в их сознательном стремлении расположить математические доказательства в такие цепочки, чтобы переход от одного звена к следующему не оставлял места сомнениям и заставлял всех с ним согласиться [10].

Появляется профессиональный ученый — человек, посвящающий свою жизнь развитию науки и получающий за это вознаграждение. Одним из самых известных ученых математиков древности являлся Евклид из Александрии, работы которого имеют непреходящее значение. До наших времен дошли все тринадцать книг его «Начал», где изложение математики той эпохи построено в виде строго логических выводов, теорем, полученных из системы определений и постулатов (аксиом). На этом основании мы считаем Евклида родоначальником аксиоматического (дедуктивного) построения математики.

Выбор системы аксиом в достаточной мере произволен и основан главным образом на интуиции и опыте. Однако система аксиом должна обладать некоторыми свойствами. В частности, она должна быть непротиворечивой, т. е. все ее аксиомы должны быть попарно совместны. Это означает, что никакие две теоремы, которые могут быть выведены из такой системы аксиом, не могут содержать взаимных противоречий.

В системе Евклида имеется одна аксиома, известная как постулат о параллельных прямых: «через данную точку, не лежащую на данной прямой, можно провести одну и только одну прямую, параллельную данной». Здесь следует заметить, что

постулируемое утверждение относится к данной прямой на всем ее протяжении. При этом предполагается, что прямая имеет неограниченное продолжение в обе стороны. Между тем очевидно, что в пределах некоторой ограниченной плоской области (например, круга) через данную точку можно провести множество прямых, не пересекающихся с данной прямой. Особенность этого утверждения в том, что экспериментально проверить его на всем протяжении прямой невозможно (радиус круга можно выбрать достаточно большим, но конечным!). Все другие аксиомы Евклида касаются конечных отрезков и фигур, чем качественно отличаются от аксиомы параллельности. В этой связи встает проблема независимости системы аксиом. Так, если они зависимы, то аксиому параллельности можно исключить из постулатов Евклида и доказывать как обычную теорему. Многие столетия математики не могли разрешить эту проблему.

Греки увидели в математике выражение глубинной сущности мира, нечто связанное с истинной и неизменной природой вещей. Они мистифицировали математику до того, что Пифагор выразил свой основной тезис фразой «все есть число». Он пришел к пониманию числа как универсальной основы всех вещей через музыку: «любое различие в звучании определяется числовым отношением», что потом было строго доказано. Например, гармоническое созвучие возникает, если заставить колебаться две одинаково натянутые струны, одна из которых вдвое длиннее другой. Музыкальный интервал между издаваемыми такими струнами тонами ныне называют октавой. Последователи Пифагора, пифагорейцы, считали, что строгие доказательства могут быть только арифметическими. Для них строгость и точность результата выражали саму сущность математики.

Первый удар по пифагорейскому учению нанесло открытие несоизмеримости отрезков, в частности диагонали квадрата с его стороной: если сторона квадрата равна единице, то длина его диагонали равна $\sqrt{2}$ (иррациональная величина). Следует отметить, что иррациональность величины $\sqrt{2}$ (несоизмеримость с единицей) была доказана в недрах самой пифагорейской школы. Отсюда следовало, что при любом варианте выбора единицы всегда найдутся величины с ней несоизмеримые, не представимые натуральным числом. Была подготовлена почва под учение об иррациональных числах. Это был подрыв гармонии между геометрией и арифметикой. Арифметическая интуиция была дискредитирована, и ее место заняла геометрическая интуиция.

После упадка античного общества центр развития математики сместился в Индию. Наиболее известным достижением древнеиндийских ученых является наша десятичная позиционная система исчисления (6 век н. э.). Позднее, с возникновением ислама, математический центр снова перемещается: теперь в страны ближнего Востока и Средней Азии. В трудах арабских математиков переплелись достижения греческих и индийских ученых. К наиболее известным работам того времени следует отнести книгу Мухаммеда ибн Муса ал-Хорезми под названием «Хисаб ал-джабр ва-л-мукабала», что, вероятно, означало науку об уравнениях. Именно ее латинский перевод породил название науки «алгебра». Эта книга наряду с «Началами» Евклида создала базу для развития математики в Западной Европе.

В эпоху почти тысячелетнего средневековья в математике не произошло существенных сдвигов. Лишь в 13 веке вследствие развития торговли в Западной Европе начинает возрождаться творческое математическое мышление. Отличительной особенностью этого этапа является ярко выраженная связь между математикой и общекультурными стремлениями эпохи Возрождения. Так, в 16 веке математики нередко были медиками и архитекторами, живописцами и купцами. В новых торговых городах образовался спрос на инженеров и «арифметиков».

Математика переменных величин

Период элементарной математики заканчивается, когда центр тяжести математических интересов переносится в область переменных величин (высшей математики). Создание новой математики переменных величин было делом ученых стран Западной Европы.

Считается, что современная математика возникла в 17 веке как наука, органически связанная с созданием математического естествознания, имеющего целью объединить течения отдельных природных явлений действием общих, математически сформулированных законов природы.

Геометрия считалась отражением пространства, а арифметика — времени. Р. Декарт заложил основы аналитической геометрии, объединив алгебру начала 17 века с геометрией древних греков. В знаменитой его работе «Теория геометрии» содержится изложение основ координатного метода, что положило начало аналитической геометрии. Теперь геометрические построения

можно было осуществлять в алгебраической форме, и наоборот. Например, любое уравнение первой степени, связывающее две переменные, описывает прямую на плоскости, а линейное уравнение, включающее три переменные — плоскость в пространстве. И по сей день координатный метод является одним из основных кирпичей в здании математики и ее приложений.

Благодаря появлению декартовой переменной величины в математику вошло движение. Стало необходимым появление дифференциального и интегрального исчисления. И такое исчисление было создано И. Ньютоном и Г. Лейбницем, что ознаменовало революцию в математике и в науке в целом.

Лейбниц ввел понятие дифференциала как бесконечно малого приращения функции dy . Эта величина не равна нулю, но столь мала, что умножая ее на любое конечное число a , не получим конечной величины. Иначе говоря, $(a \cdot dy)$ — такая же бесконечно малая, что и dy . Это в корне противоречило аксиоме Архимеда: «для любых a и b всегда найдется целое N такое, что $(a \cdot N) > b$ ». Пусть $y = x^2$. Дадим величинам x и y соответствующие бесконечно малые приращения dx и dy . Тогда $y + dy = (x + dx)^2$. Отсюда $dy = 2x dx + dx^2$. Лейбниц предложил отбрасывать величину dx^2 как несравненно малую по отношению к величине $2x dx$. Таким образом, он получал: $dy = 2x dx$ — правильный результат. Наблюдаемый здесь логический разрыв долгое время был камнем преткновения споров известных ученых. Однако, как часто бывает в науке, побеждает практика. Дело в том, что получаемые при помощи построений Лейбница правила обеспечивают правильное решение очень широкого класса задач. С тех пор дифференциальное исчисление становится одной из самых продуктивных частей математики. Логическое же обоснование этого факта было дано лишь в 19 веке.

Была проведена унификация всей системы мер на десятичной основе, появились десятичные дроби. Изобретенные в 17 веке логарифмы были ориентированы на приближенные вычисления, практически забытые со времен Архимеда. Постепенно сформировалось целое направление — приближенные методы анализа. Математика внесла в свое содержание приближенность как нечто вполне законное. Это было отступление от внутреннего изящества ради практической эффективности.

Под влиянием развития страхового дела и артиллерии формируется интерес к задачам, связанным со случайными собы-

тиями. В ответ появляется математическая наука — теория вероятностей, основателем которой считается Б. Паскаль.

В 18 веке основная деятельность математиков сосредоточилась на решении проблем математического анализа и теории вероятностей. Имена таких корифеев, как братья Бернулли, Ж. Даламбер, Ж. Лагранж, П. Лаплас, Л. Эйлер встречаются во всех учебниках по математике и известны теперь каждому студенту. Именно в это время научная работа математика становится самостоятельной профессией.

В начале 19 века в науке произошло событие, перевернувшее взгляды, господствующие в математике со времен ее возникновения. Русский ученый Н.И. Лобачевский показал, что аксиома Евклида о параллельных независима от других аксиом, т. е. невыводима из них. Следовательно, можно построить другую геометрию, столь же непротиворечивую, как и евклидова геометрия. Для этого необходимо заменить аксиому о параллельных на противоположную ей: через точку на плоскости, вне лежащей на этой плоскости прямой, можно провести не только одну параллельную этой прямой. Возможность построения неевклидовой геометрии разрешила проблему о параллельных, занимавших математиков более двух тысяч лет. Кроме того, был поставлен вопрос о коренном пересмотре представлений о математике, как о некоторой незыблемой в своей основе науке.

Благодаря появлению неевклидовой геометрии была понята возможность создания существенно новых математических теорий путем правильно выполненной абстракции от налагавшихся ранее ограничений, не имеющих внутренней логической необходимости. Было обнаружено, что подобная абстрактная теория может получить со временем более широкие, вполне конкретные применения. Важнейшее методологическое значение появления неевклидовой геометрии состоит в том, что ее открытие явилось исходным пунктом и основным стимулом в понимании природы математических знаний.

Однако только к началу 20 века в основу математики было положено понятие непротиворечивости системы высказываний, основанных на неявных определениях, что радикальным образом отличалось от всех предшествующих точек зрения. Утверждалось, что всякая формально-аксиоматическая теория может строиться по чистым законам логики, без каких-либо ссылок на внешний смысл. Одним из основателей такого мировоззрения на математику стал Д. Гильберт. Суть его форма-

лизм заключается в следующем. Математика, в отличие от, например, физики или химии, есть наука точная. Она находится над опытными науками. Математика — это набор формальных (знаковых) моделей для теоретического знания, которые связаны с опытом через другие науки. По отношению к опыту математика выступает как язык, устанавливая между эмпирическими высказываниями некоторую связь. Отсюда следует требование непротиворечивости математики и необязательность ее интуитивной ясности [11]. Любые математические понятия есть конструкции, свойства которых полностью определены включающей их системой формальных связей. Поэтому математическая теория не может быть ни истинной, ни ложной.

Единство математики обеспечивается исключительно методом, а не предметом исследования. Отсюда следует, что математика может иметь бесконечные приложения. Все, что поддается формулировке на точном языке, может быть изучено в своей логической форме с отвлечением от конкретного содержания. Теперь можно сказать, что, например, между геометрией и арифметикой отличия несущественны. Математика — логически организованная система понятий, для существования которой важна лишь ее дедуктивная, трансформирующая функция!

Д. Гильбертом была предпринята попытка обоснования всей математической теории. В основе его учения лежало положение о том, что обоснование математической теории является исключительно обоснованием ее непротиворечивости. Он считал, что математика должна быть обоснована в целом. Построение же любой математической теории должно реализовываться в два этапа. Вначале в логических символах производится запись аксиом и явно формулируются допустимые правила логики. Затем требуется доказать непротиворечивость этой системы аксиом вместе с ее логическими правилами на чисто синтаксическом уровне.

Параллельно с работами Д. Гильберта и его последователей на рубеже 19–20 веков появились работы, в которых ставилась под сомнение программа полного сведения математики к логике. К середине 20 века сформировалось так называемое интуиционистское направление в математике. Его основу составляет понимание процесса построения математического объекта. Согласно точке зрения интуиционистов, математический объект существует, если он дан интуитивно или может быть сконструирован,

мысленно построен посредством интуитивно ясных операций над интуитивно ясными элементами. Формалистская математика считает объекты заданными путем введения соответствующих непротиворечивых требований (аксиом, правил и т. п.). Интуиционистская же математика требует введения объекта через его построения из более элементарных объектов. Изъятие интуиционистами чистых доказательств (т. е. без гарантий построения соответствующего примера) приводит к радикальной реформе логики. Так, отбрасываются правила двойного отрицания, закон исключения третьего и др. В настоящее время существует целое направление в математике под названием конструктивизм, в рамках которого получены весьма интересные и важные для практики результаты.

Обоснованность математических результатов. Доказательство непротиворечивости какой-либо теории часто наталкивается на существенные трудности. В этом плане огромное значение имело появление в 1931 году метатеоремы К. Геделя. Первая его теорема о неполноте гласит, что если формальная система, содержащая арифметику, непротиворечива, то она неполна. Иначе говоря, такая система содержит в себе истинные утверждения, формулируемые в ее исходных понятиях, которые недоказуемы и неопровержимы в этой системе. В этом контексте оправдано известное изречение: «каждый прав в кругу своих понятий». Вторая теорема Геделя (о непротиворечивости) утверждает, что если система, включающая арифметику, непротиворечива, то доказательство этой непротиворечивости не может быть достигнуто в метаязыке, допускающем представление в арифметическом формализме. Появление теоремы Геделя позволило сформировать основание математики как особой формальной науки, вооруженной тонкими методами анализа математических теорий.

Вся история математики свидетельствует, что критерий обоснованности имеет исторический характер. Всякое доказательство предполагает, во-первых, некоторое неявное знание, во-вторых, содержательный язык (метаязык), на котором обсуждается доказательство. Отсюда неточности и противоречия. Таким образом, математика всегда является необоснованной на некотором уровне. При этом программы обоснования не являются бессмысленными, поскольку увеличивают строгость и надежность математики в целом. Математика не может быть окончательно обоснована, как и любое знание.

Современный эмпиризм в математике [12] утверждает, что математика не является собственно эмпирическим знанием, она имеет базовые утверждения другой природы. Вместе с тем, схема ее развития и обоснования совпадают с общей схемой развития и обоснования эмпирических наук: выдвижение возможно большего числа смелых гипотез и их последующая критика. Тем самым снимается вопрос об окончательном обосновании математического знания. Центр тяжести переносится на исследование теории роста знания вообще в его квазиэмпирической модели. Традиционное деление наук на формальные и содержательные теряет смысл.

На настоящее время сформировались следующие основные философские взгляды на математическую науку [13].

1. Ценность математических теорий состоит в их способности выполнять функцию вывода. Именно поэтому перед математикой выдвигается требование непротиворечивости.

2. Единая программа обоснования математики в настоящее время невозможна, имеется множество путей, каждый из которых имеет определенную ценность.

3. Математика столь же далека от своей окончательной обоснованности, как и всякое другое знание.

Почему теоремы, доказанные человеческим разумом в тиши кабинетов, должны быть применимы к реальному миру? Почему математика эффективна и при описании тех явлений, которые непонятны для нас? Несомненно, что аксиомы математики и принципы логики выведены из опыта, но почему следствия, вытекающие из созданных человеком аксиом и принципов, так хорошо согласуются с опытом?

Ответы на эти интересные вопросы до конца не найдены. Однако сегодня можно сделать бесспорный вывод о том, что математика и физическая реальность неразделимы! Граница между математическим и эмпирическим знанием не абсолютна. Мы непрестанно вносим коррективы в наши наблюдения и в то же время видоизменяем наши теории так, чтобы они соответствовали новым наблюдениям и экспериментальным результатам. Математика служит своего рода посредником между человеком и окружающим его внешним миром [14].

Приложения математики

Математика, возникшая как чисто прикладная наука, и в настоящее время имеет своей основной задачей изучение окру-

жающего нас реального мира. При этом одним из основных способов решения этой задачи является математическое моделирование, включающее построение и последующее изучение математической модели объекта рассмотрения. Математическое моделирование — связующее звено между математикой и другими науками.

Возможны две основные схемы математического моделирования. Первая заключается в том, чтобы подобрать из арсенала математических моделей ту, которая, по мнению специалиста, в наибольшей мере подходит для решения поставленной задачи. Если это удастся сделать, то в вашем распоряжении оказывается готовый инструментарий, для которого известны не только его сильные и слабые стороны, но и конкретные методики и рекомендации по эффективному использованию всего накопленного арсенала в этой области знаний. Очевидно, что при реализации такой схемы специалист должен обладать обширной информацией о достижениях математики в данном направлении.

По второй схеме, особенно актуальной сегодня для гуманитарных наук, исследователь сам строит математическую модель рассматриваемого объекта. При этом он должен обладать широкими математическими знаниями, чтобы решить весь комплекс задач, включающий кроме всего прочего вопросы адекватности, точности и реализуемости модели.

Всякое математическое описание явления означает известную логическую его идеализацию. Кроме того, описание происходит с определенной степенью точности: отбрасываются некоторые факторы, влиянием которых пренебрегают, применяемые для описания математические отношения могут быть не вполне адекватными реальным. В этой связи математические модели — всегда есть лишь некоторое приближенное отображение действительности. Задача состоит в том, чтобы отразить в моделях наиболее характерные и существенные для изучаемого вопроса аспекты. Что значит «наиболее существенные» — вопрос к исследователю. При разработке математической модели ему всегда приходится выбирать между точностью и сложностью.

В науке в целом, а в математике особенно, едва ли не самое важное значение имеет интуиция. Различают три вида интуиции. Во-первых, *эмпирическая интуиция*, появляющаяся на основе длительного общения с определенными объектами (интуиция опытного специалиста в данной области). Во-вторых, *конст-*

руирующая интуиция, выступающая как акт конструирования объекта в свете некоторых предъявляемых к нему требований. Здесь интуиция не предвосхищает признаков нового объекта в ряду аналогичных, а производится конструирование объектов принципиально другой природы (интуиция озарения).

Эмпирическая и конструирующая интуиции имеют динамический характер в том смысле, что они обеспечивают появление нового видения на месте первоначально смутных представлений. Интуиция одного из этих видов появляется в результате некоторого труда, следовательно, она не общезначима: интуитивно ясное для одного отнюдь не обязательно является таковым для другого.

Интеллектуальная же интуиция (третий вид интуиции) есть нечто статическое, достаточно общезначимое, заданное с самого начала как постоянный факт сознания, как некоторая предпосылка всякой другой формы очевидности. Люди имеют ясные и общезначимые представления о пространстве, времени, случайности и т. п. Вследствии этого и становится возможной языковая коммуникация, однозначное толкование конкретных высказываний и сама совместная деятельность. В частности, математический язык стал одним из основных средств понимания друг другом разных сторон человеческой деятельности. Это одно из важнейших обстоятельств необходимости математического образования каждого специалиста.

Математическому моделированию должно предшествовать качественное изучение объекта моделирования. Если качественная определенность изучаемых явлений четко выявлена, а развиваемые положения четко сформулированы, то математика становится мощным средством развития всех наук, в том числе и гуманитарных [15].

Одним из краеугольных камней эффективного моделирования в целом и качественного изучения объекта моделирования в частности является предвидение, которое в зависимости от степени конкретности и характера воздействия на ход исследуемых процессов имеет три основных формы: *гипотезу, прогноз и план*. Эти формы предвидения тесно связаны в своих проявлениях друг с другом и с исследуемым объектом, представляют собой последовательные ступени познания поведения объекта в будущем.

В рамках моделирования на этапе анализа объекта рассмотрения, формирования структуры целей необходимо общенауч-

ное предвидение в форме построения различных гипотез. В развитии последних при формировании системы задач и математических моделей определяется не только последовательность решаемых вопросов (в том числе прогнозирования, планирования и принятия решений), но и определяется список основных моделей прогнозирования, планирования и выработки решений, которые могут носить как узконаправленный, так и комплексный характер.

Гипотеза — это научно обоснованное предположение о структуре объекта, характере элементов и связей, его механизме функционирования и развития. На уровне гипотезы дается качественная характеристика объекта, выражающая общие закономерности поведения. Через прогноз гипотеза оказывает воздействие на научную организацию планирования и моделирования в целом.

Прогноз в сравнении с гипотезой имеет большую определенность и достоверность, поскольку основывается не только на качественных, но и количественных характеристиках, и потому позволяет описывать будущее также и количественно. Будущее зависит от многих, в том числе и случайных факторов, сложное переплетение которых практически учесть невозможно. Отсюда следует, что все прогнозы носят вероятностный характер.

Под планом в данном аспекте понимают заранее намеченный порядок осуществления каких-либо работ (действий, мероприятий). План и прогноз представляют собой взаимно дополняющие друг друга стадии планирования. Формы сочетания прогноза и плана могут быть различными: прогноз может предшествовать разработке плана, следовать за ним или производиться в процессе разработки плана. Вместе с тем, прогнозирование представляет собой собственную методологическую и методическую основу, которая во многом отличается от планирования.

Построение математических моделей наряду со специальными знаниями требует знакомства с основами методологии математики, базовыми понятиями этой науки и ее языка. После построения математической модели производится ее анализ, включающий анализ и обработку входной информации, численное решение задачи, анализ результатов вычислений и их интерпретацию. Здесь может применяться весь арсенал вычислительной математики, включающий как собственно вычислительные методы, так и численные методы оптимизации. В настоящее время

созданы широко доступные компьютерные пакеты прикладных задач математики, обеспечивающие эффективное решение большинства стандартных математических задач. В то же время специалист должен знать суть этих методов, разбираться в проблемах их применимости и точности.

В последние годы активно применяются математические методы в моделировании целенаправленной человеческой деятельности. Это проблемы имитации процессов функционирования социальных систем, подготовки и принятия решений, построения систем искусственного интеллекта. В гуманитарных науках положение таково, что проблемой становится не работа с существующими моделями, а создание новых моделей с применением самых разнообразных математических структур. В такой ситуации математика нужна не только как метод расчета, а как метод мышления, как язык, как средство формулирования и организации понятий.

Таким образом, приложения математики требуют широких специальных знаний и практических навыков, базирующихся на хорошем математическом образовании.

Для более подробного знакомства с историей математики мы можем рекомендовать книги [16–20].

Часть 1.

ОСНОВНЫЕ МАТЕМАТИЧЕСКИЕ ПОНЯТИЯ

Раздел 1. ВВЕДЕНИЕ В ТЕОРИЮ МНОЖЕСТВ

Глава 1.1. Множества

1.1.1. Понятие множества

Понятие **множества** относится к базовым понятиям математики. В математике нет более общего понятия, чем множество. Оно первоначально, его невозможно определить через другие понятия, так как определить — значит указать, какой частью более общих понятий является данное. Близкими к понятию «множество» можно считать понятия: собрание, совокупность, набор, комплекс, система и т. п. Замена одних слов другими не снимает самой проблемы определения. Все перечисленные понятия вводятся через описание наиболее характерных для них свойств, причем иногда приходится апеллировать к интуиции и жизненному опыту.

Основоположник теории множеств Г. Кантор (1845–1919) под множеством понимал любое собрание определенных и различимых между собой объектов, мыслимое как единое целое. Очевидно, что это высказывание не является определением, поскольку слово «множество» заменено словом «собрание». Вместе с тем здесь имеется три важных момента.

Объекты, входящие во множество, *определенные*. Это означает, что для каждого объекта можно однозначно сказать, принадлежит ли он данному множеству или нет.

Объекты, входящие во множество, *различимы между собой*. Следовательно, во множестве не может быть двух или более одинаковых объектов.

Все объекты, входящие во множество, мыслятся как *единое целое*. Этим подчеркивается, что все объекты рассматриваются в совокупности, а от свойств отдельных объектов абстрагируются.

Примеры множеств: множество всех целых чисел, буквы русского алфавита, множество людей, проживающих в данном доме, комплекс регламентирующих документов, собрание книг и т. д.

Объекты, входящие во множество, называют **элементами**, и их часто обозначают строчными буквами, а сами множества принято обозначать прописными буквами латинского алфавита. Множество, состоящее из конечного числа элементов, называется **конечным**, в противном случае множество называется **бесконечным**.

Конечное множество может быть задано просто перечислением входящих в него и разделенных запятой элементов, например $A = \{a_1, a_2, \dots, a_n\}$, где $a_1, a_2, \dots, a_n$ — объекты любой природы.

Любое множество может быть задано при помощи правила, позволяющего определить, является ли данный объект элементом множества или нет. Например, множество B решений уравнения $x^2 - 5x + 6 = 0$ можно записать так:

$$B = \{x \mid x^2 - 5x + 6 = 0\}.$$

В такой записи правило, задающее множество, отделено вертикальной чертой. Очевидно, что элементами множества B являются числа 2 и 3, то есть $B = \{2, 3\}$.

Бесконечное множество всех точек плоскости может быть задано как множество пар вещественных чисел (x, y) :

$$C = \{(x, y)\}.$$

Бесконечное множество точек на плоскости, составляющих круг с радиусом единица, можно задать так:

$$D = \{(x, y) \mid x^2 + y^2 \leq 1\}.$$

Множество может содержать и всего лишь один элемент. Например: $P = \{2\}$, множество P состоит из одного элемента — числа 2.

Множество, не содержащее вообще ни одного элемента, называется **пустым** и обычно обозначается символом $\emptyset$. Понятие «пустое множество» упрощает многие математические формулировки, так как при произвольном задании множества не всегда есть уверенность в существовании его элементов. Например, множество точек на плоскости, удовлетворяющих условию $x^2 + y^2 + c = 0$,

$$G = \{(x, y) \mid x^2 + y^2 + c = 0\},$$

при $c < 0$ — окружность, при $c = 0$ — одна точка, а при $c > 0$ — пустое множество.

Замечание: Не следует путать записи $\emptyset$ и $\{\emptyset\}$. $\emptyset$ — пустое множество и в нем нет ни одного элемента. $\{\emptyset\}$ — одноэлементное множество, единственным элементом которого является множество $\emptyset$.

Для указания того факта, что объект принадлежит данному множеству, используют знак $\in$. Например, $a_2 \in A$, $3 \in B$, $(0, 4, 0, 61) \in D$. Если же объект не принадлежит данному множеству, то пишут знак $\notin$. Например, $a_6 \notin A$, $7 \notin B$, $(4, 7) \notin D$.

1.1.2. Подмножества

Определение подмножества. Множество B называется подмножеством множества A , если каждый элемент B одновременно является и элементом множества A . Тот факт, что B является подмножеством A , записывается так: $B \subset A$ (читается: « B содержится в A » или « B — подмножество множества A »).

Пример. Пусть заданы множества $A = \{1, 3, 5, 7\}$ и $B = \{3, 5\}$. Очевидно, что B есть подмножество A , т. е. $B \subset A$.

Пример. Пусть Q — множество всех рациональных чисел, а Z — множество всех целых чисел, тогда $Z \subset Q$.

Из определения следует, что множество A есть подмножество самого себя, т. е. $A \subset A$. Говорят, что A — самое широкое подмножество A .

Пустое множество является подмножеством любого множества A , т. е. $\emptyset \subset A$. Докажем это утверждение методом «от противного». Пусть $\emptyset$ не является подмножеством A . Тогда во множестве $\emptyset$ есть элемент, который не принадлежит A . Но это невозможно, поскольку пустое множество по определению не содержит ни одного элемента. Следовательно, выражение $\emptyset \subset A$ истинно. Пустое множество является самым узким подмножеством любого множества.

Множество A и пустое множество $\emptyset$ называются **несобственными подмножествами** множества A . Все другие подмножества A называются **собственными подмножествами** A .

Пример. Если $A = \{a_1, a_2, a_3\}$, то оно имеет следующие подмножества: $\emptyset$, $\{a_1\}$, $\{a_2\}$, $\{a_3\}$, $\{a_1, a_2\}$, $\{a_1, a_3\}$, $\{a_2, a_3\}$, $\{a_1, a_2, a_3\}$. Всего 8 подмножеств.

Пусть множество A состоит из n элементов. Сколько подмножеств оно имеет? Из n элементов можно сформировать C_n^k разных множеств, каждое из которых состоит ровно из k элементов, где C_n^k — число сочетаний из n по k (см. параграф 2.3.1). Подмножества A могут состоять из $k = 0, 1, 2, \dots, n$

элементов, то есть, всего $\sum_{k=0}^n C_n^k$ подмножеств. Чему равна эта сумма? Ответить на этот вопрос поможет известная формула бинома Ньютона [1]

$$(u + v)^n = \sum_{k=0}^n C_n^k u^k v^{n-k}$$

Если в этой формуле положить $u = v = 1$, то

$$(1+1)^n = \sum_{k=0}^n C_n^k 1^k 1^{n-k} = 2^n.$$

Таким образом, мы получили, что любое множество, состоящее из n элементов, имеет ровно 2^n подмножеств. Из них ровно $2^n - 2$ являются собственными подмножествами. Элементами множества могут также выступать другие множества. В этом случае говорят не о множестве множеств, а о *системе* множеств. Частным случаем системы множеств является система всех подмножеств данного множества A и обозначается $P(A)$. Так система подмножеств множества A из предыдущего примера имеет вид $P(A) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}\}$.

Замечание. Не следует путать символы $\in$ и $\subset$. Символ $\in$ — употребляется для обозначения отношения элемента к множеству. Символ $\subset$ — употребляется для обозначения отношения множества к множеству. Пусть, например, $A = \{1, 2, 3, \{1, 2\}, \{1, 2, 3\}\}$. Тогда для элементов 1, 2, 3 пишем $1 \in A$, $2 \in A$, $3 \in A$. Для множеств $\{1, 2\}$ и $\{1, 2, 3\}$ пишем $\{1, 2\} \subset A$, $\{1, 2, 3\} \subset A$. В то же время $\{1, 3\} \notin A$, потому что в A нет элемента $\{1, 3\}$.

При работе в конкретной предметной области обычно ограничиваются некоторой совокупностью объектов. Например, при изучении вопросов дизайна одежды вряд ли изучают множество планет солнечной системы или множество сортов яблок. Зафиксированное каким-либо образом множество объектов, допустимых при данном рассмотрении, называют **базовым** или **универсумом**. Будем обозначать базовое множество буквой U . Примерами универсума являются: числа в арифметике, слова в языкознании, законы в юриспруденции и т. п.

1.1.3. Операции над множествами

Множества можно определять и при помощи операций над другими множествами.

Равенство множеств. Множества A и B считаются равными (совпадающими), если они состоят из одних и тех же элементов. Равенство множеств обозначают так: $A = B$. Если множества не равны, то пишут: $A \neq B$.

Доказательство равенства множеств состоит из двух частей:

1) для любого элемента множества A (формальная запись $\forall x$) доказывается, что он принадлежит и множеству B . Формально это записывается так:

$$\forall x: (x \in A) \rightarrow (x \in B).$$

2) для любого элемента B доказывается, что он принадлежит и множеству A . Формально это можно записать так:

$$\forall y: (y \in B) \rightarrow (y \in A).$$

Отсюда следует, что запись равенства двух множеств « $A = B$ » эквивалентна записи « $A \subset B$ и $B \subset A$ ».

Пример. Доказать, что множество $A = \{0, 2, 3\}$ равно множеству B корней уравнения $x^3 + 3x^2 + 6x = 0$, то есть $B = \{x \mid x^3 + 3x^2 + 6x = 0\}$. Для доказательства этого утверждения решим уравнение. Получим: $x_1 = 0$, $x_2 = 2$, $x_3 = 3$. Следовательно, $\{0, 2, 3\} \subset \{x \mid x^3 + 3x^2 + 6x = 0\}$ или $A \subset B$.

Затем непосредственной подстановкой убеждаемся, что любое из чисел 0, 2, 3 удовлетворяет уравнению, следовательно:

$$\{x \mid x^3 + 3x^2 + 6x = 0\} \subset \{0, 2, 3\} \text{ или } B \subset A.$$

Только теперь можно записать, что $A = B$.

Объединение (сумма) множеств. Объединением множеств A и B называется такое множество C , каждый элемент которого содержится *хотя бы* в одном из множеств A или B . Обозначается: $A \cup B$.

Пример. Если $A = \{1, 2, 3, 4\}$ и $B = \{1, 3, 5, 7, 9\}$, то $C = A \cup B = \{1, 2, 3, 4, 5, 7, 9\}$.

Можно рассматривать объединение n множеств:

$$A = \bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n,$$

при этом в A входят все элементы, которые входят *хотя бы* в одно из множеств $A_1, A_2, \dots, A_n$.

Например, множество всех действительных чисел R состоит из множества положительных чисел R^+ , множества отрицательных чисел R^- и множества $\{0\}$, содержащего один элемент — ноль, то есть $R = R^- \cup R^+ \cup \{0\}$.

Для наглядного представления соотношений между несколькими подмножествами какого-либо универсума часто используются

круги Эйлера или диаграммы Венна. Универсум представляется множеством всех точек некоторого прямоугольника, а его подмножества — соответствующими кругами. Операция объединения и другие операции иллюстрируются кругами Эйлера представленными на рис. 1.1—1.5.

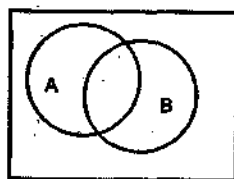


Рис. 1.1.

Объединение множеств
 $C = A \cup B$

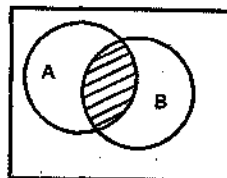


Рис. 1.2.

Пересечение множеств
 $D = A \cap B$

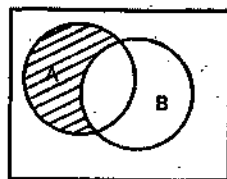


Рис. 1.3.

Разность множеств
 $G = A \setminus B$

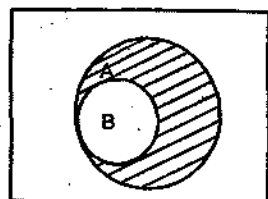


Рис. 1.4.

Дополнение к B до A
($B \subset A$); $J = A \setminus B$

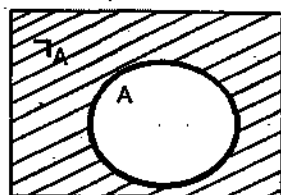


Рис. 1.5.

Дополнение A до универсума, то есть $\bar{A}$

Пересечение (умножение) множеств. Пересечением множеств A и B называется множество D, составленное из общих для множеств A и B элементов. Обозначение: $D = A \cap B$. Для множеств из примера 5 имеем: $\{1, 2, 3, 4\} \cap \{1, 3, 5, 7, 9\} = \{1, 3\}$.

Можно рассматривать пересечение n множеств:

$$A = \bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n,$$

при этом в A входят только те элементы, которые входят во все множества $A_1, A_2, \dots, A_n$.

Пересечение двух множеств иллюстрируется на рис. 1.2.

Пусть есть некоторое множество A. Говорят, что задано разбиение множества A на классы $A_1, A_2, \dots, A_m$, если

$$A = \bigcup_{i=1}^m A_i \quad \text{и} \quad A_i \cap A_j = \emptyset \quad \text{для всех } i \text{ и } j.$$

Классы — это такие подмножества разбиваемого множества, которые не имеют общих элементов, а их объединение образует исходное множество A. Следовательно, каждый элемент множества A входит в один и только в один класс. Например, разбиение всех студентов одного факультета университета на учебные группы, разбиение книги на страницы, а страницы на абзацы, разбиение уголовного кодекса на статьи и т. п.

Разность двух множеств. Разностью двух множеств A и B называется множество G, содержащее лишь те элементы из A, которые не входят в B. Обозначение: $G = A \setminus B$. Отметим, что в A могут находиться не все элементы из вычитаемого множества B (см. рис. 1.3). Например, $G = \{1, 2, 3, 4\} \setminus \{1, 3, 5, 7, 9\} = \{2, 4\}$.

Если B — подмножество A ($B \subset A$), то разность $J = A \setminus B$ называется дополнением к B до A. Например, если $A = \{1, 2, 3, 4, 5, 6\}$ и $B = \{2, 4\}$, то множество $\{1, 3, 5, 6\}$ — дополнение к B до A. Операция дополнения иллюстрируется на рис. 1.4.

Дополнение к A до универсума U имеет особое обозначение: $\bar{A} = U \setminus A$ (см. рис. 1.5).

Пример. Пусть $X = \{x \mid x \geq 0\} = \{0\} \cup \mathbb{R}^+ \equiv \mathbb{R}_0$. Такое множество называется множеством неотрицательных чисел. Тогда $\bar{X} = \{x \mid x < 0\} = \mathbb{R}^-$, это множество отрицательных чисел.

1.1.4. Свойства операций над множествами

Перечисляемые ниже свойства операций над множествами справедливы для любых множеств, поэтому их часто называют законами, часть которых имеет специальные наименования.

1. Коммутативный, или переместительный, закон имеет место, как для операции объединения, так и для операции пересечения:

$$A \cup B = B \cup A; \quad A \cap B = B \cap A.$$

2. Ассоциативный, или сочетательный, закон также имеет место и для операции объединения и для операции пересечения:

$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C,$$

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C.$$

Так как порядок выполнения операций не существен, то скобки в записи опускают.

3. Дистрибутивный, или распределительный, закон:

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$

4. Закон идемпотентности:

$$A \cap A = A, \quad A \cup A = A.$$

5. Закон поглощения:

$$A \cap (A \cup B) = A, \quad A \cup (A \cap B) = A.$$

6. Закон двойственности де Моргана:

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}, \quad \overline{(A \cup B)} = \overline{A} \cap \overline{B}.$$

$$7. A \cup U = U, \quad A \cap \emptyset = \emptyset.$$

$$8. A \cup \overline{A} = U, \quad A \cap \overline{A} = \emptyset.$$

$$9. \forall A: A \cup B = A \rightarrow B = \emptyset, \quad \forall A: A \cap B = A \rightarrow B = U.$$

$$10. \text{Если } A \cup B = U \text{ и одновременно } A \cap B = \emptyset \rightarrow B = \overline{A}.$$

$$11. \overline{\overline{U}} = \emptyset, \quad \overline{\emptyset} = U.$$

$$12. \overline{\overline{A}} = A.$$

Анализируя свойства 1–13, можно сформулировать принцип двойственности: всякое равенство, тождественно выполняемое в теории множеств, переходит также в тождественно выполняющееся равенство при замене знака объединения $\cup$ на знак пересечения $\cap$, множество универсум U на пустое множество $\emptyset$, и наоборот.

Глава 1.2. Соответствия

1.2.1. Прямое произведение множеств

Введем в рассмотрение новое базовое понятие — **кортеж**. Под кортежем будем понимать любую выделенную *упорядоченную* совокупность объектов (элементов кортежа). Синонимами понятия «кортеж» являются: упорядоченная система, «п»-ка, упорядоченная совокупность, вектор, упорядоченный набор и др. В каждой из предметных областей чаще применяется какой либо один термин, и там он, как правило, несет некоторую дополнительную смысловую нагрузку.

Элементы, составляющие кортеж, называются **компонентами**, которые в силу упорядоченности имеют номер: первый компонент, второй компонент, ... n -ый компонент. Длиной кортежа называют число компонентов в кортеже. Когда вместо термина «кортеж» употребляется термин «вектор», то говорят, соответственно, о координатах и размерности вектора.

Примеры кортежей: $N = (8, 7, 4, 4, 7)$. Это кортеж N длины 5, первый компонент которого — 8, второй — 7, третий — 4 и т. д.; $M = \{a, 2x + 8, \{1, 2, 3\}, \{a, a\}\}$, в этом случае $(2x + 8)$ — второй, а $\{a, a\}$ — четвертый компонент кортежа M .

Отличие кортежа от множества заключается в том, что компоненты кортежа упорядочены и могут полностью или частично совпадать.

Два кортежа называются равными, если они имеют одинаковую длину и все их соответствующие компоненты совпадают.

Прямое произведение множеств. Прямым произведением двух множеств A и B (обозначается $A \times B$) называется множество, состоящее из всех тех и только тех пар, первая компонента которых принадлежит A , вторая B . Если первый сомножитель имеет n элементов, а второй — m , то их прямое произведение имеет $n \cdot m$ элементов, каждый из которых — упорядоченная пара.

Например, если $A = \{1, 3\}$ и $B = \{1, 4, 5\}$, то $A \times B = \{(1, 1), (1, 4), (1, 5), (3, 1), (3, 4), (3, 5)\}$. В общем случае, если $A = \{a_1, a_2, \dots, a_n\}$ и $B = \{b_1, b_2, \dots, b_m\}$, то $A \times B = \{(a_1, b_1), (a_1, b_2), \dots, (a_1, b_m), (a_2, b_1), \dots, (a_n, b_m)\}$.

Прямым произведением n множеств $A_1, A_2, \dots, A_n$ называется множество всех кортежей длины n (« n -ок»), первый компонент

которых принадлежит A_1 , второй A_2 , ..., n -ый — A_n , т. е.

$$A_1 \times A_2 \times \dots \times A_n = \{(a_{k1}, a_{k2}, \dots, a_{kn})\},$$

где a_{ij} — i -ый элемент множества A_j .

Если все множества A_i равны между собой, то есть $A_1 = A_2 = \dots = A_n = A$, то прямое произведение множеств обозначается как A^n

$$A_1 \times A_2 \times \dots \times A_n = A \times A \times \dots \times A = A^n.$$

Например: пусть R — множество действительных чисел, тогда $R \times R = R^2$ — множество упорядоченных пар вида (x, y) , где $x, y \in R$. Геометрически R — множество точек числовой оси, тогда R^2 — множество точек плоскости, где x и y — координаты этих точек. Поскольку координатное представление точек плоскости впервые было введено Р. Декартом (1596–1650), то прямое произведение часто называют декартовым произведением множеств. Множество R называется **графиком**, если каждый его элемент является упорядоченной парой, следовательно, любое подмножество множества R^2 можно назвать графиком.

Проекцией кортежа $A = (a_1, a_2, \dots, a_i, \dots, a_n)$ на i -ую ось $(\text{Пр}_i A)$ называется i -ый компонент кортежа, т. е. $a_i \equiv \text{Пр}_i A$. Проекция точки плоскости на первую ось $(\text{Пр}_1 R^2 = x)$ называется абсциссой, на вторую ось — ординатой $(\text{Пр}_2 R^2 = y)$.

Из определения прямого произведения следует, что оно не коммутативно, т. е. $A \times B \neq B \times A$.

Пример. Пусть A — отрезок $[1, 3]$, B — отрезок $[2, 5]$. Тогда $A \times B$ — множество точек прямоугольника, заштрихованного на рис.1.6, $B \times A$ — прямоугольник, заштрихованный на рис.1.7.

Пример. Пусть A — множество, элементами которого являются буквы, цифры и все знаки операций и препинания. Такое

множество называют алфавитом. Тогда A^n — множество всех слов длины n .

Пример. Пусть $A = \{0, 1, 2, \dots, 9\}$. Тогда A^4 — множество всех четырехразрядных чисел.

Следует отметить, что природа компонентов прямого произведения обычно отличается от природы элементов сомножителей. Например, пусть Q — множество участников шахматного турнира, тогда $Q \times Q = (q_i, q_j)$, при всех $i \neq j$, есть множество пар участников, причем q_i играет белыми фигурами, q_j — черными. Если N — множество всех натуральных чисел, то $N \times N$ есть множество упорядоченных пар (n, m) , каждая из которых может определять самые различные объекты: дроби (n/m) , номера соответственно домов и квартир в почтовых адресах, номера страниц и строк в книге, месяц и день рождения и т. п.

1.2.2. Понятие соответствия

Пусть заданы два множества X и Y . Если для каждого элемента $x \in X$ указан элемент $y \in Y$, с которым сопоставляется x , то говорят, что между множествами X и Y установлено **соответствие**. Иначе говоря, соответствием называется тройка множеств $\Gamma = (X, Y, G)$, где $G \subset X \times Y$. Множество X называется областью отправления, Y — областью прибытия, G — графиком соответствия. Если $(x, y) \in G$, то множество первых проекций $(\text{Пр}_1 G)$ называется областью определения соответствия, множество вторых проекций $(\text{Пр}_2 G)$ — областью значений этого соответствия, G — **графиком соответствия**.

Два соответствия равны, тогда и только тогда, когда равны их области отправления, области прибытия и графики.

Пример. Заданы четыре разных соответствия, имеющие одинаковые области отправления и прибытия:

$$(X, Y, G_1) = (\{a, b, c\}, \{x, y, z\}, \{(a, x), (b, y), (c, z)\}),$$

$$(X, Y, G_2) = (\{a, b, c\}, \{x, y, z\}, \{(a, x), (a, y), (b, y), (c, z)\}),$$

$$(X, Y, G_3) = (\{a, b, c\}, \{x, y, z\}, \{(a, x), (b, y), (c, y)\}),$$

$$(X, Y, G_4) = (\{a, b, c\}, \{x, y, z\}, \{(b, y), (c, y)\}).$$

На рис. 1.8а, 1.8б, 1.8в, и 1.8г различия этих соответствий видны достаточно наглядно.

В соответствии (X, Y, G) множество всех $y \in Y$, которые сопоставляются элементу $x \in X$, называется **образом** x в Y .

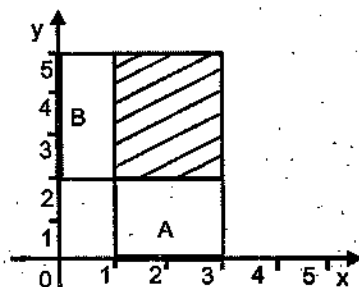


Рис. 1.6. Множество $A \times B$

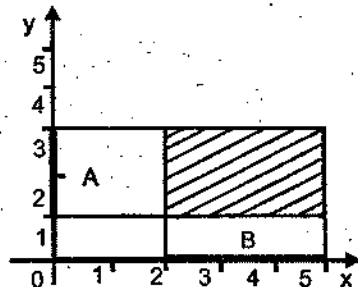


Рис. 1.7. Множество $B \times A$

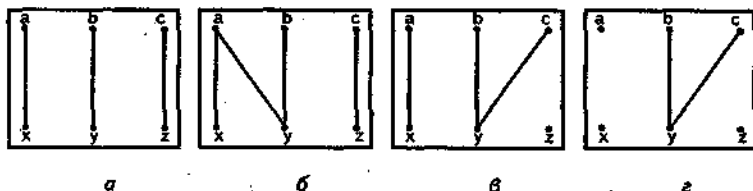


Рис. 1.8. Примеры различных соответствий

Множество же всех $x \in X$, которым сопоставляют элемент $y \in Y$, называется **прообразом** y в X .

Соответствие называется **всюду определенным**, если множество $\text{Pr}_1 G = X$, т. е. его область определения совпадает с областью отправления (в противном случае говорят о частичном соответствии). Если же $\text{Pr}_2 G = Y$, то соответствие называют **сюръективным**, или **накрывающим**. Это означает, что область значений соответствия совпадает с его областью прибытия. На рис. 1.8а и 1.8б представлено всюду определенное сюръективное соответствие. Соответствия, представленные на рис. 1.8в и 1.8г, не сюръективны, а соответствие изображенное на рис. 1.8г не всюду определенное.

Соответствие (X, Y, G) называется **функциональным** (или **однозначным**), если образом любого элемента из $\text{Pr}_1 G$ является единственный элемент из $\text{Pr}_2 G$. График такого соответствия называется **функциональным**. Это означает, что в нем нет пар с одинаковыми первыми и различными вторыми компонентами. Например, соответствие, представленное на рис. 1.8б не функционально. Соответствие называется **инъективным**, если любому элементу из $\text{Pr}_2 G$ соответствует единственный элемент из $\text{Pr}_1 G$, на рис. 1.8а изображено инъективное соответствие.

Соответствие между X и Y называется **взаимно-однозначным** (или **биективным**), если оно всюду определено, сюръективно, функционально и инъективно.

Пусть X и Y — множества вещественных чисел. В этом случае график соответствия G может быть представлен некоторой линией на плоскости.

Например. На рис. 1.9 представлено функциональное соответствие, но оно не инъективно (некоторым y соответствует более одного x), не всюду определено (G определен не для всех x), не сюръективно (G проектируется не на все y) и не биективно. На рис. 1.10 представлено не функциональное соответствие, которое не всюду определено, сюръективно и не биективно. На рис. 1.11 представлено взаимно-однозначное соответствие.

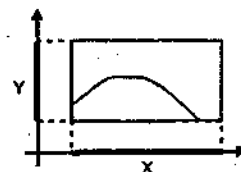


Рис. 1.9.
Функциональное
соответствие

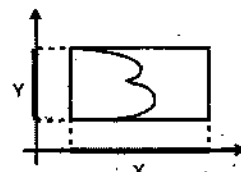


Рис. 1.10.
Не функциональное
соответствие

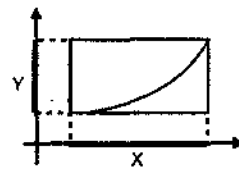


Рис. 1.11.
Взаимно-однозначное
соответствие

1.2.3. Мощность множества

Мощность множества характеризует количество элементов этого множества. Число элементов в конечном множестве A называется **кардинальным числом** и обозначается $|A|$. Подсчет элементов конечного множества заключается в установлении взаимно-однозначного соответствия между этими элементами и последовательностью натуральных чисел.

Теорема. Множества A и B имеют одинаковую мощность, если между этими множествами установлено взаимно-однозначное соответствие.

Доказательство. Пусть между множествами A и B установлено взаимно-однозначное соответствие. В силу вышеприведенного определения это означает, что оно всюду определено, сюръективно, функционально (единственность образа) и инъективно (единственность прообраза). Для доказательства теоремы требуется показать, что выполняется равенство $|A| = |B|$. Если это не так, то либо $|A| > |B|$, либо $|B| > |A|$.

Предположим, что $|A| > |B|$. В этом случае, поскольку соответствие между рассматриваемыми множествами всюду определено, в A должны найтись хотя бы два элемента, которым соответствует один элемент из B . Тем самым нарушается единственность образа. Отсюда очевидно, что мощность A не может быть больше мощности B .

Теперь предположим, что $|B| > |A|$. В силу сюръективности отображения при данном предположении в B найдутся хотя бы два элемента, соответствующие одному и тому же элементу из A . Однако в этом случае нарушается единственность прообраза и наше предположение о том, что мощность B больше мощности A ложно. Следовательно, $|A| = |B|$ и теорема доказана.

Для бесконечных множеств вводится понятие равномощности. Множества **равномощны**, если между их элементами можно установить взаимно-однозначное соответствие.

Бесконечное множество A называется **счетным**, если оно равномощно множеству всех натуральных чисел N . На практике это означает, что все элементы любого счетного множества можно сосчитать, то есть каждому элементу можно приписать отдельный номер. Примеры счетных множеств: множество целых чисел, четных чисел, рациональных чисел. Можно показать, что счетное множество образуется и при объединении счетного множества конечных множеств (например, множество слов в любом конечном алфавите) и т. д. Счетным будет и объединение счетного множества счетных множеств (множество всех векторов с натуральными компонентами). Множество A называется **не более чем счетным (дискретным)**, если оно конечно (в частности — пусто) или счетно. Счетное множество среди бесконечных множеств имеет наименьшую мощность.

Рассмотрим все вещественные числа на отрезке $[0, 1]$. Эти числа не могут быть пронумерованы, следовательно, их множество не образует счетное множество, оно несчетно (теорема Кантора [2]). По определению, множество, равномощное множеству всех вещественных чисел единичного отрезка числовой оси, имеет мощность **континуума** (непрерывного множества). Мощность множества континуума превышает мощность счетного множества. Любой конечный отрезок числовой оси равномощен единичному отрезку. Более того, любой конечный отрезок равномощен и всей числовой оси. Например, между отрезком $[-(\pi/2), +(\pi/2)]$ и множеством $(-\infty, +\infty)$ можно установить такое соответствие: $y = \operatorname{tg}(x)$.

Множества наибольшей мощности не существует. Это следует из того, что мощность любого множества A всегда строго меньше мощности множества всех его подмножеств $W(A)$.

1.2.4. Функции

Функциональное соответствие называется **функцией**. Функция — это тройка множеств (X, Y, F) , где F — такой график соответствия ($F \subset X \times Y$), что в нем нет пар с одинаковыми первыми и различными вторыми компонентами. Поскольку функция — частный случай соответствия, то все введенные для

соответствия определения и свойства справедливы и для функции. Однако ввиду важности понятия функции, она рассматривается здесь более подробно.

Обычно функции обозначают строчными латинскими буквами $f, g, h, \dots$. Если функция определена в X и принимает значения в Y , то пишут $f: X \rightarrow Y$. Первая компонента X упорядоченной пары (x, y) называется аргументом или независимой переменной, а вторая y — образом, зависимой переменной или значением функции. Отметим следующее отличие: само f — это множество упорядоченных пар, а запись $y = f(x)$ — это значение функции как второго компонента одной из таких пар.

Две функции f и g равны, если их области определения — одно и то же множество X и для любого $x \in X$ имеет место равенство $f(x) = g(x)$.

Если функция f устанавливает соответствие между множествами X и Y , то говорят, что функция f имеет тип $X \rightarrow Y$. Например, $\sin(x)$ имеет тип $R \rightarrow D$, где $D = [-1, +1]$. Функция $|x|$ имеет тип $R \rightarrow R_0$.

Если функция $f: X \rightarrow Y$ инъективна, то существует и функция $f^{-1}: Y \rightarrow X$ которая называется **обратной** к f функцией. Например, функция $\cos(x)$ отрезок $[0, \pi]$ взаимно-однозначно отражает на отрезок $[-1, +1]$, поэтому для нее существует обратная функция, обозначаемая $\cos^{-1}(x) = \arccos x$.

Функция $X_1 \times X_2 \times \dots \times X_n \rightarrow Y$ называется **n -местной функцией**. При этом считают, что функция имеет n аргументов, то есть $y = f(x_1, x_2, \dots, x_n)$, где $x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n$.

Пусть заданы две функции $f: X \rightarrow Y$ и $g: Y \rightarrow Z$. Функция $h: X \rightarrow Z$ называется **композицией** функций f и g (сложной функцией) и обозначается $h = f \circ g$ или $h(x) = g(f(x))$ где $x \in X$. Можно рассматривать и композицию нескольких функций $f_1 \circ f_2 \circ \dots \circ f_n$. Например: $\sin(\ln(x^2))$, $2^{\operatorname{tg} x}$ или $(x^4 + \operatorname{tg} x - \cos x)^3$ и т. п.

Если независимая переменная принимает только целые значения переменной, то функция называется **функцией целочисленного аргумента**. Например, $f(1), f(2), f(3), \dots$. Функции могут быть заданы и на произвольном дискретном множестве, например $\sin(k\pi/m)$, где k — произвольное целое число, m — фиксированное вещественное число.

Функцию, заданную во всех точках некоторого интервала, называют **функцией непрерывного аргумента**. Например, $\sin(x)$, где $x \in (0, \pi)$, определена для любой точки из указанного интервала.

Приведем простейшие свойства функций:

1. Функция $y = f(x)$, заданная на симметричном промежутке, называется **четной**, если для всех x из этого промежутка $f(-x) = f(x)$ и **нечетной**, если $f(-x) = -f(x)$. В противном случае говорят, что это — функция общего вида. Так функции x^2 , $\cos x$ — четные, функции x , x^3 , $\sin x$ — нечетные, функция $y = x + 3$ — общего вида.

2. Функция $y = f(x)$ называется **возрастающей** на некотором промежутке, если большему значению аргумента из этого промежутка соответствует большее значение функции, и **убывающей**, если большему значению аргумента соответствует меньшее значение функции. Возрастающая или убывающая на промежутке функция называется **строго монотонной**, если же она возрастает, убывает или сохраняет постоянное значение, то называется **нестрого монотонной**. Функция, график которой изображен на рис. 1.12, на интервале от a до b монотонно возрастает, от b до c монотонно убывает.

Функция $y = f(x)$, определенная при всех вещественных x , называется **периодической**, если существует такое постоянное число $L \neq 0$, что $f(x+L) = f(x)$. Наименьшее положительное число, обладающее этим свойством, называется **периодом** функции $f(x)$ и обозначается T . К периодическим относятся, в частности, тригонометрические функции.

Отображением называют всюду определенную функцию. При этом, если X — область отправления, Y — область прибытия, то говорят об отображении (всего!) множества X во множество Y .

Если же говорят, что множество X отображается на множество Y , то это означает отображение X не обязательно на все множество Y .

Способы задания функций

1. **Аналитический.** Аналитическим выражением называется символическое обозначение совокупности известных математических операций, которые производятся в определенной последовательности. Если f обозначает аналитическое выражение,

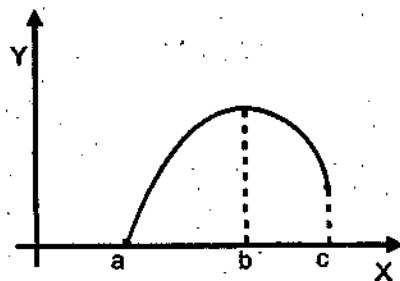


Рис. 1.12. Монотонность функции

то функция задана аналитически, например $y = \sin(4x + \ln(x^2))$. Функция может иметь разные аналитические выражения на разных подмножествах множества X , например:

$$y = |x| = \begin{cases} x, & \text{если } x \geq 0, \\ -x, & \text{если } x < 0. \end{cases}$$

2. **Табличный.** Функция определена таблицей своих значений или конечными списками пар. Например:

f:	x	1	2	3	4	5	6	7
	y	9	7	9	7	6	3	2

В этом случае вычисление значений функции сводится к непосредственному считыванию соответствующих пар. Если необходимо знать значение функции для аргументов отсутствующих в таблице, то его можно приближенно вычислить при помощи правил интерполяции или экстраполяции.

2. **Графический.** Этот способ заключается в графическом изображении пар в выбранной системе координат и может рассматриваться как обобщение табличного способа на бесконечные множества.

3. **Алгоритмический**, типичным примером которого является рекурсивный способ. Рекурсивные процедуры задают функции следующим образом: заранее определено значение функции для одного или нескольких «начальных» значений аргумента, например, $f(0)$ или $f(0)$ и $f(1)$. Значение функции при других аргументах определяется через ее значения в «предыдущих» точках. Как правило, рекурсивные процедуры задаются на множестве (подмножестве) натуральных чисел N . Например, функция $n!$ (читается « n -факториал»): $1! = 1$; $(m+1)! = m!(m+1)$. Добавим, что по определению $0! = 1$. Отличительной особенностью такого задания функции является то, что при вычислении значения функции для аргумента $x = m$, требуется предварительно вычислить значения функции во всех «предыдущих» точках. Для $m!$ — это значения функции в точках $1, 2, 3, \dots, m-1$.

Замечание. Все вышеперечисленные способы задания функции называются конструктивными. Существуют и неконструктивные способы задания, например задание функции неявно или с помощью определяющих свойств. Каждое неконструктивное определение требует доказательства существования функции с указанными свойствами.

В зависимости от элементов множеств X и Y , а также функции f тройка (X, Y, f) может иметь различные названия.

Взаимно-однозначное отображение множества $N = (1, 2, \dots, n)$ на себя, то есть функцию вида (N, N, f) , называется **подстановкой** n чисел (подстановкой n -ой степени). Подстановку часто записывают в виде двух строк. Первая содержит аргументы подстановки, а вторая — соответствующие им образы (вторые координаты). Например, подстановка четвертой степени

$$A = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{pmatrix}$$

переводит 1 в 2, 3 в 4, 5 в 6 и 7 в 8.

Числа в каждой строке подстановки расположены в определенном порядке, т. е. представляют собой некоторую **перестановку**. Меняя местами n чисел, можно получить $n!$ перестановок. Например, $(1, 2, 3, 4, 5)$, $(1, 3, 2, 4, 5)$, $(1, 2, 4, 3, 5)$, ..., $(2, 5, 3, 1, 4)$, ..., $(5, 4, 3, 2, 1)$ — всего $5! = 120$ перестановок.

Два числа образуют **инверсию** (беспорядок), если большее число стоит слева от меньшего. Например, в перестановке $(3, 1, 4, 2)$ имеется три инверсии, т. к. 3 стоит левее 1 и 2, а 4 стоит левее 2. Перестановку, в которой число инверсий равно нулю, назовем **правильной**, например, $(1, 3, 5, 7)$ или $(4, 5, 8)$ — правильные перестановки.

Количество инверсий показывает за сколько шагов данную перестановку можно привести к правильной (шагом называют обмен местами двух соседних чисел). Так, для перестановки $(3, 1, 4, 2)$ имеем: $(1, 3, 4, 2)$, $(1, 3, 2, 4)$, $(1, 2, 3, 4)$ — три шага.

Каждой подстановке можно сопоставить число инверсий в ней, которое равно сумме инверсий в верхней и нижней перестановках. Перестановка

$$B = \begin{pmatrix} 2 & 4 & 3 & 5 & 1 \\ 3 & 4 & 6 & 2 & 7 \end{pmatrix}$$

имеет $5 + 3 = 8$ инверсий. На практике часто используют подстановки, у которых первая строка является правильной. Тогда число инверсий в ней равно числу инверсий во второй строке. Например, подстановка

$$C = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 1 & 9 & 7 & 11 \end{pmatrix}$$

имеет три инверсии.

Функцию вида (X, Y, f) , где X и Y — подмножества вещественных чисел, обычно называют просто «функцией». Например, $f(x) = \sin(x)$ (читается: « f от x равно синус x »).

Функция вида (X, Y, f) , где X — множество функций, Y — множество вещественных или комплексных чисел, называется **функционалом**. Типичными примерами функционалов являются определенный интеграл $\int_a^b f(x)dx$ и связанные с ним понятия длины линии, площади плоской фигуры, объема тела (см. главу 6.5). Функционалами являются: среднее значение функции на интервале, например средняя скорость движения или среднее ускорение, скалярное произведение векторов и т. д.

Функция вида (X, Y, f) , где X, Y — множества функций, называется **оператором**. Например,

$$y(x) = \frac{d\varphi}{dx}, \text{ где } \varphi \in X, y \in Y, f = \frac{d}{dx}.$$

Глава 1.3. Отношения

1.3.1. Основные определения

Подмножество прямого произведения $M = M_1 \times M_2 \times \dots \times M_n$ называется **n -местным отношением R** на множестве M , т. е. $R \subset M$. Содержательно это означает, что рассматриваются некоторые n -ки элементов из одного и того же множества M и эти элементы находятся между собой в отношении R . Отношение может быть задано на множестве элементов любой природы. Отношения задаются в формальной или некоторой неформальной, чаще всего в высказывательной форме.

Если $n = 1$ ($M = M_1$), то отношение называется **унарным**. По существу, одноместное отношение есть подмножество M . Установить на M унарное отношение означает приписать некоторым его элементам признак R . Примеры унарных отношений: «быть студентом» на множестве людей, «быть целым числом» на множестве вещественных чисел, «быть открытым слогом» на множестве слогов и т. п.

Если $n = 2$ ($M = M_1 \times M_2$), то отношение называется **бинарным**. Эти отношения обычно записывают как $\{xRy \mid x \in M_1, y \in M_2 \text{ или } x, y \in M\}$ и говорят, что элементы x и y из множества M находятся между собой в отношении R . Например, «цвет x чаще используют при окраске, чем цвет y », «книга x находится на полке рядом с книгой y », «метод измерений x предпочтительнее метода y ».

Рассматривают тернарные ($n = 3$), тетрарные ($n = 4$), ..., n -арные отношения. Однако, мы будем в дальнейшем рассматривать как наиболее употребительные только бинарные отношения.

Для задания бинарных отношений можно применять различные формальные способы. Для отношений на конечных множествах удобно использовать списки пар или таблицы (матрицы). В этом случае, если элементы x и y находятся между собой в отношении R , то в матрице на соответствующем месте пишут 1, в противном случае пишут ноль. Например, отношение «быть больше» на множестве чисел $M = \{1, 2, 3, 4, 5, 6\}$ можно задать так:

$$c_{ij} = \begin{cases} 1, \text{ если } x_i > x_j, \text{ и } x_i, x_j \in M \\ 0 - \text{ в противном случае} \end{cases}$$

Таблица этого отношения имеет вид, представленный на рис. 1.13. Бинарные отношения удобно задавать с помощью ориентированных графов (см. главу 2.2).

x_j	1	2	3	4	5	6
x_i						
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	1	1	0	0	0	0
4	1	1	1	0	0	0
5	1	1	1	1	0	0
6	1	1	1	1	1	0

Рис. 1.13. Задание отношения «быть больше» на множестве $\{1, 2, 3, 4, 5, 6\}$

Отношение называется **полным**, если $R = M^2$. Это означает, что все элементы множества M находятся в отношении R . Например, «учиться в одной учебной группе» на множестве студентов данной учебной группы. В то время как отношение «быть сыном» не является полным на множестве всех людей. Если $M_1 = M_2 = M$, то есть $R = M^2$, то полное отношение, заданное на некотором отрезке числовой оси, геометрически соответствует квадрату со стороной M (рис. 1.14).

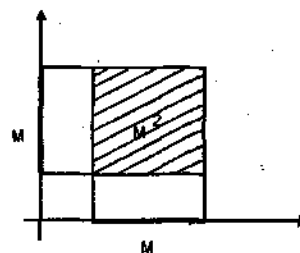


Рис. 1.14. Полное отношение

Отношение называется **тождественным**, если оно равно-сильно отношению $x = x$. Тождественное отношение обычно обозначают буквой E . Таблица такого отношения представляет собой единичную матрицу, т. е. квадратную таблицу, у которой диагональные элементы равны единице, а все остальные — нули. Примером тождественного отношения является отношение равенства на множестве вещественных чисел.

Поскольку отношения — это множества, то над ними могут выполняться все теоретико-множественные операции. В то же время могут быть определены операции специфические для отношений.

Для отношения R определена операция **обращения**. Отношение R^{-1} **обратно** к R , если $R^{-1} = \{(x, y) \mid (y, x) \in R\}$.

Переход от R к R^{-1} осуществляется взаимной перестановкой компонент каждой упорядоченной пары. Пример обратных отношений: « x_i старше x_j » и « x_j старше x_i ». Композицией двух отношений R и P , заданных на множестве M , называется отношение $Q \equiv R \cdot P$ такое, что для любых $x_i, x_k \in M$, существует хотя бы один элемент $x_j \in M$ такой, что если $(x_i, x_j) \in R$ и $(x_j, x_k) \in P$, то $(x_i, x_k) \in Q$. Можно доказать важное свойство композиции: $(R \cdot P)^{-1} = P^{-1} \cdot R^{-1}$ [4].

Пример. Пусть заданы отношения в виде таблиц

	Y_1	Y_2	Y_3	Y_4	Y_5		Z_1	Z_2	Z_3	Z_4	
$R =$	X_1	0	1	0	0	1	Y_1	1	0	1	0
	X_2	1	0	0	1	0	Y_2	1	0	1	0
	X_3	0	0	0	1	0	Y_3	0	1	0	1
							Y_4	0	1	0	0
							Y_5	1	0	0	0
							Y_6	0	0	1	1

Эти же отношения удобно представить в виде списков пар: для R это пары $\{(X_1, Y_2), (X_1, Y_5), (X_2, Y_1), (X_2, Y_4), (X_3, Y_4)\}$, соответственно, для P это пары $\{(Y_1, Z_1), (Y_1, Z_3), (Y_2, Z_1), (Y_2, Z_3), (Y_3, Z_2), (Y_3, Z_4), (Y_4, Z_2), (Y_5, Z_1), (Y_6, Z_3), (Y_6, Z_4)\}$. Композицией двух отношений R и P будет отношение $Q \equiv R \cdot P$, состоящее из следующих пар: (X_1, Z_1) — так как существует элемент Y_2 такой, что он входит в пары (X_1, Y_2) и (Y_2, Z_1) ; (X_1, Z_3) — так как существует элемент Y_5 такой, что он входит в пары (X_1, Y_5) и (Y_5, Z_3) (таким же свойством обладает и элемент Y_2), и так далее. Таблица, соответствующая композиции $R \cdot P$:

		Z_1	Z_2	Z_3	Z_4
$R \cdot P =$	X_1	1	0	1	0
	X_2	1	1	1	0
	X_3	0	1	0	0

1.3.2. Свойства отношений

Рефлексивность. Отношение R , заданное на множестве M , называется рефлексивным, если для любого $x_i \in M$ имеет место

$x_i R x_i$. Формально рефлексивность можно задать так: $E \subset R$. Это означает, что рефлексивное отношение всегда выполняется между объектом и им самим. Наиболее яркими примерами рефлексивных отношений являются самообслуживание и равенство.

Антирефлексивность. Отношение R , заданное на множестве M , антирефлексивно, если $R \cap E = \emptyset$. В антирефлексивных отношениях из условия $x_i R x_j$ следует, что $x_i \neq x_j$. Примеры антирефлексивных отношений: «быть старше», «быть меньше», «быть дочерью» и др.

Симметричность. Отношение R , заданное на множестве M , называется симметричным, если при выполнении соотношения $x_i R x_j$ одновременно выполняется и соотношение $x_j R x_i$. Формально соотношение симметрично, если $R = R^{-1}$. Например, отношения «стоять рядом на полке» на множестве книг или «быть родственниками» на множестве людей симметричны.

Асимметричность. Отношение R , заданное на множестве M , называется асимметричным, если $R \cap R^{-1} = \emptyset$. В асимметричных отношениях из двух соотношений $x_i R x_j$ и $x_j R x_i$ может выполняться не более одного (одно или не одного). Пример асимметричного отношения: «быть отцом» (если x_i — отец x_j , то x_j не может быть отцом x_i).

Антисимметричность. Отношение R , заданное на множестве M , обладает свойством антисимметричности, если $R \cap R^{-1} \subset E$. Это означает, что если для $x_i, x_j \in M$ одновременно выполняются соотношения $x_i R x_j$ и $x_j R x_i$, то $x_i = x_j$. Антисимметричны все отношения нестрогого порядка: «быть не больше», «быть не выше» и т. п.

Транзитивность. Отношение R , заданное на множестве M , транзитивно, если для любых $x_i, x_j, x_k \in M$ из выполнимости соотношений $x_i R x_j$ и $x_j R x_k$ следует $x_i R x_k$. Формально это можно записать так: $R \cdot R \subset R$. Про отношение, не обладающее таким свойством, говорят, что оно не транзитивно. Например, отношение «стоять рядом на полке» не транзитивно. Действительно, пусть тома некоторой энциклопедии стоят в порядке возрастания номеров томов. Тогда, если первый том стоит рядом со вторым, а второй рядом с третьим, то, очевидно, что первый не стоит рядом с третьим.

Все общие свойства отношений можно разбить на три группы: рефлексивности (каждое отношение рефлексивно или антирефлексивно), симметричности (отношение всегда или симметрично или асимметрично, или антисимметрично), транзитивности (отношение всегда или транзитивно или нетранзитивно).

тивности (каждое отношение транзитивно или не транзитивно). Отношениям, обладающим определенным набором свойств, присвоены специальные названия. Рассмотрим подробнее некоторые из них.

1.3.3. Отношение эквивалентности

Отношение R обладающее свойствами рефлексивности, симметричности и транзитивности называется **отношением эквивалентности**. Для эквивалентных отношений вместо записи $x_i R x_j$ обычно пишут $x_i \sim x_j$ (читается: « x_i эквивалентно x_j »). Эквивалентными отношениями являются: «быть конгруэнтными» на множестве плоских треугольников, «быть одного размера» на множестве образцов обуви, «быть родственниками» на множестве людей и т.п.

Введение отношения эквивалентности R на множестве M определяет разбиение всех элементов этого множества на классы эквивалентности $M_1, M_2, \dots, M_n$. Множество всех классов эквивалентности $\{M_1, M_2, \dots, M_n\}$ образует **фактор множество** множества M и обозначается M/R . При этом каждый элемент данного класса является полномочным представителем этого класса. Совокупность по одному и только по одному представителю каждого класса называется **системой представителей** соответствующего отношения эквивалентности. Примером введения отношения эквивалентности и образования системы представителей может служить формирование представительного органа власти на основе выборов.

1.3.4. Отношение толерантности

Отношение τ , заданное на множестве M называется **отношением толерантности**, если оно рефлексивно, симметрично и не транзитивно. Обозначение: $x_i \tau x_j \mid x_i, x_j \in M$. Очевидно, что если потребовать транзитивность всех пар элементов из M , то получим эквивалентное отношение. Следовательно, толерантность может рассматриваться как расширение эквивалентности. Эквивалентность в смысле равенства, толерантность в смысле сходства, похожести. Содержательно толерантность означает следующее. Предполагается, что объект находится в данном отношении сам с собой (рефлексивность), сходство двух объектов

не зависит от порядка сравнения (симметричность), но если первый объект сходен со вторым, а второй сходен с третьим, то не обязательно, что первый был сходен с третьим.

Толерантность позволяет формализовать интуитивные представления о сходстве объектов, их похожести в чем-то. Например, отношение $\tau =$ «быть на расстоянии не более r », заданное на множестве точек на плоскости. Рис.1.15 иллюстрирует этот пример. Точка A отстоит от B и C , а точка B от D и C не более чем на r . В то время как A находится от D на расстоянии значительно большем r .

Другие примеры толерантности: «отличаться не более чем одной буквой» на множестве слов из четырех букв (например: арка, река, рака), «совпадать по количеству цветов» (белый + синий + красный, зеленый + красный + желтый) на множестве шарфов, «совпадать по году постройки» на множестве зданий города и т. д.

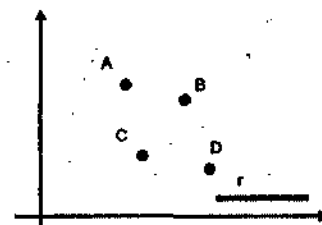


Рис. 1.15 Толерантность точек на плоскости

1.3.5. Отношение порядка

Отношение R , обладающее свойствами рефлексивности, антисимметричности и транзитивности, называется **отношением порядка**. Если на данном множестве введено отношение порядка, то это множество называется **упорядоченным**. В этом случае вместо $x_i R x_j$ пишут $x_i \geq x_j$. Множество **совершенно упорядочено**, если для любых двух элементов x_i и x_j из множества M имеет место либо $x_i \geq x_j$, либо $x_j \geq x_i$. В противном случае говорят, что множество **частично упорядочено**. Например, отношение «быть выше» на множестве деревьев — совершенно упорядочено, а отношение «быть делителем» на множестве целых чисел — частично упорядочено.

Пусть каждому элементу x из множества M по некоторому правилу f поставлено в соответствие вещественное число $f(x)$ — вес элемента x . Введение веса для каждого элемента позволяет упорядочить их по мере возрастания (убывания) весов, а затем сравнивать элементы в соответствии с присвоенным весом. Примерами упорядочения посредством введения весов являются:

присвоению каждому товару его цены, каждому станку его надежности, каждому телу его веса, объема и т. п. Как будет показано ниже, взвешивание вариантов решений посредством формирования комплексного показателя качества является одним из самых распространенных способов решения проблемы выбора на множестве разнокачественных признаков.

Если отношение обладает свойствами антирефлексивности, асимметричности и транзитивности, то оно называется **отношением строгого порядка** (обозначается $x_i > x_j$). Примером отношения строгого порядка является порядок букв в фиксированном алфавите. Упорядочение букв в алфавите позволяет, в свою очередь, упорядочить слова в словарях (лексикографическое упорядочение слов).

Более подробно с изложенным материалом можно познакомиться в изданиях [3, 4, 5].

1.4. Задачи и упражнения к разделу 1

1. Какие из приведенных множеств заданы корректно и сколько в них содержится элементов?

$$A = \{1, 3, 5, 7\};$$

$$B = \{175, 35, 7, 1\};$$

$$C = \{1, 37, 5, 17, 5, 73, 1, 5\};$$

$$D = \{7351\};$$

$$E = \{71, 35, 17, 53\}.$$

Ответ: множество C задано некорректно, так как элементы 1 и 5 включены более одного раза. Множества A, B и E содержат по четыре элемента, а множество D содержит единственный элемент.

2. Равны ли множества:

$$a) A = \{1, 2, \{1, 2\}\} \text{ и } B = \{1, 2\};$$

$$б) A = \{2, 4, 6\} \text{ и } B = \{4, 6, 2\};$$

$$в) A = \{2, \{3, 4\}, 4\} \text{ и } B = \{2, 3, \{4\}\}.$$

Ответ: равны множества в случае б).

3. Какие из приведенных соотношений заданы неверно и почему?

$$a) 7 \in \{1, 3, 5, 7\};$$

$$б) 37 \in \{1, 3, 5, 7\};$$

$$в) 71 \in \{1, 3, 5, 7\};$$

$$г) 3 \in \{1, \{3, 5\}, 7\};$$

$$д) x \in \{\operatorname{tg}(x), \cos(x)\}.$$

Ответ: Неверно задано соотношение б) так как во множество $\{1, 3, 5, 7\}$ элемент 37 не входит. По аналогичной причине неверно заданы соотношения г) и д).

$$4. \text{ Доказать, что } (A \setminus B) \cap C = (A \cap B) \setminus (A \cap C).$$

Указания: исключить в правой и левой частях равенства знак «\».

5. Упростить выражение

$$(A \cap B \cap C \cap D) \cup (A \cap C) \cup (B \cap C) \cup (C \cap D)$$

Ответ: C.

6. Образуйте множество праздничных дней в 2000 году. Пересекается ли оно с множеством пятниц этого же года?

7. Какими свойствами обладает соответствие шахматных фигур и клеток шахматной доски?

8. Какими свойствами обладают отношения:

«быть рядом» на множестве книг на полке;

«x делится на y» на множестве целых чисел;

«быть знакомым» на множестве людей;

«быть одного возраста» на множестве людей;

«находиться выше» на множестве ступеней лестницы?

9. Доказать, что множество всех равнобедренных треугольников равно множеству всех равноугольных треугольников.

Раздел 2. ВВЕДЕНИЕ В ДИСКРЕТНУЮ МАТЕМАТИКУ

Дискретная математика, или дискретный анализ, — область математики, занимающаяся изучением свойств структур конечного или не более чем счетного характера. В настоящее время существует целый ряд задач, которые не могут быть решены методами предлагаемыми непрерывной математикой. В первую очередь это задачи, связанные с развитием вычислительной техники и цифровых систем связи, изучением социальных, организационно-технических и экономических систем, математизации гуманитарных наук.

Разделение математики на дискретную и непрерывную весьма условно. Выделение дискретной математики как важного современного направления развития науки связано в первую очередь с характерными для нее объектами, специфика которых требует отказа от таких основополагающих понятий классической математики, как предел и непрерывность. А это приводит к тому, что для многих задач дискретного анализа средства классической математики оказываются часто малоприменимы.

В рамках дискретной математики обычно рассматриваются: комбинаторный анализ, теории графов и кодирования, а также ряд разделов, имеющих общематематический характер.

Глава 2.1. Математическая логика

Основу современной математической логики составляют *исчисление высказываний* (первая логическая теория) и *исчисление предикатов* (вторая логическая теория).

Исчисление высказываний оперирует утверждениями, выступающими как единое целое, не рассматривая их субъективно-предикативную структуру. Сложные высказывания получаются из более простых при помощи логических связей, формализуемых соответствующими правилами. Истинность или ложность полученного сложного высказывания зависит только от истинности или ложности составляющих высказываний и не зависит от их содержания.

Все понятия и теоремы исчисления высказываний употребляются в более широкой логической теории, называемой

исчислением предикатов. В этой теории рассматривается внутренняя структура простых высказываний: в высказывании выделяют подлежащее и сказуемое (предикат), и если на место подлежащего затем ставить другое подлежащее, то получится другое высказывание.

Тем самым предикат представляет собой высказывательную форму, определенную на множестве объектов, которые могут выступать в качестве подлежащих.

2.1.1. Алгебра высказываний

Логика высказываний — первая логическая теория — лежит в основе всех других разделов логики. Схема построения этой теории не отличается от схемы построения любой другой математической теории: берется некоторый класс объектов — первичных понятий, не интерпретируемых в данной теории, для них полагаются истинными некоторые свойства, вводятся отношения и операции над этими объектами. Так же, как и в теории множеств, основные понятия не определяются, а обычно разъясняются на примерах.

К основным понятиям математической логики относится высказывание.

Под высказыванием в классическом смысле слова будем понимать всякое утверждение (повествовательное предложение), про которое всегда определено и объективно можно сказать, является оно истинным или ложным. Например, « $5 - 3 = 2$ » или «В неделе семь дней» — истинные высказывания. Наоборот, « $5 > 8$ » или «В русском языке 35 букв» — ложные высказывания. Синонимами слова «высказывание» можно считать: логическое высказывание, булевское выражение, суждение, утверждение и т. п. Фразы: «Ура!», «Который час?» — не являются высказываниями.

Если высказывание истинное, то ему предписывается значение «истина» (другие обозначения: «1», «ДА», «И», «+», «true»). Ложному высказыванию предписывается значение «ложь» (другие обозначения: «0», «НЕТ», «Л», «-», «false»). Совокупность возможных значений высказывания образует множество истинности $\{0, 1\}$ и $\{И, Л\}$.

Введем в рассмотрение два вида высказываний: *простые и составные* (сложные). Под простым будем понимать выска-

высказывание, которое не может быть разбито на более простые высказывания. Про него всегда однозначно можно сказать, что оно истинно или ложно, не интересуясь его структурой. Из простых высказываний при помощи так называемых логических операций можно строить сложные высказывания, которые всегда только истинны или только ложны. При этом истинность или ложность сложного высказывания зависит исключительно от истинности составляющих его простых высказываний. Изучение такого рода зависимостей является основной задачей алгебры высказываний.

Рассмотрим группу простых высказываний: «В Архангельской области летний день короче зимнего» — ложь; «В Архангельской области зимой обычно лежит снег» — истина; «Реки в Архангельской области в январе покрыты льдом» — истина; «Прогулочные катера могут ходить по замерзшим рекам» — ложь; «По снегу можно кататься на лыжах» — истина. Из этих простых высказываний формально можно построить, например, такие сложные высказывания: «В январе в Архангельской области лежит снег и по рекам не ходят прогулочные катера» — истинное высказывание; «В январе в Архангельской области зимний день короче летнего и реки не покрыты льдом, поэтому можно кататься на лыжах» — ложное высказывание. Еще раз подчеркнем, что в классической математической логике абстрагируются от содержания высказываний.

Тем самым любое сколь угодно сложное и даже *содержательно неоднозначно* интерпретируемое или вовсе абсурдное высказывание *формально всегда истинно или ложно*.

В практической деятельности часто встречаются утверждения, истинность которых не может быть установлена объективно и однозначно. Эти утверждения типа «хорошая погода», «справедливый закон» и т. п., могут рассматриваться как высказывания только в том случае, если четко определена истинность исходных простых высказываний, а данное высказывание является свернутой формой простых и известен контекст, в рамках которого проводится исследование. Так, например, суждению «сейчас на море хорошая погода» может быть приписана различная истинность в зависимости от того, идет ли речь о рыбной ловле, купании, катании на весельной лодке или круизе на океанском лайнере.

Высказывания будем обозначать заглавными латинскими буквами: $A \equiv$ « $5 + 7 = 12$ », $B \equiv$ «сегодня вторник», $C \equiv$ «если

студент успешно сдал сессионные экзамены, то переводится на следующий курс и будет получать стипендию». Использование таких обозначений упрощает запись процедур оперирования высказываниями и делает формулы математической логики похожими на выражения, традиционно встречающиеся в математике.

2.1.2. Операции над высказываниями

Для построения из данных высказываний новых высказываний вводятся определенные правила, которые называются операциями над высказываниями.

Отрицание высказывания. Для каждого высказывания A может быть сформировано новое высказывание $\neg A$ — отрицание высказывания A , которое истинно, когда A ложно, и ложно, когда A истинно. Символ « $\neg$ » соответствует логическому союзу «не».

$\neg A$ читается «не A » или «не верно, что A ». Например, если $A \equiv 3 + 5 = 8$ — истинное высказывание, то $\neg A \equiv 3 + 5 \neq 8$ — ложное высказывание (отрицание A), или, если $B \equiv$ «в комнате холодно», $\neg B \equiv$ «в комнате не холодно». Отметим, что высказывание «в комнате жарко» не является отрицанием B .

Логические операции удобно задавать в виде таблиц, называемых таблицами истинности. Для операции отрицания таблица истинности имеет вид:

Отрицание — одноместная (или унарная) операция. Последующие операции — двухместные (или бинарные).

A	$\neg A$
1	0
0	1

Конъюнкция высказываний. Конъюнкцией высказываний A и B называется высказывание $C = A \wedge B$, которое истинно только в том случае, когда и A и B одновременно истинны. Выражение $A \wedge B$ читается « A и B ». Операцию конъюнкции определяет следующая таблица истинности:

Пусть $A \equiv$ «12 делится на 3» и $B \equiv$ «12 делится на 4». Тогда формула $A \wedge B$ имеет смысл: «12 делится на 3 и на 4».

A	B	$C = A \wedge B$
1	1	1
1	0	0
0	1	0
0	0	0

Операцию конъюнкции можно определить и для нескольких высказываний как связку высказываний, объединенных союзом «и». Конъюнкция из n высказываний — новое высказывание, причем высказывание

$$A = \bigwedge_{i=1}^n A_i \equiv A_1 \wedge A_2 \wedge \dots \wedge A_n$$

имеет значение «истина», если и A_1 , и A_2 , ..., и A_n истинны. Во всех других случаях эта конъюнкция имеет значение «ложь». Пусть, например, $A_1 \equiv \langle 5 > 3 \rangle$, $A_2 \equiv \langle 8 = 3 \rangle$, $A_3 \equiv \langle \text{отец старше сына} \rangle$, $A_4 \equiv \langle \text{Мурманск севернее Смоленска} \rangle$. Тогда высказывание $A_2 \wedge A_3 \wedge A_4 \equiv \langle 8 = 3, \text{ и отец старше сына, и Мурманск севернее Смоленска} \rangle$ — ложное высказывание. В то время как $A_1 \wedge A_3 \wedge A_4 \equiv \langle 5 > 3, \text{ и отец старше сына, и Мурманск севернее Смоленска} \rangle$ — истинное высказывание.

Дизъюнкция высказываний. Дизъюнкцией высказываний A и B называется высказывание $D = A \vee B$, которое ложно только тогда, когда и A и B ложны одновременно. Дизъюнкция имеет значение «истина», если хотя бы одно из высказываний, входящее в дизъюнкцию, является истинным. Выражение $A \vee B$ читается « A или B ». Пусть $A \equiv \langle 7 < 9 \rangle$ и $B \equiv \langle 3 + 5 = 8 \rangle$. Тогда $A \vee B \equiv \langle 7 < 9 \text{ или } 3 + 5 = 8 \rangle$. Таблица истинности для дизъюнкции имеет вид:

A	B	$D=A \vee B$
1	1	1
1	0	1
0	1	1
0	0	0

Пусть, например, $A \equiv \langle \text{сегодня облачно} \rangle$, $B \equiv \langle \text{идет дождь} \rangle$. Тогда $A \vee B \equiv \langle \text{сегодня облачно, или идет дождь} \rangle$. Это выражение ложно только в том случае, если сегодня и не облачно, и дождь не идет. Если же «сегодня облачно и дождь не идет» или «сегодня не облачно и дождь идет», или «сегодня облачно и идет дождь», то все эти высказывания, с точки зрения алгебры высказываний, истинны.

Операцию дизъюнкции можно определить для нескольких высказываний как связку высказываний, объединенных союзом «или»:

$$A = \bigvee_{i=1}^n A_i \equiv A_1 \vee A_2 \vee \dots \vee A_n.$$

В этом случае высказывание A истинно, если истинно хотя бы одно из высказываний, входящих в связку.

Импликация высказываний. Импликацией высказываний A и B называется высказывание $F = A \rightarrow B$, которое ложно только в том случае, когда A — истинно, а B — ложно. Во всех других случаях импликация $A \rightarrow B$ имеет значение «истина». Символ « $\rightarrow$ » соответствует логическому союзу: «если... то...». Например, A — «целое число делится на 4», B — «целое число делится на 2». Их импликация $A \rightarrow B$ есть выражение: «если целое число делится на 4, то оно делится на 2».

Таблица истинности для импликации имеет вид:

A	B	$F=A \rightarrow B$
1	1	1
1	0	0
0	1	1
0	0	1

Для иллюстрации содержательного смысла импликации рассмотрим следующий пример:

$A \equiv \langle \text{папа завтра получит премию} \rangle$, $B \equiv \langle \text{папа завтра купит сыну велосипед} \rangle$. Тогда импликация $A \rightarrow B$ может быть сформулирована так: «если папа завтра получит премию, то купит сыну велосипед». Пусть A и B истинны. Тогда папа, получив премию, покупает сыну велосипед. Естественно считать это истинным высказыванием. Когда же папа не купит сыну велосипед (B — ложно), получив премию (A — истинно), то это, мягко говоря, не логичный поступок, а импликация имеет значение «ложь». Если же папа не получит премию (A — ложно), но купит велосипед (B — истинно), то результат положителен. В том случае, если и A и B ложны, т. е. не получив премии папа не купит велосипед, обещание не нарушено: результат можно считать истинным.

Эквивалентность высказываний. Эквивалентностью высказываний A и B называется высказывание $G = A \leftrightarrow B$, которое истинно, когда высказывания A и B оба истинны или оба ложны. Символ логической эквивалентности « $\leftrightarrow$ » соответствует связке «тогда и только тогда».

Пример. Пусть $A \equiv$ «число $3n$ является четным», $B \equiv$ «число n является четным». Высказывание «число $3n$ является четным тогда и только тогда, когда n — четное число» есть эквивалентность высказываний A и B .

Эквивалентность высказываний может быть задана следующей таблицей истинности:

A	B	$G=A \leftrightarrow B$
1	1	1
1	0	0
0	1	0
0	0	1

Характерной особенностью операций над высказываниями является введение логических союзов с точно определенным смыслом, не допускающим никакой двусмысленности в толковании этих символов. В этом заключается основное отличие логического языка от естественного. Таким образом, математическая логика применима не для любых высказываний, а только для таких, которые допускают четкую оценку в двоичной системе «истина — ложь». Для преодоления такого рода ограничений в рамках нечеткой математики разрабатывается нечеткая логика.

Рассмотрим пример. Пусть заданы высказывания: $A \equiv$ «вчера шел дождь», $B \equiv$ «сегодня пасмурно», $C \equiv$ «сегодня светит солнце», $D \equiv$ «сегодня идет снег». Построим из них следующую формулу: $A \rightarrow (B \vee C)$. На естественном языке это можно записать так: «если вчера шел дождь, то сегодня пасмурно или светит солнце». Формула $C \rightarrow (B \vee D)$ может быть прочитана так: «если сегодня светит солнце, то не верно, что пасмурно или идет снег». Заметим, что если перевод с обычного языка на символический не встречает затруднений, то при переводе

с символического на естественный часто появляется некоторая «корявость» фраз.

Если в выражении встречаются различные логические операции, то возникает проблема очередности их выполнения. В качестве естественного порядка (выполняемого *поочередно слева направо*) выберем следующую последовательность: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$. Это означает, что сначала выполняются операции отрицания, затем конъюнкции и т. д. Для нарушения такого порядка служат скобки.

Рассмотрим пример. Пусть высказывания A и B имеют значения «истина», а высказывания C и D — «ложь». Тогда формула $A \rightarrow B \wedge C \leftrightarrow D \vee A$ имеет значение «ложь», т. к.:

1. $\neg D$ — «истина»;
2. $B \wedge C$ — «ложь»;
3. $(\neg D) \vee A$ — «истина»;
4. $A \rightarrow (B \wedge C)$ — «ложь»;
5. $(A \rightarrow (B \wedge C)) \leftrightarrow (\neg D \vee A)$ — «ложь».

Введя скобки, получим формулу $A \rightarrow B \wedge (C \leftrightarrow \neg D) \vee A$, которая уже имеет другое значение — «истина». Действительно:

1. $\neg D$ — «истина»;
2. $C \leftrightarrow (\neg D)$ — «ложь»;
3. $B \wedge (C \leftrightarrow \neg D)$ — «ложь»;
4. $(B \wedge (C \leftrightarrow \neg D)) \vee A$ — «истина»;
5. $A \rightarrow (B \wedge (C \leftrightarrow \neg D) \vee A)$ — «истина».

Если в выражении присутствуют арифметические операции, операции сравнения и логические операции, то порядок старшинства операций следующий:

* сначала выполняются арифметические операции (в порядке старшинства арифметических операций: первыми все операции умножения и деления, потом операции сложения и вычитания);

* $=$, $\neq$, затем операции сравнения (в том порядке, в каком они встречаются в выражении): $>$, $\geq$, $<$, $\leq$;

* наконец, логические операции, причем первой везде выполняется операция конъюнкции, а потом дизъюнкции.

Использование различных операций позволяет в удобной аналитической форме задавать различные множества.

Например, множество точек A , заштрихованное на рис. 2.1, может быть задано следующей формулой:

$$A = \{(x, y) \mid (y \geq 1 - x) \wedge (x^2 + y^2 \leq 1) \wedge (x, y) \in \mathbb{R}^2\}.$$

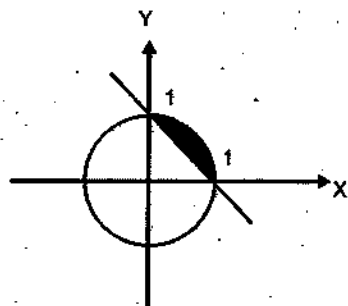


Рис. 2.1.

Множество A :

$$\{(x, y) \mid (y \geq 1 - x) \wedge (x^2 + y^2 \leq 1) \wedge (x, y) \in \mathbb{R}^2\}$$

Система операций Σ называется **полной**, если всякая формула эквивалентна некоторой формуле, в которую входят только операции из системы Σ . Можно показать [6], что система введенных пяти операций (отрицания, конъюнкции, дизъюнкции, импликации и эквивалентности) полная. Между тем, эта система операций, вообще говоря, избыточна, так как одни логические операции могут быть выражены через другие. Например, импликация и эквивалентность можно выразить через отрицание,

конъюнкцию и дизъюнкцию следующим образом:

$$A \rightarrow B \equiv \neg A \vee B$$

$$A \leftrightarrow B \equiv (A \rightarrow B) \wedge (B \rightarrow A).$$

Исключение «лишних» операций приводит к тому, что формулы становятся более длинными и менее обозримыми. Аналогичную ситуацию можно заметить, например, в арифметике, где вместо умножения можно использовать сложение (сравните $2 \times 5 = 2 + 2 + 2 + 2 + 2$). Другой пример. Находясь в рамках русского алфавита, можно было бы обойтись словами состоящими не более чем из пяти букв. Однако мы употребляем «длинные» слова (переустройство, невостребованность и т. п.), в то время как многие «короткие» слова ничего не обозначают и являются словами чисто теоретически (лыб, фкинн, сфют, алл и др.).

2.1.3. Булевы функции

Всякую формулу логики высказываний можно рассматривать как некоторую функцию: каждая буква (высказывание) может принимать одно из двух значений «истина» или «ложь», при этом сложное высказывание, заданное этой формулой, также

может быть истинным или ложным. Так, формула

$$(A \rightarrow B) \leftrightarrow \neg C \vee (B \wedge A) = f(A, B, C)$$

выражает функцию от переменных A , B и C .

Такого рода функции называются **булевыми**, а их аргументы — **булевыми переменными**. Аргументы булевых функций могут представлять собой сокращенные обозначение некоторых конкретных высказываний. Тогда функция обозначает сокращенную запись некоторого сложного высказывания. Например, $A \equiv \text{«}5 < 3\text{»}$, $B \equiv \text{«}3n \text{ делится на } 3\text{»}$, $C \equiv \text{«} \text{Мурманск севернее Смоленска} \text{»}$. В этом случае $f(A, B, C) = \neg A \rightarrow (B \wedge C) \equiv \text{«если } 5 \geq 3, \text{ то } 3n \text{ делится на } 3 \text{ и Мурманск севернее Смоленска} \text{»}$. Сравните с известной формулой физики $F = m \cdot a$, где m — масса тела, a — его ускорение, а F — сила, вызвавшая это ускорение.

Буквы в булевых функциях могут выступать в качестве переменных. Подставляя вместо них любые высказывания, можно по формуле вычислить соответствующее значение функции. Действительно, если в формуле $f(X, Y, Z) = (X \vee Y) \rightarrow (X \wedge Z)$ X — «истина», Y — «ложь», Z — «истина», то $f(X, Y, Z)$ — «истина». Если же Z будет иметь ложное значение, то $f(X, Y, Z)$ поменяет значение на противоположное и будет «ложью».

Целый ряд булевых функций обладают тем свойством, что они принимают одни и те же значения при любых значениях истинности аргументов. Такие формулы называются **тождественно истинными**. Например, при любых X и Y истинны формулы $X \vee \neg X$ и $X \rightarrow (Y \rightarrow X)$.

Тождественно ложные функции при любых значениях аргументов имеют значение «ложь». Так, формулы $X \wedge \neg X$ и $(X \rightarrow (Y \rightarrow X)) \leftrightarrow (Y \wedge \neg Y)$ всегда имеют значение «ложь».

Операции над множествами обладают рядом свойств. Поскольку эти свойства справедливы для всех высказываний, то их называют тождествами, а иногда и законами. Некоторые из них получили специальное название. Приведем основные законы без доказательств, т. к. справедливость каждого из них можно доказать, построив для левой и правой части соответствующие таблицы истинности.

1. Коммутативность:

$$A \wedge B \equiv B \wedge A,$$

$$A \vee B \equiv B \vee A.$$

2. Ассоциативность:

$$A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C,$$

$$A \vee (B \vee C) \equiv (A \vee B) \vee C.$$

3. Дистрибутивность:

$$A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C),$$

$$A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C).$$

4. Законы де Моргана:

$$\neg(A \wedge B) \equiv \neg A \vee \neg B,$$

$$\neg(A \vee B) \equiv \neg A \wedge \neg B.$$

5. Закон поглощения:

$$A \vee (A \wedge B) \equiv A \wedge (A \vee B) \equiv A.$$

6. Закон идемпотентности:

$$A \wedge A \equiv A \vee A \equiv A.$$

$$7. A \wedge \text{«истина»} \equiv A,$$

$$A \vee \text{«ложь»} \equiv A.$$

8. Закон противоречия:

$$A \wedge \neg A \equiv \text{«ложь»}.$$

9. Закон исключения третьего:

$$A \vee \neg A \equiv \text{«истина»}.$$

10. Закон двойного отрицания:

$$\neg \neg A \equiv A.$$

Эти и целый ряд других тождеств имеют фундаментальное значение, а также позволяют преобразовывать сложные выражения, например, с целью их упрощения.

Пример. Упростить выражение, используя тождественные преобразования

$$\begin{aligned} & X \wedge Y \vee \neg X \wedge (Y \vee X \wedge Z) \wedge \neg(X \wedge (\neg Y \vee Z) \vee Y \wedge Z) = \\ & X \wedge Y \vee (\neg X \wedge Y \vee \neg X \wedge X \wedge Z) \wedge (\neg X \vee \neg(\neg Y \vee Z)) \wedge \neg(Y \wedge Z) = \\ & X \wedge Y \vee \neg X \wedge Y \wedge (\neg X \vee Y \wedge Z) \wedge (\neg Y \vee Z) = \\ & X \wedge Y \vee \neg X \wedge Y \wedge (\neg X \wedge \neg Y \vee Y \wedge \neg Z \vee X \wedge Z \vee Y \wedge Z) = \\ & X \wedge Y \vee \neg X \wedge Y \wedge Z = \\ & Y \wedge (X \vee \neg X \vee Z). \end{aligned}$$

2.1.4. Методы доказательств

При построении любой теории выделяется некоторый набор высказываний, истинность которых постулируется. Такие вы-

сказывания называются аксиомами теории. Введение аксиом приводит к тому, что, исходя из них, чисто логическим путем может быть установлена истинность некоторых других высказываний, называемых теоремами. Последовательность высказываний рассматриваемой теории, каждое из которых либо является аксиомой, либо выводится из одного или более предыдущих высказываний этой последовательности по логическим правилам вывода, называется **доказательством**. Высказывание, которое можно доказать, называется **теоремой**.

Формально каждая теорема может быть выражена в форме импликации $A \rightarrow B$, где посылка A называется условием теоремы, а следствие B — заключением. Теорема верна, если выражающая ее импликация тождественно истинна, т. е. является **тавтологией**. Тавтологии рассматривают как некоторые логически истинные схемы рассуждений. В этой связи тавтологии играют роль законов, определяющих построение правильных умозаключений. Существует бесконечное множество тавтологий. Некоторые из них легли в основу методов доказательства.

Рассмотрим ряд наиболее известных методов доказательств.

Метод цепочек импликаций состоит в том, что из посылки A выстраивается цепочка из n импликаций, последним высказыванием в которой является заключение теоремы B , т. е.

$$A \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_{n-1} \rightarrow B$$

В основе этого метода лежит **закон цепного высказывания**, или **закон силлогизма**: $(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$.

Метод от противного. Используя этот метод, вместо доказательства прямого следствия «из A следует B » доказывают, что из «не B » следует «не A ». Этот метод основан на законе **контрапозиций**, имеющим следующий вид: $(A \rightarrow B) \leftrightarrow (\neg B \rightarrow \neg A)$.

Метод необходимого и достаточного. Теорема формулируется так: «Чтобы имело место A , необходимо и достаточно выполнение B ». Доказательство такого вида теоремы распадается на две части:

а) доказываются, что если имеет место A , то справедливо B (B необходимо для A);

б) если имеет место B , то имеет место и A (B достаточно для A).

Доказательство таким методом базируется на **законе тавтологии**. $(P \leftrightarrow Q) \leftrightarrow ((P \rightarrow Q) \wedge (Q \rightarrow P))$

В дальнейшем мы часто будем использовать эти методы доказательств теорем.

2.1.5. Алгебра предикатов

Мы знаем, что выражение $5 > 3$ — истинное высказывание, а выражение «в неделе восемь дней» — ложное высказывание. Истинность этих и всех других высказываний определяется объективно и однозначно. Является ли выражение « $n > 8$ » высказыванием? По форме оно похоже на высказывание, но не является таковым, поскольку невозможно сказать истинно оно или ложно. Все зависит от того, что представляет собою объект n . Выражение « $x^k + y^k = z^k$ » — также не высказывание, но может стать таковым, если определить, что такое k , x , y , z и каковы их значения. Таким образом, существуют выражения похожие по форме на высказывания, но не являющиеся ими. Такие выражения называют предикатами.

Предикатом $P(x_1, x_2, \dots, x_n)$, заданным на множествах $M_1, M_2, \dots, M_n$, называется функция P , отображающая их прямое произведение на двоичное множество, т. е. $P: M_1 \times M_2 \times \dots \times M_n \rightarrow \{0, 1\}$. Множество M называется **предметной областью** предиката, $x_1, x_2, \dots, x_n \in M$ называются **предметными переменными** или **термами**. Предикат представляет собой логическую функцию, принимающую, как и булевская функция, значение «истина» или «ложь», когда ее предметные переменные принимают определенные значения.

Рассмотрим примеры. « $x^2 - 5x + 6 = 0$ » — одноместный предикат на множестве комплексных чисел, при этом, например, если $x = 2$, то « $2^2 - 5 \cdot 2 + 6 = 0$ » — истинное высказывание, а положив $x = 7$, получим « $7^2 - 5 \cdot 7 + 6 = 0$ » — «ложь». Выражение « X — брат Y » — двухместный предикат, заданный на множестве людей. Здесь термы X и Y указывают места, на которые нужно поставить имена двух людей, чтобы получить правильно построенное высказывание. Очевидно, что X — лицо мужского пола, а Y — может выбираться из всего множества людей.

Всякий предикат $P(x_1, x_2, \dots, x_n)$ определяет отношение R , такое, что $(a_1, a_2, \dots, a_n) \in R$ тогда и только тогда, когда $P(a_1, a_2, \dots, a_n)$ — «истина». В этом случае говорят, что отношение R задается областью истинности предиката $P(x_1, x_2, \dots, x_n)$. Например, отношение $R(M, N)$ = «расстояние на плоскости между точками $M(x_1, y_1)$ и $N(x_2, y_2)$ больше величины 1» можно задать предикатом « $(x_2 - x_1)^2 + (y_2 - y_1)^2 > 1^2$ ».

Если в n -местный предикат на место одного из термов подставить определенный элемент из соответствующего множества, то предикат станет $(n-1)$ -местным. Заменяя все термы на конкретные значения из предметной области предиката, получим 0-местный предикат, т. е. высказывание. Например, « X — брат Y » — двухместный предикат, « X — брат Маши» — одноместный предикат, «Саша — брат Маши» — высказывание.

2.1.6. Логические операции над предикатами

С помощью логических операций отрицания, конъюнкции, дизъюнкции, импликации и эквивалентности из исходных предикатов могут быть построены новые предикаты.

Отрицание предиката. Пусть предикат $P(x_1, x_2, \dots, x_n)$ задан на множествах $M_1, M_2, \dots, M_n$. Предикат $R(x_1, x_2, \dots, x_n)$ называется отрицанием предиката $P(x_1, x_2, \dots, x_n)$ тогда и только тогда, когда при одних и тех же кортежах $(a_1, a_2, \dots, a_n)$, где $a_1 \in M_1, a_2 \in M_2, \dots, a_n \in M_n$, высказывание $P(a_1, a_2, \dots, a_n)$ истинно, когда $R(a_1, a_2, \dots, a_n)$ — ложно, и наоборот. Обозначение:

$$R(x_1, x_2, \dots, x_n) \equiv \neg P(x_1, x_2, \dots, x_n).$$

Например, предикат « n — четное число» есть отрицание предиката « n — нечетное число» на множестве целых чисел.

Конъюнкция предикатов. Пусть на множествах $M_1, M_2, \dots, M_n$ заданы два n -местных предиката $P(x_1, x_2, \dots, x_n)$ и $R(x_1, x_2, \dots, x_n)$. Конъюнкцией этих предикатов называется предикат

$$Q(x_1, x_2, \dots, x_n) \equiv P(x_1, x_2, \dots, x_n) \wedge R(x_1, x_2, \dots, x_n),$$

который истинен для одних и тех же кортежей только тогда, когда оба предиката, $P(x_1, x_2, \dots, x_n)$ и $Q(x_1, x_2, \dots, x_n)$, истинны.

Например, конъюнкция предикатов « $x^2 + y^2 \leq 1$ » и « $x \geq 0$ », где x, y — вещественные числа, определяет предикат «точки правой половины единичного круга» (рис. 2.2).

Дизъюнкция предикатов $P(x_1, x_2, \dots, x_n)$ и $R(x_1, x_2, \dots, x_n)$ есть новый предикат $S(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \vee R(x_1, x_2, \dots, x_n)$, который имеет значение «ложь» для тех и только тех кортежей из $M_1 \times M_2 \times \dots \times M_n$, для которых оба предиката, $P(x_1, x_2, \dots, x_n)$ и $R(x_1, x_2, \dots, x_n)$, имеют значение «ложь». На рис. 2.3 иллюстрируется дизъюнкция предиката « $x_2 + y_2 \leq 1$ » и « $x \geq 0$ » (заштрихованная область).

Импликация предикатов $P(x_1, x_2, \dots, x_n)$ и $R(x_1, x_2, \dots, x_n)$ есть новый предикат $T(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \rightarrow R(x_1, x_2, \dots, x_n)$,

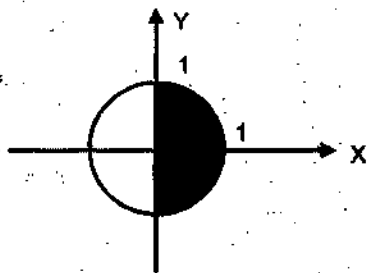


Рис. 2.2

$$Q = P(x^2 + y^2 < 1) \wedge R(x \geq 0)$$

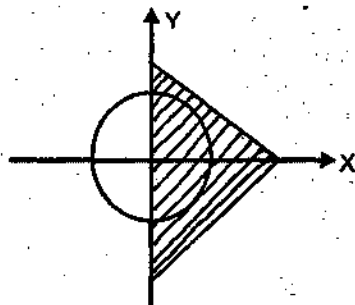


Рис. 2.3

$$Q = P(x^2 + y^2 < 1) \vee R(x \geq 0)$$

который имеет значение «ложь» для тех и только тех кортежей из $M_1 \times M_2 \times \dots \times M_n$, для которых предикат $P(x_1, x_2, \dots, x_n)$ имеет значение «истина», а предикат $R(x_1, x_2, \dots, x_n)$ имеет значение «ложь». Например, импликация « n делится на 4» $\rightarrow$ « n делится на 2» есть предикат: «если n делится на 4, то n делится на 2».

Эквивалентность предикатов $P(x_1, x_2, \dots, x_n)$ и $R(x_1, x_2, \dots, x_n)$ есть новый предикат $V(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \leftrightarrow R(x_1, x_2, \dots, x_n)$, который имеет значение «истина» для тех и только тех кортежей из $M_1 \times M_2 \times \dots \times M_n$, для которых предикат $P(x_1, x_2, \dots, x_n)$ и предикат $R(x_1, x_2, \dots, x_n)$ имеют одинаковые значения, или оба «истина» или оба «ложь». Два предиката, заданных на одних и тех же множествах, называются **равносильными**, если при всех наборах входящих в них предметных переменных эти предикаты принимают одинаковые значения. Равносильность называют также **логической эквивалентностью**. Например, эквивалентность предикатов $P(n) =$ « n делится на 6» и $R(n) =$ « n делится на 2 и n делится на 3» есть предикат $V(n) = P(n) \leftrightarrow R(n)$: «если n делится на 6, то n делится на 2 и на 3». Предикаты $P(n)$ и $R(n)$ логически эквивалентны.

Наряду с логическими операциями важную роль играют операции, называемые кванторами.

Квантор всеобщности есть операция, которая предикат $P(x)$ превращает в высказывание: «все x обладают свойством $P(x)$ ». Знак квантора всеобщности « $\forall$ ». Он заменяет фразы: «для всех», «каждый», «любой» и т. п. Обозначение $\forall x: P(x)$ читается так: «для всех x таких, что P от x ». Например, « $P(x) = x > 0$, где x — вещественное число», есть предикат « x — положи-

тельное число». Тогда $\forall x: P(x)$ есть высказывание «каждое число — положительно». Это ложное высказывание. Если же x — любое натуральное число ($x \in \mathbb{N}$), то $\forall x: P(x)$ есть выражение «каждое натуральное число — положительно» — истинное высказывание.

Квантор всеобщности можно рассматривать как обобщение серии конъюнкций единичных высказываний. Пусть M — множество очков, которое может выпасть при бросании игральной кости, т. е. $M = \{1, 2, 3, 4, 5, 6\}$, и $P(x)$ — предикат «при бросании игральной кости один раз выпадает x очков», где $x \in M$. Применение квантора всеобщности позволяет вместо сложного высказывания $P(1) \wedge P(2) \wedge P(3) \wedge P(4) \wedge P(5) \wedge P(6)$ записать равносильное ему компактное высказывание $\forall x: P(x)$, $x \in M$: «при бросании игральной кости один раз может выпасть любое из шести первых натуральных чисел».

Квантор существования есть операция, которая предикат $P(x)$ превращает в высказывание «существует хотя бы один x из M , обладающий свойством $P(x)$ ». Знак квантора существования « $\exists$ ». Он заменяет фразы: «существует, хотя бы один», «найдется», «некоторый» и т. п. Обозначение $\exists x: P(x)$ читается так: «существует хотя бы один x такой, что P от x ». Например, $P(x)$ — предикат « x — студент», где x — элемент множества жителей Москвы. Тогда выражение $\exists x: P(x)$ есть высказывание «хотя бы один житель Москвы является студентом».

Квантор существования можно рассматривать как обобщение серии дизъюнкций единичных высказываний. Если задано множество $M = \{a_1, a_2, \dots, a_n\}$ и на нем определен предикат $P(x)$, то

$$P(a_1) \vee P(a_2) \vee \dots \vee P(a_n) \equiv (\exists x \in M): P(x).$$

Кванторы обладают свойствами, являющимися аналогами законов де Моргана:

$$\neg(\forall x: P(x)) \equiv \exists x: \neg P(x),$$

$$\neg(\exists x: P(x)) \equiv \forall x: \neg P(x).$$

С помощью кванторов можно выражать ряд часто используемых на практике отношений между множествами. Например, высказывание «все объекты x из данного множества, обладающие свойством $P(x)$, обладают также и свойством $R(x)$ » формально можно записать так: $\forall x: (P(x) \rightarrow R(x))$.

Переход от $P(x)$ к $\forall x: P(x)$ или $\exists x: P(x)$ называется **квантификацией** или **связыванием переменной x** . Связанная переменная фактически не является переменной, т. е. переход от $\forall x: P(x)$ к $\forall y: P(y)$ или от $\exists x: P(x)$ к $\exists y: P(y)$ не меняет истинности выражений. Навешивание переменной на многоместный предикат уменьшает в нем число свободных переменных и превращает его в предикат от меньшего числа переменных.

Рассмотрим пример. На множестве чисел задан двухместный предикат $P(x, y) = \text{«число } x \text{ делится на число } y\text{»}$. Связывая одну переменную, можно получить следующие одноместные предикаты:

$\forall x: P(x, y) = \text{«каждое число делится на } y\text{»}$ — ложь;

$\exists x: P(x, y) = \text{«существует число, которое делится на } y\text{»}$ — истина;

$\forall y: P(x, y) = \text{«число } x \text{ делится на любое число»}$ — ложь;

$\exists y: P(x, y) = \text{«существует число, на которое делится } x\text{»}$ — истина.

Связывая обе переменные данного предиката, получим высказывания:

$\forall x, \forall y: P(x, y) = \text{«каждое число делится на любое число»}$ — ложное высказывание;

$\forall x, \exists y: P(x, y) = \text{«существует число, на которое делится любое число»}$ — истина, т. к. такое число есть 1;

$\exists x, \forall y: P(x, y) = \text{«существует число, которое делится на любое число»}$ — ложное высказывание;

$\exists x, \exists y: P(x, y) = \text{«существует число, которое делится на какое-нибудь число»}$ — истинное высказывание.

Глава 2.2. Графы

Если отвлечься от предметного содержания во всех видах изображений, то получится некоторая графическая схема, состоящая из точек и соединяющих их линий. Можно ли найти закономерности, присущие всем графическим схемам независимо от содержания отображаемых ими объектов, и использовать их в дальнейшем? На эти вопросы дает положительный ответ раздел математики, называемый теорией графов. Основным понятием этой теории является граф.

Под *графом* понимают систему объектов произвольной природы (*вершин*) и связей (*ребер*, соединяющих некоторые пары этих объектов).

Первой научной работой посвященной графам считается статья Л. Эйлера, посвященная задаче о кенигсбергских мостах и опубликованная в 1736 году. Суть задачи следующая. Четыре части города Кенигсберга (a, b, c, d на рис. 2.4а) были соединены семью мостами. Требовалось определить, можно ли, выйдя из любой части города и пройдя по каждому мосту *ровно один раз*, вернуться в исходное место. Абстрагируясь от всех других обстоятельств можно построить граф, изображенный на рис. 2.4б. Вершинами являются части города, а ребрами — мосты. Надо ответить на вопрос: можно ли, выходя из любой вершины, пройти по всем ребрам и вернуться в исходную вершину? Эйлер дал отрицательный ответ на поставленный вопрос.

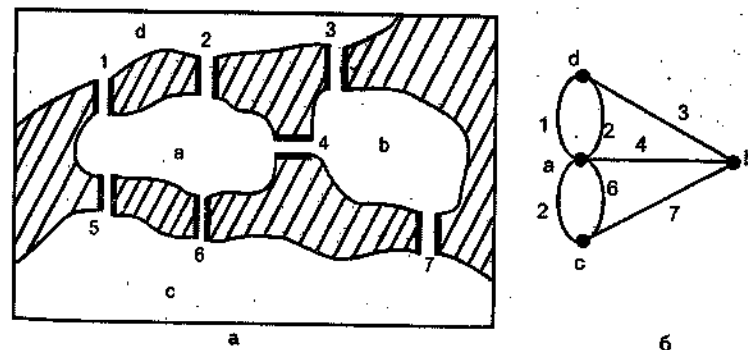


Рис. 2.4. Задача о кенигсбергских мостах

2.2.1. Определение графа

Дадим строгое определение графа. **Графом G** называется пара множеств $G = (X, E)$, где X — непустое множество вершин $\{x_1, x_2, \dots, x_n\}$, а элементами множества E являются некоторые двухэлементные подмножества X , т. е. $E = \{(x_i, x_j)\}$. Эти двухэлементные подмножества называются **ребрами**. Например,

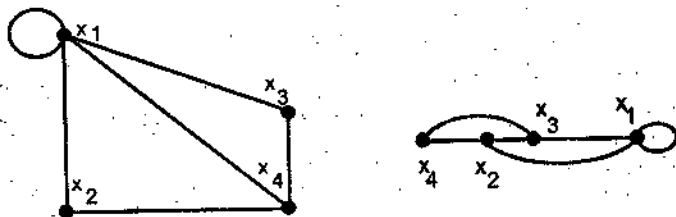


Рис. 2.5. Петля

$G = (\{x_1, x_2, x_3, x_4\}, \{(x_1, x_1), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_4), (x_3, x_4)\})$ (рис. 2.5).

Несложные графы удобно изображать в виде графических схем, где вершины — точки, а ребра — соединяющие их линии. В этих схемах длина линий, их толщина и форма, а также взаиморасположение вершин не имеют никакого значения. Так на рис. 2.5 изображен один и тот же граф G предыдущего примера. Таким образом, граф — свободная конструкция, для которой имеет значение факт наличия связей между двумя вершинами и, в некоторых случаях, характер этих связей.

Если вершины x_i и x_j принадлежат некоторому ребру (x_i, x_j) , то говорят, что это ребро **инцидентно** вершинам x_i и x_j , которые в свою очередь называются **смежными**. Если ребро инцидентно одной и той же вершине, то оно называется **петлей** (рис. 2.5). Вершина не инцидентная никакому ребру называется **изолированной**. Если в графе есть вершины, соединенные с двумя

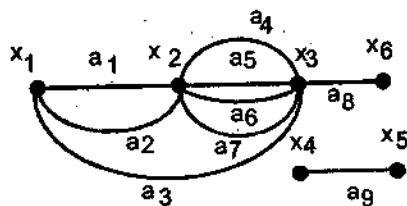


Рис. 2.6. Мультиграф

или большим числом вершин, то такой граф называют **мультиграфом** (рис. 2.6).

Число ребер инцидентных данной вершине называется **степенью** (кратностью) данной вершины. В графе, изображенном на рис. 2.6, вершина x_2 имеет степень 6, т. к. ей инцидентны ребра $a_1, a_2, a_3, a_4, a_5, a_6, a_7$, а вершина x_1 имеет степень 3. Вершина x_4 имеет степень 1, степень 0 имеет изолированная вершина. Граф без петель и кратных ребер называется **простым** или **обыкновенным**. Граф может быть задан и в виде квадратной таблицы — **матрицы смежности графа**.

Номера строк и столбцов этой матрицы совпадают с номерами вершин графа, а ее элемент c_{ij} есть число ребер, соединяющих вершины x_i и x_j . Для графа, изображенного на рис. 2.6, матрица смежности имеет вид:

$$C = \begin{vmatrix} 0 & 2 & 1 & 0 & 0 \\ 2 & 0 & 4 & 0 & 0 \\ 1 & 4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix} \quad C_1 = \begin{vmatrix} 0 & 4 & 0 & 1 & 1 \\ 4 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \end{vmatrix}$$

Нумеровать вершины графа можно произвольным образом. Например, в графе, изображенном на рис. 2.6, перенумеруем вершины следующим образом: $x_1 \rightarrow x_5, x_3 \rightarrow x_1, x_5 \rightarrow x_4, x_4 \rightarrow x_3$. При этом матрица смежности C_1 изменится (сравните матрицы C и C_1). Существенным же для графов являются лишь число вершин и ребер, а также отношения инцидентности вершин и ребер. Следовательно, графы, отличающиеся только нумерацией вершин и сохраняющие отношения их инцидентности, могут отличаться лишь начертанием. В этой связи говорят, что матрица определяет графы с точностью до изоморфизма. Такие графы называются **изоморфными**. В подавляющем большинстве приложений изоморфные графы не различаются и решения задач на изоморфных графах по существу совпадают.

Граф называется **планарным** (плоским), если существует изоморфный ему граф, который может быть изображен на плоскости без пересечения ребер. На рис. 2.7 представлена пара изоморфных графов G_1 и G_2 . В первом из них ребра пересекаются, но поскольку в изоморфном ему графе ребра не пересекаются, то оба они являются планарными.

В целом ряде задач принципиальную важность имеет ориентация отношений между вершинами. Характерным

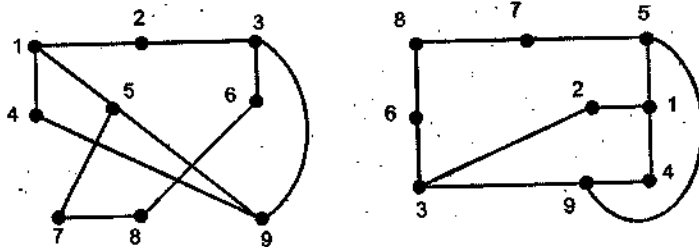


Рис. 2.7. Пара изоморфных планарных графов

примером являются отношения порядка. Например, x_i — старше x_j , но в этом случае x_j не может быть старше x_i . Следовательно, отношение «быть старше» может быть формализовано графом, у которого связи между вершинами имеют определенную ориентацию. Такие графы принято называть ориентированными.

Ориентированным (или орграфом) D , называется пара $D = (X, A)$, где X — произвольное множество вершин и A — множество упорядоченных пар вершин, называемых дугами $A \subset X \times X$. В паре (x_i, x_j) первая вершина x_i называется началом дуги, а вторая x_j — концом дуги.

На рисунках дуги изображаются стрелками (рис. 2.8).

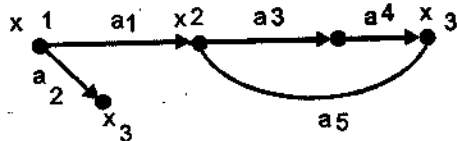


Рис. 2.8. Ориентированный граф

Чтобы лучше понять отличие ориентированного графа от неориентированного, можно привести следующую аналогию.

Если ориентированный граф описывает систему улиц города с односторонним движением, то неориентированный граф — с двусторонним движением. Очевидно, что ребро неориентированного графа, соединяющее две вершины, эквивалентно двум разнонаправленным дугам соединяющим эти же вершины. Таким образом, неориентированный граф можно рассматривать как частный случай ориентированного графа.

Ориентированные графы удобно задавать в виде таблиц — матриц инцидентий. Номера строк матрицы инцидентий равны номерам вершин, а номера столбцов — номерам дуг. Если дуга

выходит из вершины x_i , то соответствующий элемент матрицы инцидентий c_{ij} равен -1 . Если же дуга a_j входит в вершину x_i , то элемент c_{ij} равен $+1$. Все другие элементы столбца j равны нулю. Для ориентированного графа, изображенного на рис. 2.8, матрица инцидентий имеет вид:

$$C = \begin{pmatrix} 0 & -1 & -1 & 0 & 0 \\ +1 & 0 & 0 & -1 & -1 \\ +1 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 \\ 0 & +1 & 0 & +1 & 0 \end{pmatrix}$$

Матрицы часто используются для задания графов в памяти компьютеров.

Граф $G_b(X_b, E_b)$ называется **частичным графом** (суграфом) графа $G(X, E)$, если он содержит все вершины исходного графа и лишь часть ребер исходного графа, т. е. $X_b = X$ и $E_b \subset E$. **Подграфом** $G_a(X_a, E_a)$ графа $G(X, E)$ называется граф, в который входит подмножество вершин исходного графа ($X_a \subset X$) и все ребра, соединяющие вершины X_a этого графа ($E_a \subset E$). На рис. 2.9. представлены подграф и частичный граф для данного исходного графа $G(X, E)$.

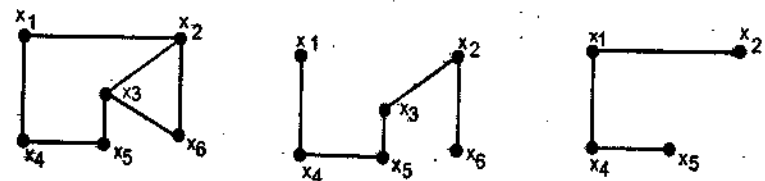


Рис. 2.9. Граф, подграф и частичный граф

2.2.2. Маршруты на графах

Пусть дан неориентированный граф $G(X, E)$. **Маршрутом** длины m называется некоторая последовательность m ребер графа $G(X, E)$, такая, что вершины двух соседних ребер последовательности совпадают. Примерами маршрутов на графе, изображенном на рис. 2.6, могут служить следующие последовательности ребер $(\alpha_1, \alpha_6, \alpha_5, \alpha_1, \alpha_3)$ и $(\alpha_2, \alpha_1, \alpha_4, \alpha_6)$. Первый

из них проходит последовательно через вершины $x_1, x_2, x_3, x_2, x_1, x_3$, соединяя x_1 с x_3 . Второй, проходя через вершины x_2, x_1, x_2, x_3, x_2 , образует **замкнутый маршрут**, т. к. приводит в ту же вершину, из которой и начался.

Две вершины графа называются **связанными**, если существует маршрут, соединяющий эти вершины. Граф, любая пара вершин которого связана, называют **связным графом**. Представленный на рис. 2.6 граф не является связным, поскольку, например, отсутствует маршрут из x_1 в x_4 .

Маршрут, все ребра которого различны, называется **цепью**. Цепь, проходящая через различные вершины, называется **простой (элементарной) цепью**. Замкнутая цепь называется **циклом**, а цикл, проходящий через различные вершины, — **простым циклом**. Так в графе, изображенном на рис. 2.6, маршрут $(\alpha_1, \alpha_2, \alpha_3)$ является простой цепью, маршрут $(\alpha_2, \alpha_3, \alpha_4, \alpha_1)$ — циклом.

Для практики большое значение имеют задачи о нахождении различного рода маршрутов. Наиболее известные из них два: эйлеров цикл и гамильтонов цикл.

Цикл, содержащий все ребра графа, называется **эйлеровым**, а граф, имеющий эйлеров цикл, называется **эйлеровым графом**. Если эйлеров граф плоский, то его можно изобразить одним росчерком пера (не отрывая пера от бумаги), причем начальная и конечная вершины совпадут. На рис. 2.10 изображен граф, обход всех ребер которого ровно один раз возможен по маршруту: $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7$.

Вопрос о существовании эйлерова графа разрешается следующей теоремой: конечный неориентированный граф эйлеров тогда и только тогда, когда он связан и степени всех его вершин четны [2]. Так граф, изображенный на рис 2.10, эйлеров, т. к.

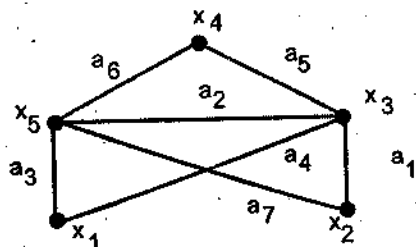


Рис. 2.10. Эйлеров граф

степени его вершин x_1, x_2 и x_4 — 2, вершин x_3 и x_5 — 4.

Простой цикл, который проходит через все вершины графа, называется **гамильтоновым** (У. Р. Гамильтон, 1805—1865).

Условий, гарантирующих существование в графе гамильтонова цикла, не установлено. Это обстоятельство существенно усложняет решение очень многих задач, решение которых сводится к нахождению гамильтонова цикла. Наиболее известна так называемая задача коммивояжера, или задача о странствующем торговце, состоящая в следующем. Пусть имеется несколько связанных между собой пунктов (городов). Выходя из фиксированного пункта, коммивояжер должен вернуться в него, посетив все пункты ровно один раз. Обычно задача усложняется введением дополнительного ограничения: например, пройти маршрут за самое короткое время или с минимальными затратами на транспорт. Отсутствие условий существования гамильтонова цикла приводит к тому, что, приступая к решению задачи, в общем случае неизвестно, имеет ли она вообще решение или нет. На рис. 2.11 и 2.12 представлены, соответственно, граф, имеющий гамильтонов цикл, и не имеющий такового.

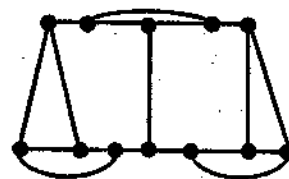


Рис. 2.11. Граф, имеющий гамильтонов цикл

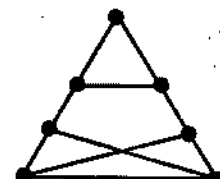


Рис. 2.12. Граф, не имеющий гамильтонова цикла

В ориентированном графе маршруты называются **ориентированными**: начальная вершина каждой последующей дуги маршрута должна совпадать с конечной вершиной предыдущей дуги. Если рассматривать изображение ориентированного графа, то движение по маршрутам на нем должно осуществляться только в направлениях, указанных стрелками. Маршрут, не содержащий повторяющихся дуг, называется **путем**. Путь, не содержащий повторяющихся вершин, называется **простым путем**. Замкнутый путь называется **контуром**. **Простой контур** — контур, не имеющий повторяющихся вершин. Если в графе содержится хотя бы один контур, то это **контурный граф**.

При помощи ориентированного графа удобно задавать бинарные отношения. Например, отношения порядка. Говорят, что на ориентированном графе $D(X, A)$ задано отношение порядка, если для любых двух вершин x_i и x_j , удовлетворяющих условию $x_i \leq x_j$, существует путь из x_i в x_j . На рисунке 2.13 представлено изображение графа, на котором определено отношение порядка « x_i не больше x_j ». Для отношения порядка характерны три свойства: рефлексивность, антисимметричность и транзитивность. С позиций теории графов эти свойства можно интерпретировать следующим образом:

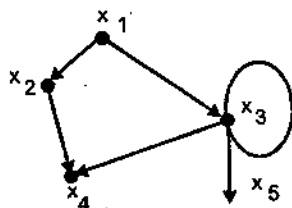


Рис. 2.13. Граф, описывающий отношение порядка

Рефлексивность означает истинность условия $x_i \leq x_i$. Следовательно, вершина x_i не больше самой себя, т. е. имеется путь из x_i в x_i , что обуславливает допустимость петли для вершины x_i . Антисимметричность $(x_i \leq x_j) \wedge (x_j \leq x_i) \rightarrow (x_i \equiv x_j)$ показывает, что если на графе имеется путь из x_i в x_j , то имеется и путь из x_j в x_i . Таким образом, в графе имеется контур.

Транзитивность отношения «быть не больше» означает истинность выражения $(x_i \leq x_j) \wedge (x_j \leq x_k) \rightarrow (x_i \leq x_k)$. Транзитивность эквивалентна тому, что вершины x_i, x_j, x_k последовательно встречаются на одном и том же пути. Отношение строгого порядка антирефлексивно и асимметрично, поэтому его можно задать на графе без петель и без контуров.

2.2.3. Деревья

Связный граф, не имеющий циклов, называется деревом, а ребра дерева называются ветвями. У дерева с n вершинами есть ровно $(n-1)$ ветка (ребро) (рис. 2.14а). Действительно, если добавить лишь одно ребро, соединяющее две вершины дерева, то в графе появится цикл (рис. 2.14б). Если же убрать хотя бы одно ребро, то граф станет несвязным (рис. 2.14в).

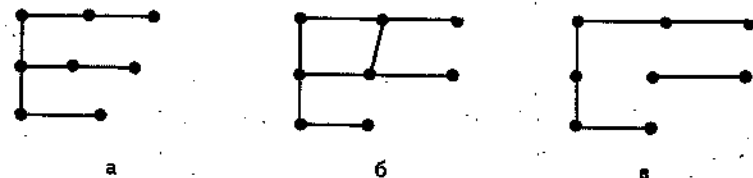


Рис. 2.14. Иллюстрация к определению дерева

Таким образом, для связи n вершин необходимо и достаточно ровно $(n-1)$ ребер.

Несвязный граф без циклов называется лесом. При этом любая связная часть леса является деревом (рис. 2.15). Могут рассматриваться и деревья с ориентированными ребрами. Ориентированное дерево называется **прадеревом** с корнем y , если существует путь между вершиной y и любой другой его вершиной (рис. 2.16).

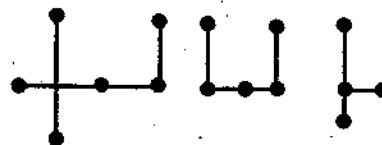


Рис. 2.15. Лес

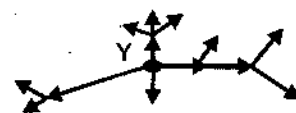


Рис. 2.16. Прадерево с корнем Y

Пара $S = (G, C)$ называется **сетью**, если $G = (X, A)$ — произвольный ориентированный граф, $C: A \rightarrow R$ — функция, ставящая каждой дуге графа в соответствие неотрицательное вещественное число $c(a_i, a_j)$, которое, в зависимости от решаемой задачи, называют по-разному: весом дуги, пропускной способностью, длиной дуги и т. п. Сеть иногда называют взвешенным графом, а сумму весов всех дуг — весом графа.

Задачи о покрывающих деревьях. Пусть задан неориентированный граф $G = (X, E)$. Дерево $D(Y, I)$ называется **покрывающим деревом графа** $G(X, E)$, если $X = Y$ и $I \subseteq E$. Таким образом, покрывающее дерево связывает все вершины данного графа, но не обязательно при этом включает все его ребра. На рис. 2.17а показан граф, для которого на рис. 2.17б построено покрывающее дерево. Очевидно, что любой связанный граф содержит хотя бы одно покрывающее дерево.

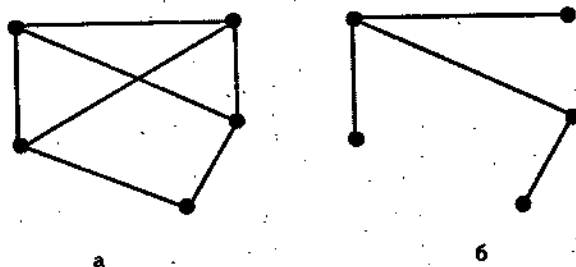


Рис. 2.17. Исходный граф и покрывающий его граф

Существует ряд стандартных задач, к которым сводятся многие приложения графов.

Задача 1. Для заданной сети $S = (G, C)$ построить покрывающее дерево минимального веса.

К этой задаче могут быть сведены многие практические проблемы. Вот одна из них. В строящемся микрорайоне необходимо разработать схему подводки газа к домам, причем затраты на строительство должны быть наименьшими. На языке теории графов задача будет формулироваться следующим образом: для заданного множества домов построить схему газоснабжения таким образом, чтобы суммарная стоимость строительства была минимальной. Для решения этой задачи удобной моделью является взвешенный граф, вершины которого соответствуют домам, ребра — возможным газовым коммуникациям между ними, веса ребер — стоимости прокладки газопровода между соответствующими домами.

Ниже мы покажем, что при рассмотрении таких задач лучше предварительно решить другую, более простую задачу.

Задача 2. Для связного графа $G(X, E)$ без петель построить покрывающее дерево.

Идея решения этой задачи следующая. Предварительно строят так называемые связки. Связка представляет собой некоторое поддерево искомого покрывающего дерева. Для построения связки последовательно выделяются отдельные вершины, объединенные некоторыми ребрами. По мере рассмотрения все новых и новых ребер исходного графа связки постепенно нарастают и объединяются до тех пор, пока не будет получена одна единая связка, включающая в себя все вершины графа — искомое покрывающее дерево. При образовании связок следует следить за тем, чтобы не образовывались циклы. Процедура

выделение вершин и ребер во время выполнения алгоритма называют окрашиванием.

Алгоритм построения покрывающего дерева.

Шаг 0. Все ребра исходного графа $G(X, E)$ неокрашены и множество связок пусто.

Шаг 1. Произвольным образом выбирается ребро. Окрашиваем это ребро. Формируем из этого ребра и инцидентных ему вершин первую связку.

Шаг 2. Произвольным образом выбираем следующее неокрашенное ребро. При этом возможно возникновение одной из четырех ситуаций:

2.1. Обе вершины выбранного нового ребра принадлежат выбранной ранее связке (в дальнейшем — одной из рассмотренных ранее связок). Исключаем выбранное ребро из дальнейшего рассмотрения. Переходим к шагу 3.

2.2. Одна вершина выбранного ребра принадлежит одной из связок, а другая не принадлежит ни одной из связок. Окрашиваем ребро и включаем его вторую концевую вершину в ту связку, в которую входит первая концевая вершина (наращиваем связку). Переходим к шагу 3.

2.3. Ни одна из вершин инцидентных выбранному ребру не принадлежит ни одной из связок. Окрашиваем это ребро и формируем новую связку из выбранного ребра и смежных ему вершин. Переходим к шагу 3.

2.4. Концевые вершины выбранного ребра принадлежат различным связкам. Окрашиваем это ребро. Обе связки, которые связывает рассматриваемое ребро, сливаем в одну связку и переходим к шагу 3.

Шаг 3. Если все ребра графа рассмотрены, то переходим к шагу 4, иначе переходим к шагу 2.

Шаг 4. Если получилась единственная связка, то она и является искомым покрывающим деревом, иначе задача не имеет решения. Конец алгоритма.

Замечание 1. Описанный алгоритм относится к так называемым «поедающим» алгоритмам, в которых судьба рассматриваемых элементов решается только один раз на протяжении всей процедуры. Поэтому, если число вершин графа конечно, то алгоритм заканчивается за конечное число шагов. Поедающие алгоритмы считаются самыми «быстрыми» из комбинаторных алгоритмов.

Замечание 2. Изложенный алгоритм позволяет построить одно из возможных покрывающих для исходного графа деревьев, поскольку очередное ребро выбирается произвольным образом.

Если алгоритм не заканчивается построением покрывающего дерева, то это означает, что граф $G(X, E)$ не связный и поставленная задача не имеет решения. Признаком такой ситуации является наличие более чем одной связки при рассмотрении всех ребер графа.

Пример. Построить покрывающее дерево для графа, изображенного на рис. 2.18.

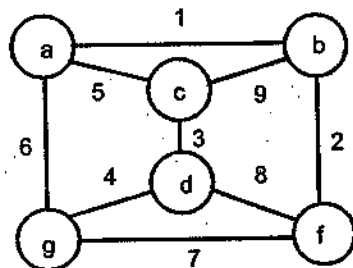
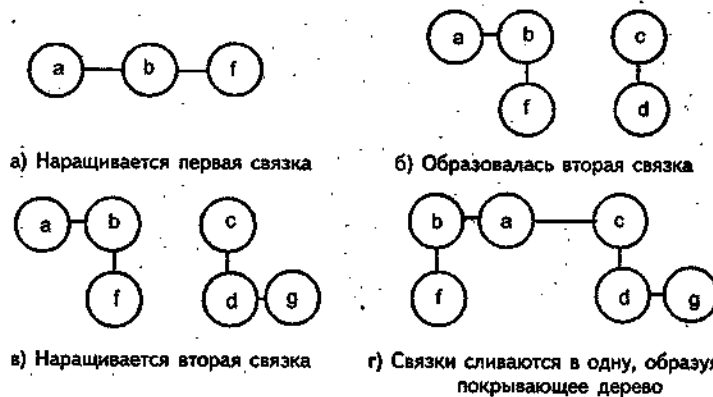


Рис. 2.18. Исходный граф для задачи о покрывающем дереве.

Решение. Произвольным образом пронумеруем ребра графа, например, так, как показано на рисунке. Согласно изложенному алгоритму, первое выбранное ребро (a, b) образует первую связку $D_1 = \{(a, b)\}$. Второе ребро (b, f) окрашивается (реализуется ситуация 2.2), а связка D_1 уже включает вершины a, b и f (рис. 2.19а). Выбор третьего ребра (c, d) реализует случай 2.3. Поэтому образовывается вторая связка $D_2 = \{(c, d)\}$ (рис. 2.19б). Четвертое выбранное ребро (d, g) , как показано на рис. 2.19в, наращивает вторую связку $D_2 = \{(c, d), (d, g)\}$. Рассматривая следующее ребро (a, c) , видим, что реализуется случай 2.4. Это требует объединения обеих связок в одну, а именно, $D = \{(a, b), (b, f), (a, c), (c, d), (d, g)\}$ (рис. 2.19г).

Шестое ребро в дерево не включается так же, как и все остальные ребра (случай). В итоге получаем покрывающее исходный граф дерево (рис. 2.19г).

Из приведенного выше алгоритма видно, что граф в общем случае может иметь более одного покрывающего дерева. Необходимость выбора наиболее подходящего дерева и приводит к задаче 1 — построению покрывающего дерева минимального веса. В таком случае можно поступить следующим



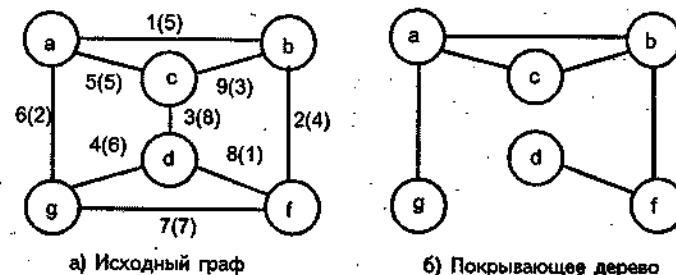
2. 19. Иллюстрации к решению задачи о покрывающем дереве

образом. В целом сохранить алгоритм решения задачи 2, но выбирать очередное ребро не произвольным образом, а строго в порядке возрастания весов.

В качестве примера построим минимальное покрывающее дерево для графа, изображенного на рис. 2.20а, цифры при ребрах в скобках означают вес соответствующего ребра.

Первым выбирается ребро (d, f) , так как оно имеет наименьший вес, равный единице. Второе ребро (a, g) образует вторую связку $\{a, g\}$. Следующим должно быть выбрано ребро (c, b) , которое образует третью связку $\{c, b\}$.

Ребро (b, f) объединяет первую и третью связку $\{d, f, c, b\}$. Следующим ребром может быть либо ребро $(a, в)$, либо (a, c) . Пусть это будет ребро (a, c) . Тогда необходимо объединить имеющиеся связки в одну $\{d, f, c, b, a, g\}$. Все вершины рассмотрены и включены в единственную связку, поэтому оставшиеся ребра можно не рассматривать.



2. 20. Решение задачи на построение покрывающего дерева

В итоге имеем дерево, изображенное на рис. 2.20б и имеющее вес 15. Сравните этот вес с весом дерева, изображенного на рис. 2.19г, вес которого составляет 27 единиц.

Покажем, что описанный алгоритм обеспечивает построение дерева минимального веса, используя метод от противного.

Пусть для исходного графа по алгоритму построили покрывающее дерево веса Q , но для него уже известно покрывающее дерево минимального веса P . Занумеруем ребра в том порядке, как их выбирали. Чтобы доказать эффективность алгоритма необходимо показать, что веса графов равны, т. е. $L(P) = L(Q)$.

Если P и Q не совпадают, то они отличаются хотя бы одним ребром. Пусть такое ребро есть (a, b) и $M(a, b)$ — цепь графа, с весом P , идущая из a в b .

Если (a, b) добавить к P , то получится цикл (т. к. P — дерево). Но Q не имеет циклов, поэтому в цикле, получившемся в P , имеется хотя бы одно ребро, не принадлежащее Q . Пусть это ребро веса U . Удалив его из P получим граф P' с тем же числом ребер, что и P .

Найдем вес графа P' :

$$L(P') = L(P) + L(a, b) - U.$$

Но по условию граф P имеет наименьший вес, следовательно: $L(P) \leq L(P')$.

Отсюда следует, что $L(a, b) \geq U$. Но ребро (a, b) , когда его выбирали, было ребром наименьшего веса, при добавлении которого к P не получалось цикла. Поэтому $U = L(a, b)$ и P' , следовательно, имеет вес равный весу графа P .

Повторяя эти рассуждения для каждого ребра, получим, что $L(Q) = L(P)$.

Глава 2.3. Основы комбинаторики

Комбинаторика (или комбинаторный анализ) — раздел математики, объектом исследования которого являются дискретные множества произвольной природы. «Комбинаторным», то есть тем, что можно исследовать методами комбинаторики, считается все, что «перечисляемо». Основной задачей комбинаторики является *определение числа способов выполнения некоторых точно определенных операций*, или, другими словами, *определение числа подчиненных тем или иным условиям комбинаций*, которые можно составить из заданной совокупности объектов. В настоящее время комбинаторные методы широко используются в таких отраслях научного значения, как теория вероятностей, математическая статистика, теория алгоритмов, кодирования и распознавания образов, математическое программирование, лингвистика и др.

2.3.1 Основные теоремы комбинаторики

2.3.1.1. Подмножества и выборки. Пусть задано произвольное множество из n объектов, которое мы обозначим за A , с элементами a . Последовательность произвольных элементов

$$\langle a_1, a_2, \dots, a_r \rangle, \quad a_i \in A, \quad i = \overline{1, r}$$

называется **выборкой объема r из A** , причем каждый элемент из множества A может встречаться в выборке произвольное число раз. Объем выборки может превосходить объем исходного множества A . Если же все компоненты r -выборки из n -множества A различны, то $r \leq n$ и r -выборка представляет собой r -подмножество A .

Если свойства выборки изменяются при транспозиции элементов (т. е. при перемене местами двух элементов), то выборка называется упорядоченной, в противном случае — неупорядоченной. Число появлений одного и того же элемента a называется его кратностью и обозначается $\chi(a)$. Типичным примером выборки может являться слово в фиксированном алфавите.

Следовательно, если каждый элемент a , m -выборки имеет кратность $\chi(a_i) = 1, i = \overline{1, m}$, то выборка является просто m -под-

множеством множества A . При этом упорядоченное m -подмножество n -множества называется m -перестановкой из n -элементов (или размещением из n элементов по m). Неупорядоченное m -подмножество n -множества называется m -сочетанием из n элементов (или сочетанием из n элементов по m).

2.3.1.2 Теорема о числе упорядоченных подмножеств. Число $A_{n,m}$ упорядоченных m -подмножеств n -множества равно:

$$A_{n,m} = n(n-1) \times (n-2) \times \dots \times (n-m+1), \quad m = \overline{1, n}.$$

Доказательство. Первый элемент m -подмножества может быть выбран из n -множества n способами. Поскольку все элементы подмножества различны, то в каждой из выборок первого элемента второй его элемент может быть выбран $n-1$ способами. В каждой из этих выборок третий элемент может быть выбран $n-2$ способами и т. д. Продолжив аналогичные рассуждения, получим утверждение теоремы.

Пример. Сколько слов, состоящих из четырех различных букв, можно составить из данных шести букв: а, в, к, л, о, с?

Поскольку в задаче нет ограничений на порядок данных букв в слове, то на первое место можно поставить любую букву (6 вариантов), на второе место можно поставить любую букву, кроме установленной на первое место (5 вариантов) и т. д. Теперь очевидно, что таких слов будет $A_{6,4} = 6 \cdot 5 \cdot 4 \cdot 3 = 360$.

Следствие. Если $m = n$, то число таких подмножеств есть:

$$P_n = A_{n,n} = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1 = n!$$

т.е. равно произведению всех первых n натуральных чисел. « $n!$ » читается: « **n -факториал**».

Пример. Сколько вариантов расположения слов допускает предложение: «Редактор вчера внимательно прочитал рукопись»?

Так как в данном предложении нет никаких грамматических ограничений на порядок слов, то на первое место можно поставить любое слово (5 вариантов), на второе любое другое, кроме выбранного (4 варианта) и т. д. Всего $P_5 = 5 \times 4 \times 3 \times 2 \times 1 = 120$.

Размещения из n элементов по n называются **перестановками**. Чтобы показать очень быстрый рост числа перестановок с ростом числа n , сравним числа $6!$, $8!$ и $10!$. Вычисляя, получим: $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6! = 720$; $8! = 40\,320$; $10! = 3\,628\,800$.

Если рассматривается произведение первых натуральных только четных или только нечетных чисел, то такие произведения называются **двойными факториалами** и обозначаются так:

$$2k \times (2k-2) \dots 6 \times 4 \times 2 = (2k)!!$$

$$(2k+1) \times (2k-1) \dots 5 \times 3 \times 1 = (2k+1)!!$$

2.3.1.3. Теорема о числе упорядоченных выборок. Число $A_{n,m}$ упорядоченных m -выборок из n -множества равно:

$$A_{n,m} = n^m, \quad m \leq n.$$

Доказательство. В отличие от m -множеств, в m -выборке n -множества элементы могут иметь произвольную кратность χ , то есть $0 \leq \chi(a_i) \leq m$, $i = \overline{1, n}$, так как любой элемент n -множества может встретиться в выборке от 0 до m раз.

Повторяя рассуждения предыдущей теоремы и учитывая, что в рассматриваемом случае число способов выбора каждого из элементов n -множества не зависит от выбора остальных элементов и равно n , получим доказываемое утверждение.

Повторим, что выборки подразумевают возможность наличия в них одинаковых элементов, а подмножество не допускает повторений элементов. Для отличия подмножеств от выборок в формулах для выборок введем подчеркивание сверху.

Пример. Сколько всего телефонных номеров можно иметь в городе, если номер имеет шесть цифр?

Каждый телефонный номер может содержать любые шесть цифр из десяти (0, 1, 2, ..., 9). При этом одинаковые цифры могут повторяться до шести раз, и, кроме того, телефонные номера различны, даже если они отличаются лишь порядком цифр. На основании этого задача сводится к подсчету количества **6-выборок** 10-множества, т. е.:

$$\underline{A}_{10,6} = 10^6$$

2.3.1.4. Теорема о числе неупорядоченных подмножеств. Число $C_{n,m}$ неупорядоченных m -подмножеств n -множества равно:

$$C_{n,m} = \frac{\underline{A}_{n,m}}{P_m} = \frac{n!}{m!(n-m)!}, \quad m = \overline{1, n}.$$

Доказательство. Согласно теореме 2.3.1.2, число упорядоченных m -подмножеств n -множества равно $\underline{A}_{n,m}$. На основании следствия к этой теореме каждое m -подмножество может быть упорядочено $P_m = m!$ способами. Поэтому если пренебречь порядком элементов, то получаем формулу (2.1). Числа $C_{n,m}$ обычно называют **биномиальными коэффициентами**.

Пример. Читатель отобрал по каталогу 8 книг. Однако в библиотеке выдают одному читателю не более 5 книг. Сколько альтернатив взять книги есть у этого читателя?

Решение. Поскольку читатель отобрал книг больше разрешенного числа, то он должен выбрать из них 5 книг. Естественно, что все книги разные и все рано в каком порядке их взять. Следовательно, каждая альтернатива есть неупорядоченное 5 — подмножество из 8 и число вариантов в выборе книг (число альтернатив) равно:

$$C_{8,5} = \frac{8!}{5!3!} = 56.$$

2.3.1.5 Теорема о числе неупорядоченных выборов. Число неупорядоченных m -выборов из n -множества равно:

$$C_{n,m} = \frac{(n+m-1)!}{m!(n-1)!}, \quad m \leq n$$

Доказательство. Пусть элементами исходного n -множества B_n являются первые натуральные числа $1, 2, \dots, n$. Тогда всякая m -выборка

$$\langle b_1, b_2, \dots, b_m \rangle \quad (2.2)$$

может быть ранжирована так, чтобы $b_1 \leq b_2 \leq \dots \leq b_m$.

Пусть теперь D есть $(n+m-1)$ — множество натуральных чисел $1, 2, \dots, n+m-1$. Тогда совокупность натуральных чисел

$$(b_1 + 0, b_2 + 1, \dots, b_m + m - 1) \quad (2.3)$$

представляет собой m -подмножество множества D . Между выборкой (2.2) и множеством (2.3) существует взаимно однозначное соответствие, следовательно, число $C_{n,m}$ неупорядоченных m -подмножеств множества B_n равно числу $C_{n+m-1,m}$ неупорядоченных m -подмножеств $(n+m-1)$ -множества D , которое, согласно предыдущей теореме, определяется равенством:

$$C_{n,m} = C_{n+m-1,m} = \frac{(n+m-1)!}{m!(n-1)!}.$$

Пример. Кости домино можно рассматривать как сочетания с повторениями по два из семи цифр: 0, 1, 2, 3, 4, 5, 6. Число всех таких сочетаний равно:

$$C_{7,2} = \frac{(7+2-1)!}{2!(7-1)!} = \frac{8!}{2!6!} = 28.$$

Набор целых чисел $(n_1, n_2, \dots, n_k)$ называется разбиением числа n , если $n = n_1 + n_2 + \dots + n_k$, например $6 = 1 + 2 + 3$.

2.3.1.6. Теорема о числе разбиений. Число $R(n; n_1, n_2, \dots, n_k)$ упорядоченных $(n_1, n_2, \dots, n_k)$ -разбиений n -множества равно:

$$R(n; n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}, \quad \sum_{i=1}^k n_i = n. \quad (2.4)$$

Доказательство. Очевидно, что существует всего одно неупорядоченное $(1, 1, \dots, 1)$ -разбиение n -множества на n штук 1-подмножества. Любая транспозиция приводит к новому упорядоченному разбиению. Поэтому число упорядоченных $(1, 1, \dots, 1)$ -разбиений n -множества равно $n!$.

Если же в первое из подмножеств входит n_1 элементов, то число $(n_1, 1, \dots, 1)$ -разбиений n -множества будет равно $n! / n_1!$, так как транспозиция элементов, входящих в n_1 -подмножество, не приводит к новому упорядоченному разбиению. Если, кроме того, второе подмножество включает в себя n_2 элементов, то по указанным выше причинам число $(n_1, n_2, 1, \dots, 1)$ -разбиений n -множества равно $n! / (n_1! n_2!)$. Повторив подобные рассуждения k раз, придем к формуле (2.4).

Замечание. Числа R в (2.4) называют также полиномиальными коэффициентами. Это название обусловлено тем, что они являются коэффициентами при произведениях степеней переменных $x_1, x_2, \dots, x_k$ в разложении полинома по степеням x_i :

$$(x_1 + x_2 + \dots + x_k)^n = \sum_{n_1+n_2+\dots+n_k=n} \frac{n!}{n_1! n_2! \dots n_k!} x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}.$$

В частном случае, когда $x_1 = x_2 = \dots = x_k = 1$, имеем важное соотношение:

$$k^n = \sum_{n_1+n_2+\dots+n_k=n} \frac{n!}{n_1! n_2! \dots n_k!}.$$

Следствие. Число $(m, n-m)$ -разбиений n -множества равно числу его упорядоченных m -подмножеств. Это следует из сравнения формул (2.1) и (2.4):

$$R(n; m, n-m) = \frac{n!}{m!(n-m)!} = C_{n,m}, \quad m = \overline{1, n}. \quad (2.5)$$

Пример. Число различных слов, которое получим, переставляя буквы слова «математика», равно:

$$R(10; 2, 3, 2, 1, 1, 1) = \frac{10!}{2! 3! 2! 1! 1! 1!} = 151200,$$

так как кратность буквы m равна двум, a — трем, t — двум, остальные буквы встречаются по одному разу.

2.3.1.7. Теорема о числе неупорядоченных подмножеств. Число N_n всех неупорядоченных подмножеств n -множества равно:

$$N_n = 2^n.$$

Доказательство. Согласно теореме 3, число неупорядоченных m -подмножеств n -множества равно:

$$C_{n,m} = \frac{n!}{m!(n-m)!}.$$

Приняв во внимание формулу для полиномиальных коэффициентов, получим при $k = 2$:

$$N_n = \sum_{m=0}^n C_{n,m} = \sum_{m=0}^n \frac{n!}{m!(n-m)!}.$$

Пример. В комнате 4 различных светильника. Сколько вариантов включения светильников может быть реализовано?

Так как в задаче речь идет лишь о том, горит светильник или нет, то мы рассматриваем неупорядоченные разбиения, т. е. применима формула

$$N_4 = 2^4 = 16 \text{ (вариантов).}$$

2.3.2. Основные правила комбинаторики

2.3.2.1. Правило суммы. Пусть существует разбиение множества изучаемых комбинаций на классы, т. е. каждая комбинация входит в один и только в один класс. Тогда полное число ком-

бинаций равно сумме количества комбинаций, входящих в каждый из классов.

Иными словами, если некоторый объект типа a можно выбрать m способами, а объект типа b можно выбрать n способами, то выбор одного из этих объектов можно осуществить N_1 способами:

$$N_1 = m + n. \quad (2.6)$$

Это правило сумм справедливо лишь в том случае, если классы разбиения не пересекаются. Если же классы разбиения пересекаются, т. е. способы выбора объекта типа a совпадают со способами выбора объекта типа b , то из формулы (2.6) следует вычесть число k таких совпадений:

$$N_1' = m + n - k. \quad (2.7)$$

Пример. Пусть a — число, делящееся на два; b — число, делящееся на три. Сколькими способами можно выбрать или a , или b , если задано множество $M_5 = \{1, 2, 3, 4, 5\}$?

Решение. Согласно формуле (2.6) имеем: $m = 2$ (числа 2 и 4) и $n = 1$ (число 3), т. е. $N_1 = 2 + 1 = 3$.

Если подобный выбор осуществляется из множества $M_6 = \{1, 2, 3, 4, 5, 6\}$, то необходимо использовать уже формулу (2.7), т. к. 6 делится и на 2, и на 3: $N_1' = 3 + 2 - 1 = 4$.

Обобщая формулы (2.6) и (2.7) на случай i классов, получим формулу сумм:

$$N_1 = \sum_{i=1}^r n_i - \frac{1}{2} \sum_{i,j=1}^r k_{ij},$$

где n_i — число способов выбора объектов i -го типа; k_{ij} — число совпадений способов выбора объектов i -го и j -го типа.

2.3.2.2. Правило произведения. Если объект типа a можно выбрать m способами и если после каждого такого выбора объект типа b можно выбрать n способами, то выбор в указанном порядке пары $\langle a, b \rangle$ можно осуществить N_2 способами: $N_2 = m \cdot n$.

Действительно, если n — число способов выбора объекта типа b не зависит от того, как выбран объект типа a , то каждый из m способов выбора a может быть скомбинирован с n способами выбора b , а это и приводит к mn способам выбора пары $\langle a, b \rangle$.

Если же при i -м способе выбора объекта типа a объект b может быть выбран p_i способами, то число N_2' способов такого выбора пары $\langle a, b \rangle$ равно:

$$N_2' = \prod_{i=1}^n p_i.$$

Пример. Сколько существует целых четырехзначных чисел, не делящихся на 5?

Целое число не делится на 5, если оно не заканчивается на 5 или на 0. Поэтому первую значащую цифру можно выбирать девятью способами (все цифры, кроме нуля), вторую и третью — десятью способами, а четвертую — лишь восемью (все цифры, кроме 0 и 5). Следовательно, искомое число есть:

$$N_2 = \prod_{i=1}^4 p_i = 9 \cdot 10 \cdot 10 \cdot 8 = 7200,$$

где $p_1 = 9$, $p_2 = p_3 = 10$, $p_4 = 8$.

2.3.3. Комбинации объектов

2.3.3.1. Размещения. Размещения — это соединения из n элементов по m , которые отличаются друг от друга либо самими элементами, либо их порядком. Число размещений из n элементов по m обозначается A_n^m и равно числу m -подмножеств исходного n -множества. Таким образом, на основании теоремы о числе упорядоченных подмножеств, имеем:

$$A_n^m = A_{n,m} = n(n-1)\dots(n-m+1) = \frac{n!}{(n-m)!}.$$

Величина $A_{n,m}$ принимается равной единице, хотя она и не имеет комбинаторного смысла. Кроме того, по определению $A_0^0 = 1$ и $A_n^m = 0$ при значениях $n < m$ или если $m < 0$, а $n > 0$.

Имеет место рекуррентное (возвратное) соотношение:

$$A_n^m = A_{n-1}^m + mA_{n-1}^{m-1}. \quad (2.8)$$

Действительно, разобьем множество m -размещений из n элементов на два класса. Пусть в первом классе отсутствует один определенный элемент, тогда число размещений в этом классе будет A_{n-1}^m . В каждом из размещений второго класса этот элемент присутствует и занимает одно из m мест. Поэтому при его включении в не содержащие его $(m-1)$ -размещения

получается A_{n-1}^{m-1} размещений, в каждое из которых этот элемент входит.

2.3.3.2. Перестановки. Число всевозможных перестановок из n элементов обозначается P_n и равно частному случаю — числу размещений из n элементов по n , т. е.:

$$P_n = A_n^n = \frac{n!}{0!} = n!,$$

так как по определению $0! = 1$.

Подсчет P при больших значениях n становится затруднительным. В этих случаях можно рекомендовать использовать приближенную формулу Стирлинга:

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \dots\right).$$

2.3.3.3. Сочетания — это соединения из n элементов по m элементов, которые отличаются друг от друга только самими элементами.

Согласно определению, m -сочетания представляют собой неупорядоченные m -подмножества из n -подмножества, и, согласно теореме п. 2.3.1.4, их число определяется равенством так:

$$C_n^m = C_{n,m} = \frac{n!}{m!(n-m)!} = \frac{A_n^m}{P_m}.$$

Если в этой формуле заменить m на разность $n-m$, то получим соотношение:

$$C_n^m = C_n^{n-m}.$$

Проводя рассуждения аналогичные сделанным при доказательстве формулы (2.8), можно получить рекуррентную формулу для сочетаний C_n^m :

$$C_n^m = C_{n-1}^{m-1} + C_{n-1}^m.$$

Принято считать, что $C_0^0 = 1$; $C_n^m = 0$ при $n < m$ или при $m < 0$ и $n > 0$.

Числа C_n^m часто условно записывают в следующем виде: $\binom{n}{m}$.

Хотя величина не имеет комбинаторного смысла, понятие чисел можно распространить на отрицательные значения n , а именно:

$$C_{-n}^m = (-1)^m C_{n+m-1}^m.$$

Пример. Абонент забыл последние три цифры телефонного номера. Какое наибольшее число вариантов номеров ему нужно перебрать, чтобы дозвониться (в этом случае необходимый номер набирается последним)?

Очевидно, что таких номеров столько, сколько можно составить размещений из десяти цифр по три, т. е.:

$$N_1 = A_{10}^3 = 720.$$

Пример. Требуется составить колонну из пяти автомашин. Сколькими способами это можно сделать?

По условиям задачи порядок следования автомобилей может быть любым, поэтому количество способов составить автоколонну из пяти машин есть число перестановок из пяти:

$$N_2 = P_5 = 125.$$

2.3.3.4. Размещения с повторениями. Размещения с повторениями — это упорядоченные m -выборки из n -множества. Таких выборок, согласно теореме 5.2, будет

$$\overline{A_n^m} = \overline{A_{n,m}} = n^m \text{ штук.}$$

2.3.3.5. Перестановки с повторениями.

Перестановки с повторениями — это упорядоченные разбиения n -множества $(n: n_1, n_2, \dots, n_k)$. На основании формулы (2.4) имеем:

$$P_{n:n_1, n_2, \dots, n_k} = R(n: n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!},$$

где n_i — число повторений в перестановке элементов i -го типа.

Очевидно, что

$$P_{n:1,1,\dots,1} = P_n = n!.$$

Пример. Сколько различных слов можно составить, переставляя буквы в слове «перепел»?

Решение. Поскольку в слове имеются три буквы е и две буквы п, перестановки будут происходить с повторениями.

Поэтому искомое число есть:

$$P_{7:3,2,1,1} = \frac{7!}{3!2!1!1!} = 420.$$

2.3.3.6. Сочетания с повторениями. Сочетания с повторениями — это неупорядоченные m -выборки из n -множества. Поэтому число таких сочетаний, согласно теореме о числе неупорядоченных выборок, равно:

$$\overline{C_n^m} = \overline{C_{n,m}} = \frac{(n+m-1)!}{m!(n-1)!} = C_{n+m-1}^m$$

Имеет место также следующее рекуррентное соотношение:

$$C_n^m = C_n^{m-1} + C_{n-1}^m.$$

Пример. Число целых неотрицательных решений уравнения

$$x_1 + x_2 + \dots + x_m = n$$

равно числу сочетаний из m элементов по n элементов с повторениями. Это, в частности, означает, что уравнение

$$x_1 + x_2 + x_3 = 5$$

имеет $C_{5+4-1}^5 = C_8^5 = 56$ решений.

Для дальнейшего изучения материала по темам второго раздела можно рекомендовать литературу [6, 7, 8, 9, 10].

2.4. Задачи и упражнения к разделу 2.

1. Какие из формул тождественно истинны и какие тождественно ложны?

а) $x^2 + y^2 = 1 \vee x^2 + y^2 \neq 1 \quad \forall x, y \in R;$

б) $(A \wedge B \rightarrow C) \rightarrow (A \rightarrow (B \rightarrow C));$

в) $(A \rightarrow B) \wedge \neg B \rightarrow \neg A;$

г) $A \wedge \neg A.$

2. Докажите равносильность выражений:

а) $A \rightarrow B \equiv \neg(A \wedge \neg B);$

б) $\neg(A \rightarrow \neg B) \equiv A \wedge B;$

в) $A \vee C \rightarrow B \equiv (A \rightarrow B) \wedge (C \rightarrow B).$

3. Запишите в виде формул высказывания:

- «Картину написал Репин или Серов, но не Шишкин»;
- «Подозреваемого можно выбрать из группы присутствующих, если совпадает голос, или группа крови, или рост»;
- «Если завтра не будет дождя и будет тепло, то можно идти купаться»;
- «Можно установить, что человек вырос в данной местности, если его диалект и привычки такие же, что и у местных жителей»;
- «Если произведение трех чисел равно нулю, то одно из этих чисел равно нулю».

4. Постройте переключательную схему по формуле:

- $((X_1 \vee X_2) \vee \neg X_1) \rightarrow \neg(X_1 \wedge X_2)$;
- $(X_1 \wedge X_2) \rightarrow \neg(X_2 \vee X_3 \wedge X_1) \wedge X_2$;
- $\neg X_1 \vee X_2 \equiv (X_1 \wedge \neg X_2) \rightarrow X_3$.

5. Составьте таблицу истинности для следующих формул:

- $(X \rightarrow \neg X) \equiv X$;
- $(X_1 \vee X_2) \vee \neg X_2 \equiv X_1$;
- $((X_1 \wedge X_2) \wedge \neg X_2) \equiv X_1$;
- $((X_1 \rightarrow X_2) \wedge X_1) \rightarrow (X_1 \wedge X_2)$.

6. Для вещественных x и y , для которых:

- $\{(x, y) | (x+2)^2 + (y-1)^2 < 1 \vee x > 0\}$;
- $\{(x, y) | y > x^2 - 1 \wedge y > x - 1 \wedge x > 0 \wedge y > 0\}$,

построить соответствующие геометрические области.

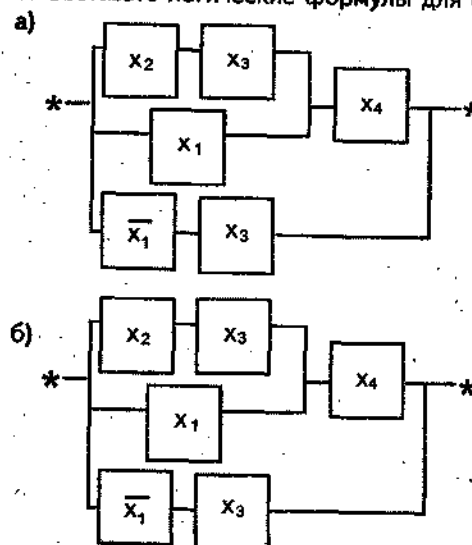
7. Запишите в виде формул теоремы:

- Каждое число делящееся на 4, делится и на 2;
- Некоторые числа делящиеся на 2 не делятся на 4.

8. Запишите словами следующие логические выражения и определите их истинность, если предикат D означает « x , квадрат числа x_2 », заданный на множестве целых чисел:

- $\forall x_1: D(x_1, x_2)$;
- $\exists x_1: D(x_1, x_2)$;
- $\forall x_1, \forall x_2: D(x_1, x_2)$;
- $\exists x_1 \forall x_2: D(x_1, x_2)$.

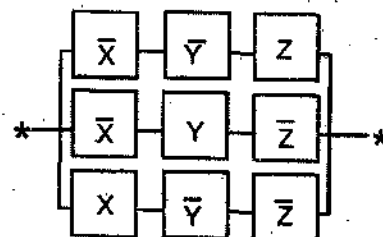
9. Составьте логические формулы для следующих схем:



10. Построить схему, соответствующую логической формуле A , заданной следующей таблицей истинности:

X	Y	Z	A
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Ответ:



11. Постройте граф возрастных отношений в вашей семье.

12. Постройте граф движения общественного транспорта в вашем районе и найдите кратчайшие пути между выделенными пунктами.

13. Задан граф на рис. 2.21.

а) Построить матрицу смежностей.

б) Построить различные маршруты из а в h.

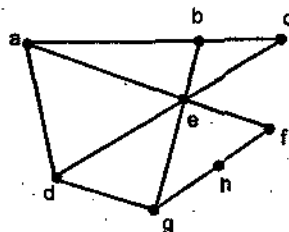


Рис. 2.21

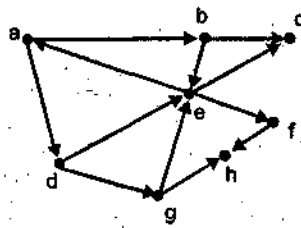


Рис. 2.22

14. Задан ориентированный граф, изображенный на рис. 2. 22.

а) Построить матрицу инцидентий.

б) Построить маршруты из а в f.

15. Для графа из примера 13 построить различные подграфы.

16. Для выигрыша в игральном автомате должны выпасть подряд четыре различные цифры. Сколько вариантов выигрышей при таких условиях возможно?

Решение. Существует всего десять цифр ($n = 10$). Все цифры в выигрышной ситуации должны быть различны, и порядок их следования безразличен. Следовательно, речь идет о числе N — числе возможных неупорядоченных 4-подмножествах из 10-множества.

$$N = C_{10}^4 = \frac{10!}{4! \cdot 6!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 21.$$

17. В колоде десять карт, из которых три — тузы. Наудачу последовательно вынимается, запоминаются и возвращаются в колоду четыре карты. После каждого возвращения карты колода перемешивается. Сколько возможно случаев, когда среди вынутых карт окажется хотя бы один туз?

Решение. Нас интересуют случаи, когда из четырех вынутых карт окажется один, два или три туза. В отличие от предыдущего случая нас интересуют сочетания с повторениями, поскольку после возвращения каждой вынутой карты исходная ситуация повторяется и следующей может появиться любая из десяти карт. Следовательно:

$$N = C_3^1 C_7^3 + C_3^2 C_7^2 + C_3^3 C_7^1 = 490.$$

18. Группу командировочных из восьми человек требуется расселить в три комнаты, из которых две трехместные и одна двухместная. Сколько вариантов расселения возможно?

Решение. Очевидно, что в задаче речь идет о разбиении 8-множества. Поэтому $8!$.

$$N = R(10; 3, 3, 2) = \frac{8!}{3! 3! 2!} = 560.$$

19. Дано n точек, из которых никакие три не лежат на одной прямой. Сколько прямых можно провести, соединяя точки попарно?

Ответ: $\frac{n(n-1)}{2}$.

20. Комиссия состоит из девяти человек. Материалы комиссии хранятся в сейфе. Сколько замков должен иметь сейф и сколько ключей для них нужно подготовить, чтобы доступ к сейфу был возможен тогда и только тогда, когда соберутся не менее шести членов комиссии?

Ответ: число замков — C_9^5 , число ключей — $4C_9^5$.

21. Сколько вариантов трехцветных полосатых флагов можно составить, если имеется возможность использовать пять цветов, но одна из полос должна быть красной?

Ответ: 36 вариантов.

22. Сколько различных слов можно получить, переставляя буквы в следующих исходных словах: а) академия, б) электротехника, в) молокопродукт?

Ответ: а) $\frac{8!}{2}$, б) $\frac{14!}{8}$, в) $\frac{13!}{48}$.

23. Сколько четырехзначных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, если:

Глава 3.1. Понятие пространства

- а) ни одна цифра не повторяется;
 б) все цифры могут повторяться;
 в) числа должны быть четными?
 Ответ: а) 300, б) 1080, в) 540.

24. Сколькими способами можно разделить 12 предметов между тремя студентами, чтобы каждому досталось ровно по четыре предмета?

Ответ: $\frac{12!}{4! \times 4! \times 4!} = 3150$.

25. Сколькими способами можно упорядочить множество $\{1, 2, 3, \dots, 2n\}$ так, чтобы каждое четное число стояло на четном месте?

Ответ: $(n!)^2$.

Под пространством понимают любую логически мыслимую форму или структуру, служащую средой, в которой реализуются те или иные конструкции и осуществляются определенные операции. Пространство можно определить как некоторое множество имеющее структуру. Пространство считается заданным, если между объектами множества установлены вполне определенные соотношения и (или) определены некоторые операции.

Поскольку понятие пространства сформировалось в результате абстрагирования и обобщения трехмерного евклидова пространства, то в пространстве обычно фиксируются отношения сходные по формальным признакам с этим пространством. Наиболее характерным среди них является расстояние.

Исторически первым сформировалось понятие *трехмерного геометрического пространства*, которое в дальнейшем обобщалось и трансформировалось. Один из вариантов обобщения — увеличение размерности объектов составляющих пространство вплоть до объектов бесконечной размерности. Другим примером обобщения служит переход от числовых последовательностей как элементов пространства к объектам, имеющим самую различную природу.

Возможность перехода от трехмерных векторов к многомерным появилась тогда, когда вектор стали рассматривать как упорядоченную последовательность n чисел. Примером многомерного пространства может выступать так называемое фазовое пространство. Каждая точка фазового пространства характеризуется упорядоченным набором параметров, описывающих состояние объекта рассмотрения. Так, экономическое состояние предприятия может характеризоваться стоимостью основных фондов, количеством работников, объемом выпускаемой продукции, ее себестоимостью и т. п., которые в совокупности можно рассматривать как точку в n -мерном пространстве, а изменения экономического состояния — как траекторию (годограф) движения в этом пространстве состояний.

Интересным примером может служить трехмерное *цветовое пространство*, состоящее из векторов, компоненты которого суть интенсивности красного, зеленого и синего цветов. Изменяя

интенсивности только этих трех цветов и накладывая их потом друг на друга, можно получать цветовую палитру с неограниченным числом различных оттенков. На этом принципе, например, основана работа цветных электронно-лучевых трубок в телевизорах и мониторах компьютеров. Формально переход от одного цвета или оттенка к другому можно описать перемещением от точки к точке в трехмерном цветовом пространстве. При этом изменение цвета можно измерять количественно, используя операции над векторами.

Дальнейшее абстрагирование привело к введению в рассмотрение не векторных пространств. Здесь элементами пространства являются собственно рассматриваемые объекты (люди, слова, законы и т. п.), операции над которыми выполняются по специально установленным правилам. Формализация операций над такого рода объектами имеет свою специфику и часто не похожа на отношения, используемые в классических геометрических пространствах. Так, появление компьютерных баз данных породило специфические объекты (поля, записи, массивы, файлы и т. д.) и операции над ними (введение информации, ее редактирование, отображение и т. п.).

Глава 3.2. Геометрическое пространство

3.2.1. Числовая ось

Возьмем прямую и зафиксируем на ней две точки, первой сопоставим число 0 и назовем точкой начала отсчета, второй сопоставим число 1. Выбор второй точки указывает положительное направление на прямой и масштаб для измерения длин. Положительные и отрицательные числа изобразятся совокупностью равноотстоящих точек (рис. 3.1), причем положительные числа отмечаются по выбранному направлению, а отрицательные — в противоположном. Прямая, для которой указано направление, начало отсчета и масштаб, называется **числовой осью**. Если каждый из единичных отрезков разделить

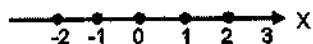


Рис. 3.1. Числовая ось

на n равных частей, то точки деления будут изображать дроби со знаменателем n . Проведя подобную операцию со всеми натуральными числами, по-

лучим изображение всех **рациональных чисел**, то есть чисел типа m/n . Следовательно, каждому рациональному числу поставлена в соответствие определенная точка числовой оси, называемая иногда **числовой точкой**. Можно доказать, что на любом сколь угодно малом интервале числовой оси всегда находятся рациональные точки. Этот факт фундаментальной важности выражается так: *рациональные точки расположены на числовой оси всюду плотно*.

Если число m меньше числа n , оно расположено на числовой оси левее, чем n . Каждая пара точек m и n , вместе со всеми точками между ними, называется **отрезком** числовой оси (или **сегментом**) и обозначается $[m, n]$. Если же рассматриваются только промежуточные точки между m и n , то говорят о **промежутке** (или **интервале**) числовой оси (m, n) .

Расстояние от точки 0 до точки m есть **положительное число**, которое называется **абсолютной величиной** числа m и обозначается $|m|$. Расстояние между точками m и n есть **положительное число**, которое называется **длиной отрезка** $[m, n]$ и обозначается $|m, n|$.

Пусть отрезок $B = [x_1, y_1]$ находится внутри отрезка $A = [x_2, y_2]$. При сравнении этих отрезков возможны два принципиально разных случая.

Если существуют такие два числа n и m , что длины отрезков A и B удовлетворяют соотношению $n|x_2, y_2| = m|x_1, y_1|$, то говорят, что отрезки A и B **соизмеримы**. Это означает, что они обладают некоторой общей мерой длины, т. е. отрезок длиной A/n содержится в отрезке A ровно n раз, а в отрезке B — ровно m раз.

Для иллюстрации второго случая рассмотрим пример. Возьмем квадрат со стороной равной 1. Его диагональ, согласно теореме Пифагора, имеет длину x , удовлетворяющую уравнению $x^2 = 1^2 + 1^2 = 2$. Отсюда $x = \sqrt{2}$. Если бы x было соизмеримо с 1, то можно было бы найти такие два целых числа p и q , что $x = p/q$. В этом случае $p^2 = 2q^2$. Однако можно доказать, что такого равенства быть не может [4]. Вместе с тем, при помощи циркуля на числовой оси от 0 можно отложить отрезок равный

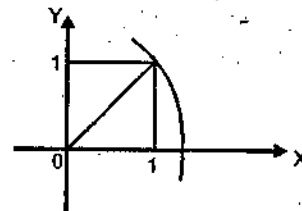


Рис. 3.2. Иллюстрация к понятию иррационального числа

диагонали квадрата. Построенная таким образом точка (правая граница отрезка x) существует на числовой оси и не является рациональной. Такие точки, а следовательно, и числа, несоизмеримые с единицей, называются **иррациональными**.

На основании вышеизложенного можно заключить, что между точками числовой оси, ее отрезками и вещественными числами существует взаимно-однозначное соответствие. Тем самым устанавливается четкая связь между геометрическими объектами — отрезками прямой и вещественными числами.

3.2.2. Системы координат на плоскости

На плоскости наиболее часто используются декартова и полярная системы координат.

Прямоугольная (декартова) система координат вводится следующим образом: выбираются две взаимно перпендикулярные числовые оси, которые пересекаются в точке O , называемой **началом системы координат**. Горизонтальную ось, имеющую положительное направление слева направо, называют **осью абсцисс**, а вертикальную, направленную снизу вверх, — **осью ординат**.

Каждой точке плоскости M можно сопоставить ориентированный отрезок OM берущий начало в точке O и оканчивающийся в точке M (рис. 3.3). Такой отрезок называют **радиус-вектором** точки M .

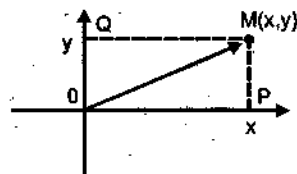


Рис. 3.3. Декартова система координат

Зафиксируем произвольную точку M .

Отрезок OM спроектируем ортогонально на обе оси. Получим ориентированный отрезок числовой оси абсцисс OP , координату точки P обозначим за x , и ориентированный отрезок числовой оси ординат OQ , координату

точки Q обозначим за y . Числа x и y называются **координатами** точки M в декартовой системе координат. Принципиально важным является то, что положение любой точки плоскости M вполне определяется заданием координат этой точки — упорядоченной пары чисел (x, y) . Задать точку в

фиксированной системе координат означает указать значения ее координат.

На плоскости расстояние между двумя точками измеряется по прямой и полагается равным длине соединяющего их отрезка. В фиксированной системе координат расстояние d между двумя точками на плоскости $M(x_1, y_1)$ и $N(x_2, y_2)$ вычисляется по следующей формуле (теорема Пифагора (Пифагор, 570–496 г. до н. э.)):

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

Пример. Найти расстояние d между двумя точкам $M(-3, 4)$ и $N(5, 2)$. Согласно вышеприведенной формуле, имеем:

$$d = \sqrt{(-3 - 5)^2 + (4 - 2)^2} = \sqrt{64 + 4} = \sqrt{68}$$

Введение системы координат позволяет геометрические объекты определять при помощи формул. Наука, занимающаяся исследованием геометрических форм с помощью алгебраического анализа, называется **аналитической геометрией**. В аналитической геометрии найти геометрическую фигуру или тело означает записать уравнение, решение которого определяет координаты искомой фигуры (тела). Рассмотрим с точки зрения аналитической геометрии некоторые наиболее важные для приложений линии на плоскости.

3.2.2.1. Прямая линия на плоскости

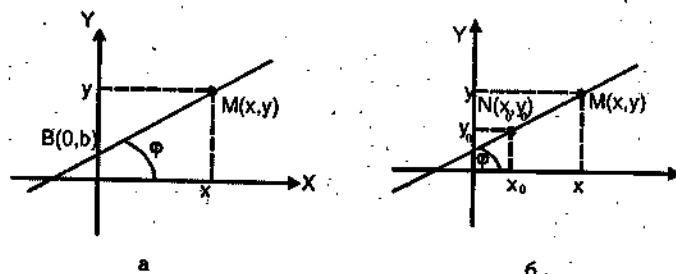


Рис. 3.4. Прямая линия на плоскости

Пусть, как показано на рис. 3.4а, в фиксированной системе координат прямая линия пересекает ось ординат в точке $B(0, b)$ под углом ϕ . Выберем на прямой произвольную точку $M(x, y)$,

такая точка называется текущей. Проекции направленного отрезка BM на оси координат соответственно равны $пр. BM = x$, $пр. BM = y - b$. При скольжении точки M по прямой проекции изменяются, однако, их отношение, равное

$$\operatorname{tg} \varphi = \frac{y - b}{x} \quad (3.1)$$

сохраняется для всех точек нашей прямой и не выполняется для точек, не принадлежащих прямой. Тангенс угла φ называется **угловым коэффициентом** и обозначается k . Выразив y , получим «уравнение прямой линии с угловым коэффициентом».

$$y = x \operatorname{tg} \varphi + b \quad \text{или} \quad y = kx + b, \quad (3.2a)$$

Если $b=0$, то прямая проходит через начало координат. Если $k=0$ (рис. 3.5a), то прямая проходит параллельно оси абсцисс и ее уравнение $y = b$.

Если вместо точки B дана другая фиксированная точка $N(x_0, y_0)$ (рис. 3.4в), то, проводя аналогичные рассуждения, легко получить «уравнение прямой, проходящей через данную точку»:

$$k = \operatorname{tg} \varphi = \frac{y - y_0}{x - x_0} \quad \text{или} \quad y - y_0 = k(x - x_0) \quad (3.2b)$$

Любое из уравнений (3.2) можно привести к виду $Ax + By + C = 0$. Например, для уравнения (3.2a) $A = k$, $B = -1$, $C = b$, т. е. прямая в прямоугольной системе координат может быть описана линейным уравнением первой степени. Если $B \neq 0$, то и линейное уравнение можно привести к виду (3.2b):

$$y = -\frac{A}{B}x - \frac{C}{B}, \quad k = -\frac{A}{B}, \quad b = -\frac{C}{B}.$$

Если $B = 0$, то получим уравнение

$$Ax + C = 0, \quad \text{или} \quad x = -\frac{C}{A}.$$

Это уравнение определяет прямую параллельную оси ординат и пересекающую ось абсцисс в точке $x = a$ ($a = -\frac{C}{A}$) (рис. 3.5б).

Следовательно, уравнение $Ax + By + C = 0$ описывает только прямые линии на плоскости и называется **общим уравнением прямой** на плоскости. Верно и обратное утверждение: каждому уравнению первой степени с двумя неизвестными соответствует в прямоугольной системе координат одна и только одна прямая.

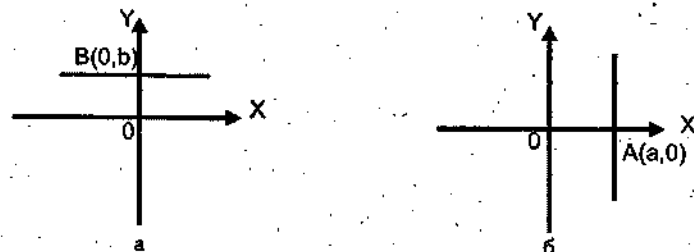


Рис. 3.5. Прямые линии на плоскости, параллельные осям координат

Напомним, что в евклидовой геометрии через две точки на плоскости можно провести только одну прямую и бесконечное множество кривых линий.

3.2.2.2. Кривые второго порядка

Для практики большое значение имеют линии, которые в прямоугольной системе координат описываются алгебраическими уравнениями второй степени:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (3.3)$$

причем хотя бы один из коэффициентов A, B, C должен быть не равен нулю. Рассмотрим наиболее важные частные случаи уравнения (3.3).

Окружность. Окружностью называется геометрическое место точек, равноудаленных от данной точки, называемой центром окружности.

Уравнение окружности радиуса R с центром в точке $M(a, b)$ имеет вид:

$$(x - a)^2 + (y - b)^2 = R^2 \quad (3.4)$$

Если раскрыть скобки, то мы увидим, что уравнение (3.4) получается из уравнения (3.3), если $A = C = 1$, $B = 0$, $D = -2a$, $E = -2b$, $F = -R^2 + a^2 + b^2$.

Пусть задано уравнение $x^2 + y^2 - 4x = 0$. Является ли это уравнение уравнением окружности, и если да, то каков ее радиус и координаты центра? Попробуем привести данное уравнение к виду (3.4). Выделим полный квадрат относительно x , прибавляя и вычитая число 4.

$$x^2 + y^2 - 4x = (x^2 - 4x + 4) + y^2 - 4 = 0$$

или

$$(x - 2)^2 + y^2 = 2^2.$$

Сравнивая (3.5) с (3.4), видим, что заданное уравнение есть уравнение окружности радиусом $R = 2$ и с центром в точке M с координатами $a = 2, b = 0$.

Эллипс. Эллипс — замкнутая кривая, для всех точек которой сумма расстояний до двух фиксированных точек, называемых фокусами эллипса, одинакова и равна, по определению, $2a$. Для эллипса, представленного на рис. 3.6, сумма расстояний MF_1 и MF_2 равна сумме расстояний NF_1 и NF_2 и равна $2a$. Каноническое уравнение эллипса, центр симметрии которого находится в начале координат, а фокусы лежат на оси OX симметрично относительно оси OY [11],

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (3.6)$$

Параметры a и b называются его **полуосями**. Уравнение (3.6) получим из (3.3), если $B=D=E=0, A=1/a^2, C=1/b^2, F=-1$. Очевидно, что окружность — частный случай эллипса, у которого $a = b = R$, а центр находится в начале координат.

Гипербола. Гипербола — неограниченная кривая, для всех точек которой разность расстояний до двух фиксированных точек, называемых фокусами, есть величина постоянная (рис. 3.7). Разность $MF_2 - MF_1$ равна разности $NF_2 - NF_1$ и, по определению, равна $2a$. Каноническое уравнение гиперболы, центр симметрии которой совпадает с началом координат, а фокусы лежат на оси OX симметрично оси OY [11],

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \quad (3.7)$$

Параметры a и b называются полуосью и мнимой полуосью гиперболы. Уравнение (3.7) получим из (3.3) если $B=D=E=0, A=1/a^2, C=(-1)/b^2, F=-1$. Особенность гиперболы — наличие асимптот — прямых, к которым неограниченно приближается кривая при $x \rightarrow \infty$. Уравнения асимптот: $y = \pm \frac{b}{a}x$.

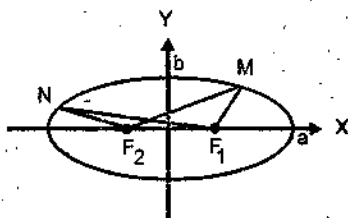


Рис. 3.6. Эллипс

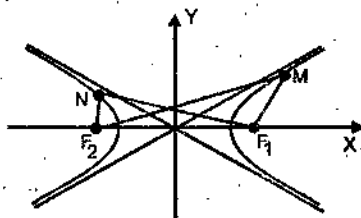


Рис. 3.7. Гипербола

Парабола. Парабола — неограниченная кривая, все точки которой (рис.3.8) равноудалены от точки, называемой фокусом, и прямой, называемой директриссой. Для параболы, изображенной на рис. 3.8, расстояния $MK = MF, NF = NL$ и $DO = OF$. Каноническое уравнение параболы, ось которой совпадает с осью OX , а вершина лежит в начале координат [11],

$$y^2 = 2px. \quad (3.8)$$

Уравнение (3.8) получим из (3.3), если $A=B=E=F=0, C=1, D=-2p$.

Сделав поворот и сдвиг системы координат (см. раздел 3.4.2), любое уравнение (3.3) можно привести только к одному из трех уравнений второй степени (3.6), (3.7), (3.8) или к уравнению вида $a^2 x^2 = b^2 y^2$, которому соответствуют две прямые. Это означает, что уравнениями второй степени можно описать только эллипс (и его частный случай окружность), гиперболу или параболу. Эти кривые называют кривыми второго порядка. Все другие кривые в прямоугольной системе координат будут описываться уравнениями более высокого порядка. Важным свойством линий второго порядка является то, что все они могут быть получены (рис.3.9) как сечения конуса плоскостью, пересекающего его под различными углами.

На плоскости можно ввести бесконечное число систем координат. В каждой из этих систем точка M будет иметь свои координаты. Для сравнения объектов на плоскости обычно

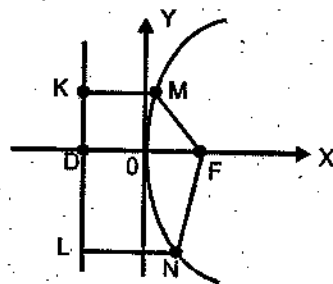


Рис. 3.8. Парабола

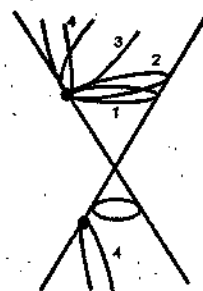


Рис. 3.9. Конические сечения:

1. Окружность
2. Эллипс
3. Парабола
4. Гипербола

фиксируют одну систему координат и в ней ведут все рассмотрение. Если же необходимо вводить несколько систем координат, то устанавливают правила пересчета координат одной и той же точки при переходе из одной системы в другую.

Рассмотрим следующую задачу. Заданы две системы прямоугольных координат X_1Y_1 и X_2Y_2 (рис. 3.10а). Зная координаты точки $M(x_1, y_1)$ в одной из систем, найти ее же координаты (x_2, y_2) в другой системе координат.

Решение задачи удобно провести в два этапа: вначале совмещаются начала координат, причем сохраняются старые направления осей (рис. 3.10б), потом одна из систем поворачивается так, чтобы совпали направления осей координат.

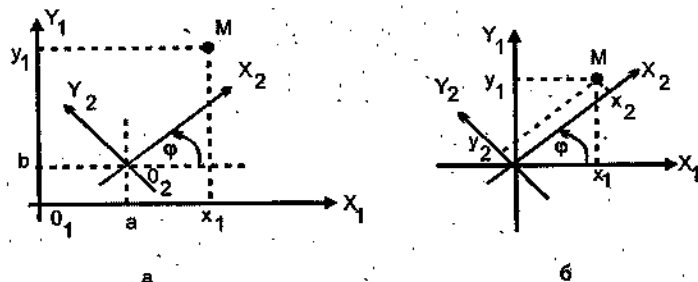


Рис. 3.10. Преобразования координат

Параллельный перенос системы координат. В первой системе координат точка O имеет координаты $(0, 0)$, точка O_2 — (a, b) , а точка M — (x_1, y_1) . Рассматривая проекции этих точек на оси координат первой системы, имеем:

$$x_1 = a + x_2, \quad y_1 = b + y_2. \quad (3.9)$$

Формулы (3.9) означают, что первая координата равна второй плюс соответствующая координата начала второй системы. Чтобы получить координаты во второй системе, необходимо провести обратные действия. Это приведет к зависимостям

$$x_2 = x_1 - a, \quad y_2 = y_1 - b. \quad (3.10)$$

Поворот системы координат с общей точкой начала. Пусть оси OX_1 и OX_2 повернуты на угол φ . Из рис. 3.10б следуют соотношения:

$$\begin{aligned} x_1 &= x_2 \cos \varphi - y_2 \sin \varphi, \\ y_1 &= x_2 \sin \varphi + y_2 \cos \varphi. \end{aligned} \quad (3.11)$$

В общем случае связь между координатами точки в различных прямоугольных системах координат выражается линейными соотношениями:

$$\begin{aligned} x_1 &= x_2 \cos \varphi - y_2 \sin \varphi + a, \\ y_1 &= x_2 \sin \varphi + y_2 \cos \varphi + b \end{aligned} \quad \text{или} \quad (3.12)$$

$$\begin{aligned} x_2 &= x_1 \cos \varphi + y_1 \sin \varphi - a, \\ y_2 &= -x_1 \sin \varphi + y_1 \cos \varphi - b. \end{aligned}$$

Пример. Как изменятся координаты точки $M(-2, 3)$, если система будет повернута на 30° и сдвинута вверх на две единицы?

Применяя формулы (3.12) для $x_1 = -2$, $y_1 = 3$, угла $\varphi = 30^\circ$, $a = 0$ и $b = 2$, имеем:

$$x_2 = -2 \cos 30^\circ + 3 \sin 30^\circ = -2 \frac{\sqrt{3}}{2} + 3 \frac{1}{2} = \frac{3}{2} - \sqrt{3}$$

$$y_2 = 2 \sin 30^\circ + 3 \cos 30^\circ - 2 = 2 \frac{1}{2} + 3 \frac{\sqrt{3}}{2} - 2 = \frac{3\sqrt{3}}{2} - 1$$

Для определения положения точек на плоскости часто применяется так называемая **полярная система координат**. Выберем на плоскости фиксированную точку O , называемую **полюсом**, и исходящую из нее ось OP , называемую **полярной осью**. На полярной оси выбрана единица масштаба. В этой системе, как показано на рис. 3.11, положение точки M на плоскости вполне задается отрезком OM , называемым **полярным радиусом** точки M , равным расстоянию отрезка OM , и углом φ , который составляет полярный радиус с полярной осью, считая против часовой стрелки от полярной оси ($-\pi < \varphi \leq \pi$).

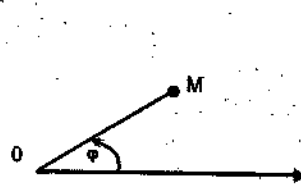


Рис. 3.11. Полярная система координат

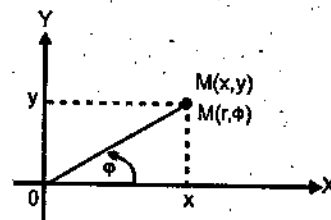


Рис. 3.12. Связь полярной и декартовой систем координат

Пусть на плоскости заданы прямоугольная и полярная системы координат. При этом начало координат и положительная часть оси абсцисс прямоугольной системы совпадают, соответственно, с полюсом и осью полярной системы координат, как показано на рис. 3.12. Точка М имеет в прямоугольной системе координаты x и y , а в полярной — r и φ , которые связаны очевидным соотношением:

$$\begin{aligned} x &= r \cos \varphi, \\ y &= r \sin \varphi. \end{aligned} \quad (3.13)$$

Формулы (3.13) выражают координаты точки М в прямоугольной системе через ее же координаты в полярной системе.

Возводя оба уравнения из (3.13) и складывая их почленно, получим:

$$x^2 + y^2 = r^2(\cos^2 \varphi + \sin^2 \varphi) = r^2$$

или

$$r = \sqrt{x^2 + y^2}. \quad (3.14)$$

С другой стороны, из равенств (3.13) следует, что

$$\operatorname{tg} \varphi = \frac{y}{x}. \quad (3.15)$$

Формулы (3.14) и (3.15) выражают полярные координаты точки М через ее координаты в прямоугольной системе координат.

3.2.3. Системы координат в пространстве

В трехмерном пространстве чаще всего используются декартова, цилиндрическая и сферическая системы координат.

Прямоугольная (декартова) система координат возникает, если взяты три одинаковые взаимно перпендикулярные числовые оси — оси координат, которые пересекаются в точке О, называемой началом системы координат. Первую ось ОХ называют осью абсцисс, вторую ось ОУ — осью ординат, третью ОZ — осью аппликат. Через каждые две координатные оси проходит координатная плоскость.

Существуют две, не сводящиеся друг к другу, системы координат: **правая система координат** и **левая система координат**. Различить системы координат можно следующим образом: если посмотреть из любой положительной точки оси

ОZ на ось ОУ и ось ОХ окажется справа, то это правая система координат, если слева — левая (сравните рис. 3.13а и рис. 3.13б).

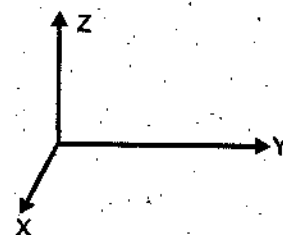


Рис. 3.13 а. Правая система координат

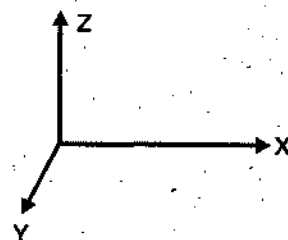


Рис. 3.13 б. Левая система координат

Каждой пространственной точке М можно сопоставить ориентированный отрезок ОМ, берущий начало в точке начала координат и оканчивающийся в точке М (рис. 3.14). Такой отрезок называют **радиус-вектором** точки М.

Спроектируем, сделав прямоугольное проектирование, точку М на оси координат. Тогда каждой точке М соответствуют три точки на осях (на рис. 3.14 Р, Q, R), их координаты называют **координатами** точки М. Они однозначно определяют положение этой точки в выбранной системе координат. Наоборот, задав на каждой из осей координат по одной точке, например, Р, Q, и R, мы определим одну и только одну точку в пространстве (на рис. 3.14 точка М). Эта точка получается при пересечении трех взаимно перпендикулярных плоскостей РМ₁ММ₂, QМ₂ММ₃, RМ₃ММ₁, проходящих, соответственно, через точки Р, Q и R параллельно осям координат.

Определение по данной точке М ее координат и построение точки в пространстве по ее координатам — две основные задачи метода координат в пространстве.

Пусть заданы две точки в пространстве: М(х₁, у₁, z₁) и N(х₂, у₂, z₂). Расстоянием между двумя точками в пространстве

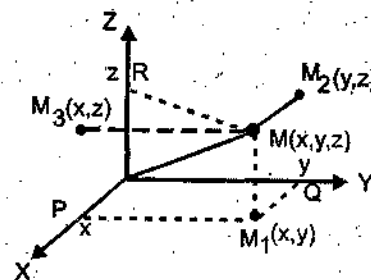


Рис. 3.14. Метод координат в пространстве

называется число d , равное длине отрезка прямой соединяющей эти точки:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}. \quad (3.16)$$

Например, расстояние между двумя точками $M(2, -1, 3)$ и $N(-2, -1, 0)$, согласно (3.16), равно:

$$d = \sqrt{(2 - (-2))^2 + (-1 - (-1))^2 + (3 - 0)^2} = \sqrt{25} = 5.$$

Если на плоскости основными объектами рассмотрения были линии и образуемые ими плоские фигуры, то в пространстве главным образом рассматриваются поверхности и образуемые ими тела.

Всякая поверхность может рассматриваться как некоторое множество точек в пространстве, между координатами которых установлены определенные соответствия. Например, связь между координатами задана уравнением

$$F(x, y, z) = 0, \quad (3.17)$$

которое получено из геометрических свойств поверхности.

Наиболее простой вид уравнения (3.17) — уравнение, линейное относительно всех неизвестных $Ax + By + Cz + D = 0$, которое описывает **плоскость в пространстве**. Если $D = 0$, то уравнение $Ax + By + Cz = 0$ описывает плоскость, проходящую через начало координат.

Другой пример — сфера. Сфера есть геометрическое место точек в пространстве, равноудаленных от одной точки называемой центром. Величина удаления точек сферы от центра есть расстояние от точки центра до точек сферы. Следовательно, используя (3.16), можно записать уравнение сферы:

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = R^2, \quad (3.18)$$

где R — радиус сферы или расстояние от произвольной точки сферы (x, y, z) до ее центра — фиксированной точки с координатами (x_0, y_0, z_0) .

Линию в пространстве образует пересечение двух поверхностей. Отсюда следует, что пространственную линию можно описать системой двух уравнений:

$$\begin{cases} F_1(x, y, z) = 0 \\ F_2(x, y, z) = 0. \end{cases}$$

Пример. Найти линию, образуемую пересечением плоскости $z = 1$ со сферой $(x - 3)^2 + (y + 5)^2 + (z - 2)^2 = 16$.

Решение. Искомая линия находится как решение системы этих уравнений:

$$\begin{cases} (x - 3)^2 + (y + 5)^2 + (z - 2)^2 = 16 \\ z = 1. \end{cases}$$

Решение этой системы есть уравнение окружности:

$$(x - 3)^2 + (y + 5)^2 = 15.$$

Следовательно, плоскость пересекает сферу по окружности.

Пересечение трех поверхностей может давать просто точку в пространстве. Математически это соответствует единственному решению системы трех уравнений:

$$\begin{cases} F_1(x, y, z) = 0 \\ F_2(x, y, z) = 0 \\ F_3(x, y, z) = 0. \end{cases} \quad (3.19)$$

Если система (3.19) несовместна, то это означает, что поверхности, описываемые данными уравнениями, не пересекаются в одной точке.

К наиболее распространенным относятся следующие поверхности.

1. **Цилиндрические поверхности** — это поверхности, описываемые прямой, называемой **образующей**,двигающейся параллельной фиксированной заданной прямой и пересекающей некоторую линию L , называемую **направляющей** цилиндрической поверхности. Направляющая линия не обязательно замкнута.

В частности, если образующая параллельна оси OZ , то уравнение такой цилиндрической поверхности описывается уравнением, не содержащим z :

$$F(x, y) = 0.$$

В этом случае вид функции F определяет направляющую линию цилиндра. Так (рис. 3.16а, б, в) в пространстве:

уравнение $(x - a)^2 + (y - b)^2 = R$ описывает круговой цилиндр,

уравнение $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ описывает эллиптический цилиндр,

уравнение $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ описывает гиперболический цилиндр.

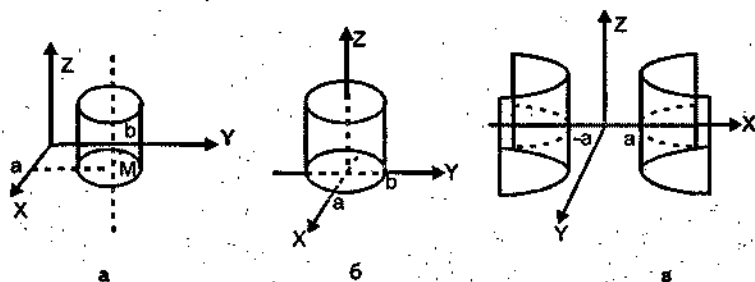


Рис. 3.15. Цилиндры в пространстве

Пример. Какую поверхность определяет следующее уравнение $x^2 + z^2 - 4x - 4z + 4 = 0$?

Решение. Вопросы подобного рода решаются путем выделения полных квадратов в левой части уравнения. Если прибавить к обеим частям заданного уравнения число 4, то, переставляя местами члены уравнения, получим:

$$x^2 - 4x + 4 + z^2 - 4z + 4 = 4$$

или

$$(x - 2)^2 + (z - 2)^2 = 4.$$

Это уравнение описывает круговой цилиндр, вытянутый вдоль оси OY (координата y отсутствует).

2. Конические поверхности. Поверхность, описываемая прямой (образующая конической поверхности) проходящей через данную точку, называемую вершиной, и пересекающей данную линию (направляющую конуса), называется **конической поверхностью**.

Наиболее простой формулой описывается конус, имеющий вершину в начале координат, а его образующая описывает вокруг оси координат некоторую замкнутую кривую, например, как показано на рис. 3.16, эллипс. Уравнения такого конуса имеет вид

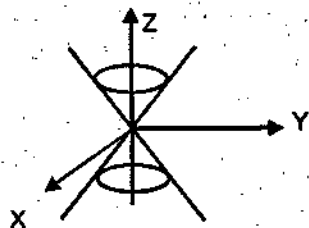


Рис. 3.16. Конус

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0.$$

Пример. Найти уравнение поверхности, получаемой при вращении прямой вокруг оси OX.

$$\begin{cases} x + 2y = 4 \\ z = 0 \end{cases}$$

Решение. Вершиной конуса будет являться точка пересечения его образующей с осью OX с координатами $x = 4, y = 0, z = 0$. Произвольная фиксированная точка образующей прямой $A(x, y, 0)$ при вращении вокруг оси OX описывает окружность, задаваемую уравнением $Y^2 + Z^2 = y^2$, где X, Y, Z — произвольные точки поверхности искомого конуса, соответствующие сечению $x = X$. Подставляя значения x и y в уравнение образующей прямой, имеем искомое уравнение конуса:

$$X + 2\sqrt{Y^2 + Z^2} = 4,$$

или, после преобразования:

$$4Y^2 + 4Z^2 - (X - 4)^2 = 0.$$

3. Поверхности вращения. Пусть в плоскости YOZ лежит кривая, уравнение которой $F(y, z) = 0$. Если вращать эту кривую вокруг оси OZ, то образуется поверхность вращения описываемая уравнением

$$F(\sqrt{x^2 + y^2}, z) = 0. \quad (3.20)$$

При анализе поверхностей вращения в каждом конкретном случае необходимо указывать, в какой плоскости лежит образующая кривая и вокруг какой оси она вращается. Так, например, эллипсоид вращения, описываемый уравнением

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1$$

образован вращением вокруг оси OZ эллипса, лежащего в плоскости XOZ (рис. 3.16а). Если этот же эллипс вращать вокруг оси OX, то уравнение соответствующего эллипсоида вращения (рис. 3.16б) имеет вид:

$$\frac{y^2 + z^2}{c^2} + \frac{x^2}{a^2} = 1.$$

Пример. Записать уравнение эллипсоида вращения, полученного от вращения эллипса вокруг оси OY, если на его поверхности лежат точки $A(3, 0, 0)$ и $B(0, 2, 0)$.

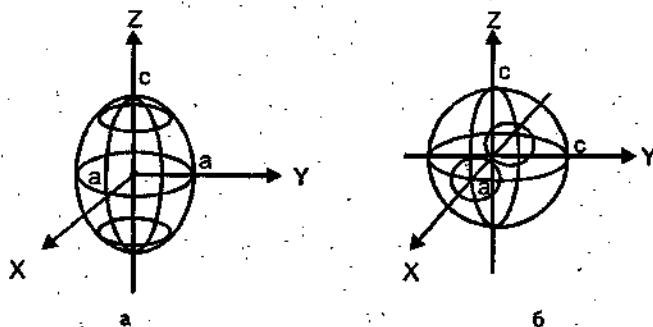


Рис. 3.16. Эллипсоиды вращения

Решение. Заданные точки лежат в координатной плоскости XOY и определяют вершины эллипса ($a = 3$, $b = 2$), вращение которого образует искомый эллипсоид. Принимая во внимание предыдущие рассуждения, запишем уравнение эллипсоида вращения:

$$\frac{x^2 + z^2}{9} + \frac{y^2}{4} = 1.$$

Существует огромное количество поверхностей, которые можно описать аналитически, а затем исследовать их свойства и применять для решения практических задач, используя математические методы и компьютерную технику.

Глава 3.3. Матрицы

3.3.1. Основные определения

Матрицей называется совокупность чисел или объектов другой природы (элементов), расположенных в виде прямоугольной таблицы. В дальнейшем, если специально не оговорено, будем рассматривать матрицы, состоящие только из чисел. Матрицы обозначаются различными способами:

$$A \equiv \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \equiv \left\| \begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{matrix} \right\|$$

или, сокращенно, $A = (a_{ij})_{m,n}$.

Числа a_{ij} , составляющие матрицу, называются ее элементами. Первый индекс указывает номер строки, второй — номер столбца. Если матрица состоит из m строк и n столбцов, то говорят, что размерность матрицы есть m на n . Количество элементов в такой матрице равно произведению $m \times n$.

Матрица называется **прямоугольной**, если $m \neq n$. Если $m = n$, то матрица называется **квадратной** и число n — порядком матрицы. Матрица, содержащая один столбец, называется **столбцовая матрица** или **матрица-столбец**. Матрица, состоящая из одной строки, — **матрица-строка**. Матрица, все элементы которой равны нулю, называется **нулевой матрицей**. Нулевая матрица может быть любого размера.

Рассмотрим квадратную матрицу порядка n , т. е. $A = (a_{ij})_{n,n}$. Совокупность элементов a_{11} , a_{22} , ..., a_{nn} , у которых оба индекса совпадают, образуют **главную диагональ**. Квадратная матрица называется **диагональной**, если все элементы, стоящие вне главной диагонали, равны нулю: $a_{ij} = 0$ при $i \neq j$. Диагональная матрица обозначается так:

$$A = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \equiv \text{diag}(a_{11}, a_{22}, \dots, a_{nn}).$$

Диагональная матрица, у которой все диагональные элементы равны единице, называется **единичной** и обозначается

$$I = E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Квадратная матрица называется **треугольной**, если все элементы, стоящие ниже (или выше) главной диагонали, равны нулю:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

Каждой квадратной матрице ставится в соответствие число, называемое **детерминантом** или **определителем**. Обозначается символами $\det A$ или $\Delta(A)$, или $|a_{ij}|$. Определитель квадратной матрицы A представляет собой числовую функцию, аргументы которой есть элементы этой матрицы. Общее выражение определителя матрицы n -го порядка задается в следующем виде:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum_{\sigma} (-1)^{\nu} a_{1\alpha_1} a_{2\alpha_2} \dots a_{n\alpha_n} \quad (3.21)$$

В правой части выражения (3.21) стоит сумма всех произведений, которые можно образовать из n элементов матрицы по правилу: по одному из каждой ее строки и из каждого ее столбца. Это означает, что в каждом произведении среди всех первых и среди всех вторых индексов не должно быть одинаковых. Первые индексы располагают в возрастающем порядке. Отсюда следует, что каждое произведение в (3.21) образует подстановку n -ой степени:

$$\sigma = \begin{pmatrix} 1, & 2, & 3, & \dots, & n \\ \alpha_1, & \alpha_2, & \dots, & \dots, & \alpha_n \end{pmatrix}$$

Поскольку число подстановок из n чисел равно $n!$, то в (3.21) $n!$ слагаемых вида $a_{1\alpha_1} a_{2\alpha_2} \dots a_{n\alpha_n}$. Число ν равно числу инверсий соответствующей перестановки $\alpha_1, \alpha_2, \dots, \alpha_n$.

Пример. Вычислить определитель матрицы

$$A = \begin{vmatrix} 3 & 2 & -1 \\ 5 & -4 & 8 \\ 1 & 7 & -2 \end{vmatrix}$$

Решение. В соответствии с определением (3.21) имеем:

$$\det A = \begin{vmatrix} 3 & 2 & -1 \\ 5 & -4 & 8 \\ 1 & 7 & -2 \end{vmatrix} = (3 \cdot (-4) \cdot 4) + ((-1) \cdot 5 \cdot 7) + (1 \cdot 2 \cdot 8) - ((-1) \cdot (-4) \cdot 1) - (3 \cdot 8 \cdot 7) - (2 \cdot 5 \cdot 4) = -279.$$

Следует помнить, что матрица представляет собой сложный единый объект (таблицу), а ее определитель — число.

Определитель единичной матрицы равен единице: $\det E = 1$.

Квадратная матрица называется **невырожденной** (неособенной), если $\det A \neq 0$, и **вырожденной** (особенной), если $\det A = 0$. Легко проверить, что определитель треугольной матрицы равен произведению диагональных элементов:

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11} a_{22} \begin{vmatrix} a_{33} & \dots & a_{3n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{nn} \end{vmatrix} = \dots = a_{11} a_{22} a_{33} \dots a_{nn}.$$

Отсюда следует, что если все диагональные элементы в треугольной матрице отличны от нуля, то матрица невырожденная.

3.3.2. Действия над матрицами

Равенство матриц. Две матрицы $A = (a_{ij})_{m,n}$ и $B = (b_{ij})_{k,q}$ называются равными, если они одинаковых размеров ($m = k$, $n = q$) и их соответствующие элементы равны ($a_{ij} = b_{ij}$).

Сложение матриц. Складывать можно лишь матрицы одинакового размера. Суммой двух матриц $A = (a_{ij})_{m,n}$ и $B = (b_{ij})_{m,n}$ называется матрица $C = (c_{ij})_{m,n}$ тех же размеров. Элементы матрицы C равны суммам соответствующих элементов матриц A и B , т. е. $C = A + B$, если $c_{ij} = a_{ij} + b_{ij}$.

Пример.

$$\begin{pmatrix} 2 & -5 & 1 & 4 \\ 3 & 0 & 7 & -3 \\ 8 & 2 & 5 & -6 \end{pmatrix} + \begin{pmatrix} 1 & 8 & -5 & -9 \\ 0 & 6 & -4 & 2 \\ 0 & 3 & 1 & -3 \end{pmatrix} = \begin{pmatrix} 3 & 3 & -4 & -5 \\ 3 & 6 & 3 & -1 \\ 8 & 5 & 6 & -9 \end{pmatrix}$$

Умножение матрицы на число. Произведением матрицы $A = (a_{ik})_{m,n}$ на число α называется матрица $C = (c_{ij})_{m,n}$, элементы которой получаются из соответствующих элементов матрицы A умножением на число α : $C = \alpha \times A$, где $c_{ij} = \alpha \cdot a_{ij}$.

Пример.

$$(-3) \cdot \begin{pmatrix} 2 & -5 & 1 & 4 \\ 3 & 0 & 7 & -3 \\ 8 & 2 & 5 & -6 \end{pmatrix} = \begin{pmatrix} -6 & 15 & -3 & -12 \\ -9 & 0 & -21 & 9 \\ -24 & -6 & -15 & 18 \end{pmatrix}$$

Легко проверить, что операции сложения матриц и умножения матриц на число обладают следующими свойствами:

- 1) $A + B = B + A$ (перестановочный закон);
- 2) $(A + B) + C = A + (B + C)$ (сочетательный закон);
- 3) $(\alpha + \beta)A = \alpha A + \beta A$ (распределительный закон);
- 4) $(A + B) = A + B$;
- 5) $A + (-A) = 0$;
- 6) $A + 0 = A$,

где $A, B, C, 0$ матрицы одного размера, α, β — числа.

Замечание. Пусть A, B — квадратные матрицы порядка $n > 1$. Линейные операции над матрицами не переносятся на их определители, т. е. $\det(A + B) \neq \det A + \det B$, $\det(\alpha \cdot A) \neq \alpha \cdot \det A$, но $\det(\alpha A) = \alpha^n \cdot \det A$.

Произведение матриц. Произведение матриц $A \cdot B$ определено в том и только в том случае, если число столбцов матрицы A равно числу строк матрицы B . Произведением матрицы A размера

$m \cdot k$ на матрицу B размера $k \cdot n$ называется матрица $C = A \cdot B$ размера $m \cdot n$, элементы которой определяются по формуле

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{ik}b_{kj} = \sum_{r=1}^k a_{ir} \cdot b_{rj} \quad (3.22)$$

Сумма в (3.22) представляет собой скалярное произведение вектора-строки i матрицы A на вектор-столбец j матрицы B . Поэтому говорят, что умножение матриц производится по правилу «строка на столбец», т. е. для получения элемента c_{ij} необходимо скалярно умножить строку i матрицы A на столбец j матрицы B .

Пример. Требуется перемножить матрицы

$$A = \begin{pmatrix} 0 & 1 \\ 3 & 4 \end{pmatrix} \text{ и } B = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$$

Поскольку это матрицы квадратные и одного размера, то умножение таких матриц возможно всегда. В соответствии с (3.22) имеем:

$$A \cdot B = \begin{pmatrix} 0 & 1 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0+0 & 0+1 \\ 3+0 & 9+4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 9 & 13 \end{pmatrix};$$

$$B \cdot A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 0+9 & 1+12 \\ 0+3 & 0+4 \end{pmatrix} = \begin{pmatrix} 9 & 13 \\ 3 & 4 \end{pmatrix}.$$

Из примера следует важное правило: «Произведение матриц не перестановочно», т. е. в общем случае $A \cdot B \neq B \cdot A$.

Пример. Найти произведение матриц

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 2 & -2 \\ 3 & -3 & 2 \end{pmatrix} \text{ и } B = \begin{pmatrix} 0 & 1 \\ 3 & 4 \\ 2 & 1 \end{pmatrix}$$

В соответствии с правилом умножения матриц произведение $B \cdot A$ не существует, а произведение

$$A \cdot B = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 2 & -2 \\ 3 & -3 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 3 & 4 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} -3 & -3 \\ 2 & 7 \\ -5 & -7 \end{pmatrix}$$

Можно проверить, что умножение матриц удовлетворяет свойствам:

$$1) A \cdot (B \cdot C) = (A \cdot B) \cdot C;$$

$$2) \alpha \cdot (A \cdot B) = (\alpha \cdot A) \cdot B = A \cdot (\alpha \cdot B);$$

$$3) (A + B) \cdot C = A \cdot C + B \cdot C;$$

$$4) C \cdot (A + B) = C \cdot A + C \cdot B,$$

где α — число, A, B, C — матрицы, для которых произведения существуют;

$$5) \text{ если } A, B \text{ — квадратные матрицы одного порядка, то } \det(A \cdot B) = \det A \cdot \det B.$$

Транспонирование матриц. Рассмотрим произвольную матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Матрица

$$A^t = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

полученная из матрицы A заменой строк столбцами, называется **транспонированной по отношению к A** . Например,

$$A = \begin{pmatrix} 0 & 1 \\ 3 & 4 \\ 2 & 1 \end{pmatrix}, A^t = \begin{pmatrix} 0 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix}.$$

Можно доказать, что:

$$1) (A^t)^t = A;$$

$$2) (A + B)^t = A^t + B^t;$$

$$3) (\alpha \cdot A)^t = \alpha A^t;$$

$$4) (A \cdot B)^t = B^t \cdot A^t,$$

где A и B — матрицы, α — число;

$$5) \text{ Если } A \text{ — квадратная матрица, то } \det A = \det A^t.$$

Квадратная матрица называется **симметрической**, если $A^t = A$ и **кососимметрической**, если $A^t = -A$. У симметрической матрицы $a_{ij} = a_{ji}$, т. е. все элементы, стоящие ниже главной диагонали, равны соответствующим элементам, расположенным выше диагонали. Пример симметрической матрицы:

$$A = \begin{pmatrix} 3 & 2 & 4 & 2 \\ 2 & 5 & 6 & 3 \\ 4 & 6 & 4 & 1 \\ 2 & 3 & 1 & 7 \end{pmatrix}.$$

У кососимметрической матрицы $a_{ij} = -a_{ji}$, следовательно, $a_{ii} = 0$. Пример кососимметрической матрицы:

$$A = \begin{pmatrix} 0 & 2 & 4 & 2 \\ -2 & 0 & -3 & 1 \\ -4 & 3 & 0 & 3 \\ -2 & -1 & -3 & 0 \end{pmatrix}.$$

Произведение матрицы A на транспонированную матрицу A^t есть симметрическая матрица. Для доказательства этого факта проверим, что $C = A \cdot A^t$ — симметрическая матрица. Действительно, $C^t = (A \cdot A^t)^t = (A^t)^t A^t = A \cdot A^t = C$.

Пусть в определителе матрицы A размером $m \times n$ выделено k ($k \leq n$) произвольных строк и столько же произвольных столбцов. Элементы расположенные на пересечении этих строк и столбцов образуют определитель k -го порядка, называемый **минором k -го порядка матрицы A** .

Например, для матрицы

$$A = \begin{pmatrix} 2 & -5 & 0 & 1 \\ -7 & 3 & 5 & 3 \\ 4 & 1 & -8 & 4 \end{pmatrix}$$

все элементы матрицы являются минорами первого порядка:

$|2|, |-5|, |0|, |1|, \dots, |4|$
(здесь знак $|-5|$ означает минор первого порядка, а не модуль числа -5);
определители

$$\begin{vmatrix} 2 & -5 \\ -7 & 3 \end{vmatrix}, \begin{vmatrix} 2 & 0 \\ -7 & 0 \end{vmatrix}, \begin{vmatrix} 3 & 3 \\ 1 & 4 \end{vmatrix}, \begin{vmatrix} 3 & 5 \\ 1 & -8 \end{vmatrix}, \dots, \begin{vmatrix} -7 & 3 \\ 4 & 4 \end{vmatrix} -$$

миноры второго порядка, а миноры третьего порядка для этой матрицы следующие:

$$\begin{vmatrix} 2 & -5 & 0 \\ -7 & 3 & 5 \\ 4 & 1 & -8 \end{vmatrix}, \begin{vmatrix} 2 & 0 & 1 \\ -7 & 5 & 3 \\ 1 & -8 & 4 \end{vmatrix}, \begin{vmatrix} -5 & 0 & 1 \\ 3 & 5 & 3 \\ 1 & -8 & 4 \end{vmatrix}, \begin{vmatrix} 2 & -5 & 1 \\ -7 & 3 & 3 \\ 4 & 1 & 4 \end{vmatrix}.$$

Число r называется **рангом** матрицы A (обозначается $\text{rg}A$), если у нее существует хотя бы один минор порядка r не равный нулю, а все миноры большего порядка, если они существуют, равны нулю. Очевидно, что ранг матрицы не может превышать наименьшее из чисел m, n . Если ранг матрицы равен r , то ее минор порядка r не равный нулю называется **базисным минором**. Базисных миноров у матрицы может быть несколько. Миноры, образованные строками и столбцами с одинаковыми номерами, называются **главными**.

Пример. Найти ранг и базисный минор матрицы

$$A = \begin{vmatrix} 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 0 & -1 & 1 \end{vmatrix}.$$

Все четыре минора третьего порядка этой матрицы равны нулю. Если же построить определитель из строк 1, 3 и столбцов 1, 3, то он не равен нулю; т. е.

$$\begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix} = -1 \neq 0,$$

следовательно, $\text{rg}A = 2$.

Элементарными называются следующие преобразования матрицы:

- 1) перестановка двух строк (столбцов);
- 2) умножение одной из строк (столбцов) на число не равное нулю;

3) прибавление к одной из строк (столбцов) другой строки (столбца), умноженной предварительно на произвольный не равный нулю множитель.

Все матрицы, полученные из данной путем элементарных преобразований, — разные матрицы. Вместе с тем все они обладают некоторыми одинаковыми свойствами, например имеют одинаковый ранг. По этой причине матрицы полученные в результате элементарных преобразований называют **эквивалентными**. При переходе к эквивалентной матрице вместо знака « $=$ » часто используют знак соответствия « $\div$ ».

3.3.3. Обратная матрица

Пусть A — квадратная матрица n -го порядка, а E — единичная матрица того же порядка. Матрица B называется **обратной** по отношению к матрице A , если выполняются равенства:

$$A \cdot B = B \cdot A = E.$$

Матрицу, обратную к матрице A , обозначают символом A^{-1} .

Алгебраическим дополнением элемента a_{ij} матрицы A называется определитель, получаемый из определителя исходной матрицы A путем изъятия из него строки i и столбца j и взятый со знаком $(-1)^{i+j}$. Например, если матрица A имеет определитель

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

то алгебраическим дополнением элемента a_{23} этой матрицы, поскольку $(-1)^{2+3} = -1$, является определитель

$$A_{23} = - \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}$$

Составим матрицу $\bar{A}$, называемую **присоединенной** (или **союзной**) к матрице A . Элементами присоединенной матрицы являются алгебраические дополнения соответствующих элемен-

тов исходной матрицы A , расположенные, как в транспонированной по отношению к ней матрице.

$$\bar{A} = \begin{vmatrix} A_{11} & A_{21} & A_{31} & \dots & A_{n1} \\ A_{12} & A_{22} & A_{32} & \dots & A_{n2} \\ A_{13} & A_{23} & A_{33} & \dots & A_{n3} \\ \dots & \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & A_{3n} & \dots & A_{nn} \end{vmatrix}$$

Присоединенная матрица обладает свойством:

$$A \cdot \bar{A} = \bar{A} \cdot A = E \cdot \det A. \quad (3.23)$$

Теорема существования обратной матрицы. Для того, чтобы для матрицы A существовала обратная, необходимо и достаточно, чтобы матрица A была невырожденной, т. е. $\det A \neq 0$.

Доказательство необходимости. Пусть A^{-1} существует, тогда $A \cdot A^{-1} = E$. Вычислим определитель от правой и левой части этого равенства: $\det A \cdot \det A^{-1} = 1$. Отсюда следует, что $\det A$ не может равняться 0.

Доказательство достаточности. Пусть $\det A \neq 0$. Построим присоединенную матрицу $\bar{A}$, при ее помощи некоторую матрицу $\bar{B} = \frac{1}{\det A} \bar{A}$. Вычислим произведение $A \cdot \bar{B}$, применяя (3.23):

$$A \cdot \bar{B} = A \cdot \frac{1}{\det A} \bar{A} = \frac{1}{\det A} \cdot \det A \cdot E = E$$

Поскольку произведение матрицы A на какую-то другую матрицу B равно единичной матрице, то B — обратная по отношению к A матрица. Теорема доказана.

Теорема единственности обратной матрицы. Пусть для матрицы A существует обратная по отношению к ней матрица A^{-1} и для некоторой матрицы X выполняется равенство $A \cdot X = E$. Тогда $X = A^{-1}$, т. е. X — обратная к A матрица.

Доказательство. Дано $A \cdot X = E$. Умножим это равенство слева на A^{-1} . Получим

$$A^{-1} \cdot (A \cdot X) = (A^{-1} \cdot A) \cdot X = E \cdot X = X$$

$$A^{-1} \cdot E = A^{-1} \rightarrow X = A^{-1}$$

Что и требовалось доказать.

Свойства обратной матрицы непосредственно следуют из ее определения. Вот основные из них:

- 1) $(A^{-1})^{-1} = A$
- 2) $(AB)^{-1} = B^{-1} \cdot A^{-1}$
- 3) $(A^T)^{-1} = (A^{-1})^T$

$$4) \det A^{-1} = \frac{1}{\det A}$$

Докажем, например, свойство 2, вычислив произведение

$$(A \cdot B) \cdot (B^{-1} \cdot A^{-1}) = A \cdot (B \cdot B^{-1}) \cdot A^{-1} = A \cdot A^{-1} = E$$

Следовательно, $B^{-1} \cdot A^{-1}$ — обратная по отношению к AB матрица.

Пример. Найти обратную матрицу для

$$A = \begin{vmatrix} 2 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{vmatrix}$$

Вначале вычислим определитель этой матрицы: $\det A = 6$, следовательно, обратная матрица A^{-1} существует.

Теперь построим присоединенную к A матрицу $\bar{A}$. Для этого вычислим алгебраические дополнения для всех элементов A :

$$A_{11} = \begin{vmatrix} 1 & 4 \\ 0 & 3 \end{vmatrix} = 3, \quad A_{12} = -\begin{vmatrix} 0 & 4 \\ 0 & 3 \end{vmatrix} = 0, \quad A_{13} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} = 0,$$

$$A_{21} = 0, \quad A_{22} = 6, \quad A_{23} = 0, \quad A_{31} = -3, \quad A_{32} = -8, \quad A_{33} = 2.$$

Присоединенная по отношению к A матрица

$$\bar{A} = \begin{vmatrix} 3 & 0 & -3 \\ 0 & 6 & -8 \\ 0 & 0 & 2 \end{vmatrix}$$

Обратная к A матрица

$$A^{-1} = \frac{\bar{A}}{\det A} = \begin{vmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & \frac{1}{3} \end{vmatrix}$$

Глава 3.4. Линейные пространства

Геометрические векторы — это направленные отрезки в пространстве или на плоскости. Свойства и отношения, характерные для геометрических векторов, используются для описания многих понятий в окружающей нас действительности. Вместе с тем, наиболее существенные из этих свойств и отношений присущи также другим математическим объектам, рассматривая которые мы приходим к понятию линейного пространства. Множество же геометрических векторов часто оказывается частным случаем линейного пространства и может служить во многих ситуациях его наглядной иллюстрацией.

3.4.1. Определение линейного пространства

Пусть на каком-либо множестве объектов определены действия сложения и умножения на число. Это означает, что указанные действия имеют смысл и результатами их действий являются элементы того же самого множества. Например, сложение определено на множестве матриц одной размерности, а для матриц различной размерности сложение не имеет смысла. Умножение на вещественное число во множестве целых чисел не определено, так как результатом такого умножения может оказаться нецелое число (объект другого множества).

Линейным пространством называется множество, на котором определены операции сложения и умножения на число, удовлетворяющие следующим условиям:

1. Коммутативность: $X + Y = Y + X$;
2. Ассоциативность: $(X + Y) + Z = X + (Y + Z)$;
3. Существует нулевой элемент 0 такой, что для любого X $X + 0 = X$;
4. Для любого элемента X существует противоположный ему элемент $(-X)$ такой, что $X + (-X) = 0$;
5. Пусть c и d — числа, тогда:
 $c \cdot (X + Y) = c \cdot X + c \cdot Y$; $(c + d)X = cX + dX$;
6. $c \cdot (d \cdot X) = (c \cdot d) \cdot X$; $1 \cdot X = X$;
7. Кроме того: $0 \cdot X = 0$; $(-1) \cdot X = -X$; $(-c) \cdot X = -(c \cdot X)$

Примерами линейного пространства являются: пространство действительных чисел, все множество геометрических векторов

на плоскости, пространство матриц фиксированной размерности, пространство решений однородных линейных систем и др.

Будем рассматривать в качестве элементов линейного пространства n -мерные векторы как упорядоченные наборы из n вещественных чисел, называемых **координатами** или **компонентами векторов**, т. е.

$$X \equiv (x_1, x_2, x_3, \dots, x_n).$$

Линейное пространство таких n -мерных векторов обычно называют **n -мерным вещественным векторным пространством** и обозначают R^n . Числа x_i ($i = 1, n$) — составляющие вектора, n — размерность вектора.

Сумма двух n -мерных векторов X и Y есть новый n -мерный вектор Z , составляющие которого равны суммам соответствующих составляющих складываемых векторов, т. е.

$$Z = X + Y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \dots, x_n + y_n).$$

Складывать можно лишь векторы одинаковой размерности. Например,

$$(2, -1, 0, 4) + (-5, 0, 2, 2) = (-3, -1, 2, 6).$$

Но операция $(3, -1, 9) + (6, 2)$ не определена. В то время как, например, $(3, -1, 9) + (6, 0, 2) = (9, 1, 11)$. При этом, однако, следует помнить, что $(6, 2) \neq (6, 0, 2)$.

Произведение вектора X на число λ есть новый вектор Y , составляющие которого равны произведению каждой составляющей вектора X на это число, т. е.

$$Y = \lambda \cdot X = (\lambda \cdot x_1, \lambda \cdot x_2, \lambda \cdot x_3, \dots, \lambda \cdot x_n).$$

Пусть заданно m векторов $X_1, X_2, \dots, X_m$. Из этих векторов в линейном пространстве может быть построена следующая комбинация: каждый из векторов умножается на произвольное число, а затем полученные результаты складываются, т. е.

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_m X_m = \sum_{i=1}^m \lambda_i X_i. \quad (3.24)$$

Выражение (3.24) называется **линейной комбинацией векторов**.

Скалярным произведением векторов X и Y называется число равное сумме парных произведений соответствующих составляющих этих векторов, т. е.

$$(X, Y) \equiv \sum_{i=1}^n x_i \cdot y_i.$$

Например, скалярное произведение двух 5-мерных векторов $X = (2, -4, 0, -5, 3)$ и $Y = (-1, 0, 6, -2, 2)$ равно:

$$(X, Y) = 2 \cdot (-1) + (-4) \cdot 0 + 0 \cdot 6 + (-5) \cdot (-2) + 3 \cdot 2 = 14.$$

Вещественное векторное пространство, в котором введено скалярное произведение, называется **евклидовым пространством**. Из определения следуют основные свойства скалярного произведения:

1. $(X, Y) = (Y, X)$;
2. $(X + Y, Z) = (X, Z) + (Y, Z)$;
3. $(cX, Y) = (X, cY) = c(X, Y)$;
4. $(X, X) \geq 0$.

Нормой (или длиной) вектора X называется число

$$|X| = \sqrt{\sum_{i=1}^n x_i^2}.$$

Два вектора называются **ортогональными**, если их скалярное произведение равно нулю, т. е. $(X, Y) = 0$.

Система векторов называется **ортогональной**, если любые два вектора из этой системы ортогональны. Примером ортогонального векторного пространства может служить система трех взаимно перпендикулярных векторов (декартова система координат в трехмерном геометрическом пространстве).

3.4.2. Линейная независимость векторов

В задачах, где рассматриваются геометрические векторы, зачастую бывает важным выяснить, являются ли **векторы компланарными**, т. е. лежат ли они в одной и той же плоскости.

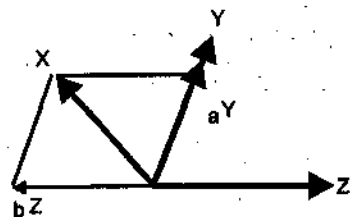


Рис. 3.17. Представление вектора в виде линейной комбинации других векторов

Если все векторы лежат в одной плоскости, то речь идет о плоской задаче. Если они не лежат в одной и той же плоскости, то рассматривается стереометрическая задача.

Рассмотрим три трехмерных вектора X, Y, Z , лежащих в одной плоскости. В этом случае один из них, на-

пример X , можно, как показано на рис. 3.17, разложить по направлению двух других.

Аналитически это означает, что вектор X можно представить как линейную комбинацию векторов Y и Z .

$$X = a \cdot Y + b \cdot Z,$$

где a, b — числовые множители.

Если же X, Y, Z не компланарны (не лежат в одной плоскости), то ни один из них в виде линейной комбинации представить нельзя.

Пусть задана система из m штук n -мерных векторов:

$$X_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$$

$$X_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$$

$$\dots$$

$$X_m = (x_{m1}, x_{m2}, x_{m3}, \dots, x_{mn}).$$

Векторы называются **линейно зависимыми**, если существуют такие не все равные нулю числа $\lambda_1, \lambda_2, \dots, \lambda_n$, что выполняется равенство

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_n X_n = 0. \quad (3.25)$$

В развернутом виде через проекции векторов эту формулу можно записать так:

$$\lambda_1 x_{11} + \lambda_2 x_{12} + \lambda_3 x_{13} + \dots + \lambda_n x_{1n} = 0$$

$$\lambda_1 x_{21} + \lambda_2 x_{22} + \lambda_3 x_{23} + \dots + \lambda_n x_{2n} = 0$$

$$\dots$$

$$\lambda_1 x_{n1} + \lambda_2 x_{n2} + \lambda_3 x_{n3} + \dots + \lambda_n x_{nm} = 0. \quad (3.26)$$

Если равенство (3.25) выполняется только тогда, когда все $\lambda_i = 0$, то $X_1, X_2, X_3, \dots, X_n$ называются **линейно независимыми**. Иначе говоря, векторы линейно независимы, если их линейная комбинация равна нулю тогда и только тогда, когда все ее коэффициенты равны нулю.

Содержательно линейная зависимость совокупности векторов означает, что в ней имеются «лишние векторы». Эти «лишние» векторы не несут информации, которая не могла бы быть получена из всех векторов совокупности, состоящей только из линейно независимых векторов.

Теорема. Для того, чтобы векторы $X_1, X_2, X_3, \dots, X_m$ были линейно зависимы, необходимо и достаточно, чтобы хотя бы один из них являлся линейной комбинацией других.

Доказательство необходимости. Пусть $X_1, X_2, X_3, \dots, X_m$ векторы линейно зависимы, т. е. существуют $\lambda_1, \lambda_2, \dots, \lambda_m$, не все равные нулю, такие, что:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_m X_m = 0.$$

Предположим, что, например, $\lambda_1 \neq 0$. Тогда:

$$X_1 = -\frac{\lambda_2}{\lambda_1} X_2 - \dots - \frac{\lambda_m}{\lambda_1} X_m.$$

т. е. X_1 есть линейная комбинация остальных векторов.

Доказательство достаточности. Пусть вектор X_1 есть линейная комбинация других векторов:

$$X_1 = c_2 X_2 + c_3 X_3 + \dots + c_m X_m.$$

Тогда:

$$-X_1 + c_2 X_2 + c_3 X_3 + \dots + c_m X_m = 0,$$

т. е. $X_1, X_2, X_3, \dots, X_m$ — линейно зависимы. Теорема доказана.

Теорема. Для того, чтобы векторы $X_1, X_2, \dots, X_n$ в пространстве R^n были линейно независимы, необходимо и достаточно, чтобы определитель, составленный из компонент этих векторов, был отличен от нуля:

$$\det A = \begin{vmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & X_{n3} & \dots & X_{nn} \end{vmatrix} \neq 0.$$

Доказательство необходимости. Пусть векторы $X_1, X_2, \dots, X_n$ линейно независимы, тогда равенство $\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n = 0$ выполняется только при нулевых λ_i , т. е. однородная алгебраическая система (3.26) имеет только нулевые решения. Как известно (см., например [9]), это возможно лишь в том случае, когда определитель системы не равен нулю ($\det A \neq 0$).

Доказательство достаточности. Пусть $\det A \neq 0$. Тогда однородная система (3.26) имеет только нулевое решение и, следовательно, $X_1, X_2, \dots, X_n$ линейно независимы. Теорема доказана.

Пример. Дана система векторов

$$e_1 = (1, 0, \dots, 0)$$

$$e_2 = (0, 1, \dots, 0)$$

(3.27)

.....

$$e_n = (0, 0, \dots, 1).$$

Линейно зависимы или линейно независимы эти векторы?

Для ответа на этот вопрос составим определитель:

$$\Delta = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix} = \det E = 1 \neq 0.$$

Следовательно, по доказанной теореме, данные векторы линейно независимы.

Замечание. Система векторов, содержащая нулевой вектор, всегда линейно зависима.

3.4.3. Базис системы векторов

Дана система векторов (конечная или бесконечная). Совокупность n -мерных векторов $e_1, e_2, \dots, e_n$ называется **базисом** (системой координат) векторного пространства R^n , если они линейно независимы и любой вектор этого пространства является их линейной комбинацией:

$$X = x_1 e_1 + x_2 e_2 + x_3 e_3 + \dots + x_n e_n. \quad (3.28)$$

Равенство (3.28) называется разложением вектора X по базису $e_1, e_2, \dots, e_n$, а компоненты вектора $x_1, x_2, \dots, x_n$ называются его координатами в этом базисе.

Имеют место следующие утверждения.

1. Каждый вектор системы может быть разложен по базису единственным способом.

2. Два различных базиса одной и той же системы векторов содержат одинаковое количество векторов.

3. Система n линейно независимых векторов $e_1, e_2, \dots, e_n$ образует базис в пространстве R^n .

4. В пространстве R^n система, состоящая более чем из n векторов, линейно зависима.

Теперь становится ясным, что если рассматривается некоторый вектор X в фиксированной системе координат, имеющей базис,

то речь идет уже о системе из $n + 1$ векторов (X и базис) и возможно представление этого вектора в виде (3.28).

Определение. **Размерностью** конечномерного линейного пространства R^n называется число векторов его базиса.

Пример. Можно ли принять за базис пространства R^3 следующие вектора (матрицы-столбцы):

$$a) \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix} \quad б) \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} ?$$

Составим из этих столбцов определители и вычислим их значения:

$$a) \begin{vmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 3 & 4 \end{vmatrix} = 0, \quad б) \begin{vmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 0 & 2 & 1 \end{vmatrix} = 2 \neq 0.$$

Следовательно, вектора (матрицы-столбцы) «а» базис не образуют, а столбцы «б» — образуют.

Важное практическое значение имеет система векторов специального вида (3.27). Все они имеют единичную длину и попарно ортогональны. В этой связи система (3.27) часто выбирается за систему координат в n -мерном векторном пространстве (аналог декартовой системы координат в трехмерном пространстве). Такая система координат (базис) называется **стандартной**.

3.4.4. Преобразование координат

Пусть задано два базиса в пространстве R^n :

1. $[e_1, e_2, \dots, e_n]$;
2. $[e'_1, e'_2, \dots, e'_n]$.

Любой вектор можно разложить по базисным векторам 1 и 2 системы:

$$X = x_1 e_1 + x_2 e_2 + \dots + x_n e_n;$$

$$X = x'_1 e'_1 + x'_2 e'_2 + \dots + x'_n e'_n,$$

или в компактной форме:

$$X = \sum_{i=1}^n x_i e_i = \sum_{i=1}^n x'_i e'_i. \quad (3.29)$$

Найдем формулы преобразования координат одного и того же вектора X при переходе от одного базиса к другому. Для этого разложим каждый вектор второго базиса (как $(n + 1)$ -ый вектор в n -мерном пространстве) по элементам первого базиса:

$$e'_1 = a_{11} e_1 + a_{12} e_2 + \dots + a_{1n} e_n$$

$$e'_2 = a_{21} e_1 + a_{22} e_2 + \dots + a_{2n} e_n$$

$$\dots \dots \dots$$

$$e'_n = a_{n1} e_1 + a_{n2} e_2 + \dots + a_{nn} e_n,$$

или, в компактной форме:

$$e'_k = \sum_{i=1}^n a_{ik} e_i, \quad k = \overline{1, n}. \quad (3.30)$$

Из координат a_{ik} векторов e'_k в первой системе базисных векторов построим векторы-столбцы типа

$$X_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \dots \\ a_{nk} \end{pmatrix},$$

из которых сформируем матрицу, называемую **матрицей перехода** от базиса $\{e_i\}$ к базису $\{e'_k\}$.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Так как базисные векторы $[e'_1, e'_2, \dots, e'_n]$ линейно независимы, то матрица A будет невырожденной ($\det A \neq 0$). Подставим в (3.29) вместо e'_k формулу (3.30) и изменим порядок суммирования:

$$X = \sum_{i=1}^n x_i e_i = \sum_{k=1}^n x'_k \cdot \left(\sum_{i=1}^n a_{ik} e_i \right) = \sum_{i=1}^n e_i \cdot \sum_{k=1}^n a_{ik} x'_k.$$

Отсюда, в силу единственности разложения вектора по базису, имеем:

$$x_i = \sum_{k=1}^n a_{ik} x'_k, \quad i = \overline{1, n}.$$

Следовательно, (в развернутой форме) координаты вектора X в первом базисе $(x_1, x_2, \dots, x_n)$ связаны с координатами этого же вектора во втором базисе $(x'_1, x'_2, \dots, x'_n)$ следующим образом:

$$\begin{aligned} x_1 &= a_{11}x_1' + a_{12}x_2' + \dots + a_{1n}x_n' \\ x_2 &= a_{21}x_1' + a_{22}x_2' + \dots + a_{2n}x_n' \\ &\dots\dots\dots \end{aligned} \quad (3.31)$$

$$X_n = a_{n1}X_1' + a_{n2}X_2' + \dots + a_{nn}X_n'.$$

Запишем формулы перехода (3.31) в матричной форме:

$$X = A \cdot X', \quad (3.32)$$

$$\text{где } X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad X' = \begin{pmatrix} x_1' \\ x_2' \\ \dots \\ x_n' \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Поскольку $\det A \neq 0$, то существует обратная к A матрица A^{-1} , и

$$X' = A^{-1} \cdot X. \quad (3.33)$$

Следовательно, переход от одного базиса к другому осуществляется при помощи умножения вектора на соответствующую матрицу.

Пример. Пусть задан вектор $X = e_1 + 4e_2 - e_3$. При помощи матрицы перехода

$$A = \begin{bmatrix} 5 & -1 & -2 \\ 2 & 3 & 0 \\ -2 & 1 & 1 \end{bmatrix}$$

перешли к новому базису e_1', e_2', e_3' , т. е.:

$$e_1' = 5e_1 - e_2 - 2e_3,$$

$$\mathbf{e}_2' = 2\mathbf{e}_1 + 3\mathbf{e}_2,$$

$$\mathbf{e}_3' = -2\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3;$$

Найти вектор X в новом базисе, т. е. вектор X' .

Для решения этой задачи сначала найдем обратную к A матрицу A^{-1} .

$$\det A = \begin{vmatrix} 5 & -1 & -2 \\ 2 & 3 & 0 \\ -2 & 1 & 1 \end{vmatrix} = 15 - 4 + 0 - 12 + 0 + 2 = 1.$$

Так как определитель матрицы A равен единице, то присоединенная к A матрица A равна обратной A^{-1} и имеет вид:

$$\overline{\mathbf{A}} = \mathbf{A}^{-1} = \begin{vmatrix} 3 & -1 & 6 \\ -2 & 1 & -4 \\ 8 & -3 & -17 \end{vmatrix}$$

Теперь, в соответствии с (3.31), найдем вектор X в новом базисе:

$$X = \begin{bmatrix} 3 & -1 & 6 \\ -2 & 1 & -4 \\ 8 & -3 & -17 \end{bmatrix} \times \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} -7 \\ 6 \\ -21 \end{bmatrix},$$

или в виде разложения по новому базису:

$$\mathbf{X} = -7\mathbf{e}'_1 + 6\mathbf{e}'_2 - 21\mathbf{e}'_3.$$

3.4.5. Линейные преобразования n-мерного пространства

Будем говорить, что задано преобразование L , если каждому вектору $X \in R^n$ соответствует определенный вектор Y из того же пространства. Вектор Y называется **образом** вектора X и обозначается $Y = LX$. Преобразование L называется **линейным**, если выполняются следующие два условия:

$$\begin{aligned} L(X+Y) &= LX + LY && \text{для } \forall X, Y \in R^n; \\ L(\lambda X) &= \lambda LX && \text{для } \forall X \in R^n, \lambda - \text{постоянное число.} \end{aligned}$$

Примерами линейных преобразований являются: поворот пространства вокруг какой-либо точки или оси; равномерное растяжение пространства вдоль одной или нескольких осей и т. п.

Любое линейное преобразование можно записать с помощью матриц. Действительно, пусть L — линейное преобразование: $Y = LX$. Выберем базис в пространстве R^n : $[e_1, e_2, \dots, e_n]$. Преобразованием L базисные векторы $e_1, e_2, \dots, e_n$ переводятся

в векторы $Le_1, Le_2, \dots, Le_n$, где

$$Le_i = \sum_{j=1}^n a_{ji} e_j. \quad (3.34)$$

Матрица

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

где столбцы являются координатами векторов Le_i , называется **матрицей преобразования** для данного базиса.

Таким образом, каждому линейному преобразованию соответствует определенная матрица. И наоборот, каждая квадратная матрица размера $n \times n$ однозначно определяет линейное преобразование в векторном пространстве R^n .

Пусть в пространстве задано два базиса $\{e_k\}$ и $\{e'_k\}$. Координаты любого вектора X связаны формулами (3.32) и (3.33),

$$X = A \cdot X' \quad \text{и} \quad X' = A^{-1} X,$$

где $X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$ — координаты вектора X в базисе $\{e_k\}$;

$X' = \begin{pmatrix} x'_1 \\ x'_2 \\ \dots \\ x'_n \end{pmatrix}$ — координаты вектора X в базисе $\{e'_k\}$;

$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$ — матрица перехода от базиса $\{e_k\}$ к базису $\{e'_k\}$.

Рассмотрим линейное преобразование L . В базисе $\{e_k\}$ ему соответствует матрица M :

$$Y = M \cdot X, \quad (3.35)$$

а в базисе $\{e'_k\}$ — матрица M' :

$$Y' = M' \cdot X', \quad (3.36)$$

где Y, Y' — столбцы из координат вектора Y в базисе $\{e_k\}$ и $\{e'_k\}$ соответственно.

Определим связь между матрицами M и M' . Применяя формулы преобразования координат, получим:

$$Y' = A^{-1} \cdot Y$$

Подставим сюда формулу (3.32), а затем формулу (3.33), имеем:

$$Y' = A^{-1} \cdot M \cdot X = A^{-1} \cdot M \cdot A \cdot X'.$$

Сравнивая это равенство с (3.34), получим:

$$M' = A^{-1} \cdot M \cdot A \quad (3.37)$$

Определение. Матрица B называется **подобной** матрице M , если существует такая невырожденная матрица A , что

$$B = A^{-1} \cdot M \cdot A.$$

Из равенства (3.37) следует, что матрицы M и M' подобны. Следовательно, одному и тому же линейному преобразованию в различных базисах соответствуют подобные матрицы. Поскольку в данном пространстве можно построить неограниченное число различных базисов, то матрица может иметь бесконечное количество подобных матриц.

Свойства подобных матриц:

1. Определители подобных матриц равны; если $B = A^{-1} \cdot M \cdot A$, то $\det B = \det M$.
2. Если матрица A подобна матрице B , то и матрица B подобна матрице A .
3. Если A подобна B и C подобна B , то A и C — подобные матрицы.

3.4.6. Системы линейных алгебраических уравнений

Одной из центральных задач вычислительной математики является решение системы m линейных уравнений с n неизвестными.

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\
 &\dots\dots\dots \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m
 \end{aligned}
 \tag{3.38}$$

В более компактном виде систему можно записать так:

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = \sum_{j=1}^n a_{ij}x_j = b_i, \quad i = \overline{1, m},$$

или в матричной форме $A \cdot X = B$,

$$\text{где } A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Матрица A называется **основной (базовой) матрицей системы**, X — матрица-столбец неизвестных, B — матрица-столбец свободных членов.

Линейные системы часто возникают на практике, например при решении математических проблем в экономике и статистике. Даже при решении нелинейных задач, проводят процедуру линеаризации и сводят задачу к решению системы уравнений вида (3.38).

Система (3.38) имеет m уравнений, n неизвестных, коэффициенты при неизвестных a_{ij} (элементы матрицы A) и свободные члены b_i .

Если все свободные члены (не имеющие в своем составе неизвестных) равны нулю, то система называется **однородной**, в противном случае система **неоднородна**.

Среди уравнений системы могут быть линейно зависимые, т. е. такие, которые можно представить как результат сложения некоторых других уравнений системы, предварительно умноженные на какие-либо числа (линейной комбинацией других уравнений системы). Это означает, что зависимые уравнения не содержат никакой дополнительной информации, влияющей на значения искомых неизвестных. Примером могут служить одни и те же финансовые расчёты, записанные для различных денежных единиц, или геометрические расчёты, записанные в различных системах измерения.

Линейные системы, полученные одна из другой путем элементарных преобразований (перестановкой двух уравнений, умножением одного из них на число не равное нулю, почленным сложением двух уравнений) называются **эквивалентными** (или **равносильными**). Исключение зависимых уравнений из системы также приводит к эквивалентной системе уравнений. Все эквивалентные системы имеют одинаковые решения. Число линейно независимых уравнений в системе (3.38) называется **рангом** этой системы.

Система (3.38) называется **совместной**, если она имеет хотя бы одно решение, и **несовместной**, если у нее не существует ни одного решения.

Если к основной матрице системы приписать столбец свободных членов, то получится **расширенная матрица** системы уравнений:

$$(A|B) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right).$$

Теорема. Для того, чтобы линейная система (3.38) являлась совместной, необходимо и достаточно, чтобы ранг расширенной матрицы этой системы был равен рангу ее основной матрицы, т. е. $\text{rg} A = \text{rg}(A|B)$.

Пример. Определить совместность системы:

$$x_1 + 2x_2 - x_3 + x_4 = 2$$

$$x_1 + 3x_2 + 2x_3 = 1$$

$$-x_1 - x_2 + 4x_3 - 2x_4 = 2.$$

Составим расширенную матрицу системы и проведем с ней ряд элементарных преобразований, не меняющих ранг матрицы:

$$(A|B) = \left(\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 2 \\ 1 & 3 & 2 & 0 & 1 \\ -1 & -1 & 4 & -2 & 2 \end{array} \right)$$

Первую строку оставим без изменения, а во второй и третьей строках с помощью элементарных преобразований (от второй строки отнимем первую, а к третьей прибавим первую строку) в первом столбце получим нули, т. е.:

$$(A|B) = \left\| \begin{array}{cccc|c} 1 & 2 & -1 & 1 & 2 \\ 0 & 1 & 3 & -1 & -1 \\ 0 & 1 & 3 & -1 & 4 \end{array} \right\|.$$

Вычитая из третьей строки вторую, получим:

$$(A|B) = \left\| \begin{array}{cccc|c} 1 & 2 & -1 & 1 & 2 \\ 0 & 1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 5 \end{array} \right\|.$$

Теперь очевидно, что $\text{rg} A = 2$ и $\text{rg}(A|B) = 3$. Следовательно, система несовместна.

Исследуем вопрос о числе решений такой системы. Совместная система называется **определенной**, если она имеет единственное решение, и **неопределенной**, если у нее существует по крайней мере два различных решения.

Пусть задана совместная система линейных уравнений (3.38), т. е. система, у которой $\text{rg} A = \text{rg}(A|B) = r$. Возможны следующие случаи.

1. Если $r < n$, то исходная система заведомо имеет $(n - r)$ линейно зависимых уравнений и их можно исключить из системы. Те уравнения, коэффициенты которых образуют минор порядка r не равный нулю, являются линейно независимыми и называются **базисными**. После исключения лишних уравнений систему исследуют снова (см. пункт 2 и 3).

Пример. Исследовать систему уравнений:

$$\begin{aligned} x_1 + 2x_2 &= 4 \\ -x_1 + 5x_2 &= 3 \\ x_1 + 9x_2 &= 11 \end{aligned}$$

Здесь $m = 3$, $n = 2$. Составим расширенную матрицу и упростим ее путем проведения элементарных преобразований (добавим ко второй строчке первую и вычтем из третьей первую и т. д.):

$$(A|B) = \left\| \begin{array}{cc|c} 1 & 2 & 4 \\ -1 & 5 & 3 \\ 1 & 9 & 11 \end{array} \right\| + \left\| \begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 7 & 7 \\ 0 & 7 & 7 \end{array} \right\| + \left\| \begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array} \right\|.$$

Видим, что $\text{rg} A = \text{rg}(A|B) = 2$. Система совместна, базисные уравнения — первое и второе. Третье уравнение является их линейной комбинацией и может быть отброшено. Поэтому эквивалентная система имеет вид:

$$\begin{aligned} x_1 + 2x_2 &= 4 \\ -x_1 + 5x_2 &= 3. \end{aligned}$$

Решение этой системы очевидно: $x_1 = 2$, $x_2 = 1$. Сделав подстановку найденных решений в исходную систему, убеждаемся, что оно удовлетворяет всем трем исходным уравнениям.

2. Если $r = n = \text{rg} A = \text{rg}(A|B)$, то система имеет единственное решение. Эквивалентная система из предыдущего примера иллюстрирует именно эту ситуацию.

3. Если $r < n$, то система имеет бесчисленное множество решений.

Пример. Исследовать систему

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 3 \\ 3x_1 + 2x_2 - x_3 &= 5. \end{aligned}$$

Построим расширенную матрицу:

$$\left\| \begin{array}{ccc|c} 1 & -2 & 1 & 3 \\ 3 & 2 & -1 & 5 \end{array} \right\|.$$

В этом случае $\text{rg} A = \text{rg}(A|B) = 2$. Система совместна, но т. к. $r < n$, то она имеет бесконечное число решений. Действительно, переписав исходную систему в виде

$$\begin{aligned} x_1 - 2x_2 &= 3 - x_3 \\ 3x_1 + 2x_2 &= 5 + x_3 \end{aligned}$$

и положив $x_3 = k$, получим решение системы:

$$x_1 = 2, x_2 = 0,5 \cdot (k - 1), x_3 = k,$$

где k — произвольное число.

Выбрав, например, $k = 1$, получим такое решение:

$$X_1 = \{x_1 = 2, x_2 = 0, x_3 = 1\},$$

если $k = 5$, то

$$X_2 = \{x_1 = 2, x_2 = 2, x_3 = 5\}$$

и т. д.

Для практики особо важен тот случай, когда число уравнений равно числу неизвестных. Пусть дана алгебраическая система n уравнений с n неизвестными (их иногда называют квадратными системами):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned} \tag{3.39}$$

или, в матричном виде,

$$A \cdot X = B. \tag{3.40}$$

Если матрица A невырожденная ($\det A \neq 0$), то тогда существует обратная по отношению к ней матрица A^{-1} . Умножим равенство (3.40) на A^{-1} слева и выполним операции с матрицами. Получим:

$$A^{-1}(A \cdot X) = A^{-1}B \rightarrow (A^{-1} \cdot A) \cdot X = A^{-1}B \rightarrow I \cdot X = A^{-1}B \rightarrow X = A^{-1} \cdot B.$$

Таким образом, решение квадратной системы алгебраических уравнений в матричной форме сводится к построению матрицы обратной по отношению к матрице системы и последующему умножению ее справа на матрицу свободных членов:

$$X = A^{-1} \cdot B. \quad (3.41)$$

Пример: Решить систему алгебраических уравнений

$$\begin{aligned} x_1 - x_2 - x_3 &= 1 \\ -x_1 + 2x_2 + x_3 &= -2 \\ 2x_1 + 3x_2 &= 0,5. \end{aligned}$$

Решение. Вычислим определитель матрицы системы:

$$\det A = \begin{vmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ 2 & 3 & 0 \end{vmatrix} = 2.$$

Найдем алгебраические дополнения элементов матрицы A :

$$\begin{aligned} A_{11} &= -3, A_{12} = 2, A_{13} = -7, A_{21} = -3, A_{22} = 2, \\ A_{23} &= -5, A_{31} = 1, A_{32} = 0, A_{33} = 1. \end{aligned}$$

Присоединенная матрица и обратная матрица, соответственно, равны:

$$\bar{A} = \begin{vmatrix} -3 & -3 & 1 \\ 2 & 2 & 0 \\ -7 & -5 & 1 \end{vmatrix}, \quad A^{-1} = \begin{vmatrix} -3/2 & -3/2 & 1/2 \\ 1 & 1 & 0 \\ -7/2 & -5/2 & 1/2 \end{vmatrix}.$$

По формуле (3.41) получим решение системы:

$$X = \begin{vmatrix} -3/2 & -3/2 & 1/2 \\ 1 & 1 & 0 \\ -7/2 & -5/2 & 1/2 \end{vmatrix} \times \begin{vmatrix} 1 \\ -2 \\ 0,5 \end{vmatrix} = \begin{vmatrix} 1,75 \\ 1,00 \\ 1,75 \end{vmatrix},$$

т. е. $x_1 = 1,75$, $x_2 = -1$, $x_3 = 1,75$.

Система называется **однородной**, если свободные члены равны нулю:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= 0 \\ \dots &\dots \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= 0. \end{aligned} \quad (3.42)$$

Всякая однородная система совместна, так как *всегда имеет хотя бы нулевое решение*: $X_0 = \{x_1=0, x_2=0, x_3=0, \dots, x_n=0\}$, в чем легко убедиться, подставив в (3.42) вместо неизвестных нули. Такое решение называется тривиальным.

Выясним, при каких условиях однородная система имеет ненулевое (нетривиальное) решение.

Теорема. Для того, чтобы однородная система имела ненулевые решения, необходимо и достаточно, чтобы ранг этой системы был меньше числа неизвестных ($r < n$).

Доказательство необходимости. Пусть система (3.40) имеет ненулевые решения. Предположим противное, что $r = n$. Тогда система имеет единственное решение и других решений, кроме нулевого, не имела бы. Следовательно, $r < n$.

Доказательство достаточности. Пусть ранг системы меньше числа неизвестных ($r < n$). В этом случае система имеет бесконечно много различных решений и, следовательно, имеет ненулевые решения.

Следствие. Любая однородная система, у которой число уравнений меньше числа неизвестных, имеет нетривиальное решение. Действительно, если $m < n$, то ранг меньше числа неизвестных ($r < n$).

Теорема. Для того, чтобы квадратная однородная система имела ненулевое решение, необходимо и достаточно, чтобы ее определитель был равен нулю.

Действительно, условие предыдущей теоремы $r < n$ эквивалентно условию равенства нулю определителя системы.

Пример. Исследовать и найти решение системы

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &= 0 \\ 2x_1 - 2x_2 + x_3 &= 0. \end{aligned}$$

Решение. В данном примере $m = 2$, $n = 3$, $m < n$. Рассмотрим минор

$$\begin{vmatrix} 3 & 2 \\ 2 & -2 \end{vmatrix} = -10,$$

приходим к заключению, что $\text{rg} A = 2$. Одна переменная — «лишняя», и ее можно выбирать произвольно. Пусть $x_3 = k$,

тогда $x_1 = 0$ и $x_2 = 0.5 \times k$. Заметим, что за базисные переменные можно выбрать x_1 и x_3 , положив $x_2 = k$, но нельзя выбрать x_2 и x_3 , т. к. их нельзя выразить через x_1 .

3.4.7. Метод жордановых исключений

При построении компьютерных алгоритмов для решения задач линейной алгебры и в целом ряде других случаев широкое применение находит так называемый метод жордановых исключений.

Рассмотрим систему m линейных функций y , зависящих от $n + m$ переменных:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1s}x_s + \dots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2s}x_s + \dots + a_{2n}x_n \\ &\dots \\ y_r &= a_{r1}x_1 + a_{r2}x_2 + \dots + a_{rs}x_s + \dots + a_{rn}x_n \\ &\dots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{ms}x_s + \dots + a_{mn}x_n \end{aligned} \quad (3.43)$$

Переменные $x_1, x_2, \dots, x_n$ (их называют свободными или независимыми) можно выбирать произвольным образом, при этом зависимые переменные $y_1, y_2, \dots, y_m$ (их называют базисными) в соответствии с (3.43) примут вполне определенные значения.

Если, например, $a_{rs} \neq 0$, то свободную переменную x_s можно поменять местами с базисной переменной y_r . Это реализуется следующим образом: r -ое уравнение решается относительно x_s :

$$x_s = \frac{1}{a_{rs}} (y_r - a_{r1}x_1 - a_{r2}x_2 - \dots - a_{rn}x_n),$$

затем найденное выражение для x_s подставляется в (3.43). Получается система уравнений вида:

$$\begin{aligned} y_1 &= \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1s}y_r + \dots + \alpha_{1n}x_n \\ y_2 &= \alpha_{21}x_1 + \alpha_{22}x_2 + \dots + \alpha_{2s}y_r + \dots + \alpha_{2n}x_n \\ &\dots \\ x_s &= \alpha_{s1}x_1 + \alpha_{s2}x_2 + \dots + \alpha_{ss}y_r + \dots + \alpha_{sn}x_n \\ &\dots \\ y_m &= \alpha_{m1}x_1 + \alpha_{m2}x_2 + \dots + \alpha_{ms}y_r + \dots + \alpha_{mn}x_n \end{aligned} \quad (3.44)$$

Шаг жордановых исключений — формализованная схема пересчета коэффициентов a_{ij} в коэффициенты α_{ij} . В результате реализации этой схемы одна свободная переменная меняется местами с одной базисной переменной, т. е. осуществляется переход от системы (3.43) к системе (3.44).

Запишем исходную систему (3.43) в виде специальной таблицы.

	x_1	x_2	...	x_s	...	x_n
$y_1 =$	a_{11}	a_{12}	...	a_{1s}	...	a_{1n}
$y_2 =$	a_{21}	a_{22}	...	a_{2s}	...	a_{2n}
...	...	...	...	...	...	...
$y_r =$	a_{r1}	a_{r2}	...	a_{rs}	...	a_{rn}
...	...	...	...	...	...	...
$y_m =$	a_{m1}	a_{m2}	...	a_{ms}	...	a_{mn}

При замене свободной переменной x_s на базисную переменную y_r все коэффициенты системы и, следовательно, таблицы изменятся.

Элемент a_{rs} , стоящий на пересечении строки r и столбца s , называется разрешающим. Строка и столбец, на пересечении которых стоит разрешающий элемент, называются, соответственно, разрешающей строкой и разрешающим столбцом.

Алгоритм одного шага жордановых исключений следующий:

- 1) разрешающий элемент заменяется единицей;
- 2) остальные элементы разрешающей строки меняют знак на противоположный;
- 3) остальные элементы разрешающего столбца остаются без изменений;
- 4) прочие (обыкновенные) элементы вычисляются по формуле $b_{ij} = a_{ij} \cdot a_{rs} - a_{is} \cdot a_{rj}$;
- 5) все (!) получившиеся элементы делятся на разрешающий элемент.

Проделав все эти операции, получим новую жорданову таблицу эквивалентную системе (3.44).

	x_1	x_2	...	y_r	...	x_n
$y_1 =$	b_{11}	b_{12}	...	a_{1r}	...	b_{1n}
$y_2 =$	b_{21}	b_{22}	...	a_{2r}	...	b_{2n}
...	...	...	...	...	...	...
$x_s =$	$-a_{r1}$	$-a_{r2}$	...	1	...	$-a_{rn}$
...	...	...	...	...	...	...
$y_m =$	b_{m1}	b_{m2}	...	a_{mr}	...	b_{mn}

 $\cdot \frac{1}{a_{rs}}$

Проиллюстрируем применение метода жордановых исключений к решению некоторых задач.

Задача 1. Задана квадратная матрица

$$A = \begin{pmatrix} 1 & -2 & 3 \\ -1 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix}$$

Построить обратную к ней матрицу A^{-1} .

Решение. Запишем равенство

$$Y = A \cdot X, \quad (3.45)$$

где X и Y — матрицы-столбцы соответствующего размера. Умножая слева это равенство на A , получим

$$X = A^{-1} \cdot Y. \quad (3.46)$$

Сравнивая (3.45) и (3.46), заметим, что обратная матрица получится, если все свободные переменные, входящие в X поменять местами с базисными переменными, входящими в Y . Проведем три шага жордановых исключений, помечая разрешающие элементы знаком (*).

$$\begin{array}{c}
 \begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline y_1 & 1* & -2 & 3 \\ y_2 & -1 & 1 & 2 \\ y_3 & 2 & -1 & -1 \end{array} \rightarrow \begin{array}{c|ccc} & y_1 & x_2 & x_3 \\ \hline x_1 & 1 & 2 & -3 \\ y_2 & -1 & 1* & 5 \\ y_3 & 2 & 3 & -7 \end{array} \rightarrow \\
 \rightarrow \begin{array}{c|ccc} & y_1 & y_2 & x_3 \\ \hline x_1 & -1 & -2 & 7 \\ x_2 & -1 & -1 & 5 \\ y_3 & -1 & -3 & 8* \end{array} \rightarrow \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & -1 & 5 & 7 \\ x_2 & -3 & 7 & 5 \\ x_3 & 1 & 3 & 1 \end{array} \cdot \frac{1}{8}
 \end{array}$$

Искомая обратная матрица —

$$A^{-1} = \begin{pmatrix} -1/8 & 5/8 & 7/8 \\ -3/8 & 7/8 & 5/8 \\ 1/8 & 3/8 & 1/8 \end{pmatrix}$$

Проверить правильность полученного результата можно, вычислив произведение прямой и обратной матрицы, которое должно быть равно единичной матрице, т. е.

$$A \cdot A^{-1} = \begin{pmatrix} 1 & -2 & 3 \\ -1 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} -1/8 & 5/8 & 7/8 \\ -3/8 & 7/8 & 5/8 \\ 1/8 & 3/8 & 1/8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Задача 2. Решить при помощи жордановых исключений систему уравнений

$$\begin{aligned}
 2x_1 + 2x_2 - x_3 + x_4 &= 4 \\
 4x_1 + 3x_2 - x_3 + 2x_4 &= 6 \\
 8x_1 + 5x_2 - 3x_3 + 4x_4 &= 12 \\
 3x_1 + 3x_2 - 2x_3 + 2x_4 &= 6.
 \end{aligned}$$

Решение. В матричной форме система имеет вид:

$$A \cdot X = B \quad \text{или} \quad 0 = A \cdot X - B.$$

Запишем последнее равенство в форме жордановой таблицы и проведем замену x_4 на 0 из первой строки.

$$\begin{array}{c|cccc|c} & x_1 & x_2 & x_3 & x_4 & -1 \\ \hline 0 = & 2 & 2 & -1 & 1* & -4 \\ 0 = & 4 & 3 & -1 & 2 & -6 \\ 0 = & 8 & 5 & -3 & 4 & -12 \\ 0 = & 3 & 3 & -2 & 2 & -6 \end{array} \rightarrow \begin{array}{c|cccc|c} & x_1 & x_2 & x_3 & 0 & -1 \\ \hline x_4 & -2 & -2 & 1 & 1 & 4 \\ 0 = & 0 & -1 & 1 & 2 & 2 \\ 0 = & 0 & -3 & 1 & 4 & 4 \\ 0 = & -1 & -1 & 0 & 2 & 2 \end{array}$$

Величины, стоящие в верхней строчке таблицы, при переходе к традиционной записи системы уравнений умножаются на все элементы соответствующего столбца. Поэтому столбцы со стоящими сверху нулями из таблицы удаляются. Проведем дальнейшие шаги жордановых исключений.

	x_1	x_2	x_3	-1
$x_4 =$	-2	-2	1	4
$0 =$	0	-1	1*	2
$0 =$	0	-3	1	4
$0 =$	-1	-1	0	2

 $\rightarrow$

	x_1	x_2	-1
$x_4 =$	-2	-1	2
$x_3 =$	0	1	-2
$0 =$	0	-2	2
$0 =$	-1	-1*	2

 $\rightarrow$

	x_1	-1
$x_4 =$	-1	0
$x_3 =$	-1	0
$0 =$	2*	-2
$x_2 =$	-1	2

 $\rightarrow$

	-1
$x_4 =$	-1
$x_3 =$	-1
$x_1 =$	1
$x_2 =$	1

В последней таблице читаем ответ: $x_1 = 1, x_2 = 1, x_3 = -1, x_4 = -1$, в правильности которого можно убедиться подстановкой найденных значений неизвестных в исходную систему уравнений.

Для более глубокого изучения материала по темам этого раздела можно воспользоваться литературой [11,12].

3.5. Задачи и упражнения к разделу 3.

1. Какая линия определяется параметрическими уравнениями?

а)
$$\begin{cases} x = \cos t \\ y = \cos^2 t \end{cases}$$

Ответ: парабола $y = x^2$.

б)
$$\begin{cases} x = \sin t \\ y = \operatorname{cosect} \end{cases}$$

Ответ: $y = 1/x$.

2. Составить уравнение прямой, отсекающей на оси ординат отрезок $b = -3$, образующий с положительным направлением оси OX угол $\pi/6$.

Ответ: $x - \sqrt{3}y - 3\sqrt{3} = 0$.

3. Составить уравнение окружности радиусом 0.5 с центром в точке с координатами (2, -1).

Ответ: $(x - 2)^2 + (y + 1)^2 = 0.25$.

4. Составить уравнение эллипса, проходящего через точки

$M(\frac{5}{2}, \frac{\sqrt{6}}{4})$ и $N(-2, \frac{\sqrt{15}}{5})$.

Ответ: $\frac{x^2}{10} + \frac{y^2}{1} = 1$

5. Какую линию определяет уравнение $x^2 - 9y^2 + 2x + 36y - 44 = 0$?

Ответ: гиперболу, т. к. уравнение приводится к виду $(x + 1)^2 - 9(y - 2)^2 = 9$.

6. Найти координаты центра и радиус сферы, заданной уравнением

$x^2 + y^2 + z^2 - x + 2y + 1 = 0$.

Ответ: поскольку уравнение приводится к виду

$(x - 0.5)^2 + (y + 1)^2 + z^2 = 0.25$,

то центр сферы лежит в точке (0.5, -1, 0) и радиус сферы равен 0.5.

7. Найти определитель матрицы

$$A = \begin{vmatrix} 1 & -2 & 2 & 1 \\ 3 & 2 & 1 & -2 \\ 0 & -2 & 0 & 1 \\ 0 & 4 & 3 & 2 \end{vmatrix}$$

Ответ: $\det A = -34$.

8. Найти матрицу обратную к данной:

а)
$$A = \begin{vmatrix} 0 & -1 & 2 \\ 2 & 2 & 1 \\ 1 & 2 & 1 \end{vmatrix}$$

Ответ:

$$A^{-1} = \begin{vmatrix} 0 & 1 & -1 \\ -0.2 & -0.4 & 0.8 \\ 0.4 & -0.2 & 0.4 \end{vmatrix}$$

6)

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ -1 & 3 & -4 \end{pmatrix}.$$

Ответ: обратная матрица не существует, т. к. $\det A = 0$.

9. Найти ранг матрицы:

а)

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ -1 & 1 & -3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Ответ: $\text{rg} A = 2$.

б)

$$A = \begin{pmatrix} 1 & 0 & 2 & 0 & -3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 4 & 0 & -6 & 0 \end{pmatrix}.$$

Ответ: $\text{rg} A = 2$.

10. Найти произведение матриц

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 0 & 4 \\ 1 & 2 & 3 \end{pmatrix} \text{ и } B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -1 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Ответ:

$$A \cdot B = \begin{pmatrix} 8 & 0 & 7 \\ 14 & -6 & 4 \\ 13 & 5 & 7 \end{pmatrix}.$$

11. Может ли система линейных уравнений иметь ровно два различных решения? Ровно три решения?

12. Доказать, что если сумма двух решений линейной системы является решением этой же системы, то она однородна.

13. Решить системы уравнений:

а)

$$\begin{aligned} x_1 - x_2 &= 1 \\ x_1 + 2x_2 &= 4 \\ -2x_1 + x_2 &= 0 \end{aligned}$$

Ответ: система несовместна.

б)

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 2 \\ x_1 - x_2 + x_3 - x_4 &= 0 \end{aligned}$$

Ответ: $x_1 = 1 - x_3$, $x_2 = 1 - x_4$.

в)

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_2 + 3x_3 &= 0 \end{aligned}$$

Ответ: $x_1 = -0.5x_2$, $x_3 = -0.5x_2$.

14. Могут ли быть линейно независимыми два вектора на плоскости? Три вектора на плоскости? Пять векторов в 4-мерном пространстве?

Ответ: да; нет; нет.

15. Дано линейное преобразование

$$x' = x + y + z$$

$$y' = x + y$$

$$z' = x$$

и даны три точки в системе координат (x, y, z) : $M = (3, -2, -1)$, $N = (1, -1, 1)$, $K = (-1, -2, -3)$. Определить координаты этих же точек в системе (x', y', z') .

Ответ: $M' = (0, 1, 3)$, $N' = (1, 0, 1)$, $K' = (-6, -3, -1)$.

Раздел 4. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Глава 4.1. Пределы

4.1.1. Предел числовой последовательности

Числовую функцию $a_n = f(n)$, заданную на множестве натуральных чисел, называют **числовой последовательностью**. Говорят, что последовательность задана, если каждому натуральному числу n по некоторому закону f поставлено в соответствие число $f(n)$. В том случае, когда закон f задан формулой, говорят об **аналитическом задании** последовательности, при этом, как правило, записывают аналитическое выражение для члена a_n , называемого общим членом последовательности. Например, последовательность квадратов целых чисел:

$$1, 4, 9, 16, 25, \dots, n^2, \dots \equiv \{n^2\}.$$

Часто закон, по которому задается последовательность, позволяет построить очередной член последовательности по известным предыдущим. Такой способ задания называется **рекуррентным**. Хорошо известная арифметическая последовательность (арифметическая прогрессия)

$$a, a + d, a + 2d, a + 3d, \dots, a + nd, \dots$$

может быть задана при помощи рекуррентного соотношения $a_n = a_{n-1} + d$.

Геометрическая последовательность (геометрическая прогрессия)

$$c, cq, cq^2, cq^3, \dots, cq^n, \dots$$

может быть задана при помощи следующего рекуррентного соотношения:

$$b_n = q \cdot b_{n-1}.$$

Для полного задания последовательности, кроме рекуррентного соотношения, необходимо задать некоторое число первых членов последовательности. Так для арифметической и геометрической последовательностей, из бесконечного числа членов достаточно задать лишь их первые члены a и c , а также и параметры прогрессий d и q .

Рассмотрим далее три последовательности:

$$1, 0, -1, -2, -3, \dots, -n, \dots \quad (4.1)$$

$$2, 4, 6, \dots, 2 \cdot n, \dots \quad (4.2)$$

$$1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots \quad (4.3)$$

При увеличении числа n члены первой последовательности неограниченно убывают, члены второй — неограниченно возрастают, а члены третьей последовательности убывают начиная с числа 1 до числа 0. Следовательно, числовые последовательности могут быть ограничены либо сверху, либо снизу, либо и сверху и снизу.

Формально последовательность называется **ограниченной сверху**, если существует число M (верхняя граница), такое, что $a_n \leq M$ для всех n . Последовательность (4.1) ограничена сверху, например, числом 2. Поскольку для всех n кроме того выполняется условие $a_n \leq a_{n-1}$, то это **убывающая последовательность**.

Последовательность (a_n) называется **ограниченной снизу**, если существует такое число m (нижняя граница), что $a_n \geq m$ для всех n . Если для всех членов последовательности выполняется условие $a_n \geq a_{n-1}$, то она называется **возрастающей**. Последовательность (4.2) является ограниченной снизу возрастающей последовательностью.

Последовательность (a_n) называется **ограниченной**, если она ограничена и сверху и снизу. Очевидно, что (4.3) — ограниченная убывающая последовательность.

Рассмотрим последовательность

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots, \frac{n}{n+1}, \dots = \{a_n\}.$$

Члены этой последовательности по мере возрастания номера члена приближаются к числу 1. Говорят, что последовательность (4.4) **сходится** к числу 1.

Формально **сходимость последовательности** определяется следующим образом. Последовательность $\{a_n\}$ сходится к числу A , если для любого сколь угодно малого числа $\varepsilon > 0$ можно указать такое натуральное число $n_0(\varepsilon)$, что для всех n больше n_0 , то есть для $n > n_0$, выполняется неравенство $|a_n - A| < \varepsilon$.

Проиллюстрируем это определение на примере последовательности (4.4). Построим отклонение общего члена последовательности от 1.

$$|a_n - 1| = \left| \frac{n}{n+1} - 1 \right| = \left| \frac{-1}{n+1} \right| = \frac{1}{n+1}$$

С возрастанием n это отклонение, уменьшаясь, стремится к нулю. Пусть $\varepsilon = 3/500$. Вычислим отклонение для a_{166} :

$$|a_{166} - 1| = \left| \frac{166}{167} - 1 \right| = \frac{1}{167} = \frac{3}{501} < \frac{3}{500}$$

Тем самым установлено, что начиная с $n = 166$ ($n_0 = 165$) выполняется неравенство $|a_n - 1| < 3/500$.



4.1. Геометрический смысл сходимости последовательности

Геометрический смысл сходимости последовательности проиллюстрирован на рис. 4.1. Каждому члену числовой последовательности (поскольку это число) соответствует точка на числовой оси. Если последовательность имеет пределом

точку A , то для всех номеров последовательности, начиная с некоторого n_0 , члены последовательности находятся внутри отрезка $(A - \varepsilon, A + \varepsilon)$, называемого ε -окрестностью числа A .

Если ε очень мало, то число n_0 может быть весьма большим. Следовательно, много членов последовательности окажутся вне ε -окрестности, однако их всегда будет лишь конечное число. Все остальные члены последовательности, начиная с номера n_0 и более, попадают в ε -окрестность. Таким образом, если последовательность сходится к A , то какую бы окрестность точки A ни взять, почти все числа a_n попадают в выбранную окрестность.

Отсюда следует, что добавление или исключение конечного числа членов такой последовательности не влияет на ее сходимость. Если последовательность $\{a_n\}$ сходится к A , то пишут:

$$\lim_{n \rightarrow \infty} a_n = A$$

(читается: «предел a_n при n стремящемся к бесконечности равен A »). В этом случае говорят, что число A есть **предел** последовательности $\{a_n\}$ или иначе, при неограниченном увеличении номера общий член последовательности стремится к величине A .

Если последовательность не имеет конечного предела или не имеет предела вообще, то ее называют **расходящейся**. Так, последовательность

$$-1, 1, -1, 1, \dots, (-1)^n, \dots$$

расходится, так как в этом случае $\lim_{n \rightarrow \infty} a_n$ не существует.

Последовательность (4.2) тоже является расходящейся, так как по мере возрастания n члены последовательности становятся больше любого наперед заданного числа, то есть a_n стремится к бесконечности. В этом случае пишут: $\lim_{n \rightarrow \infty} a_n = \infty$. Заметим, что, хотя здесь предел формально и существует, он не является конечным числом.

4.1.2. Предел функции и его свойства

Пусть $f(x)$ — функция непрерывного аргумента. Число A называется **пределом функции** $y = f(x)$ при $x \rightarrow x_0$, если для каждого сколь угодно малого числа $\varepsilon > 0$ можно указать зависящее от ε число $\delta(\varepsilon) > 0$ такое, что для всех x , удовлетворяющих неравенству $|x_0 - x| < \delta$, имеет место неравенство $|A - f(x)| < \varepsilon$. В формализованной форме это записывается так:

$$\lim_{x \rightarrow x_0} f(x) = A$$

(читается: $f(x)$ стремится к A при x стремящемся к x_0). Наличие у функции $f(x)$ предела A в точке x_0 содержательно означает, что, как только независимая переменная x достаточно близко приблизится к значению x_0 , функция $f(x)$ будет сколь угодно близка к A .

Справедливы следующие свойства пределов.

1. Если предел функции существует, то он единственен.
2. Предел постоянной величины равен самой постоянной.
3. Если при $x \rightarrow x_0$ существуют конечные пределы функций $f(x)$ и $g(x)$, то:

$$\lim_{x \rightarrow x_0} [af(x) + bg(x)] = a \lim_{x \rightarrow x_0} f(x) + b \lim_{x \rightarrow x_0} g(x),$$

где a и b — числа.

$$4. \lim_{x \rightarrow x_0} [f(x) \cdot g(x)] = \lim_{x \rightarrow x_0} f(x) \cdot \lim_{x \rightarrow x_0} g(x).$$

$$5. \lim_{x \rightarrow x_0} f(x)/g(x) = \lim_{x \rightarrow x_0} f(x) / \lim_{x \rightarrow x_0} g(x), \text{ если } \lim_{x \rightarrow x_0} g(x) \neq 0.$$

В качестве примера вычислим два предела.

$$\lim_{x \rightarrow (-1)} \frac{2x+1}{3+x} = \frac{2 \lim_{x \rightarrow -1} x + 1}{3 + \lim_{x \rightarrow -1} x} = \frac{2(-1)+1}{3+(-1)} = -\frac{1}{2}$$

$$\lim_{x \rightarrow 2} \frac{x-2}{x^2-4} = \lim_{x \rightarrow 2} \frac{x-2}{(x+2)(x-2)} = \lim_{x \rightarrow 2} \frac{1}{x+2} = \frac{1}{4}$$

4.1.3 Бесконечно малые и бесконечно большие величины

Если при $x \rightarrow x_0$ функция $f(x)$ стремится к 0, то ее называют **бесконечно малой величиной** (или просто бесконечно малой) в окрестности точки x_0 . Бесконечно малые на практике часто обозначают греческими буквами α , β , γ . Примеры бесконечно малых величин: $\alpha = 2x - 6$ при $x \rightarrow 3$, $\beta = e^x$ при $x \rightarrow \infty$, $\gamma = \sin x$ при $x \rightarrow 0$ и т. п.

Замечание. Следует отличать очень малую величину от бесконечно малой. Очень малая — это некоторая фиксированная величина (константа), которой в рассматриваемой ситуации, как правило, пренебрегают. Бесконечно малая (переменная величина!) всегда меньше любой наперед заданной очень малой величины.

Произведение бесконечно малых есть величина бесконечно малая более высокого порядка малости. На практике бесконечно малыми высшего порядка по сравнению с бесконечно малыми первого обычно пренебрегают, т. к. $\alpha \gg \alpha^2$.

Если $f(x)$ при $x \rightarrow x_0$ в пределе стремится к ∞ , то она называется **бесконечно большой величиной** (в окрестности точки x_0). К таким относятся, например, $y = 1/x$ при $x \rightarrow 0$, $y = \ln(x)$ при $x \rightarrow \infty$ и при $x \rightarrow 0$, $y = \operatorname{tg}(x)$ при x стремящемся к $\pi/2$ слева и др.

Между бесконечно малой и бесконечно большой величинами существует простое соотношение: если $f(x)$ — бесконечно малая величина не равная тождественно нулю, то $1/f(x)$ — бесконечно большая величина, и наоборот. Поэтому, если $2x-6$ является бесконечно малой величиной в окрестности точки $x = 3$, то $1/(2x-6)$ — бесконечно большая в окрестности той же точки.

Замечание. Бесконечность (обозначаемая знаком ∞) не является числом. Это означает, что величина $f(x)$ может стать больше (по модулю) любого наперед заданного фиксированного числа.

Функция $f(x)$ называется **непрерывной** в точке x_0 , если она определена в этой точке и некоторой ее окрестности, а также:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0). \quad (4.5)$$

Точки, в которых равенство (4.5) не выполняется, называются **точками разрыва функции**.

Все элементарные функции, а также любая их суперпозиция непрерывны в своей области определения. На практике непрерывные функции часто называют «хорошими функциями».

Глава 4.2. Производная функции одной переменной

4.2.1. Определение производной

Значение функции в каждой допустимой фиксированной точке есть число. Изменяя значения аргумента, получим в общем случае различные значения функции. Как сильно изменится значение функции при данном изменении аргумента? Поиск ответа на этот вопрос приводит к понятию производной.

Пусть задана некоторая функция $y = f(x)$. Выберем произвольное допустимое значение аргумента x и вычислим $f(x)$. Затем, не выходя из области определения, изменим x на малую величину Δx . Вычислим $f(x + \Delta x)$ и образуем отношение

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \equiv \frac{\Delta f(x)}{\Delta x}. \quad (4.6)$$

Если существует конечный предел отношения

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

то функция $f(x)$ называется дифференцируемой в точке x , а значение предела называется производной от функции $f(x)$ в точке x и обозначается

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \equiv f'(x) \equiv \frac{df(x)}{dx} \equiv f'_x. \quad (4.7)$$

Значение производной зависит от выбранного значения точки x . Следовательно, производная — это функция от того же аргумента, что и $f(x)$. Операция вычисления производной называется дифференцированием функции.

Рассмотрим геометрический смысл производной. Если изобразить на рисунке график функции $f(x)$, то величина отношения (4.6) равна тангенсу угла наклона секущей графика к его оси абсцисс (рис. 4.2).

Если $\Delta x \rightarrow 0$, то точка N стремится к точке M и секущая MN стремится занять положение касательной к $f(x)$ в точке M . Следовательно, значение производной $f'(x)$ в любой точке x области определения функции равно тангенсу наклона касательной к графику $y = f(x)$ к оси абсцисс в точке с координатами x и $f(x)$.

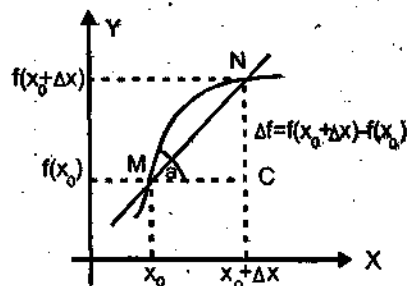


Рис. 4.2. Геометрический смысл производной

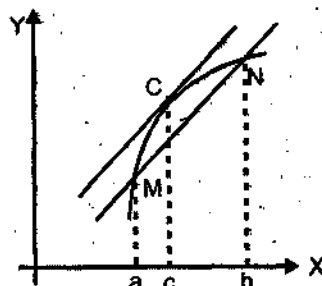


Рис. 4.3. Геометрический смысл и формулы конечных приращений

Геометрический смысл производной позволяет пояснить важную формулу конечных приращений:

$$f(b) - f(a) = f'(c)(b - a). \quad (4.8)$$

Действительно (рис. 4.3), на отрезке $[a, b]$ всегда найдется точка c , в которой касательная параллельна секущей MN . В этом случае из построения следует формула (4.8), позволяющая выразить приращение функции через приращение аргументов и производную.

Для практического применения математического анализа важен еще и механический смысл производной. Если пройденный путь есть известная функция времени $s = f(t)$, то ее производная $f'(t)$ равна скорости движения в каждый момент времени t . В общем случае производная описывает скорость изменения функции при изменении аргумента независимо от физического смысла величины, описываемой этой функцией.

4.2.2. Правила дифференцирования

Нахождение производных — одна из наиболее распространенных операций математического анализа. Для того, чтобы уметь вычислять любые производные, необходимо помнить производные основных элементарных функций и знать правила дифференцирования. В таблице 4.1 представлены основные элементарные функции и их производные. Вывод этих зависимостей можно найти в любом учебнике по высшей математике, например в [13].

Отметим, что функция $f(x) = e^x$ (эта функция называется экспонента) — единственная в математике функция, которая не изменяется при дифференцировании.

Основные правила дифференцирования.

1. Производная линейной комбинации функций:

$$(c_1 f_1(x) + c_2 f_2(x))' = c_1 f_1'(x) + c_2 f_2'(x).$$

Например: $(6 \sin x - 2 \ln x)' = 6 \cos x - \frac{2}{x}$.

2. Производная произведения функций:

$$(f(x)g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Например: $(\ln x \cdot \cos x)' = \frac{1}{x} \cdot \cos x - \ln x \cdot \sin x$.

3. Производная частного двух функций:

$$\left(\frac{f(x)}{g(x)} \right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}.$$

Например:

$$(\operatorname{tg} x)' = \frac{\sin x}{\cos x} = \frac{\cos x \times \cos x - (-\sin x) \times \sin x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

4.1. Таблица производных основных элементарных функций

Функция $f(x)$	Производная $f'(x)$	Функция $f(x)$	Производная $f'(x)$
c (const)	0	$\ln x$	$\frac{1}{x}$
x^a (a — любое число)	$a x^{a-1}$	$\log_a x$	$\frac{1}{x \ln a}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$	a^x	$a^x \ln a$
$\frac{1}{x}$	$-\frac{1}{x^2}$	e^x	e^x
$\cos x$	$-\sin x$	$\operatorname{arctg} x$	$\frac{1}{1+x^2}$
$\sin x$	$\cos x$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
$\operatorname{tg}(x)$	$\frac{1}{\cos^2(x)}$	$\operatorname{ctg}(x)$	$-\frac{1}{\sin^2(x)}$

4. Дифференцирование сложной функции.

Пусть значения дифференцируемой функции $g(x)$ являются аргументами другой функции $f(x)$. В этом случае говорят о сложной функции $f(g(x))$ или суперпозиции функций f и g . Рассмотрим приращение сложной функции, вызванное изменением аргумента x на Δx :

$$\Delta f = f(g(x+\Delta x)) - f(g(x)).$$

В соответствии с определением (4.7), вычислим производную этой сложной функции:

$$\begin{aligned} \frac{df(g(x))}{dx} &= \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x} = \\ &= \lim_{\Delta x \rightarrow 0} \left[\frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)} \cdot \frac{g(x+\Delta x) - g(x)}{\Delta x} \right] = \\ &= \lim_{\Delta x \rightarrow 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta g(x)} \cdot \lim_{\Delta x \rightarrow 0} \frac{g(x+\Delta x) - g(x)}{\Delta x} = \end{aligned}$$

так как у дифференцируемой функции (см. формулу (4.8)) при $\Delta x \rightarrow 0$, и приращение $\Delta g(x) \rightarrow 0$

$$= \lim_{\Delta g \rightarrow 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta g(x)} \cdot \lim_{\Delta x \rightarrow 0} \frac{g(x+\Delta x) - g(x)}{\Delta x} = \frac{df}{dg} \cdot \frac{dg}{dx}$$

В компактной форме производную от сложной функции можно записать так:

$$\frac{df(g(x))}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}, \text{ или } f'_x = f'_g \cdot g'_x.$$

Вычислим производную от функции $f(x) = \sin 7x$. Здесь функция $g(x) = 7x$. В соответствии с (4.9) имеем: $f'(x) = \cos 7x \cdot (7x)' = 7 \cos 7x$.

Формула (4.9), называемая **правилом цепочки**, может быть обобщена на случай большего числа аргументов.

$$y'(x) = [f(g(h(\dots(v(x)\dots))))]' = f'_g \cdot g'_h \cdot \dots \cdot v'_x \quad (4.10)$$

Согласно правилу цепочки, результатом дифференцирования является произведение производных от отдельных функций. При этом отдельные функции формируют в последовательности обратной вычислению сложной функции.

В качестве примера вычислим производную от сложной функции $y = \ln(\sin(x^2))$. Эта сложная функция из следующих отдельных функций: $f = \ln x$, $g = \sin x$, $h = x^2$. Итак, следуя формуле (4.10), получим:

$$y' = \frac{1}{\sin(x^2)} \cdot \cos(x^2) \cdot 2x = 2x \operatorname{ctg} x^2.$$

4.2.3. Дифференциал

Дифференциалом $df(x)$ функции $f(x)$ в точке x называется произведение производной от функции $f(x)$ в этой точке на величину приращения аргумента Δx :

$$df(x) = f'(x)\Delta x.$$

По определению, для независимой переменной $\Delta x = dx$. Действительно, если взять функцию $f(x) = x$, то ее дифференциал $dx = df(x) = f'(x)\Delta x = (x)' \cdot \Delta x = 1 \cdot \Delta x = \Delta x$, т. е. $dx = \Delta x$.

Поэтому дифференциал функции $f(x)$ записывают чаще так:

$$df(x) = f'(x)dx. \quad (4.11)$$

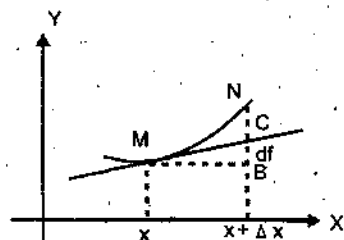


Рис. 4.4. Геометрический смысл дифференциала

Геометрический смысл дифференциала поясняет рис. 4.4. Производная $f'(x)$ численно равна тангенсу угла наклона касательной к графику функции $f(x)$. Дифференциал равен изменению ординаты, вызванному изменением аргумента, если двигаться не по графику функции, а по касательной к нему в точке M .

Замена истинного приращения функции

$$\Delta f(x) = f(x + \Delta x) - f(x)$$

на дифференциал

$$df(x) = f'(x) \cdot \Delta x$$

равносильна замене части графика функции на соответствующую часть касательной к этому графику. В этой связи говорят, что

дифференциал определяет главную линейную часть приращения функции.

4.2.4. Производные высших порядков

Производная $f'(x)$ является функцией того же аргумента x , что и исходная функция. Поэтому ее можно опять дифференцировать, т. е. вычислять предел:

$$\lim_{\Delta x \rightarrow 0} \frac{f'(x + \Delta x) - f'(x)}{\Delta x}.$$

Если этот предел существует и конечен, то он называется **второй производной от функции $f(x)$ в точке x** . Принятые обозначения:

$$(f'(x))' = f''(x) = \frac{d^2 f(x)}{dx^2}.$$

Подобным образом можно ввести производные n -го порядка

$$f^{(n)}(x) = (f^{(n-1)}(x))'$$

В качестве примера рассмотрим производные от степенной функции $y = x^n$.

$$y' = nx^{n-1}, \quad y'' = n(n-1)x^{n-2}, \quad y''' = n(n-1)(n-2)x^{n-3}, \dots, \\ y^{(k)} = n(n-1)(n-2) \dots (n-k+1)x^{n-k} \text{ при } (k \leq n).$$

Аналогично **дифференциалом второго порядка** называется дифференциал от дифференциала. Вычислим его, считая приращение Δx постоянным:

$$d(df(x)) = (df(x))' \Delta x = (f'(x)\Delta x)' \Delta x = f''(x) (\Delta x)^2.$$

Или, в других обозначениях:

$$d^2 f(x) = d(df(x)) = f''(x) dx^2.$$

Аналогичным образом можно определить дифференциал k -го порядка:

$$d^k f(x) = d(d^{k-1} f(x)) = \dots = f^{(k)}(x) dx^k.$$

4.2.5. Формула Тейлора

Пусть функция $f(x)$ в интервале (a, b) имеет производные до $(n+1)$ -го порядка включительно. Поставим задачу: найти полином первой степени ($y = a_1x + a_0$) наиболее близкий к функции $f(x)$ в окрестности некоторой точки $x_0 \in (a, b)$. При этом значение исходной функции и полинома в точке x_0 , а также наклоны соответствующих им линий (т. е. производные) должны совпадать:

$$f(x_0) = a_1x_0 + a_0 \quad (4.12)$$

и, одновременно,

$$f'(x_0) = a_1. \quad (4.13)$$

Сравнивая (4.12) с (4.13), имеем:

$$a_0 = f(x_0) - x_0 f'(x_0).$$

Отсюда искомым уравнением приближающего полинома первой степени будет:

$$y = f(x_0) + f'(x_0) \cdot (x - x_0)$$

Если хотим получить приближение функции полиномом второго порядка $y = a_2x^2 + a_1x + a_0$, то потребуем, чтобы в точке x_0 совпадали функции, их первые и вторые производные.

$$f(x_0) = a_2x_0^2 + a_1x_0 + a_0;$$

$$f'(x_0) = 2a_2x_0 + a_1;$$

$$f''(x_0) = 2a_2.$$

Находя из этой системы коэффициенты — $a_2 = f''(x_0)/2$, $a_1 = f'(x_0) - x_0 f''(x_0)$, $a_0 = f(x_0) - x_0 f'(x_0) + x_0^2 f''(x_0)/2$, — получим приближающий полином второй степени:

$$y = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2} (x - x_0)^2.$$

Продолжая подобные рассуждения, можно получить приближающий полином n -ой степени следующего вида:

$$y = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2} \cdot (x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!} \cdot (x - x_0)^n. \quad (4.13)$$

В точке x_0 значения этого полинома и функции совпадают. В окрестности точки x_0 замена функции полиномом (4.14) дает некоторую ошибку. Замечательным фактом является возможность вычислить эту ошибку, которая в большинстве практических случаев мала. Формула, обеспечивающая возможность точной замены данной функции полиномом, известная как **формула Тейлора**, такова:

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{f''(x_0)}{2!} \cdot (x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!} \cdot (x - x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!} \cdot (x - x_0)^{n+1}, \quad (4.15)$$

где $a < x < b$, $a < x_0 < b$, $x_0 < c < x$.

Выражение

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}$$

называется **остаточным членом формулы Тейлора в форме Лагранжа**. Величина $R_n(x)$ определяет погрешность, возникающую при замене функции полиномом степени n из (4.14). Форма Лагранжа позволяет при вычислениях найти оценку сверху для $|R_n(x)|$.

На практике часто рассматривается частный случай формулы Тейлора, когда $x_0 = 0$.

$$f(x) = f(0) + f'(0) \cdot x + \frac{f''(0)}{2!} \cdot x^2 + \dots + \frac{f^{(n)}(0)}{n!} \cdot x^n + \frac{f^{(n+1)}(c)}{(n+1)!} \cdot x^{n+1}, \quad (4.16)$$

где $0 < c < x$.

Формула (4.16) называется **формулой Маклорена**.

Формулы Тейлора и Маклорена широко применяются при приближенных вычислениях функций, в том числе в компьютерных вычислениях.

Пример. Вычислить значение числа e .

Решение. Построим формулу Тейлора для функции $f(x) = e^x$ в окрестности точки $x_0 = 0$. Прежде всего, вычислим производные:

$$f(x) = e^x, \quad f'(x) = e^x, \quad f''(x) = e^x, \quad \dots, \quad f^{(n)}(x) = e^x.$$

Отсюда $f(0) = f'(0) = f''(0) = \dots = f^{(n)}(0) = 1$.

Из (4.16) для $f(x) = e^x$ имеем:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \frac{e^{\xi x}}{(n+1)!} + R_n(x).$$

Эта формула получена для e^x . Если в правой части положить $x = 1$, то:

$$e = 2 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} + \frac{e^{\xi}}{(n+1)!} + R_n(x).$$

В зависимости от требований задачи эта формула позволяет получить сколь угодно точные значения величины e . Так:

для $n = 2$ $e \approx 2.5$, ошибка не превышает величины 0.23,
 для $n = 3$ $e \approx 2.667$, ошибка не превышает величины 0.052,
 для $n = 10$ $e \approx 2.7182818$ и ошибка не более, чем $4.3 \cdot 10^{-7}$.

Глава 4.3. Функции многих переменных

4.3.1. Частные производные

Если каждому упорядоченному набору возможных значений аргументов $x_1, x_2, \dots, x_n$ по закону f поставлено в соответствие единственное число, то говорят, что задана **функция n переменных**:

$$y = f(x_1, x_2, \dots, x_n).$$

Для задания такой функции необходимо, кроме того, указать n -мерную область D возможных значений независимых переменных.

Каждая точка в области D может рассматриваться и как n -мерный вектор в выбранной системе координат. Так, если $n = 2$, то D — часть плоскости, а $f(x, y)$ определяет некоторую поверхность в пространстве (рис 4.5). Проекция этой поверхности на плоскость XOY и есть область D .

Если $n = 3$, то D — некоторая объемная область и $f(x, y, z)$ рассматривается как функция от точки в трехмерном пространстве. Физически, например, функцию $f(x, y, z)$ можно интерпретировать как плотность вещества в объемной области D .

Для простоты и наглядности обычно подробно рассматриваются функции двух переменных. Полученные при этом

свойства остаются верными и для функций произвольного числа переменных.

Пусть задана некоторая функция двух переменных $f(x, y)$. Зафиксируем один из аргументов, выбрав, например, $y = y_0$. Тогда получим функцию одной переменной $f(x, y_0)$ — функцию переменной x .

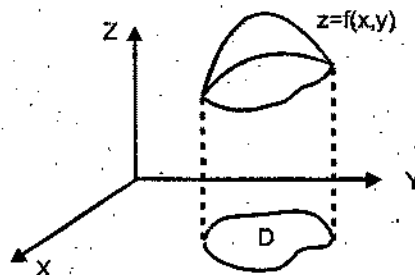


Рис. 4.5. Функция двух переменных

Производная от такой функции по x есть предел:

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y_0) - f(x, y_0)}{\Delta x}$$

Если этот предел существует и конечен, то его значение называется частной производной от функции $f(x, y)$ по переменной x при $y = y_0$. Так как y_0 любое фиксированное число из области допустимых значений, то его можно заменить на просто y . Тогда:

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} = \frac{\partial f(x, y)}{\partial x} = f'_x(x, y)$$

Частная производная от $f(x, y)$ по y определяется и обозначается аналогичным образом:

$$\lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y} = \frac{\partial f(x, y)}{\partial y} = f'_y(x, y).$$

То есть, при вычислении частной производной от функции двух переменных $f(x, y)$ по x второй аргумент y выступает как величина постоянная. Если же вычисляется частная производная по y , то x принимается постоянной величиной.

Пример. Вычислить частные производные f'_x и f'_y от функции $f(x, y) = x^3 y^2 + \sin x - 4y$

В соответствии с определением, имеем:

$$f'_x(x, y) = 3x^2 y^2 + \cos x \quad \text{и} \quad f'_y(x, y) = 2x^3 y - 4.$$

Частная производная от $f(x, y)$ тоже является функцией двух переменных и от нее вновь можно вычислять частные производные, и так далее.

Функция двух переменных имеет следующие **вторые производные**:

- вторая производная от $f(x, y)$ по x дважды:

$$f''_{xx} = \frac{\partial^2 f(x, y)}{\partial x^2} = \frac{\partial}{\partial x} \frac{\partial f(x, y)}{\partial x};$$

- вторая производная от $f(x, y)$ по y дважды:

$$f''_{yy} = \frac{\partial^2 f(x, y)}{\partial y^2} = \frac{\partial}{\partial y} \frac{\partial f(x, y)}{\partial y};$$

- вторая смешанная производная от $f(x, y)$ по x и по y :

$$f''_{xy} = \frac{\partial^2 f(x, y)}{\partial x \partial y} = \frac{\partial}{\partial x} \frac{\partial f(x, y)}{\partial y};$$

- вторая смешанная производная от $f(x, y)$ по y и по x

$$f''_{yx} = \frac{\partial^2 f(x, y)}{\partial y \partial x} = \frac{\partial}{\partial y} \frac{\partial f(x, y)}{\partial x} = f''_{xy}.$$

Т. е. для функций, имеющих непрерывные частные производные второго порядка, смешанные производные второго порядка совпадают.

Пример. Вычислить вторые производные для функции $f(x, y) = x^3 y^2 + \sin x - 4y$.

Применяя правила дифференцирования, получим:

$$f''_{xx} = 6xy^2 - \sin x, \quad f''_{yy} = 2x^3, \quad f''_{xy} = 6x^2 y = f''_{yx}.$$

Теперь не представляет труда решение задачи о вычислении производных любого порядка. Например, для функции $f(x, y) = 2x^4 \cdot \ln y - \cos(x + y^3) + x^3$ требуется вычислить четвертую производную, причем одну по x и три по y .

В соответствии с правилами дифференцирования сложных функций и функций многих переменных имеем:

$$\begin{aligned} \frac{\partial^4 f(x, y)}{\partial x \partial y^3} &= ((2x^4 \ln y - \cos(x + y^3) + x^3)'_y)'_{yy} = (8x^3 \ln y + \sin(x + y^3) + 3x^2 y)'_{yy} = \\ &= \left(\frac{8x^3}{y} + 3y^2 \cos(x + y^3) \right)'_{yy} = \left(\frac{-8x^3}{y^2} + 6y \cos(x + y^3) - 9y^4 \sin(x + y^3) \right)'_y = \\ &= 16 \frac{x^3}{y^3} + 6 \cos(x + y^3) - 18y^3 \sin(x + y^3) - 36y^3 \sin(x + y^3) - 27y^6 \cos(x + y^3) = \\ &= 16 \frac{x^3}{y^3} + (6 - 27y^6) \cos(x + y^3) - 54y^3 \sin(x + y^3). \end{aligned}$$

Производные от функций большего числа переменных вычисляются по тем же правилам. Пусть имеем функцию четырех переменных x, y, z, t :

$$f(x, y, z, t) = xz^3 t^2 + yz^2 \cos(y^3 - t).$$

Вычислим вторую смешанную производную по аргументам z и t :

$$\begin{aligned} \frac{\partial^2 f(x, y, z, t)}{\partial z \partial t} &= ((xz^3 t^2 + yz^2 \cos(y^3 - t))'_z)'_t = (xt^2 3z^2 + y \cos(y^3 - t) 2z)'_t = \\ &= 3xz^2 2t + 2yz(-\sin(y^3 - t)(-1)) = 6xtz^2 + 2yz \sin(y^3 - t). \end{aligned}$$

4.3.2. Полный дифференциал

В некоторых случаях необходимо вычислить приращение функции, $\Delta f(x, y) = f(x + \Delta x, y + \Delta y) - f(x, y)$, возникшее при приращении двух ее аргументов. Если значение $f(x, y)$ известно, но неизвестно $f(x + \Delta x, y + \Delta y)$, то вычисление $\Delta f(x, y)$ становится сложной задачей. Приближенное значение приращения функции удобно вычислять при помощи полного дифференциала.

Полным дифференциалом функции $f(x, y)$ называется выражение:

$$df(x, y) = \frac{\partial f(x, y)}{\partial x} \Delta x + \frac{\partial f(x, y)}{\partial y} \Delta y = \frac{\partial f(x, y)}{\partial x} dx + \frac{\partial f(x, y)}{\partial y} dy.$$

Напомним, что для независимых переменных по определению $\Delta x = dx$, $\Delta y = dy$. Полное приращение функции двух переменных, вызванное приращением ее аргументов, равно полному дифференциалу плюс бесконечно малая функция более высокого порядка малости, чем приращения аргументов, т. е.

$$\Delta f(x, y) = df(x, y) + \alpha(\Delta x, \Delta y).$$

В этой связи на практике при небольших изменениях аргументов приращение функции заменяют на ее полный дифференциал.

Пример. Найдем дифференциал функции $z = x \sin(x^2 \sqrt{y})$:

$$z'_x = \sin(x^2 \sqrt{y}) + x \cos(x^2 \sqrt{y}) 2x$$

$$z'_y = x^3 \cos(x^2 \sqrt{y}) \frac{1}{2\sqrt{y}}$$

$$dz = \left[\sin(x^2 \sqrt{y}) + x \cos(x^2 \sqrt{y}) \right] dx + \left[\frac{x}{2\sqrt{y}} \cos(x^2 \sqrt{y}) \right] dy.$$

Пример. Найдем для функции $f(x, y) = xy$ приращение и соответствующий полный дифференциал в точке $(x = 4, y = 3)$ при $\Delta x = 0.2$ и $\Delta y = 0.1$.

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = y \Delta x + x \Delta y = 3 \cdot 0.2 + 4 \cdot 0.12 = 1.08$$

$$\Delta f = (x + \Delta x)(y + \Delta y) - xy = (4 + 0.2) \cdot (3 + 0.1) - 12 = 1.02$$

Следовательно, разность (рассогласование) между Δf и df составит 0.06.

Глава 4.4. Интегрирование

4.4.1. Неопределенный интеграл

Операция нахождения производной от функции называется дифференцированием. Обратная дифференцированию операция — отыскание функции по ее производной называется интегрированием. Функция $F(x)$, производная которой равна некоторой функции $f(x)$, т. е. $F'(x) = f(x)$, называется первообразной для $f(x)$. Так, например, если $f(x) = x^n$, то ее первообразная

есть $F(x) = \frac{x^{n+1}}{n+1}$, если же $f(x) = \sin 2x$, то ее первообразная

$$F(x) = -0.5 \cos 2x.$$

Поскольку производная любой постоянной величины равна нулю, то $-0.5 \cos 2x$ является первообразной не только для $\sin 2x$, но и для $\sin 2x + 4$, и для $\sin 2x - \sqrt{3}$, и вообще для любой функции вида $\sin 2x + C$, т. е. первообразные одной и той же функции отличаются на постоянную величину.

Обобщая вышесказанное, можно сказать, что функция $f(x)$ имеет бесконечное множество первообразных вида $F(x) + C$.

Множество всех первообразных для функции $f(x)$ называется неопределенным интегралом от этой функции и обозначается так:

$$\int f(x) dx = F(x) + C, \quad (4.17)$$

где $\int$ — знак интеграла, читается «интеграл», $f(x)$ — подынтегральная функция от переменной интегрирования x , $f(x)dx$ — подынтегральное выражение. C — постоянная интегрирования.

Часто вместо слов «вычислить неопределенный интеграл» говорят «взять неопределенный интеграл».

Из определения интеграла следует, что:

$$\left(\int f(x) dx \right)' = (F(x) + C)' = F'(x) = f(x). \quad (4.18)$$

Отсюда, вспомнив, что $df = F'(x)dx$, получим:

$$d \left(\int f(x) dx \right) = \left(\int f(x) dx \right)' dx = f(x) dx.$$

Рассмотрим выражение $\int F'(x) dx$, где $F(x)$ — первообразная для функции $f(x)$ (т. е. $F'(x) = f(x)$). Тогда из определения неопределенного интеграла следует:

$$\int F'(x)dx = \int f(x)dx = F(x) + C,$$

или

$$\int dF(x) = F(x) + C \quad (4.19)$$

Формулы (4.17–4.19) наглядно иллюстрируют то обстоятельство, что операции дифференцирования и интегрирования взаимно обратны с точностью до постоянной. В этой связи по аналогии с таблицей формул дифференцирования элементарных функций 4.1 можно построить таблицу основных интегралов.

Таблица основных интегралов

$\int dx = x + C$	$\int e^x dx = e^x + C$
$\int x^a dx = \frac{x^{a+1}}{a+1} + C \quad a \neq -1$	$\int \frac{1}{a^2 + x^2} dx = \frac{1}{a} \operatorname{arctg} \left(\frac{x}{a} \right) + C$
$\int \frac{dx}{x} = \ln x + C$	$\int \frac{1}{\sqrt{a^2 - x^2}} dx = \operatorname{arcsin} \left(\frac{x}{a} \right) + C$
$\int \sin(x) dx = -\cos(x) + C$	$\int \frac{1}{a^2 - x^2} dx = \frac{1}{2a} \ln \left \frac{x-a}{x+a} \right + C$
$\int \cos(x) dx = \sin(x) + C$	$\int \frac{1}{\cos^2 x} dx = \operatorname{tg} x + C$
$\int a^x dx = \frac{a^x}{\ln a} + C$	$\int \frac{1}{\sin^2 x} dx = -\operatorname{ctg} x + C$

Справедливость этих формул проверяется по формуле (4.18) непосредственным дифференцированием.

В различных справочниках, например в [14], приводятся обширные таблицы неопределенных интегралов. Но, при решении практических задач, для того, чтобы иметь возможность воспользоваться таблицами, вначале необходимо преобразовать исходный интеграл до вида, который есть в таблице. Для этого надо знать основные свойства неопределенного интеграла и использовать тождественные преобразования подынтегральной функции.

Основные свойства неопределенного интеграла.

1. Постоянный множитель можно выносить за знак неопределенного интеграла:

$$\int cf(x)dx = c \int f(x)dx.$$

2. Неопределенный интеграл от суммы функций равен сумме неопределенных интегралов от этих функций:

$$\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx.$$

Замечание. Если каждый из суммируемых неопределенных интегралов содержит свою постоянную интегрирования, то для всей суммы записывается одна постоянная интегрирования.

Пример. Найти $\int \frac{1}{\sin^2(x)\cos^2(x)} dx$.

Запишем стоящую в числителе единицу в тригонометрическом виде ($1 = \sin^2 x + \cos^2 x$) и разделим почленно числитель на знаменатель, а далее получим табличные интегралы:

$$\begin{aligned} \int \frac{1}{\sin^2 x \times \cos^2 x} dx &= \int \frac{\sin^2 x + \cos^2 x}{\sin^2 x \times \cos^2 x} dx = \\ &= \int \left(\frac{1}{\cos^2 x} + \frac{1}{\sin^2 x} \right) dx = \int \frac{dx}{\cos^2 x} + \int \frac{dx}{\sin^2 x} = \\ &= \operatorname{tg} x + C_1 - \operatorname{ctg} x + C_2 = \operatorname{tg} x - \operatorname{ctg} x + C. \end{aligned}$$

4.4.2. Методы интегрирования

Для вычисления неопределенных интегралов часто используют так называемые стандартные методы интегрирования. Перечислим основные из них.

Метод интегрирования по частям. Пусть на некотором промежутке существуют производные функций $u = u(x)$ и $v = v(x)$. По правилу дифференцирования произведения, имеем:

$$d(u \cdot v) = du \cdot v + u \cdot dv.$$

Проинтегрируем это равенство, учитывая, что $\int d(u \cdot v) = u \cdot v$:

$$u \cdot v = \int du \cdot v + \int u \cdot dv, \quad (4.20)$$

или

$$\int u dv = uv - \int v du. \quad (4.21)$$

Эта формула носит название **формулы интегрирования по частям**. Ее применение полезно в тех случаях, когда подынтегральное выражение можно представить в виде произведения двух функций $f(x) = u(x) \cdot v'$ и выражение $v du = v \cdot u' \cdot dx$ для взятия интеграла проще, чем подынтегральное выражение $u \cdot dv = u \cdot v' \cdot dx$.

Пример. Найти $\int x \sin(x) dx$. Использование формулы интегрирования по частям позволяет вместо исходного нетабличного интеграла вычислить интеграл от $\sin x$. Покажем это, приведя схему записи удобную при использовании метода интегрирования по частям.

$$\begin{aligned} \int x \sin(x) dx &= \left| \begin{array}{ll} u = x & dv = \sin x dx \\ du = dx & v = -\cos x \end{array} \right| = -x \cos x + \\ &+ \int \cos(x) dx = -x \cos x + \sin x + C. \end{aligned}$$

Метод замены переменной. Добиться упрощения подынтегрального выражения можно при помощи **метода замены переменной интегрирования**. Суть этого метода заключается в замене переменной интегрирования x на некоторую функцию $\varphi(t)$ с тем, чтобы преобразовать исходный интеграл к более простому виду.

$$\int f(x) dx = \int f(\varphi(t)) \varphi'(t) dt. \quad (4.22)$$

Формула (4.22) называется **формулой замены переменной** под знаком неопределенного интеграла. Априорных рекомендаций по эффективному применению этой формулы не существует и все зависит от интуиции и опыта исследователя.

Алгоритм метода замены переменной следующий. Вначале необходимо найти замену переменной интегрирования $x = \varphi(t)$, записать интеграл с новой переменной интегрирования t , вычислить его, а затем вновь вернуться к исходной переменной интегрирования.

Пример. Найти $\int \frac{e^{\arctg(x)}}{1+x^2} dx$.

Решение. Особенностью данного интеграла является то обстоятельство, что его подынтегральное выражение содержит множитель $\frac{dx}{1+x^2}$, который является дифференциалом функции $\arctg x$. Поэтому в данном интеграле целесообразно ввести замену переменной: $t = \arctg x$. Отсюда $dt = d(\arctg(x)) =$

$$= \frac{dx}{1+x^2}, \text{ и } e^{\arctg x} = e^t. \text{ Подставляя в исходный интеграл, имеем:}$$

$$\int \frac{e^{\arctg(x)}}{1+x^2} dx = \int e^t dt = e^t + C = e^{\arctg x} + C.$$

Пример. Найти $\int \sin^3 x \cdot \cos^2 x dx$

Решение. Здесь уместна замена $t = \cos x$, т. к. $dt = -\sin x dx$, и $\sin^3 x dx = \sin^2 x \sin x dx$. Поэтому:

$$\begin{aligned} \int \sin^3 x \cdot \cos^2 x \cdot dx &= \int (1 - \cos^2) \cdot \cos^2 x \cdot (\sin x \cdot dx) = -\int (1 - t^2) \cdot t^2 \cdot dt = \\ &= \int (t^4 - t^2) \cdot dt = \frac{t^5}{5} - \frac{t^3}{3} + C = \frac{\cos^5 x}{5} - \frac{\sin^3 x}{3} + C. \end{aligned}$$

Глава 4.5. Определенный интеграл

4.5.1. Определение определенного интеграла

Пусть на отрезке $[a, b]$ задана неотрицательная непрерывная функция $f(x)$. На плоскости XOY , как показано на рис. 4.6, график этой функции, отрезок оси абсцисс и прямые $x = a$ и $y = b$ образуют так называемую криволинейную трапецию. Требуется найти площадь такой криволинейной трапеции S .

Разделим отрезок $[a, b]$ на произвольные n частей и проведем координаты, соответствующие точкам деления $x_0 = a < x_1 < x_2 < \dots < x_i < x_{i+1} < \dots < x_n = b$. Криволинейная трапеция стала состоять из n «узких» криволинейных трапеций (полосок) шириной $\Delta x_i = x_{i+1} - x_i$. Каждую i -ю полоску заменим на соответствующий прямоугольник, высота которого равна $f(x_i)$, а площадь $S_i = f(x_i) \cdot \Delta x_i$, $i = 1, n$.

Интуитивно ясно, сумма площадей полученных прямоугольников приближенно равна площади исходной криволинейной трапеции.

$$S \approx \sum_{i=1}^n f(x_i) \Delta x_i.$$

Увеличим число разбиений n . При этом каждый раз обязательно должна уменьшаться длина наибольшего из разбиений Δx_i , т. е. $\max \Delta x_i \rightarrow 0$ при $n \rightarrow \infty$. Погрешность такого приближения будет стремиться к нулю, т. е.

$$\lim_{\substack{n \rightarrow \infty \\ \max \Delta x_i \rightarrow 0}} \sum_{i=1}^n f(x_i) \Delta x_i = S \quad (4.23)$$

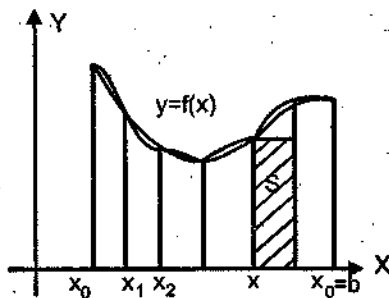


Рис. 4.6. Геометрический смысл определенного интеграла

Для обозначения предельного значения этой суммы Лейбниц ввел символ « $\int$ » как стилизацию начертания буквы S — начальной буквы латинского слова Summa. К построению суммы вида (4.23), которые называют **интегральными суммами**, приходят при решении многих задач: о массе плоской пластины, о длине дуги кривой

линии, о координатах центра тяжести тел, о площади поверхности тел вращения и др.

Введем определение определенного интеграла. Если независимо от способа разбиения отрезка $[a, b]$ на части, для функции $f(x)$ существует конечный предел интегральной суммы при $n \rightarrow \infty$ и $\max \Delta x_i \rightarrow 0$, то этот предел называется **определенным интегралом** функции $f(x)$ от a до b , а сама функция $f(x)$ — **интегрируемой** на $[a, b]$.

Обозначение:

$$\lim_{\substack{n \rightarrow \infty \\ \max \Delta x_i \rightarrow 0}} \sum_{i=1}^n f(x_i) \cdot \Delta x_i \equiv \int_a^b f(x) dx.$$

Читается: «Интеграл от a до b от функции $f(x)$ де $икс$ ». Числа a и b называются соответственно нижним и верхним пределами интеграла. Если $a = b$, то, как следует из определения,

$$\int_a^a f(x) dx = 0.$$

Интеграл существует для всех непрерывных функций, т. е. функция $f(x)$ интегрируема на отрезке $[a, b]$, если она непрерывна на этом отрезке.

4.5.2. Основные свойства определенного интеграла

1. *Определенный интеграл — есть число!* Его значение зависит только от вида функции $f(x)$ и пределов интегрирования, но не от переменной интегрирования, которую можно обозначить любой буквой:

$$\int_a^b f(x) dx \equiv \int_a^b f(t) dt.$$

2. Постоянный множитель можно выносить за знак интеграла:

$$\int_a^b c \cdot f(x) dx = c \int_a^b f(x) dx.$$

3. Определенный интеграл от алгебраической суммы нескольких функций равен алгебраической сумме интегралов от слагаемых. Для суммы двух функций имеем:

$$\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$$

4. Аддитивность. Если функция $f(x)$ интегрируема на отрезках $[a, c]$ и $[c, b]$, $a < c < b$, то она интегрируема и на отрезке $[a, b]$, причем выполняется равенство:

$$\int_a^c f(x) dx + \int_c^b f(x) dx = \int_a^b f(x) dx. \quad (4.24)$$

Свойство аддитивности имеет наглядный геометрический смысл: оно выражает свойство аддитивности площади, например, плоских фигур (рис.4.7).

5. Если $f(x)$ — нечетная функция, т. е. $f(-x) = -f(x)$, то:

$$\int_{-a}^a f(x) dx = 0,$$

если $f(x)$ — четная функция, т. е. $f(-x) = f(x)$, то:

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx.$$

6. По определению $\int_a^b f(x) dx = - \int_b^a f(x) dx$.

7. Теорема о среднем. Если $f(x)$ непрерывна на отрезке $[a, b]$, то на этом отрезке найдется такая точка c , что справедливо равенство:

$$\int_a^b f(x) dx = f(c) \cdot (b - a). \quad (4.25)$$

Эта формула имеет ясный геометрический смысл (рис.4.8): площадь криволинейной трапеции численно равна площади прямоугольника с тем же основанием, что и трапеция, причем высота прямоугольника равна ординате $f(c)$ в некоторой точке c , лежащей между a и b .

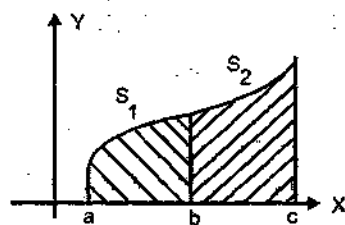


Рис. 4.7. Аддитивность определенного интеграла

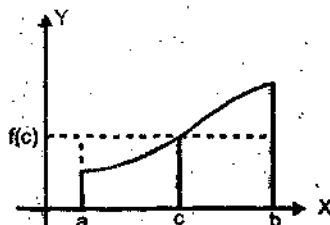


Рис. 4.8. Геометрическая интерпретация теоремы о среднем

4.5.3. Вычисление определенного интеграла

Определенный интеграл вычисляется по формуле Ньютона-Лейбница

$$\int_a^b f(x) dx = F(b) - F(a), \quad (4.26)$$

где $F(x)$ — первообразная для функции $f(x)$. Формула читается так: *определенный интеграл равен разности значений первообразной на концах промежутка.*

Сформулируем и докажем теоремы, обосновывающие справедливость этой основной формулы математического анализа.

Теорема о существовании первообразной. Если функция

$f(x)$ непрерывна на отрезке $[a, b]$, то функция $\Phi(x) = \int_a^x f(t) dt$

дифференцируема в любой внутренней точке x этого отрезка, причем $\Phi'(x) = f(x)$, то есть функция $\Phi(x)$ является первообразной функции $f(x)$.

Доказательство. Найдем производную функции $\Phi(x)$, являющейся интегралом с переменным верхним пределом. Для этого вначале выберем Δx столь малым, чтобы точка $x + \Delta x$ лежала внутри отрезка $[a, b]$, и построим приращение функции $\Phi(x)$.

$$\begin{aligned} \Delta \Phi(x) &= \Phi(x + \Delta x) - \Phi(x) = \int_a^{x+\Delta x} f(t) dt - \int_a^x f(t) dt = \\ &= \int_a^x f(t) dt + \int_x^{x+\Delta x} f(t) dt - \int_a^x f(t) dt = \int_x^{x+\Delta x} f(t) dt. \end{aligned}$$

К полученному интегралу применим теорему о среднем:

$$\Delta \Phi(x) = \int_x^{x+\Delta x} f(t) dt = f(\xi) \Delta x, \text{ где } \xi \in [x, x+\Delta x].$$

Следовательно, $\frac{\Delta \Phi}{\Delta x} = f(\xi)$. Поскольку $f(x)$ непрерывна и $\xi \rightarrow x$

при $\Delta x \rightarrow 0$, то $\lim_{\Delta x \rightarrow 0} f(\xi) = f(x)$. Поэтому производная

$$\Phi'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta \Phi}{\Delta x} = \lim_{\xi \rightarrow x} f(\xi) = f(x).$$

Мы показали, что функция, являющаяся интегралом от функции $f(x)$ с постоянным нижним и переменным верхним пределом x , есть одна из первообразных функций от функции $f(x)$. Этот факт, установленный Ньютоном и Лейбницем, показывает, что дифференциальное и интегральное исчисление представляет собой нечто единое, и известен как основная теорема математического анализа.

Теперь выведем основную формулу математического анализа.

Теорема. Если функция $f(x)$ непрерывна на отрезке $[a, b]$, то

$$\int_a^b f(x) dx = F(b) - F(a), \quad (4.27)$$

Доказательство. В силу непрерывности на отрезке $[a, b]$ функция $f(x)$ интегрируема и на основании основной теоремы математического анализа имеет первообразную $\Phi(x) =$

$= \int_a^x f(t) dt$. Первообразных у функции бесконечно много, и все

они отличаются друг от друга на произвольную постоянную (обозначим ее через C). Поэтому, если $F(x)$ — первообразная для $f(x)$, то:

$$\Phi(x) = \int_a^x f(t) dt = F(x) + C. \quad (4.28)$$

Константу C легко выразить через значение первообразной F в точке a . Действительно принимая во внимание, что

$$\Phi(a) = \int_a^a f(t) dt = 0$$

из (4.28) получим: $-F(a) = C$. Поскольку $\Phi(b) = \int_a^b f(t) dt$, запи-

сав (4.27) для точки b с переменной интегрирования x , получаем основную формулу математического анализа — формулу Ньютона-Лейбница (4.27).

Эта формула дает следующее правило: для вычисления определенного интеграла необходимо найти первообразную подынтегральной функции, а затем вычислить разность значений первообразной на верхнем и нижнем пределе.

Пример. Вычислить интеграл $\int_2^5 x^3 dx$.

Решение. Одна из первообразной функции x^3 есть функция $x^4/4$. Поэтому:

$$\int_2^5 x^3 dx = \frac{(5)^4}{4} - \frac{(2)^4}{4} = \frac{625}{4} - \frac{16}{4} = \frac{609}{4}.$$

Формулу (4.27) при вычислении определенного интеграла удобно записывать, используя знак подстановки в следующем виде:

$$\int_a^b f(x) dx = F(x) \Big|_{x=a}^{x=b} = F(b) - F(a)$$

При вычислении определенного интеграла используются те же основные приемы, что и при вычислении неопределенного интеграла.

Формула интегрирования по частям.

$$\int_a^b u dv = u \cdot v \Big|_{x=a}^{x=b} - \int_a^b v du \quad (4.29)$$

Пример. Вычислить $\int_0^1 x e^{-x} dx$.

Решение. Прямому вычислению данного интеграла препятствует наличие множителя x в подынтегральном выражении. Поскольку производная от x по x равна 1, то целесообразно, используя (4.29), положить $u = x$. Тогда:

$$\begin{aligned} \int_0^1 x e^{-x} dx &= \left| \begin{array}{ll} u = x & dv = e^{-x} dx \\ du = dx & v = -e^{-x} \end{array} \right| = -x e^{-x} \Big|_{x=0}^{x=1} + \int_0^1 e^{-x} dx = \\ &= -e^{-1} - e^{-x} \Big|_0^1 = -2e^{-1} + 1 = \frac{e-2}{e}. \end{aligned}$$

Замена переменной в определенном интеграле. Во многих случаях подынтегральное выражение можно упростить, если заметить, что его часть является дифференциалом некоторой функции. Тогда, по аналогии с (4.22), можно записать:

$$\begin{aligned} \int_a^b f(x) dx &= \int_a^b f(\varphi(t)) \varphi'(t) dt = F(\varphi(t)) \Big|_{\alpha}^{\beta} = F(\varphi(\beta)) - F(\varphi(\alpha)) = \\ &= F(b) - F(a), \end{aligned}$$

где $x = \varphi(t)$, $\varphi(\alpha) = a$, $\varphi(\beta) = b$, т. е. $[\varphi(\alpha); \varphi(\beta)] = [a, b]$.

Пример. Вычислить $\int_1^e \frac{\ln^2 x}{x} dx$.

Решение. Положив $\ln x = t$, имеем $dx/x = dt$. Если $x = 1$, то $t = 0$, если $x = e$, то $t = 1$. Тогда:

$$\int_1^e \frac{\ln^2 x}{x} dx = \int_0^1 t^2 dt = \frac{1}{3} t^3 \Big|_0^1 = \frac{1}{3} (1^3 - 0^3) = \frac{1}{3}.$$

4.5.4. Приложения определенного интеграла

В целом ряде практически важных задач оказывается, что искомая величина может быть приближенно вычислена по формуле, аналогичной по структуре интегральной сумме вида (4.23). Если оказывается, что предел такой интегральной суммы существует и конечен, то точное значение искомой величины вычисляется как определенный интеграл. Рассмотрим некоторые частные случаи.

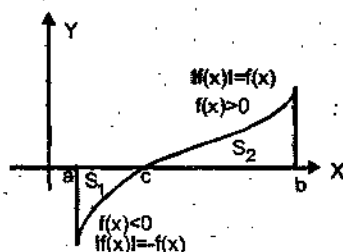


Рис. 4.9. Вычисление площади при помощи определенного интеграла

Это означает, что если на отрезке $[c, d] \subset [a, b]$ функция $f(x) < 0$, то на этом отрезке берется отрицательное значение функции.

Пример. Вычислить площадь, ограниченную осью абсцисс и синусоидой на отрезке $[0, 2\pi]$.

Поскольку $\sin(x) \geq 0$ на отрезке $[0, \pi]$ и $\sin(x) \leq 0$ на $[\pi, 2\pi]$, то искомая площадь S равна:

$$S = \int_0^{2\pi} |\sin x| dx = \int_0^{\pi} \sin x dx + \int_{\pi}^{2\pi} -\sin x dx = -\cos x \Big|_0^{\pi} + \cos x \Big|_{\pi}^{2\pi} = -(\cos \pi - \cos 0) + (\cos 2\pi - \cos \pi) = -(-1 - 1) + (1 + 1) = 4.$$

В более общем случае требуется вычислить площадь плоской фигуры, ограниченной несколькими кривыми линиями. В этом случае искомая площадь есть алгебраическая сумма нескольких криволинейных трапеций. Например, как показано на рис. 4.10:

$$S_{A_1ABB_1} = S_{cbB_1} + S_{aABb} - S_{aA_1c}.$$

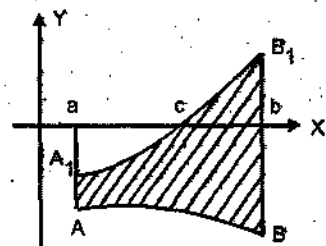


Рис. 4.10. Вычисление площади плоской фигуры

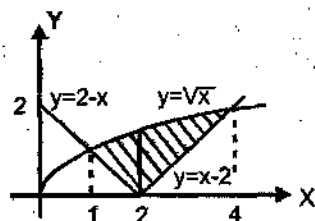


Рис. 4.11.

Пример. Вычислить площадь фигуры, ограниченной функциями $y = |x - 2|$ и $y = \sqrt{x}$. Как следует из рис. 4.11, если провести вертикальную прямую $x = 2$, то искомая площадь может быть вычислена как сумма площадей двух смежных криволинейных трапеций:

$$S = S_1 + S_2 = \int_1^2 (\sqrt{x} - (2 - x)) dx + \int_2^4 (\sqrt{x} - (x - 2)) dx = \left[\frac{x^{3/2}}{3/2} + \frac{x^2}{2} - 2x \right]_1^2 + \left[\frac{x^{3/2}}{3/2} - \frac{x^2}{2} + 2x \right]_2^4 = \frac{13}{6}.$$

Вычисление длины дуги. Пусть некоторая гладкая плоская кривая описывается функцией $f(x)$ и отрезку оси абсцисс $[a, b]$ отвечает дуга AB .

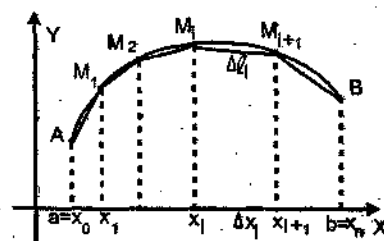


Рис. 4.12. Длина дуги

Произвольным образом разобьем эту дугу, как показано на рис. 4.12, на n частей точками $M_0, M_1, \dots, M_n$. Соединив каждые две соседние точки, получим вписанную в дугу AB ломаную линию.

Длину звена ломанной Δl_i находим по формуле:

$$\Delta l_i = \sqrt{(f(x_{i+1}) - f(x_i))^2 + (x_{i+1} - x_i)^2} = \\ = \sqrt{(\Delta f(x_i))^2 + (\Delta x_i)^2}.$$

При бесконечном увеличении числа точек разбиения, проводимого произвольным образом, длина ломаной будет неограниченно приближаться к длине дуги. На основании этих построений может быть доказана формула для вычисления точного значения длины дуги L плоской кривой:

$$L = \lim_{n \rightarrow \infty} \sqrt{\Delta^2 f_i(x) + \Delta^2 x_i} = \int_a^b \sqrt{1 + (f'(x))^2} dx. \quad (4.30)$$

Если кривая задана в параметрическом виде — $x = \varphi(t)$, $y = \psi(t)$ ($\alpha \leq t \leq \beta$), — то длина кривой вычисляется по формуле:

$$L = \int_{\alpha}^{\beta} \sqrt{(\varphi'(t))^2 + (\psi'(t))^2} dt. \quad (4.31)$$

Пример. Найти длину дуги кривой $y^2 = x^3$, заданной на отрезке от $x = 0$ до $x = 1$ ($y \geq 0$).

Решение. Дифференцируя уравнение кривой, найдем $y = 3/2 \sqrt{x}$. Подставляя затем этот результат в (4.30), получим:

$$L = \int_0^1 dx \sqrt{1 + \frac{9}{4}x} = \frac{4}{9} \cdot \frac{2}{3} \left(1 + \frac{9}{4}x \right)^{3/2} \Big|_0^1 = \frac{8}{27} \left(\frac{13}{4} - 1 \right) = \frac{2}{3}$$

Пример. Найти длину дуги кривой $\varphi = a \cos^3 t$, $\psi = a \sin^3 t$, если t изменяется 0 до 2π .

Решение. Вначале находим производные по t :

$$\varphi' = -3a \cos^2 t \sin t, \quad \psi' = 3a \sin^2 t \cos t$$

Поскольку заданная кривая симметрична относительно осей координат, то достаточно найти длину дуги в пределах $0 \leq t \leq \pi/2$, а затем результат умножить на 4. Подставляя в формулу (4.31), имеем:

$$L = 4 \int_0^{\pi/2} \sqrt{9a^2(\cos^4 t \sin^2 t + \sin^4 t \cos^2 t)} dt = 4 \int_0^{\pi/2} 3a \sqrt{\sin^2 t \cos^2 t} dt = \\ = 4 \int_0^{\pi/2} 3a \sin t \cos t dt = 3a \int_0^{\pi/2} \sin 2t d(2t) = 3a(-\cos 2t) \Big|_0^{\pi/2} = \\ = -3a(\cos \pi - \cos 0) = 6a.$$

Вычисление объемов тел. Пусть для некоторого тела высотой h площадь сечения (перпендикулярного оси OZ) S может быть выражена как функция высоты этого тела $S = S(z)$. В этом случае объем тела V вычисляется по формуле:

$$V = \int_0^h S(z) dz.$$

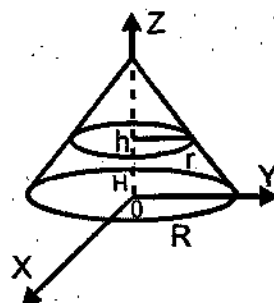


Рис. 4.13. Вычисление объема конуса

Пример. Вычислить объем конуса высотой H и радиусом основания R (см. рис. 4.13).

Решение. Каждое сечение конуса плоскостью параллельной плоскости XY есть окружность радиуса r . Из геометрии известно, что радиус сечения r выражается через расстояние сечения до вершины h и высоту конуса H — $r = \frac{hR}{H}$ —

следовательно, площадь сечения $S(h)$ равна:

$$S(h) = \pi \cdot \frac{R^2}{H^2} h^2.$$

Искомый объем теперь можно вычислить так:

$$V = \int_0^H \pi \frac{R^2}{H^2} h^2 \cdot dh = \frac{\pi R^2}{H^2} \cdot \frac{h^3}{3} \Big|_0^H = \frac{1}{3} \pi \cdot R^2 \cdot H.$$

Глава 4.6. Несобственные интегралы

Для существования определенного интеграла необходимо, чтобы промежуток интегрирования был конечен и подынтегральная функция ограничена. Когда не выполняется одно или оба эти условия, приходят к понятию несобственного интеграла.

4.6.1. Несобственные интегралы с бесконечными пределами

Пусть функция $f(x)$ определена и непрерывна для всех x таких, что $a \leq x < +\infty$. Рассмотрим интеграл $\int_a^b f(x) dx$, который имеет смысл при всех $b > a$. При изменении величины b этот интеграл будет вести себя как непрерывная функция от b . Если при бесконечном возрастании величины b существует конечный пре-

дел $\lim_{b \rightarrow \infty} \int_a^b f(x) dx$, то он называется **несобственным интегралом от функции $f(x)$ с бесконечным верхним пределом**. Таким образом, по определению:

$$\int_a^{+\infty} f(x) dx = \lim_{b \rightarrow \infty} \int_a^b f(x) dx. \quad (4.32)$$

Если предел в (4.32) бесконечен или не существует, то говорят, что **несобственный интеграл не существует или расходится**.

Аналогичным образом определяются несобственные интегралы с бесконечным нижним пределом

$$\int_{-\infty}^b f(x) dx = \lim_{a \rightarrow -\infty} \int_a^b f(x) dx$$

и несобственный интеграл с двумя бесконечными пределами

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^{+\infty} f(x) dx.$$

Из определений несобственных интегралов непосредственно следует схема их вычисления: вначале находится первообразная $F(x)$ для подынтегральной функции $f(x)$, затем рассматривается

разность пределов первообразных в точках верхнего и нижнего пределов интегрирования, т. е.:

$$\int_a^b f(x) dx = \lim_{b \rightarrow +\infty} F(b) - \lim_{a \rightarrow -\infty} F(a).$$

Важный пример. Установить, при каких значениях сходится

и при каких расходится интеграл $I = \int_1^{+\infty} \frac{dx}{x^p}$

Решение:

$$I = \int_1^{+\infty} \frac{dx}{x^p} = \lim_{b \rightarrow +\infty} \frac{1}{1-p} x^{1-p} \Big|_1^b = \lim_{b \rightarrow +\infty} \frac{1}{1-p} (b^{1-p} - 1), \text{ если } p \neq 1$$

$$I = \lim_{b \rightarrow +\infty} \int_1^b \frac{dx}{x^p} = \lim_{b \rightarrow +\infty} (\ln x) \Big|_1^b = \infty, \text{ если } p = 1.$$

Вывод: сходимость интеграла I зависит от значения параметра p :

если $p > 1$, то $\int_1^{+\infty} \frac{dx}{x^p} = \frac{1}{1-p}$, т. е. интеграл сходится;

если $p < 1$, то $\int_1^{+\infty} \frac{dx}{x^p} = \infty$, т. е. интеграл расходится;

если $p = 1$, то $\int_1^{+\infty} \frac{dx}{x^p} = \ln x \Big|_1^{+\infty} = \infty$, интеграл расходится.

4.6.2. Несобственные интегралы от разрывных функций

Пусть функция $f(x)$ определена и непрерывна на $[a, b]$ за исключением точки $c \in [a, b]$. Рассмотрим три случая.

1. Функция терпит разрыв в точке $c = b$. Интеграл от функции $f(x)$ с точкой разрыва на верхнем пределе определяется так:

$$\int_a^c f(x) dx = \lim_{b \rightarrow c-0} \int_a^b f(x) dx.$$

Запись $b \rightarrow c - 0$ означает стремление b к c слева.

Пример. Вычислить интеграл:

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} = \lim_{x \rightarrow 1-0} \arcsin x \Big|_0^{1-0} = \frac{\pi}{2} - 0 = \frac{\pi}{2}.$$

2. Функция терпит разрыв в точке $c = a$. Тогда, по аналогии с предыдущим случаем, интеграл с точкой разрыва на нижнем пределе определяется так:

$$\int_c^b f(x) dx = \lim_{a \rightarrow c+0} \int_a^b f(x) dx.$$

Запись $a \rightarrow c+0$ означает стремление a к точке c справа.

Пример. Исследовать интеграл $I = \int_0^1 \frac{dx}{x}$. Здесь подынтеграль-

ная функция в точке $x = 0$ не существует, а справа от этой точки существует, поэтому:

$$\int_0^1 \frac{dx}{x} = \lim_{a \rightarrow 0+0} \int_a^1 \frac{dx}{x} = \lim_{a \rightarrow 0+0} |\ln x|_a^1 = \lim_{a \rightarrow 0+0} (\ln 1 - \ln a) = +\infty.$$

Таким образом, данный интеграл расходится (не существует).

3. Функция имеет разрыв во внутренней точке отрезка $[a, b]$, т. е. $a < c < b$.

$$\int_a^b f(x) dx = \lim_{c \rightarrow b-0} \int_a^c f(x) dx + \lim_{c \rightarrow a+0} \int_c^b f(x) dx.$$

Пример. Вычислить интеграл $\int_{-1}^1 \frac{dx}{\sqrt[3]{x}}$.

$$\int_{-1}^1 \frac{dx}{\sqrt[3]{x}} = \lim_{c \rightarrow 0-0} \frac{3}{2} x^{\frac{2}{3}} \Big|_{-1}^c + \lim_{c \rightarrow 0+0} \frac{3}{2} x^{\frac{2}{3}} \Big|_c^1 = -\frac{3}{2} + \frac{3}{2} = 0.$$

Глава 4.7. Ряды

Еще из элементарной математики известно, как сложить два произвольных числа, а также три, четыре, и т. д. В математическом же анализе часто приходится рассматривать задачу сложения бесконечного числа слагаемых. Что понимать под суммой бесконечного числа чисел, функций, матриц и других объектов, как вычислять такие суммы? На эти и ряд других вопросов дает ответ теория рядов.

4.7.1. Числовые ряды

4.7.1.1. Основные определения теории числовых рядов. Пусть задана бесконечная числовая последовательность, где u_n — общий член последовательности:

$$u_1, u_2, u_3, \dots, u_n, \dots \quad (4.33)$$

Числовым рядом называется бесконечная последовательность чисел, соединенная знаком плюс, т. е. выражение:

$$u_1 + u_2 + u_3 + \dots + u_n + \dots = \sum_{n=1}^{\infty} u_n \quad (4.34)$$

числа $u_1, u_2, u_3, \dots, u_n, \dots$ называются членами ряда и являются элементами заданной последовательности (4.33). Например, числовой ряд

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots + \frac{1}{n(n+1)} + \dots = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \quad (4.35)$$

имеет общий член $u_n = \frac{1}{n(n+1)}$.

Сходимость и сумма ряда. Частичной суммой S_n называется сумма первых n членов ряда, т. е.

$$S = u_1 + u_2 + u_3 + \dots + u_n.$$

Частичные суммы ряда образуют новую последовательность — последовательность частичных сумм:

$$S_1, S_2, S_3, \dots, S_n, \dots$$

Если существует конечный предел последовательности частичных сумм

$$\lim_{n \rightarrow \infty} S_n = S < \infty,$$

то ряд (4.34) называется **сходящимся**, а число S — **суммой** ряда. В этом случае пишут:

$$\sum_{n=1}^{\infty} u_n = S.$$

Если предел последовательности частичных сумм бесконечен или не существует, то ряд (4.34) называется **расходящимся**.

Пример. Определить сходимость ряда:

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots + \frac{1}{n(n+1)} + \dots$$

Решение. Вначале запишем частичную сумму заданного ряда:

$$S_n = \sum_{k=1}^n \frac{1}{k(k+1)} = 1 - \frac{1}{n+1}.$$

Рассмотрим предел частичных сумм:

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n+1} \right) = \lim_{n \rightarrow \infty} 1 - \lim_{n \rightarrow \infty} \frac{1}{n+1} = \lim_{n \rightarrow \infty} 1 = 1.$$

Следовательно, ряд (4.35) сходится и его сумма равна 1.

Пример. Дан числовой ряд

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} + \dots = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}},$$

исследовать сходимость ряда.

Решение.

$$S_n = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} > n \frac{1}{\sqrt{n}} = \sqrt{n}.$$

Величина $\sqrt{n}$ бесконечно возрастает с ростом n . Поэтому предел последовательности S_n при $n \rightarrow \infty$ равен бесконечности и ряд расходится.

Пример. Определить сходимость следующего ряда:

$$1 - 1 + 1 - 1 + (-1)^{n+1} + \dots$$

Решение. Четная частичная сумма этого ряда $S_{2n} = 0$, а нечетная —

$S_{2n+1} = 1$. Это означает, что предел $\lim_{n \rightarrow \infty} S_n$ не существует. Следовательно, данный ряд расходится.

4.7.1.2. Необходимое условие сходимости ряда. Для сходящихся числовых рядов всегда выполняется одно условие: его общий член стремится к нулю. Дадим строгую формулировку необходимого условия сходимости ряда.

Теорема о необходимом условии сходимости числового ряда. Если числовой ряд сходится, то его общий член при $n \rightarrow \infty$ стремится к нулю, т. е.:

$$\lim_{n \rightarrow \infty} u_n = 0. \quad (4.36)$$

Доказательство. Рассмотрим две соседние частичные суммы ряда (4.34):

$$S_{n-1} = u_1 + u_2 + u_3 + \dots + u_{n-1},$$

$$S_n = u_1 + u_2 + u_3 + \dots + u_{n-1} + u_n.$$

Из сходимости ряда следует, что:

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} S_{n-1} = S$$

С другой стороны,

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} (S_{n-1} + u_n) = \lim_{n \rightarrow \infty} S_{n-1} + \lim_{n \rightarrow \infty} u_n,$$

т. е. $S = S + \lim_{n \rightarrow \infty} u_n$,

откуда и следует (4.36).

Введенное условие сходимости является лишь необходимым, но не достаточным. Это означает, что существуют расходящиеся ряды, у которых $\lim_{n \rightarrow \infty} u_n = 0$.

Пример. Покажем, что ряд

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} + \dots$$

удовлетворяет необходимому условию сходимости ряда, но является расходящимся.

Действительно, необходимое условие выполняется, так как

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0.$$

Чтобы доказать расхождение ряда, рассмотрим его n -ю частичную сумму:

$$S_n = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} \geq \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \dots + \frac{1}{\sqrt{n}} = \frac{n}{\sqrt{n}} = \sqrt{n}.$$

Очевидно, что ряд расходится, поскольку

$$\lim_{n \rightarrow \infty} S_n \geq \lim_{n \rightarrow \infty} \sqrt{n} = \infty.$$

4.7.1.3. Основные свойства сходящихся числовых рядов.

Свойство 1. Добавление или отбрасывание конечного числа членов не изменяет сходимости ряда.

Доказательство. Пусть A — сумма отброшенных (добавленных) членов ряда, а S_n — частичная сумма исходного ряда (4.34). Тогда частичная сумма ряда с отброшенными (добавленными) членами имеет вид: $S^* = S_n \pm A$.

Поскольку A — конечное число, то

$$\lim_{n \rightarrow \infty} S^* = \lim_{n \rightarrow \infty} (S_n \pm A) = \lim_{n \rightarrow \infty} S_n \pm A.$$

Следовательно, если существует $\lim_{n \rightarrow \infty} S_n = S$, то существует и

$$\lim_{n \rightarrow \infty} S^* = S \pm A.$$

Свойство 2. Если ряд $u_1 + u_2 + u_3 + \dots + u_n + \dots$ (4.37) сходится и имеет сумму S , то ряд $cu_1 + cu_2 + \dots + cu_n + \dots$, получаемый из предыдущего умножением всех членов на одно и то же число c , также сходится и имеет сумму $c \cdot S$.

Доказательство. Рассмотрим частичную сумму ряда (4.37):

$$\sigma_n = cu_1 + cu_2 + cu_3 + \dots + cu_n = c \cdot S_n.$$

Поэтому:

$$\lim_{n \rightarrow \infty} \sigma_n = \lim_{n \rightarrow \infty} c \cdot S_n = c \cdot \lim_{n \rightarrow \infty} S_n = c \cdot S.$$

Свойство 3. Сходящиеся ряды можно почленно складывать и вычитать.

Доказательство. Пусть

$$u_1 + u_2 + u_3 + \dots + u_n + \dots = S,$$

$$v_1 + v_2 + v_3 + \dots + v_n + \dots = \Phi,$$

тогда ряд

$$(u_1 \pm v_1) + (u_2 \pm v_2) + \dots + (u_n \pm v_n) + \dots$$

также сходится и имеет сумму $S \pm \Phi$, так как

$$\lim_{n \rightarrow \infty} (S_n \pm \Phi_n) = \lim_{n \rightarrow \infty} S_n \pm \lim_{n \rightarrow \infty} \Phi_n = S \pm \Phi.$$

4.7.2. Признаки сходимости числовых рядов с положительными членами

Числовой ряд называется рядом с положительными членами или просто **положительным рядом**, если все числа $u_1, u_2, u_3, \dots, u_n \geq 0$. Рассмотрим некоторые достаточные признаки сходимости для положительных рядов.

4.7.2.1. Первый признак сравнения. Пусть даны три ряда:

— ряд, сходимость которого надо определить:

$$u_1 + u_2 + u_3 + \dots + u_n + \dots; \quad (4.38)$$

— сходящийся ряд:

$$v_1 + v_2 + v_3 + \dots + v_n + \dots; \quad (4.39)$$

— расходящийся ряд:

$$w_1 + w_2 + w_3 + \dots + w_n + \dots \quad (4.40)$$

Тогда:

а) если, начиная с некоторого номера n , выполняется условие

$$u_n \leq v_n, \quad (4.41)$$

то из сходимости ряда (4.39) следует сходимость ряда (4.38);

б) если, начиная с некоторого номера n , выполняется условие

$$u_n \geq w_n, \quad (4.42)$$

то из расходимости ряда (4.40) следует расходимость ряда (4.38).

Доказательство пункта а.

Найдем частичные суммы:

$$S_n = u_1 + u_2 + u_3 + \dots + u_n; \quad \Phi_n = v_1 + v_2 + v_3 + \dots + v_n.$$

В силу (4.41) имеем $S_n \leq \Phi_n$.

По условию задачи ряд (4.38) сходится, т. е. $\lim_{n \rightarrow \infty} \Phi_n = \Phi$, следовательно:

$$\Phi > \Phi_n \geq S_n.$$

Это означает, что последовательность частичных сумм S_n возрастает (в силу положительности ряда (4.38)) и ограничена сверху величиной Φ . Поэтому ряд (4.39) существует и конечен, а ряд (4.38) сходится.

Доказательство пункта б.

Обозначив

$$W_n = w_1 + w_2 + w_3 + \dots + w_n + \dots,$$

в силу (4.42) имеем $S_n \geq W_n$.

По условию ряд (4.40) расходится, т. е. $\lim_{n \rightarrow \infty} W_n = \infty$, следовательно

$$\lim_{n \rightarrow \infty} S_n \geq \lim_{n \rightarrow \infty} W_n = \infty$$

и ряд (4.38) расходится.

4.7.2.2. Второй признак сравнения. Пусть даны два ряда

$$\begin{aligned} u_1 + u_2 + u_3 + \dots + u_n + \dots \\ v_1 + v_2 + v_3 + \dots + v_n + \dots \end{aligned} \quad (4.44)$$

и можно указать такие постоянные числа $k_1 > 0$ и $k_2 > 0$, что, начиная с некоторого достаточно большого n ,

$$k_1 \leq \frac{u_n}{v_n} \leq k_2. \quad (4.45)$$

Тогда ряды (4.43) и (4.44) одновременно сходятся или одновременно расходятся.

Доказательство. Из (4.45) следует, что

$$k_1 v_n \leq u_n \leq k_2 v_n. \quad (4.46)$$

Если ряд (4.43) сходится, то из левого неравенства (4.46), согласно первому признаку сравнения, вытекает сходимость ряда

$$k_1 v_1 + k_1 v_2 + k_1 v_3 + \dots + k_1 v_n + \dots$$

Из сходимости же этого ряда, в соответствии со свойством 2, вытекает и сходимость ряда (4.44).

Предположим теперь, что ряд (4.43) расходится. В этом случае расходится и ряд

$$\frac{u_1}{k_2} + \frac{u_2}{k_2} + \frac{u_3}{k_2} + \dots + \frac{u_n}{k_2} + \dots$$

Из правой части (4.45) следует, что

$$\frac{u_n}{k_2} \leq v_n.$$

Следовательно, согласно первому признаку сравнения, ряд (4.44) также расходится.

Следствие (предельный признак сравнения). Если для рядов (4.43) и (4.44) выполняется условие

$$\lim_{n \rightarrow \infty} \frac{u_n}{v_n} = r < \infty, \text{ где } r \neq 0, \quad (4.47)$$

то эти ряды сходятся или расходятся одновременно.

Доказательство. Соотношение (4.47) означает, что, начиная с некоторого n , все отношения вида u_n/v_n будут близки к r . Поэтому будут выполняться, например, неравенства

$$\frac{r}{2} \leq \frac{u_n}{v_n} \leq 2r$$

откуда, согласно доказанной выше теореме, и вытекает данное следствие.

Для сравнения обычно используются следующие эталонные ряды.

Геометрический ряд

$$a + aq + aq^2 + \dots + aq^{n-1} + \dots$$

Геометрический ряд сходится при условии $0 < q < 1$. В противоположном случае ($q \geq 1$) ряд расходится. Например, ряд

$$1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{n-1}} + \dots$$

сходится ($q = 1/2$), а ряд

$$1 + 2 + 4 + \dots + 2^{n-1} + \dots$$

расходится ($q = 2$). Можно также сказать, что этот ряд расходится потому, что не выполнено необходимое условие сходимости ряда.

Обобщенным гармоническим рядом называется ряд

$$a + \frac{a}{2^p} + \frac{a}{3^p} + \dots + \frac{a}{n^p} + \dots$$

Этот ряд сходится при $p > 1$ и расходится при $p \leq 1$.

Например, ряд

$$1 + \frac{1}{2^3} + \frac{1}{3^3} + \dots + \frac{1}{n^3} + \dots \text{ сходится, а ряд}$$

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}} + \dots \text{ расходится.}$$

Обобщенный гармонический ряд при $p = 1$ называют просто гармоническим рядом:

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \dots = \sum_{n=1}^{\infty} \frac{1}{n}.$$

Гармонический ряд расходится!

Пример. Исследовать сходимость ряда

$$\frac{1}{2 \cdot 3} + \frac{2}{3 \cdot 3^2} + \frac{3}{4 \cdot 3^3} + \dots + \frac{n}{(n+1)3^n} + \dots = \sum_{n=1}^{\infty} \frac{n}{(n+1)3^n}.$$

Решение. Сравним общий член этого ряда с геометрическим

рядом $\sum_{n=1}^{\infty} \frac{1}{3^n}$, который сходится.

Так как

$$\frac{n}{(n+1)3^n} < \frac{1}{3^n},$$

то, по первому признаку сравнения, исследуемый ряд сходится.

Пример. Исследовать сходимость ряда

$$\frac{1}{\ln 2} + \frac{1}{\ln 3} + \frac{1}{\ln 4} + \dots + \frac{1}{\ln n} + \dots$$

Решение. Сравнивая общий член этого ряда с общим членом гармонического ряда

$$\frac{1}{\ln n} > \frac{1}{n},$$

закключаем, что этот ряд также расходится (по первому признаку сравнения).

Пример. Исследовать сходимость ряда

$$\frac{1}{2 + \ln 2} + \frac{1}{3 + \ln 3} + \frac{1}{4 + \ln 4} + \dots + \frac{1}{n + \ln n} + \dots = \sum_{n=2}^{\infty} \frac{1}{n + \ln n}.$$

Решение. Воспользуемся следствием из второго признака сравнения, используя гармонический ряд

$$\lim_{n \rightarrow \infty} \frac{1/(n + \ln n)}{1/n} = \lim_{n \rightarrow \infty} \frac{1}{1 + \frac{\ln n}{n}} = 1.$$

Сравниваемые ряды ведут себя одинаково, поэтому заключаем, что исследуемый ряд расходится (т. к. гармонический ряд расходится).

Пример. Исследовать сходимость ряда

$$2 \cdot 1 + \frac{1}{2 \cdot 2} + 2 \frac{1}{3} + \frac{1}{2 \cdot 4} + 2 \frac{1}{5} + \dots + \frac{2}{2n-1} + \frac{1}{2 \cdot 2n} + \dots$$

Решение. Рассмотрим отношение членов этого ряда к соответствующим членам гармонического ряда:

- при нечетном n имеем $\frac{2}{n} : \frac{1}{n} = 2$;
- при четном n имеем $\frac{1}{2n} : \frac{1}{n} = \frac{1}{2}$.

Следовательно, отношение u_n/v_n ни к какому пределу не стремится. Однако при всех n оно заключено между $1/2$ и 2 . Поэтому, согласно второму признаку сравнения, исследуемый ряд ведет себя так же, как и гармонический, т. е. расходится.

4.7.2.3. **Признак Даламбера.** Пусть дан положительный ряд

$$u_1 + u_2 + u_3 + \dots + u_n + \dots \quad (4.48)$$

Если отношения последующего члена ряда u_n к предыдущему u_{n-1} , начиная с некоторого значения $n = N$, удовлетворяют неравенству

$$\frac{u_n}{u_{n-1}} \leq q < 1, \quad n > N, \quad (4.49)$$

то ряд (4.48) сходится.

Если же, начиная с некоторого N , имеем

$$\frac{u_n}{u_{n-1}} > 1, \quad n > N, \quad (4.50)$$

то ряд (4.48) расходится.

Доказательство. Пусть имеет место соотношение (4.49), которое выполняется для всех n . Тогда:

$$u_n \leq u_{n-1}q, \quad u_{n-1} \leq u_{n-2}q, \dots, \quad u_2 \leq u_1q.$$

Отсюда, подводя почленную подстановку, получаем:

$$u_n \leq u_1 q^{n-1}.$$

Это неравенство означает, что общий член ряда (4.48) не превосходит соответствующего члена сходящегося ($q < 1$) геометрического ряда. В силу первого признака сравнения ряд (4.48) сходится.

Пусть имеет место соотношение (4.50). Тогда

$$u_1 < u_2 < u_3 < \dots < u_{n-1} < u_n < \dots,$$

т. е. члены ряда не убывают по мере возрастания n . Следовательно, не выполнено необходимое условие сходимости ряда и ряд (4.48) расходится.

На практике удобнее пользоваться предельным признаком Даламбера, формулировку которого дадим в виде следствия.

Следствие. (Предельный признак Даламбера). Если

$$\lim_{n \rightarrow \infty} \frac{u_n}{u_{n-1}} = p$$

то при $p < 1$ ряд (4.48) сходится, при $p > 1$ этот ряд расходится.

Пример. Исследовать сходимость ряда

$$\frac{2}{1!} + \frac{4}{2!} + \frac{8}{3!} + \dots + \frac{2^n}{n!} + \dots = \sum_{n=1}^{\infty} \frac{2^n}{n!}.$$

Решение. Рассмотрим предел отношения:

$$\lim_{n \rightarrow \infty} \frac{u_n}{u_{n-1}} = \lim_{n \rightarrow \infty} \frac{2^n (n-1)!}{2^{n-1} n!} = \lim_{n \rightarrow \infty} \frac{2}{n} = 0 < 1.$$

Следовательно, исследуемый ряд сходится.

Замечание. Если $\lim_{n \rightarrow \infty} \frac{u_n}{u_{n-1}} = 1$, то признак Даламбера не дает

ответа на вопрос о сходимости ряда. В этих случаях надо привлекать другие признаки сходимости ряда.

4.7.3. Знакопередающие ряды

Знакопередающими рядами называются ряды вида

$$u_1 - u_2 + u_3 - u_4 + \dots \dots, \quad (4.51)$$

где все $u_n > 0$.

Сходимость таких рядов исследуется по теореме Лейбница: если в знакопередающемся ряде $u_1 - u_2 + u_3 - u_4 + \dots \dots$, $u_n > 0$ все члены таковы, что $u_1 > u_2 > u_3 > u_4 \dots$ и $\lim_{n \rightarrow \infty} u_n = 0$, то ряд (4.51) сходится, его сумма положительна и не превосходит первого члена ряда u_1 .

Доказательство. Возьмем сумму четного числа первых членов S_{2m} , т. е. $n = 2m$, которая положительна.

$$S_{2m} = (u_1 - u_2) + (u_3 - u_4) + \dots \dots + (u_{2m-1} - u_{2m}) > 0,$$

так как выражение в каждой скобке больше нуля.

S_{2m} возрастает при росте m , т. к.

$$S_{2m} = S_{2(m-1)} + (u_{2m-1} - u_{2m}) > S_{2(m-1)}.$$

С другой стороны:

$$S_{2m} = u_1 - (u_2 - u_3) - (u_4 - u_5) \dots \dots - (u_{2m-2} - u_{2m-1}) - u_{2m} < u_1.$$

т. е. при росте m S_{2m} возрастает и ограничена сверху,

следовательно, имеет предел $S = \lim_{m \rightarrow \infty} S_{2m}$.

Нечетные суммы $n = 2m + 1$ будут иметь тот же предел; т. к.

$$S_{2m+1} = S_{2m} + u_{2m+1}, \text{ то}$$

$$\lim_{m \rightarrow \infty} S_{2m+1} = \lim_{m \rightarrow \infty} S_{2m} + \lim_{m \rightarrow \infty} u_{2m+1} = S + 0 = S.$$

Четные и нечетные суммы ряда имеют тот же предел, следовательно, ряд сходится. Теорема доказана.

По знакопередающемуся ряду можно построить соответствующий ему положительный ряд $u_1 + u_2 + u_3 + u_4 + u_5 + \dots$. Если такой положительный ряд сходится, то знакопередающийся ряд называют **абсолютно сходящимся**, в противном случае ряд называют **условно сходящимся**. В абсолютно сходящемся ряде члены ряда можно переставлять без потери сходимости; в условно сходящемся ряде перестановка членов ряда запрещена, т. к. она может привести к потере сходимости.

4.7.4. Степенные ряды

4.7.4.1. Основные определения. **Функциональным рядом** называется ряд, членами которого являются функции от аргумента x :

$$u_1(x) + u_2(x) + u_3(x) + \dots + u_n(x) + \dots = \sum_{n=1}^{\infty} u_n(x). \quad (4.52)$$

Если в членах ряда (4.52) зафиксировать значение аргумента $x = x_0$, то получим числовой ряд

$$u_1(x_0) + u_2(x_0) + u_3(x_0) + u_n(x_0) + \dots \quad (4.53)$$

Если при $x = x_0$ числовой ряд (4.53) сходится, то x_0 называется точкой сходимости ряда (4.52).

Областью сходимости функционального ряда называется множество всех точек сходимости этого ряда.

Если значение x_0 принадлежит области сходимости ряда (4.52), то можно говорить о сумме этого ряда в точке $x = x_0$:

$$u_1(x_0) + u_2(x_0) + u_3(x_0) + u_n(x_0) + \dots = S(x_0).$$

Таким образом, значение суммы функционального ряда зависит от значения x_0 переменной x , т. е. сумма функционального ряда сама является функцией переменной x .

Степенным рядом по степеням x называется функциональный ряд вида

$$C_0 + C_1 x + C_2 x^2 + \dots + C_n x^n + \dots = \sum_{n=0}^{\infty} C_n x^n, \quad (4.54)$$

где $C_0, C_1, \dots, C_n, \dots$ не зависят от переменной x и называются коэффициентами этого ряда.

Степенной ряд (4.54) всегда сходится, по крайней мере в точке $x = 0$. При любых конкретных $x = x_0$ ряд (4.54) превращается в числовой ряд

$$C_0 + C_1 x_0 + C_2 x_0^2 + \dots + C_n x_0^n + \dots \quad (4.55)$$

Степенной ряд (4.54) сходится в точке x_0 абсолютно, если сходится ряд, образованный из модулей членов числового ряда (4.55):

$$|C_0| + |C_1 x_0| + |C_2 x_0^2| + \dots + |C_n x_0^n| + \dots$$

4.7.4.2. Нахождение области сходимости степенного ряда. Найдем область сходимости ряда (4.54), используя признак Даламбера для положительных числовых рядов. По этому

признаку ряд (4.55) сходится, если:

$$\lim_{n \rightarrow \infty} \left| \frac{C_n x_0^n}{C_{n-1} x_0^{n-1}} \right| = |x_0| \cdot \lim_{n \rightarrow \infty} \left| \frac{C_n}{C_{n-1}} \right| < 1.$$

Следовательно, по признаку Даламбера ряд (4.54) заведомо сходится при

$$|x| < \frac{1}{\lim_{n \rightarrow \infty} \left| \frac{C_n}{C_{n-1}} \right|} = \lim_{n \rightarrow \infty} \frac{|C_{n-1}|}{|C_n|}$$

и расходится при

$$|x| > \frac{1}{\lim_{n \rightarrow \infty} \left| \frac{C_n}{C_{n-1}} \right|} = \lim_{n \rightarrow \infty} \frac{|C_{n-1}|}{|C_n|}.$$

Величина

$$R = \lim_{n \rightarrow \infty} \frac{|C_{n-1}|}{|C_n|} \quad (4.56)$$

называется **радиусом сходимости** ряда (4.54). Ряд заведомо сходится в интервале $|x| < R$ или $-R < x < R$, который называется интервалом сходимости.

Признак Даламбера ничего не говорит о сходимости ряда в точках $x = \pm R$. В этих точках сходимость ряда исследовать отдельно.

Исследовать степенной ряд на сходимость означает найти его интервал сходимости и установить сходимость или расходимость ряда в граничных точках интервала, т. е. при $x = R$ и $x = -R$.

Пример. Исследовать на сходимость степенной ряд

$$\frac{x}{2} + \frac{4x^2}{2^2} + \frac{9x^3}{2^3} + \dots + \frac{n^2 x^n}{2^n} + \dots = \sum_{n=0}^{\infty} \frac{(nx)^2}{2^n}.$$

Решение. Используя формулу (4.56), имеем:

$$R = \lim_{n \rightarrow \infty} \frac{|C_{n-1}|}{|C_n|} = \lim_{n \rightarrow \infty} \frac{(n-1)^2 2^n}{2^{n-1} n^2} = 2.$$

Интервал сходимости данного ряда характеризуется неравенством $|x| < 2$. Исследуем сходимость ряда в граничных точках $x = \pm 2$. Очевидно, что:

$$\sum_{n=0}^{\infty} \frac{n^2 (\pm 2)^n}{2^n} = \sum_{n=0}^{\infty} (\pm 1)^n n^2.$$

Оба ряда при $x = 2$ и $x = -2$ расходятся, т. к. не выполняется необходимое условие сходимости числовых рядов. Следовательно, область сходимости данного степенного ряда совпадает с интервалом сходимости.

Пример. Найти область сходимости следующего ряда:

$$1 + x + 2^2 x^2 + 3^3 x^3 + \dots + n^n x^n + \dots = 1 + \sum_{n=1}^{\infty} n^n x^n.$$

Решение. По формуле (4.56) найдем

$$R = \lim_{n \rightarrow \infty} \frac{(n-1)^{n-1}}{n^n} = 0.$$

Следовательно, ряд сходится только в одной точке $x = 0$.

Пример. Найти область сходимости следующего ряда:

$$1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Решение. Так как

$$R = \lim_{n \rightarrow \infty} \frac{n!}{(n-1)!} = \lim_{n \rightarrow \infty} n = \infty,$$

то ряд сходится при всех конечных значениях x , т. е. $-\infty < x < \infty$.

4.7.4.3. Основные свойства степенных рядов.

1) Во всех точках, лежащих внутри интервала сходимости, сумма степенного ряда является непрерывной функцией переменной x :

$$\sum_{n=0}^{\infty} C_n x^n = S(x)$$

2) Степенной ряд можно почленно интегрировать внутри интервала сходимости:

$$\sum_{n=0}^{\infty} \int_{x_1}^{x_2} C_n x^n dx = \int_{x_1}^{x_2} S(x) dx.$$

3) Внутри интервала сходимости степенной ряд можно почленно дифференцировать:

$$\sum_{n=0}^{\infty} C_n (x_n)' = S'(x).$$

При почленном интегрировании и дифференцировании степенных рядов их интервалы сходимости не меняются.

Пример. Найти сумму ряда

$$S(x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots + \frac{x^n}{n} + \dots \quad (4.57)$$

Решение. Найдем сначала интервал сходимости этого ряда n :

$$R = \lim_{n \rightarrow \infty} \frac{n}{n-1} = 1.$$

Следовательно, интервал сходимости ряда $(-1, +1)$. Продифференцировав (4.57), имеем:

$$S'(x) = 1 + x + x^2 + \dots + x^n + \dots$$

Правая часть этого выражения — геометрический ряд с $q = x$, который сходится при $|x| < 1$. Поэтому, используя формулу суммы сходящейся геометрической прогрессии, получим:

$$S'(x) = \frac{1}{1-x}.$$

Отсюда сумму исходного ряда найдем интегрированием:

$$S(x) = \int \frac{dx}{1-x} = -\ln(1-x) + C.$$

Найдем C . Из (4.57) следует, что $S(0) = 0$. Следовательно,

$$0 = -\ln(1-0) + C, \quad C = 0.$$

Таким образом, $S(x) = -\ln(1-x) = \ln \frac{1}{1-x}$.

4.7.3.4. Ряды по степеням $(x-a)$. Наряду со степенными рядами относительно переменной x часто рассматривают степенные ряды по переменной $(x-a)$, т. е. ряды вида:

$$C_0 + C_1(x-a) + C_2(x-a)^2 + \dots + C_n(x-a)^n + \dots \quad (4.58)$$

Очевидно, что этот ряд подстановкой $y = (x-a)$ превращается в ряд типа (4.54). Поэтому, если степенной ряд (4.54) имеет интервал сходимости $R < x < R$, то соответствующий ряд вида (4.58) имеет интервал сходимости $(a-R) < x < (a+R)$, центр которого расположен в точке $x = a$.

Пример. Найти интервал сходимости следующего ряда:

$$(x^3-1) + \frac{(x^3-1)^2}{2} + \frac{(x^3-1)^3}{3} + \dots + \frac{(x^3-1)^n}{n} + \dots = \sum_{n=1}^{\infty} \frac{(x^3-1)^n}{n}.$$

Решение. Найдем радиус сходимости ряда. В соответствии с (4.56) имеем:

$$R = \lim_{n \rightarrow \infty} \left| \frac{C_{n-1}}{C_n} \right| = \lim_{n \rightarrow \infty} \frac{n}{n-1} = 1.$$

Отсюда $-1 < (x^3-1) < 1$, или $0 < x^3 < 2$.

Следовательно, интервал сходимости ряда $0 < x < \sqrt[3]{2}$.

4.7.3.5. Ряд Тейлора. Пусть функция $f(x)$ в точке $x = a$ имеет производные любого порядка. Предположим, что имеется сходящийся степенной ряд

$$\begin{aligned} C_0 + C_1(x-a) + C_2(x-a)^2 + \dots + C_n(x-a)^n + \dots = \\ = \sum_{n=0}^{\infty} C_n(x-a)^n, \end{aligned} \quad (4.59)$$

сумма которого равна функции $f(x)$, т. е.

$$S(x) = \sum_{n=0}^{\infty} C_n(x-a)^n = f(x). \quad (4.60)$$

Найдем коэффициенты такого ряда. Очевидно, что $f(a) = C_0$. Продифференцировав (4.60) в точке $x = a$, имеем $C_1 = f'(a)$. Продифференцировав (4.60) в точке $x = a$ дважды, получим $C_2 = f''(a)/2$. Продолжая дифференцирование равенства (4.60) в точке $x = a$, можно убедиться, что коэффициенты ряда (4.59) находятся по формуле

$$C_n = \frac{f^{(n)}(a)}{n!}.$$

Степенной ряд вида

$$f(a) + \frac{f'(a)}{1}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \dots$$

называется **рядом Тейлора** для функции $f(x)$.

Теорема. Для того, чтобы бесконечно дифференцируемая функция $f(x)$ являлась суммой составленного для нее ряда Тейлора, необходимо и достаточно, чтобы остаточный член в формуле Тейлора (4.15) для $f(x)$ стремился к нулю при $n \rightarrow \infty$.

Если для всех значений x из некоторой окрестности точки $x = a$ имеет место равенство

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \dots,$$

то функция $f(x)$ называется **разложимой** в ряд Тейлора в окрестности точки $x = a$ (или по степеням $(x - a)$).

В частном случае при $a = 0$ ряд Тейлора имеет вид

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + \dots$$

и его называют рядом Маклорена.

Пример. Разложить в ряд Тейлора функцию $f(x) = e^x$.

Поскольку $e^x = (e^x)' = (e^x)'' = \dots = (e^x)^{(n)}$, то при $x = 0$ для $f(x) = e^x$ имеем:

$$f(0) = 1; f'(0) = 1; f''(0) = 1; \dots; f^{(n)}(0) = 1; \dots$$

Следовательно, ряд Тейлора функции $y = e^x$ в окрестности точки $x = 0$ имеет вид:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \quad (4.61)$$

Ряд (4.61) сходится на всей числовой оси к функции $y = e^x$.

Пример. Разложить в ряд Тейлора функцию $f(x) = \sin x$.

Решение. Для функции $f(x) = \sin x$ имеем:

$$f'(x) = \cos x = \sin\left(x + \frac{\pi}{2}\right);$$

$$f''(x) = -\sin x = \sin\left(x + \pi\right);$$

$$\dots$$

$$f^{(n)}(x) = \sin\left(x + n \frac{\pi}{2}\right).$$

Следовательно, ряд Тейлора для $\sin x$:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \dots;$$

или

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

Аналогично получается разложение для функции $\cos x$:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots + (-1)^{n+1} \frac{x^{2n}}{(2n)!} + \dots,$$

или

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}.$$

Подобным образом можно получить разложения в ряд Тейлора и многих других функций.

Глава 4.8. Дифференциальные уравнения

4.8.1. Основные понятия теории дифференциальных уравнений

Пусть x — независимая переменная, $y(x)$ — неизвестная функция этой переменной. $y', y'', \dots, y^{(n)}$ — производные неизвестной функции. Уравнение, связывающее независимую переменную x с функцией $y(x)$ и ее производными до порядка n включительно, называется **обыкновенным дифференциальным уравнением**:

$$F(x, y, y', y'', \dots, y^{(n)}) = 0.$$

Порядком дифференциального уравнения называется наивысший порядок производной, входящей в уравнение.

Дифференциальное уравнение n -го порядка может не содержать некоторые из величин $x, y, y', \dots, y^{(n-1)}$ или даже все эти величины, но оно обязательно содержит n -ю производную $y^{(n)}$.

Пример. $y' + 2y = 0$ — уравнение 1-го порядка, так как наивысший порядок производной равен единице.

Пример. $y^{(4)} - y' = 0$ — уравнение 4-го порядка: входят производные 1-го и 4-го порядков, наивысший порядок производной равен 4.

Решение дифференциального уравнения — это функция $y = y_0(x)$, которая, будучи подставлена в уравнение, обращает его в тождество:

$$F(x, y_0(x), y_0'(x), \dots, y_0^{(n)}(x)) \equiv 0.$$

Пример. Пусть дано уравнение $y'' + y = 0$. Покажем, что функция $y = \sin x$ является решением этого уравнения.

Имеем $y' = (\sin x)' = \cos x$, $y'' = (\cos x)' = -\sin x$. Подставим в уравнение вместо y и y' функции $\sin x$ и $-\sin x$:

$$-\sin x + \sin x \equiv 0.$$

Покажем, что функция $y = C_1 \cos x + C_2 \sin x$, где C_1 и C_2 — произвольные постоянные, также является решением данного уравнения. Имеем:

$$y' = -C_1 \sin x + C_2 \cos x; y'' = -C_1 \cos x - C_2 \sin x.$$

Подставим в уравнение выражения y и y'' :

$$-C_1 \cos x - C_2 \sin x + (C_1 \cos x + C_2 \sin x) \equiv 0.$$

График решения $y = y(x)$ называется **интегральной кривой**. Задача нахождения решений дифференциального уравнения называется задачей **интегрирования дифференциального уравнения**.

Рассмотрим уравнение n -го порядка, разрешенное относительно старшей производной:

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)}). \quad (4.62)$$

Предполагаем, что функция f определена, однозначна, непрерывна и дифференцируема в некоторой области изменения своих аргументов. Задача нахождения решения $y = y(x)$, удовлетворяющего заданным начальным условиям — при $x = x_0$ $y = y_0, y' = y_0', \dots, y_0^{(n-1)} = y_0^{(n-1)}$, (4.63)

где $x_0, y_0, y_0', \dots, y_0^{(n-1)}$ суть заданные числа (начальные данные), — называется задачей Коши.

Начальные условия можно записать и так:

$$y|_{x=x_0} = y_0, y'|_{x=x_0} = y_0', \dots, y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}.$$

Дадим определения общего и частного решений уравнения n -го порядка $y^{(n)} = f(x, y, y', \dots, y^{(n-1)})$, правая часть которого есть функция определенная, непрерывная и дифференцируемая в некоторой области G изменения переменных $x, y, y', \dots, y^{(n-1)}$.

Функция $y = \varphi(x, C_1, C_2, \dots, C_n)$, (4.64)
зависящая от аргумента x и n произвольных постоянных $C_1, C_2, \dots, C_n$, называется **общим решением уравнения (4.62)** в области G , если она удовлетворяет двум условиям:

1) функция (4.64) является решением уравнения (4.62) при любых значениях произвольных постоянных $C_1, C_2, \dots, C_n$;

2) каковы бы ни были начальные условия (4.63), существует единственный набор постоянных $C_1^0, C_2^0, \dots, C_n^0$, такой, что функция $y = \varphi(x, C_1^0, C_2^0, \dots, C_n^0)$ является решением уравнения (4.62) и удовлетворяет начальным условиям (4.63).

Чтобы найти решение уравнения (4.62) с начальными данными $x_0, y_0, y_0', \dots, y_0^{(n-1)}$ из области G , если известно общее решение (4.64) поступают следующим образом:

1) составляют систему уравнений:

$$\begin{cases} y_0 = \varphi(x_0, C_1, C_2, \dots, C_n) \\ y_0' = \varphi'(x_0, C_1, C_2, \dots, C_n) \\ \dots \\ y_0^{(n-1)} = \varphi^{(n-1)}(x_0, C_1, C_2, \dots, C_n); \end{cases} \quad (4.65)$$

2) решая систему (4.65), находят $C_1^0, C_2^0, \dots, C_n^0$;

3) подставляют найденные значения произвольных постоянных в общее решение (4.64) и получают искомое решение:

$$y = \varphi(x, C_1^0, C_2^0, \dots, C_n^0),$$

которое является искомым единственным решением задачи.

Если общее решение уравнения (4.62) задано в неявном виде

$$\Phi(x, y, C_1, C_2, \dots, C_n) = 0, \quad (4.66)$$

то оно называется **общим интегралом** дифференциального уравнения.

Всякое решение, получаемое из общего решения (4.64) при конкретных значениях постоянных $C_1 = C_1^0, C_2 = C_2^0, \dots, C_n = C_n^0$, называется **частным решением** уравнения (4.62).

Пример. Дано уравнение $y'' + y = 0$. Найти его частное решение, удовлетворяющее начальным условиям:

$$y|_{x=0} = 0, \quad y'|_{x=0} = 1.$$

Решение. Выше было показано, что функция $y = C_1 \cos(x) + C_2 \sin(x)$, где C_1, C_2 — произвольные постоянные, является решением данного уравнения. Это общее решение. Для нахождения частного решения используем начальные условия: $x_0 = 0, y_0 = 0, y_0' = 1$.

Составим систему типа (4.65):

$$\begin{cases} y = C_1 \cos x + C_2 \sin x; \\ y' = -C_1 \sin x + C_2 \cos x. \end{cases}$$

Для заданных начальных условий имеем:

$$\begin{cases} 0 = C_1 \cos(0) + C_2 \sin(0) \\ 1 = -C_1 \sin(0) + C_2 \cos(0) \end{cases} \Rightarrow \begin{cases} 0 = C_1; \\ 1 = C_2; \end{cases} \Rightarrow \begin{cases} C_1^0 = 0; \\ C_2^0 = 1. \end{cases}$$

Найденные значения C_1^0, C_2^0 подставим в общее решение: $y = C_1^0 \cos x + C_2^0 \sin x \Rightarrow y = \sin x$.

Итак, искомое частное решение уравнения — $y = \sin(x)$.

4.8.2. Дифференциальные уравнения первого порядка

Решение дифференциальных уравнений — одна из важнейших и сложных задач математики. Большинство дифференциальных

уравнений решаются только численными методами при помощи вычислительной техники. Здесь рассмотрим лишь наиболее простые уравнения первого порядка, которые могут быть решены аналитически.

4.8.2.1. Основные определения. Дифференциальное уравнение первого порядка — это уравнение, в которое входят независимая переменная, неизвестная функция и первая производная этой функции. Общий вид дифференциального уравнения первого порядка:

$$F(x, y, y') = 0. \quad (4.67)$$

Здесь F — заданная функция трех аргументов. Она может не зависеть от x или y (или от обеих переменных), но должна содержать y' .

Если уравнение (4.67) разрешить относительно y' , то получим:

$$y' = f(x, y), \quad (4.68)$$

где f — заданная функция от x и y .

Решение дифференциального уравнения (4.68) — это функция $y = y_0(x)$, которая, будучи подставлена в это уравнение, обращает его в тождество:

$$y_0'(x) \equiv f(x, y_0).$$

Пример. Дано уравнение $y' + y \operatorname{ctg} x - 2 \cos x = 0$.

Покажем, что функция $y = \sin x$ является его решением. Для этого подставим в данное уравнение вместо y и y' функции $\sin x$ и $(\sin x)' = \cos x$. Имеем:

$$\cos x + \sin x \operatorname{ctg} x - 2 \cos x = \cos x + \cos x - 2 \cos x \equiv 0.$$

Уравнение обратилось в тождество.

Функция $y = \varphi(x, C)$ называется общим решением уравнения (4.68), если она является решением этого уравнения при всех значениях произвольной постоянной C .

Если общее решение задано в неявном виде $\varphi(x, y, C) = 0$, то оно называется общим интегралом. Частное решение уравнения (4.68) — это решение, которое получается из общего (4.69) при конкретном значении C .

Для дифференциального уравнения (4.68) задача Коши формулируется так: среди всех решений уравнения (4.68) найти решение $y = y(x)$, удовлетворяющее условию:

$$y|_{x=x_0} = y_0, \quad (4.70)$$

где x_0, y_0 — заданные числа.

Условие (4.70) называется начальным условием, а числа x_0, y_0 — начальными значениями.

4.8.2.2. Решение дифференциальных уравнений с разделяющимися переменными. Уравнение с разделяющимися переменными — это уравнение, правая часть которого $f(x, y)$ есть произведение двух сомножителей $f(x)$ и $g(y)$, каждый из которых зависит только от одной переменной:

$$y' = f(x)g(y). \quad (4.71)$$

Уравнения с разделяющимися переменными интегрируются следующим образом: y' заменим на $\frac{dy}{dx}$, затем умножим обе

части (4.71) на $\frac{dx}{g(y)}$.
Получим:

$$\frac{dy}{dx} = f(x)g(y) \Rightarrow \frac{dy}{g(y)} = f(x)dx. \quad (4.72)$$

Дифференциалы переменных x и y , и соответствующие функции стоят отдельно, т. е. переменные отделены.

Если обозначить

$$G(y) = \int \frac{dy}{g(y)}, \quad F(x) = \int f(x)dx,$$

то уравнение (4.72) можно переписать в виде:

$$dG(y) = dF(x).$$

Так как из равенства дифференциалов двух функций следует, что сами функции отличаются на произвольное постоянное слагаемое, то:

$$G(y) = F(x) + C,$$

или

$$\int \frac{dy}{g(y)} = \int f(x)dx + C. \quad (4.73)$$

Выражение (4.73) представляет собой общий интеграл уравнения (4.71). При этом нужно помнить, что решение уравнения

y — это функция от x , т. е. $y = y(x)$. Вычислив интегралы в (4.73), получим решение исходного уравнения.

Пример. Решить уравнение $xy' + y = 0$.

Решение. Разрешим уравнение относительно y' : $y' = -\frac{y}{x}$

Здесь $f(x) = -1/x$, а $g(y) = y$.

Заменяем в этом уравнении y на dy/dx и умножим обе части уравнения на dx/y :

$$\frac{dy}{dx} = -\frac{y}{x} \Rightarrow \frac{dy}{y} = -\frac{dx}{x}.$$

Получили уравнение с разделенными переменными. Интегрируя, находим:

$$\int \frac{dy}{y} = -\int \frac{dx}{x} + C_1,$$

где C_1 — произвольная постоянная.

Отсюда следует: $\ln |y| = -\ln |x| + C_1$.

В данном случае удобно вместо C_1 написать $\ln C_2$ ($C_2 > 0$).

Тогда $\ln |y| = -\ln |x| + \ln C_2$,

или

$$|y| = \frac{C_2}{|x|} \Rightarrow y = \pm \frac{C_2}{x}.$$

Так как $\pm C_2$ принимает любые значения, то обозначая $\pm C_2 = C$, окончательно получим:

$$y = \frac{C}{x},$$

где C — произвольная постоянная.

Эта формула и дает общее решение заданного уравнения. Найдем теперь частное решение, удовлетворяющее начальному условию $y|_{x=1} = 1/2$. Для этого в равенство $y = C/x$ подставим вместо x и y значения 1 и $1/2$. Получим $1/2 = C/1$. Отсюда следует, что $C = 1/2$. Таким образом, искомое частное решение имеет вид: $y = 1/(2x)$.

Если уравнение задано в виде $M(x) \cdot N(y) \cdot dx + P(x) \cdot Q(y) \cdot dy = 0$, то для отделения переменных достаточно разделить обе части на произведение $N(y)P(x)$:

$$\frac{M(x)}{P(x)} dx + \frac{Q(y)}{N(y)} dy = 0.$$

Отсюда получим интеграл:

$$\int \frac{M(x)}{P(x)} dx + \int \frac{Q(y)}{N(y)} dy = C.$$

Пример. Найти общее решение уравнения $(x + xy^2)dx + (y + x^2y)dy = 0$.

Решение. Для отделения переменных представим уравнение в виде:

$$x(1 + y^2)dx + y(1 + x^2)dy = 0.$$

Разделим обе части уравнения на $(1+y^2)(1+x^2)$:

$$\frac{x}{1+x^2} dx + \frac{y}{1+y^2} dy = 0.$$

Переменные разделились. Интегрируем это уравнение:

$$\int \frac{x dx}{1+x^2} + \int \frac{y dy}{1+y^2} = C.$$

Вычислим неопределенный интеграл $\int \frac{x dx}{1+x^2}$, сделав замену $1+x^2 = t$, $dt = 2x dx$, т. е. $x dx = dt/2$:

$$\int \frac{x dx}{1+x^2} = \int \frac{dt}{2t} = \frac{1}{2} \ln |t| = \frac{1}{2} \ln |1+x^2| = \frac{1}{2} \ln(1+x^2).$$

(Вместо $|1+x^2|$ можно написать $1+x^2$, так как $(1+x^2) > 0$; аналогично и для $(1+y^2)$).

Аналогично вычисляется $\int \frac{y dy}{1+y^2}$. Кроме того, вместо C удобнее взять $\ln |C|$. В итоге получаем:

$$\frac{1}{2} \ln(1+x^2) + \frac{1}{2} \ln(1+y^2) = \ln |C|.$$

или $\ln(1+x^2) + \ln(1+y^2) = 2 \ln |C|$.

Так как $2 \ln C = \ln C^2$, то окончательно находим общий интеграл уравнения в виде:

$$(1+x^2)(1+y^2) = C^2.$$

4.8.2.3. Решение линейных дифференциальных уравнений первого порядка. Линейным дифференциальным уравнением первого порядка называется уравнение, линейное относительно искомой функции $y(x)$ и ее производной $y'(x)$. В общем случае оно имеет вид:

$$A(x)y' + B(x)y + C(x) = 0,$$

где коэффициенты A, B, C — заданные непрерывные функции от x .

Предполагая, что в некотором интервале изменения x коэффициент $A(x)$ не обращается в нуль, делим обе части этого уравнения на $A(x)$:

$$y' + \frac{B(x)}{A(x)}y + \frac{C(x)}{A(x)} = 0.$$

Обозначая

$$\frac{B(x)}{A(x)} = p(x), \quad \frac{C(x)}{A(x)} = -f(x),$$

запишем уравнения в виде:

$$y' + p(x)y = f(x). \quad (4.74)$$

Если $f(x) \equiv 0$, то уравнение называется линейным уравнением без свободного члена (правой части) или линейным однородным уравнением. Итак, $y' + p(x)y = 0$ — линейное однородное уравнение (оно же уравнение с разделяющимися переменными).

Если $f(x) \neq 0$, то уравнение (4.74) называется линейным неоднородным уравнением.

Например, уравнение $y' - y \cos^2 x = x^2$ является линейным неоднородным уравнением. Однородное по отношению к нему будет уравнение $y' - y \cos^2 x = 0$. Уравнение (4.74) можно интегрировать разными методами. Мы рассмотрим метод Бернулли. Он состоит в следующем. В уравнении (4.74) делаем замену:

$$y = u(x) \cdot v(x). \quad (4.75)$$

Дифференцируя по правилу дифференцирования произведения двух функций, имеем:

$$y' = u'(x) \cdot v(x) + u(x) \cdot v'(x). \quad (4.76)$$

Подставим в уравнение (4.74) вместо y и y' их выражения из (4.75) и (4.76), получим:

$$u'(x)v(x) + u(x)v'(x) + p(x)u(x)v(x) = f(x),$$

или

$$u'v + u[v' + p(x)v] = f(x). \quad (4.77)$$

Так как одну из функций в (4.75) можно выбрать произвольно, то функцию v выберем таким образом, чтобы коэффициент при u обратился в нуль, т. е.:

$$v' + p(x)v = 0. \quad (4.78)$$

Уравнение (4.78) относительно функции $v(x)$ является уравнением с разделяющимися переменными (и линейным однородным). Поэтому из (4.78) имеем:

$$\frac{dv}{dx} + pv = 0 \Rightarrow \frac{dv}{dx} = -pv \Rightarrow \frac{dv}{v} = -pdx.$$

Интегрируя, находим

$$\int \frac{dv}{v} = -\int pdx + C \Rightarrow \ln(v) = -\int pdx + C.$$

Так как функция v — это любая функция, удовлетворяющая (4.78), то полагаем $C = 0$.

Итак,

$$\ln(v) = -\int pdx \Rightarrow v = e^{-\int pdx}.$$

Представляя найденную функцию $v(x)$ в уравнение (4.77), получим:

$$u'(x) \cdot e^{-\int pdx} = f(x).$$

Отсюда следует:

$$u' = e^{\int pdx} \cdot f(x) \Rightarrow du = e^{\int pdx} \cdot f(x)dx.$$

Интегрируя, получим:

$$u = \int e^{\int pdx} \cdot f(x)dx + C,$$

где C — произвольная постоянная. Для того, чтобы найти $y(x)$, умножим найденную $u(x)$ на $v(x)$:

$$y = e^{-\int pdx} \left(\int e^{\int pdx} \cdot f(x)dx + C \right) \quad (4.79)$$

Формула (4.79) дает общее решение дифференциального уравнения (4.74).

Задача. Найти общее решение уравнения

$$y' - \frac{2xy}{x^2 + 3} = (x^2 + 3) \cos x.$$

Решение. Это линейное неоднородное уравнение, где:

$$p(x) = -\frac{2x}{x^2 + 3}, \quad f(x) = (x^2 + 3) \cos x.$$

Выполнив замену $y = u \cdot v$, получаем $y' = u'v + uv'$.

Заданное дифференциальное уравнение перепишем в виде

$$u'v + uv' - \frac{2x}{x^2 + 3} uv = (x^2 + 3) \cos x,$$

или

$$u'v + u \left[v' - \frac{2x}{x^2 + 3} v \right] = (x^2 + 3) \cos x.$$

Приравняем выражение в скобках нулю:

$$v' - \frac{2x}{x^2 + 3} v = 0.$$

Получили уравнение для функции $v(x)$ — уравнение с разделяющимися переменными. Интегрируем его:

$$\begin{aligned} \frac{dv}{dx} - \frac{2x}{x^2 + 3} v = 0 &\Rightarrow \frac{dv}{v} = \frac{2x}{x^2 + 3} dx \Rightarrow \int \frac{dv}{v} = \int \frac{2x}{x^2 + 3} dx \Rightarrow \\ &\Rightarrow \ln(v) = \int \frac{2x dx}{x^2 + 3} \Rightarrow \ln(v) = \ln(x^2 + 3) \Rightarrow v = x^2 + 3. \end{aligned}$$

Подставляя функцию $v(x)$ в уравнение, найдем уравнение для функции $u(x)$:

$$u'(x^2 + 3) = (x^2 + 3) \cos x.$$

Отсюда следует:

$$u' = \cos(x) \text{ или } u = \int \cos(x) dx + C \Rightarrow u = \sin(x) + C.$$

Теперь находим общее решение заданного уравнения $y(x)$:

$$y = uv \Rightarrow y = (\sin(x) + C)(x^2 + 3).$$

Более подробно с дифференциальными уравнениями можно познакомиться, например, в [15]. Для практических самостоятельных занятий по материалу разделов 3 и 4 рекомендуется [16–32].

4.9. Задачи и упражнения к разделу 4

1. Найти производные функции:

а) $f(x) = 5(\operatorname{tg} x - x)$;

б) $f(x) = \frac{1}{e^x + 1}$

в) $f(x) = (2x^3 + 5)^4$;

г) $f(x) = \operatorname{tg} \ln x$;

д) $f(x) = \ln(x + \sqrt{x^2 + 1})$.

2. Найти производные второго порядка от функций:

а) $f(x) = x^2 \times (2 \ln x - 3)$;

б) $f(x) = x \sin 3x - 2 \cos 3x$;

в) $f(x) = (2x + 3)^3 \cdot \sqrt{2x + 3}$.

3. Найти дифференциал функций:

а) $f(x) = \ln \frac{x-6}{x+6}$;

б) $f(x) = \operatorname{arctg}(x^2)$;

в) $f(x) = \ln(x + \sqrt{x^2 + 4})$.

4. Найти интегралы:

а) $\int x \cdot \sqrt{x} \cdot dx$

б) $\int (2 \operatorname{tg} x + 3 \operatorname{ctg} x)^2$

в) $\int \sqrt{\sin x} \cdot \cos x \cdot dx$.

5. Вычислить определенные интегралы:

а) $\int_0^1 x \cdot e^{-x} \cdot dx$;

б) $\int_1^e \frac{\ln^2 x}{x} dx$;

в) $\int_0^1 \sqrt{r^2 - x^2} dx$.

6. Вычислить несобственные интегралы:

а) $\int_0^{+\infty} x e^{-x} dx$;

б) $\int_{-\infty}^0 \frac{dx}{4 + x^2}$;

в) $\int_{-1}^1 \frac{dx}{\sqrt[3]{x^2}}$.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ

Раздел 1. МАТЕМАТИКА СЛУЧАЙНОГО

7. Определить сходимость ряда

а) $\frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{(n+1)^2} + \dots;$

б) $\ln(1+1) + \ln\left(1+\frac{1}{2}\right) + \ln\left(1+\frac{1}{3}\right) + \dots + \ln\left(1+\frac{1}{n}\right) + \dots;$

в) $1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} + \dots$

8. Определить область сходимости ряда

а) $1 + \frac{x}{1^2} + \frac{x^2}{2^2} + \dots + \frac{x^n}{n^2} + \dots;$

б) $x + \frac{x}{3!} + \frac{x}{5!} + \dots + \frac{x}{(2n+1)!} + \dots;$

в) $x + (2x)^2 + (3x)^3 + \dots + (nx)^n + \dots;$

9. Разложить в ряд Маклорена следующие функции:

а) $f(x) = \ln(1+x);$

б) $f(x) = (1+x)^n;$

в) $f(x) = \frac{1}{1+x}.$

10. Разделить переменные в уравнении $y' = \frac{x^2}{y + \sin y}.$

11. Решить уравнения:

а) $xy' + y = y^2;$

б) $y' = y^3;$

в) $y' = \sqrt{\frac{1-y^2}{1-x^2}}.$

12. Решить уравнение $(1+e^x)yy' = e^x$ и найти частное решение, удовлетворяющее начальному условию $y|_{x=0} = 0.$

Ответ: $y^2 = \ln(1+e^x).$

13. Решить уравнение $xy' = y + e^x.$

В результате деятельности человека или процессов протекающих вокруг нас происходят различные события (факты, явления). Появление того или иного события обусловлено наличием целого ряда условий и обстоятельств. Познание закономерностей окружающего нас мира позволяет выявить связи между событиями и условиями, которые в значительной мере определяют их появление.

Однако еще давно было замечено, что при повторении одной и той же ситуации многие события каждый раз (от случая к случаю) происходят по-разному. При этом различия результатов могут быть существенными или весьма незначительными, иногда ожидаемое событие вообще не происходит. Такие события называют случайными.

Типичными примерами случайных событий являются многие погодные факторы: будет ли сегодня дождь, выпадет ли снег и т. д. Действительно, когда на небе появляются густые облака, мы ждем дождя. Однако даже при точном, на наш взгляд, повторении условий, при которых дождь всегда начинался, в данном случае он может и не пойти.

Случайные события сопутствуют огромному большинству явлений. На практике, к сожалению, невозможно учесть всю совокупность сопутствующих условий, факторов, взаимосвязей и вместо реального явления часто строят его модель, включающую только самые главные, основные, решающие факторы и связи, пренебрегая при этом остальными (второстепенными). Такой подход позволяет построить модели, для исследования которых можно применить математический аппарат, обеспечивающий выявление главных закономерностей, позволяющих достаточно точно предсказывать результаты опыта. По мере развития науки число учитываемых факторов растет, явление изучается все более точно, прогноз становится точнее.

Вместе с тем такая классическая схема далеко не всегда применима. Дело в том, что если число «второстепенных» фак-

торов очень велико или их нельзя в принципе зарегистрировать и учесть, а влияние таких факторов заметно, то классическая схема уже себя не оправдывает. Для учета таких второстепенных факторов необходимо применение принципиально других математических методов, которые развиваются в рамках теории вероятностей и математической статистики.

Добавим также, что одной из причин появления теории вероятностей была попытка осмыслить результаты всевозможных известных еще с древности игр для разработки стратегии наиболее выгодного поведения. Именно поэтому чаще всего удобно приводить примеры, связанные с играми: картами, монетами, костями (кубики с нанесенными цифрами от единицы до шести), рулеткой, урной с разноцветными шарами и т. п. Результаты, полученные для игр, потом оказались очень полезными для описания совсем других явлений.

Вторым истоком понятий для теории вероятностей явилась необходимость влиять на точность попаданий при стрельбе. Обработка результатов стрельб привела к понятию геометрической вероятности.

Глава 1.1. Основные понятия теории вероятностей

Теория вероятностей есть математическая наука, изучающая закономерности в случайных явлениях независимо от их конкретной природы и дающая методы количественной оценки влияния случайных факторов на различные явления.

Основной принцип теории вероятностей состоит в том, что при изучении массовых явлений возникают своеобразные новые закономерности. Так, при одном бросании игральной кости может выпасть любое из первых шести чисел, то же самое можно сказать и про результаты других бросаний. Однако при большом числе бросаний можно с уверенностью ожидать, что выпадение, например, числа три будет наблюдаться в одной шестой части всех бросаний, а выпадение чисел 5 или 6 — в одной трети всех попыток.

1.1.1. Случайные события

1.1.1.1. Событие. В теории вероятностей событием A называют все то, что может произойти при осуществлении

некоторого комплекса условий G . Событие наступает в результате реализации различных процессов, которые будем называть опытами (экспериментами). Примеры событий:

- A_1 — появление цифры при бросании монеты;
- A_2 — выпадение четного числа очков при игре в кости;
- A_3 — выход из строя компьютера после пяти часов работы;
- A_4 — замерзание воды при сильном морозе;
- A_5 — в перечне месяцев года после января непосредственно следует апрель.

Все эти события отличаются в первую очередь тем, что возможность их появления различна. Одно событие (A_1) происходит всегда, другое (A_5) никогда не наступает, остальные могут произойти или не произойти в результате проведения одного опыта.

Если при реализации условий G событие A всегда происходит, то оно называется **достоверным**. Если же это событие при заданных условиях никогда не наступает, то его называют **невозможным**. Так, если сегодня пятница, то завтра по календарю непременно наступит суббота. Во всех других случаях наступление субботы невозможно. Другой пример. При бросании одной игральной кости невозможно выпадение целого числа большего шести. В то же время гарантировано выпадение натурального числа меньшего семи.

Если в результате опыта при реализации определенного комплекса условий данное событие может наступить или не наступить, то оно называется **случайным**. Условия проведения такого опыта часто называют случайным опытом (экспериментом). Очевидно, что после бросания игральной кости число три может выпасть, но оно может и не выпасть. Через пять часов после включения компьютер может быть исправным, но может и выйти из строя. Стреляя в мишень, можно попасть с первого выстрела, можно с третьего, но можно не попасть и никогда.

Пусть при данных условиях проводится случайный опыт, в результате которого обязательно наступает одно и только одно из возможных событий, неразложимых на более простые. Такие события называются **элементарными событиями**. Множество Ω всех элементарных событий w_i образует **пространство элементарных событий**. Например, при бросании игральной кости множество элементарных событий — исходов, образует пространство из шести элементарных исходов w_i : выпало одно очко, выпало два очка и т. д.

$$\Omega = \{w_1 = 1, w_2 = 2, w_3 = 3, w_4 = 4, w_5 = 5, w_6 = 6\}.$$

Наряду с элементарными рассматриваются так называемые составные, или разложимые, события. Событие B называется **составным**, если можно указать по меньшей мере два таких элементарных события w_1 и w_2 , что из существования каждого из них в отдельности следует существование события B . Этот факт записывается в виде:

$$B = \{w_1, w_2\}.$$

Используя введенную терминологию, случайным событием A называют любое подмножество S пространства элементарных событий ($S \subset \Omega$). Содержательно это означает, что появление любого из элементарных событий, входящих в S , влечет за собой появление события A .

Например, при бросании игральной кости составное событие $A = \{\text{число очков четное}\}$ можно записать так: $A = \{2, 4, 6\}$, подразумевая при этом: если выпадет число 2 или 4 или 6, то наступит событие A .

Таким образом, случайным событием A можно называть такой элемент подмножества всех элементарных событий F ($A \in F$), про которое мы можем сказать, наступило это событие в результате опыта или нет.

1.1.1.2. Правила действий над событиями.

Объединением (суммой) событий $A_1, A_2, \dots, A_n$ называется событие A , состоящее в наступлении хотя бы одного из этих событий. Обозначается одним из следующих способов:

$$A = A_1 \cup A_2 \cup \dots \cup A_n \equiv A_1 + A_2 + \dots + A_n.$$

Пример. В урне 6 шаров, которые отличаются лишь номером. Событие A_i — наугад выбрать шар под номером i . Событие

$$A = A_1 + A_3 + A_5$$

состоит в том, что будет выбран шар с номером 1 или 3, или 5, т. е. с нечетным номером.

Пересечением (произведением) событий $A_1, A_2, \dots, A_n$ называется событие B , состоящее в обязательном наступлении всех этих событий. Обозначается пересечение событий так:

$$B = A_1 \cap A_2 \cap \dots \cap A_n \equiv A_1 \cdot A_2 \cdot \dots \cdot A_n.$$

Пример. В урне 12 шаров, среди которых одна половина белых с номерами от 1 до 6, а другая черных — с такими же номерами.

Пусть A — «вынуть белый шар», B — «вынуть шар с нечетным номером». Тогда событие

$$C = A \cdot B$$

означает «выбрать белый шар с нечетным номером».

Как следует из определения, пересечение событий не изменится, если поменять местами сомножители:

$$C = B \cdot A.$$

Разностью событий A и B (обозначается $A - B$) называется событие, заключающееся в наступлении события A и одновременном ненаступлении события B . Для предыдущего примера событие

$$D = A - B$$

означает «выбрать белый шар с четным номером».

Дополнение (противоположное к A) — это событие $\bar{A}$, состоящее в ненаступлении события A . Так, если A — «вынуть белый шар», то $\bar{A}$ — «вынуть небелый шар».

Достоверным можно считать событие A_d , состоящее из всех элементарных событий, т. е.:

$$A_d = \Omega = \{w_i\}.$$

Невозможным же событием в используемой здесь терминологии можно считать пустое событие $\emptyset$, т. е. событие противоположное по отношению к достоверному:

$$\bar{A}_d = \emptyset.$$

События $A_1, A_2, \dots, A_n$ называются **несовместными**, если в результате одного опыта никакие два из них не могут произойти одновременно. Содержательно это означает, что среди событий $A_1, A_2, \dots, A_n$ нельзя найти такую пару событий A_i и A_j , в которой обнаружилось бы хотя бы по одному общему элементарному событию. Формально несовместность событий A_i и A_j определяется следующим образом:

$$A_i \cdot A_j = \emptyset, \quad i, j = \overline{1, n}.$$

Например, при однократном бросании игральной кости выпадение четного и нечетного числа — несовместные события. Несовместными являются также промах и попадание при стрельбе по мишени.

Совокупность событий $A_1, A_2, \dots, A_n$ называется **полной группой несовместных событий**, если:

$$(A_1 + A_2 + \dots + A_n = \Omega) \wedge (A_i \cdot A_j = \emptyset).$$

Примером полной группы несовместных событий является пространство элементарных событий. Другим характерным примером является пара двух противоположных событий $(A, \bar{A})$. Например, выпадение герба и цифры в результате подбрасывания монеты, работоспособность компьютера и его неисправность в данный момент времени, попадание и непопадание в мишень при одном выстреле и т. д.

События A и B называются **независимыми**, если появление одного из них не изменяет шансы появления другого. Например, одновременно бросаются две игральные кости. Появление на одной из них трех очков ни коим образом не зависит от того, какое количество очков появилось на верхней грани другой кости.

Если появление одного события влияет на появление другого, то такие события называются **зависимыми**. Рассмотрим пример. В урне 4 пронумерованных подряд и одинаковых на ощупь шара. Вынимается один шар, а затем второй. Событие A — первый вынутый шар имеет четный номер. Событие B — вынутый второй шар имеет четный номер. Очевидно, что эти события зависимы, т. к. от того, какой шар был вынут в первом опыте, зависят шансы наступления второго события. Действительно, если первым вынули четный шар, то шансы вынуть четный шар и во втором опыте (событие $A \cdot B$) будут вдвое меньше, чем если бы первым был вынут нечетный шар (событие $\bar{A} \cdot B$).

1.1.2. Понятие вероятности

1.1.2.1. Априорно-частотный подход. Пусть в урне 10 одинаковых по размеру шаров, из них: два белых, три черных и пять красных. Случайный опыт заключается в том, что вынимается наугад один шар. При этом шар может оказаться белым или черным, или красным. Интуитивно ясно, что появление красного шара имеет шансов больше, чем появление белого или черного шара. Но насколько шансы вынуть красный шар больше, чем другие альтернативы?

Для количественного сравнения событий по степеням их возможности целесообразно связать определенное число, которое тем больше, чем более возможно событие. Такой ко-

личественной мерой степени объективной возможности реализации события является **вероятность события**.

Вернемся к примеру с разноцветными шарами в урне. Будем в каждом опыте вытаскивать шар, фиксировать его цвет, опускать вынутый шар обратно в урну и тщательно перемешивать там шары. В результате каждого проведенного опыта можно вытащить шар любого из трех возможных цветов (A_1 — вынут белый шар, A_2 — вынут черный шар, A_3 — красный шар).

Пусть проведено n рассматриваемых опытов, в результате которых белый шар вынимался m_1 раз, черный — m_2 , а красный — m_3 раз. Частота появления каждого из возможных событий есть отношение:

$$p^* = \frac{m_i}{n}, \text{ где } i = \overline{1, 3}.$$

В общем случае **частотой** появления события A в n опытах называется отношение числа появлений этого события (m) к числу проведенных опытов n :

$$p^*(A) = \frac{m}{n}.$$

Если проводить такой опыт значительное число раз (n — велико), то окажется, что примерно в половине случаев был вынут красный шар, в двадцати процентах случаев белый, а в тридцати — черный. По мере увеличения числа проведенных опытов уверенность в соотношении шансов возможных событий, соответственно 5 : 2 : 3, будет подтверждаться со все большей точностью.

Этот факт, как и огромное количество подобных примеров, сопровождавших историю человеческой деятельности, наталкивает на мысль о том, что при увеличении числа опытов, проведенных в одинаковых условиях, частота появления события A все меньше отличается от некоторого определенного значения. Это значение и есть **вероятность события A** :

$$p(A_i) = \lim_{n \rightarrow \infty} \frac{m_i}{n}.$$

Введенная таким образом вероятность события носит название **статистической, или априорно-частотной**. Она фиксирует связь между частотой события как непосредственно измеряемой величиной и вероятностью как формальной характеристикой случайного события (теорема Я. Бернулли).

1.1.2.2. Аксиоматический подход. На основе обобщения свойств частоты случайного события, в основном трудами русских ученых, была построена теория вероятностей как математическая наука. В основу этой теории положено аксиоматическое определение вероятности.

Числовая функция $p(A)$ называется **вероятностью события** A , если она удовлетворяет следующим аксиомам:

1. Вероятность $p(A)$ есть неотрицательное число, заключенное между нулем и единицей: $0 \leq p(A) \leq 1$;
2. Вероятность достоверного события равна единице;
3. Вероятность невозможного события равна нулю;
4. Вероятность суммы двух несовместных событий A и B равна сумме вероятностей этих событий $p(A + B) = p(A) + p(B)$.

Тройка (Ω, F, P) , где Ω — множество элементарных событий, F — допустимое множество составных событий (множество подмножеств Ω), P — множество вероятностей случайных событий, называется **вероятностным пространством**, понятие о котором введено А. Н. Колмогоровым. Построение вероятностного пространства является основным этапом в создании математической модели того или иного эксперимента.

Непосредственной проверкой можно установить, что частота случайного события удовлетворяет сформулированным аксиомам. Тем самым подтверждается естественная связь между статистическим (базирующимся на эмпирических знаниях) и аксиоматическим (отражающим формальную точку зрения) подходами к определению вероятности события. Важно подчеркнуть, что сформулированных аксиом достаточно, чтобы построить огромное непротиворечивое здание, каким является теория вероятностей.

1.1.2.3. Классический подход. Исторически первым и наиболее доступным для понимания является метод непосредственного подсчета вероятности события. Рассмотрим полную группу несовместных равновозможных событий. Примерами таких групп являются число очков при бросании игральных костей, число попаданий в мишень при выстрелах проводимых в одинаковых условиях, появление шара с заданным номером при наличии в урне нескольких неразличимых на ощупь шаров. События, образующие такую группу, называются случаями, а соответствующая схема логических построений — **системой случаев** или **схемой урн**.

Пусть среди всех возможных в рассматриваемой схеме исходов n нас интересует только вполне определенное их m подмножество. Случаи, входящие в m подмножество будем называть благоприятными. Например, в урне два белых, три черных и пять красных одинаковых на ощупь шаров. Будем считать благоприятным выбор белого шара. Таких благоприятных случаев два. Появление же черного или красного шара — случай неблагоприятный. Очевидно, что их восемь.

Обозначим через A любой из благоприятных случаев. Вероятность события A вычисляется как отношение числа благоприятных случаев к общему числу случаев:

$$P(A) = \frac{m}{n}.$$

Эта формула называется классической формулой для вычисления вероятностей, когда опыт сводится к системе случаев.

Схема применения классической формулы следующая:

1. Удостоверяются в том, что рассматриваемые условия опыта можно интерпретировать как схему урн (возможные исходы образуют полную группу несовместных равновозможных событий);
2. Выбирается интересующее нас возможное событие A (благоприятное событие);
3. Вычисляется число возможных исходов (n) и число благоприятных исходов (m);
4. Вычисляется искомая вероятность $P(A)$.

Задача. Имеется колода из 36 карт. Наугад вынимаются три карты. Найти вероятность того, что среди них окажется ровно один валет.

Решение. Колода карт относится к одному из наиболее известных примеров схемы урн. В выбранных трех картах может быть три валета, два, один или ни одного валета. Всего в колоде четыре валета. Поэтому реализовать благоприятный случай A — выбрать один из них (любой!) можно C_4^1 различными способами, две другие карты (не валеты) можно выбрать различными способами. Следовательно, $m = C_4^1 \cdot C_{32}^2$ (здесь логическое «и»). Общее же число исходов есть число сочетаний из 36 по 4, т. е. $n = C_{36}^4$. Искомая вероятность вычисляется по формуле:

$$P(A) = \frac{C_4^1 \cdot C_{32}^2}{C_{36}^4} = \frac{31 \cdot 16 \cdot 4}{35 \cdot 3 \cdot 17 \cdot 33} = \frac{1984}{58905}.$$

При использовании классической формулы широко применяются результаты комбинаторики.

Рассмотрим задачу. Замок имеет четырехзначный цифровой шифр. Наугад выбираются четыре цифры. Какова вероятность открыть при этом замок, если известно, что в коде все цифры различны?

Решение. Поскольку в шифре замка важен не только набор цифр, но и их порядок, то число благоприятных событий равно 1. Общее же число возможных упорядоченных комбинаций из четырех различных цифр определяется по формуле $n = A_{10}^4$. Таким образом, искомая вероятность:

$$P(A) = \frac{1}{A_{10}^4} = \frac{1}{5040}$$

1.1.2.4. Геометрические вероятности. Пусть пространство элементарных событий является несчетным, но выполняются следующие условия:

1. Любые два элементарных события несовместны;
2. Все события являются равновероятными.

В таких опытах вероятности некоторых событий можно вычислить геометрически как отношение длин отрезков, площадей фигур, объемов соответствующих областей.

Задача. Поезда в метро идут с интервалом в три минуты. Чему равна вероятность того, что пассажир будет ждать поезда более двух минут?

Решение. Считаем, что моменты появления пассажира в интервале между поездами равновероятны. Если все время ожидания принять за интервал длиной три единицы $I_0 = 3$, то событию A — «время ожидания более двух минут», соответствует одна единица длины $I(A) = 1$. Следовательно, искомая вероятность равна отношению длин временных интервалов:

$$P(A) = \frac{I(A)}{I_0} = \frac{1}{3}.$$

1.1.3. Основные теоремы теории вероятностей

Изложенные выше прямые методы вычисления вероятностей не всегда эффективны, а иногда и просто невозможны. На практике искомую вероятность часто находят косвенными методами, позволяющими по известным вероятностям одних событий находить вероятности других событий, с ними связанных.

В основе этих косвенных методов лежат основные теоремы теории вероятностей.

1.1.3.1. Теорема умножения вероятностей. Пусть рассматриваются два события A и B , для которых известны вероятности их появления $P(A)$ и $P(B)$. Вероятность произведения двух событий $P(A \cdot B)$ равна произведению вероятности одного из них на условную вероятность другого события $P(B|A)$, т. е. вычисленную при условии, что первое событие имело место:

$$P(A \cdot B) = P(A) \cdot P(B|A), \quad (1.1)$$

Докажем эту теорему для системы случаев [1]. Пусть в данном опыте возможно n исходов, из них событию A благоприятны m исходов и событию B — k исходов. Предположим, что события A и B совместны, т. е. существуют исходы, благоприятные событию A и B одновременно и таких исходов s . Тогда:

$$P(A \cdot B) = \frac{s}{n}, \quad P(A) = \frac{m}{n}.$$

Вычислим условную вероятность $P(B|A)$, т. е. вероятность наступления события B в предположении, что событие A наступило. Если известно, что событие A произошло, то из ранее возможных n случаев осталось лишь m благоприятных событию A . Из них лишь s благоприятны событию B . Следовательно:

$$P(B|A) = \frac{s}{m}.$$

Подставляя в (1.1) найденные выражения, получим тождество, доказывающее теорему умножения вероятностей:

$$\frac{s}{n} = \frac{m}{n} \cdot \frac{s}{m} = \frac{s}{n}.$$

Доказанную теорему можно распространить на произведение n событий:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cdot A_2) \cdot \dots \cdot P(A_n|A_1 \cdot A_2 \cdot \dots \cdot A_{n-1}). \quad (1.2)$$

Для независимых событий (когда $P(B) = P(B|A)$) формула (1.1) имеет упрощенный вид:

$$P(A \cdot B) = P(A) \cdot P(B),$$

т. е. вероятность произведения двух независимых событий равна произведению вероятностей этих событий.

Для вероятности произведения n независимых событий формула (1.2) приобретает вид:

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \dots \cdot P(A_n).$$

Задача. Механизм собирается из трех одинаковых деталей и считается неработоспособным, если все три детали вышли из строя. В сборочном цехе осталось 15 деталей, из которых 5 нестандартных (бракованных). Какова вероятность того, что собранный из взятых наугад оставшихся деталей механизм будет неработоспособным.

Решение. Обозначим искомое событие через A , а выбор первой нестандартной детали через A_1 , второй — через A_2 и третьей — через A_3 . Событие A произойдет, если произойдет и событие A_1 , и событие A_2 , и событие A_3 . Однако события A_1 , A_2 , A_3 зависимы и, если вероятность выбрать первую нестандартную деталь равна $5/15$, то вероятность выбрать вторую нестандартную деталь равна $4/14$, третью нестандартную деталь — $3/13$. Следовательно, используя формулу (1.2), имеем:

$$P(A) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1, A_2) = \frac{5}{15} \cdot \frac{4}{14} \cdot \frac{3}{13} = 0.022.$$

1.1.3.2. Теорема сложения вероятностей. Вероятность суммы двух совместных событий равна сумме вероятностей этих событий минус вероятность произведения этих же событий:

$$P(A+B) = P(A) + P(B) - P(A \cdot B), \quad (1.3)$$

Доказательство. Для наступления события A необходимо наступление попарно несовместных событий либо AB , либо $A\bar{B}$. Для наступления события B аналогично: или BA , или $B\bar{A}$. Следовательно, по аксиоме 4, имеем:

$$P(A) = P(AB) + P(A\bar{B}) \Rightarrow P(A\bar{B}) = P(A) - P(AB), \quad (1.4)$$

$$P(B) = P(BA) + P(B\bar{A}) \Rightarrow P(B\bar{A}) = P(B) - P(BA). \quad (1.5)$$

Для наступления хотя бы одного из событий A или B (сумма событий $A+B$) необходимо наступление одного из событий AB , $B\bar{A}$, $A\bar{B}$, которые попарно несовместны. Поэтому:

$$P(A+B) = P(AB) + P(A\bar{B}) + P(B\bar{A}). \quad (1.6)$$

Подставив $P(A\bar{B})$ и $P(B\bar{A})$ из (1.4) и (1.5) в (1.6), т. к. $P(AB) = P(BA)$, имеем:

$$P(A+B) = P(AB) + P(A) + P(B) - 2P(AB),$$

откуда непосредственно следует искомая формула (1.3).

Задача. Производится два независимых выстрела в одну и ту же мишень. Вероятность попадания при первом выстреле 0.6, а при втором — 0.8. Найти вероятность попадания в мишень при двух выстрелах.

Решение. Обозначим попадание при первом выстреле как событие A_1 , при втором — как событие A_2 . Попадание в мишень предполагает хотя бы одно попадание: или только при первом выстреле, или только при втором, или и при первом, и при втором. Следовательно, в задаче требуется определить вероятность суммы двух совместных событий A_1 и A_2 .

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 \cdot A_2).$$

Поскольку события независимы, то $P(A_1 A_2) = P(A_1) \cdot P(A_2)$. Поэтому:

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1) \cdot P(A_2) = 0.6 + 0.8 - 0.6 \cdot 0.8 = 0.92.$$

Из формулы (1.3) можно получить формулу для вероятности суммы трех событий:

$$P(A+B+C) = P(A) + P(B) + P(C) - P(A \cdot B) - P(A \cdot C) - P(B \cdot C) + P(A \cdot B \cdot C).$$

При увеличении числа слагаемых формула вероятности суммы всё более усложняется. На практике при решении некоторых задач целесообразен переход к противоположному событию. При этом используют известное правило: вероятность суммы противоположных событий всегда равна единице, т. е.:

$$P(A + \bar{A}) = P(A) + P(\bar{A}) = 1.$$

В общем же случае:

$$P(A_1 + A_2 + A_3 + \dots + A_n) = 1 - P(\bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3 \cdot \dots \cdot \bar{A}_n). \quad (1.7)$$

Задача. Ремонтное ателье обслуживает пять клиентов. Вероятность вызова на обслуживание от каждого клиента равна 0.2. Какова вероятность того, что в данный момент ателье занято обслуживанием клиентов.

Решение. Очевидно, что событие A «быть занятым обслуживанием клиентов» есть сумма событий $A = A_1 + A_2 + A_3 + A_4 + A_5$, где A_i — «быть занятым обслуживанием i -го клиента». Противоположное событие «не быть занятым обслуживанием клиентов» определяется так: $\bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3 \cdot \bar{A}_4 \cdot \bar{A}_5$.

Следовательно, можно применить формулу (1.7):

$$P(A_1 + A_2 + A_3 + A_4 + A_5) = 1 - P(\bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3 \cdot \bar{A}_4 \cdot \bar{A}_5) = 1 - (1 - 0.2)^5 = 0.672.$$

1.1.3.3. Формула полной вероятности. Рассмотрим следующую ситуацию. Приобретается компьютер, который может быть изготовлен с полным комплексом тестирования или по сокращенному варианту. Известно, какая часть компьютеров изготавливается по сокращенному варианту, а также вероятность безотказной работы компьютера в течение гарантийного срока при различных вариантах изготовления. Какова вероятность того, что приобретенный компьютер безотказно проработает гарантийный срок?

Сложность решения данного вопроса обусловлена тем, что неизвестно, какой вариант изготовления реализован при изготовлении именно приобретенного компьютера. Для решения подобного класса задач наиболее удобно применение формулы полной вероятности, являющейся обобщением теорем умножения и сложения вероятностей.

Сформулируем в общем виде задачу, приводящую к формуле полной вероятности. Пусть некоторое событие A может произойти с какой-то вероятностью только как следствие каждого из событий $H_1, H_2, \dots, H_n$, образующих полную группу несовместных событий. Такие события H_i называются гипотезами. Заданы вероятности гипотез $P(H_i)$ и условные вероятности наступления события A с каждой из гипотез $P(A|H_i)$. Вероятность наступления события A дается формулой полной вероятности:

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A|H_i). \quad (1.8)$$

Задача. Имеется три одинаковых урны. В первой два белых и один черный шар, во второй три белых и один черный шар, в третьей урне два белых и два черных шара. Из выбранной наугад урны выбирается один шар. Какова вероятность того, что он окажется белым?

Решение. Гипотезой H_i будем считать выбор i -ой урны. Все урны считаются одинаковыми, следовательно, вероятность выбрать i -ую урну есть $P(H_i) = 1/3, i=1,2,3$. Зная состав шаров в каждой из урн, легко определить вероятности вынуть белый шар из каждой урны: $P(A|H_1) = 2/3, P(A|H_2) = 3/4, P(A|H_3) = 1/2$. Подставляя найденные значения в (1.8), имеем:

$$P(A) = \frac{1}{3} \left(\frac{2}{3} + \frac{3}{4} + \frac{1}{2} \right) = \frac{23}{36}.$$

1.1.3.4. Формула апостериорной вероятности. Рассмотрим ситуацию в некотором смысле противоположную предыдущей. Пусть событие A уже произошло. Требуется определить вероятность того, что событие произошло именно совместно с гипотезой H_i . Ответ на подобный вопрос дает формула апостериорной вероятности или формула Байеса:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{i=1}^n P(H_i)P(A|H_i)}. \quad (1.9)$$

Задача. Пусть сохраняются условия предыдущей задачи, но из урны уже вынут белый шар. Требуется определить вероятность, что шар вынут из первой урны.

Решение. Поскольку все вероятности необходимые для применения формулы (1.9) найдены, то имеем:

$$P(H_1|A) = \frac{1/3 \cdot 2/3}{1/3 \cdot (2/3 + 3/4 + 1/2)} = \frac{8}{23}.$$

1.1.3.5. Схема опытов Бернулли. Если проводится n одинаковых и независимых опытов, в каждом из которых событие A может произойти с вероятностью p , то говорят, что реализуется схема опытов Бернулли.

Задача. Определить вероятность того, что в схеме опытов Бернулли событие A появится ровно m раз ($m \leq n$).

Решение. Пусть событие B заключается в том, что событие A в n опытах происходит подряд m раз, а затем $(n - m)$ раз происходит событие $\bar{A}$, т. е.:

$$B = A_1 \cdot A_2 \cdot A_3 \cdot \dots \cdot A_m \cdot \bar{A}_{m+1} \cdot \dots \cdot \bar{A}_n.$$

Вероятность такого события B равна:

$$P(B) = p_m (1-p)^{n-m}.$$

Поскольку в задаче не требуется учитывать порядок следования этих событий, то необходимо учесть все события типа В. Число событий же, подобных В и отличающихся лишь порядком наступления событий А и $\bar{A}$, равно числу сочетаний из n по m. Следовательно, искомая формула, носящая имя Бернулли, имеет вид:

$$P_{m,n} = C_n^m \cdot p^m \cdot (1-p)^{n-m} \quad (1.10)$$

В качестве примера рассмотрим следующую задачу. Вероятность прибытия англоязычной группы туристическое агентство оценивает величиной 0.6. Прибывает шесть групп. Найти вероятность того, что: а) из прибывших ровно четыре англоговорящие группы; б) прибудет не более двух таких групп.

Решение. Ответ на вопрос а) получаем прямой подстановкой данных в формулу (1.10):

$$P_{4,6} = C_6^4 p^4 (1-p)^2 = 15(0.6)^4 (0.4)^2 = 0.311.$$

Чтобы получить ответ на вопрос б), необходимо вычислить сумму вероятностей прибытия двух, одной или ни одной групп:

$$\begin{aligned} P_{0,6} + P_{1,6} + P_{2,6} &= (1-p)^6 + C_6^1 p(1-p)^5 + C_6^2 p^2(1-p)^4 = \\ &= 0.4^6 + 6 \cdot 0.6 \cdot (0.4)^5 + 15 \cdot (0.6)^2 \cdot (0.4)^4 = 0.201 \end{aligned}$$

Глава 1.2. Распределения случайных величин

1.2.1. Случайные величины

Случайное событие является в первую очередь качественной характеристикой опыта. Однако, при рассмотрении результатов опыта часто помимо самого факта наступления или ненаступления исследуемого события интересуются некоторыми количественными характеристиками. Например, исследуя вопросы безотказности работы компьютера, представляет интерес время его безотказной работы. При анализе спроса на тот или иной товар необходимо прогнозировать объемы возможных продаж. Рассматривая проблему мягкой посадки космических кораблей, важно оценивать скорость приземления в момент касания поверхности земли и т. п. Количественными характеристиками случайных событий являются случайные величины.

1.2.1.1. Определение случайной величины. Величина, которая в результате опыта может принимать одно и только одно определенное значение, до опыта неизвестное и зависящее от причин, которые нельзя учесть заранее, называется **случайной величиной**. Названия случайных величин обычно обозначают заглавными буквами: X, Y, Z, ..., а их возможные значения — прописными буквами: x, y, z, ...

Рассмотрим примеры.

1. Проводятся выборы в представительные органы власти. Случайным событием является факт выбора (или невыбора) конкретного кандидата. При этом несомненный интерес представляет количество поданных за него голосов — случайная величина, количественно характеризующая результаты выборов.

2. Исследуется надежность нового прибора. Случайным событием является факт прохождения тестовых испытаний (удовлетворяет — не удовлетворяет оговоренному уровню требований). Случайной величиной является процент приборов прошедших испытание. В рамках испытаний может рассматриваться та или иная группа характеристик. Например, факты безотказной работы не менее заданного срока могут измеряться конкретным параметром — временем безотказной работы.

3. Проводится социальное обследование, в котором изучается социальный портрет жителей данного города. Случайно вы-

бранному жителю задаются вопросы о его социальной принадлежности, профессии, возрасте, семейном положении, количестве детей, обеспеченности жильем и т. п. Случайным событием является факт опроса того или иного жителя. Количественным же результатом опроса выступает вектор значений, каждый компонент которого есть количественная мера соответствующей характеристики опрашиваемого.

Эти и многие другие примеры приводят к выводу, что случайные величины обычно имеют конкретный физический смысл и могут быть как скалярными, так и векторными.

Формальное определение: случайной величиной называется измеримая функция $x = x(\omega)$, отображающая пространство элементарных событий Ω на множество действительных чисел R .

Если множество Ω конечно или счетно, то случайная величина называется **дискретной**. Примеры дискретной величины: количество уголовных дел, рассматриваемых данным судом за определенное время; количество клиентов; количество диалектов в данном языке; количество избирателей округа, которые примут участие в предстоящих выборах, количество телевизоров, которые необходимо проверить до выявления первого неисправного и т. п. Итак, дискретной считают такую случайную величину, возможные значения которой можно пронумеровать.

Непрерывной называют такую случайную величину, возможные значения которой непрерывно заполняют некоторый промежуток числовой оси, т. е. множество значений имеет мощность континуума. Например, время выхода из строя работающего компьютера, ошибка указателя скорости автомобиля, вес выбранного яблока и т. п.

1.2.2. Законы распределения случайных величин

1.2.2.1. Понятие о законах распределения случайных величин.
Для решения основной задачи, решаемой при помощи теории вероятностей, — «Найти вероятность события A » недостаточно задание пространства элементарных событий Ω , как множества теоретически возможных значений рассматриваемой случайной величины. Принципиально необходимым является знание связи между возможными значениями исхода опыта и соответствующими вероятностями.

Всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими вероятностями называется **законом распределения** этой случайной величины. Знание закона распределения позволяет заранее (до опыта!) установить, какое значение случайной величины будет появляться чаще, а какое реже и насколько.

Для дискретной случайной величины законом распределения выступает правило сопоставления каждому возможному значению x_i случайной величины X вероятности его появления: $p_i = P(X=x_i)$. Последнее выражение читается так: «вероятность того, что случайная величина X примет значение x_i ». Для дискретной случайной величины закон распределения проще всего задавать в виде таблицы

X	x_1	x_2	x_3		x_n	...
P	P_1	P_2	P_3		P_n	...

Для примера, вернемся к задаче п.1.1.3.5. Вероятность прибытия одной англоязычной группы туристов оценивается величиной $p = 0.6$. В случае прибытия шести групп возможно прибытие различного числа таких групп (от 0 до 6). Используя формулу (1.10), можно предсказать вероятность наступления любого из этих исходов. Так, вероятность прибытия двух англоязычных групп равна $P(X=2) = 0.138$, в два раза меньше, чем прибытие трех таких групп — $P(X=3) = 0.276$. Следовательно, формула

$$p_m = P(X=m) = C_6^m \cdot (0.6)^m \cdot (0.4)^{6-m}$$

есть закон распределения случайной величины m : «число англоговорящих туристических групп при ожидаемом прибытии шести, каждая из которых прибывает из неизвестной заранее страны».

Закон распределения для непрерывной случайной величины X есть всякое соотношение, сопоставляющее с каждой измеримой областью ее возможных значений ΔX соответствующую вероятность $P(x \in \Delta X)$. Например, если среднее время безотказной работы компьютеров данного типа равно s часов, то вероятность проработать без отказа не менее t часов, т. е. $P(T \geq t)$, часто может быть выражена законом распределения в форме:

$$P(T \geq t) = e^{-st}, \text{ при } t \geq 0.$$

1.2.2.2. Функция распределения. Существует много различных форм законов распределения. Наиболее универсальной из форм является функция распределения.

Функцией распределения $F(x)$ или интегральным законом распределения случайной величины X называется вероятность выполнения неравенства $X < x$, т. е.:

$$F(x) = P(X < x).$$

Функция распределения определена для случайных величин любого типа: дискретных и непрерывных.

Определение функции распределения имеет наглядную геометрическую интерпретацию. Если рассматривать X как случайную точку на числовой оси, то $F(x)$ есть вероятность того, что случайная точка попадет левее выбранной величины x (рис. 1.1).

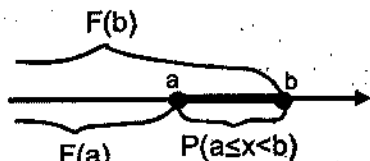


Рис. 1.1. Геометрическая интерпретация функции распределения

Свойства функции распределения непосредственно вытекают из ее определения:

1) $F(x)$ — неотрицательная функция, значения которой заключены между 0 и 1, т. е.:

$$0 \leq F(x) \leq 1;$$

$$2) F(-\infty) = 0;$$

$$3) F(+\infty) = 1;$$

4) $F(x)$ — неубывающая функция своего аргумента, т. е., если $x_1 > x_2$, то $F(x_1) \geq F(x_2)$;

$$5) P(a \leq X < b) = F(b) - F(a)$$

(1.11)

Формула (1.11) дает прямое решение некоторых практических задач. Рассмотрим две из них.

1. Вероятность попадания при одном выстреле равна 0.4. Производится 4 выстрела. Какова вероятность того, что будет менее двух попаданий?

Решение. Поиск ответа задачи заключается в определении значения функции распределения для значения аргумента $x = 2$. Действительно, $F(2)$ есть вероятность того, что произойдет менее двух попаданий. Следовательно:

$$P(2) = F(2) = \sum_{i \leq 2} P(x_i) = P(0) + P(1) = 0.648.$$

2. Время безотказной работы прибора имеет функцию распределения:

$$F(t) = 1 - e^{-0.02t}.$$

Вероятность того, что прибор безотказно проработает не менее 30 и не более 40 часов находится следующим образом:

$$P(30 \leq t < 40) = F(40) - F(30) = 1 - e^{-0.8} - 1 + e^{-0.6} = 0.1$$

1.2.2.3. Функция плотности вероятности. Непрерывную случайную величину удобнее описывать законом распределения, который называют функцией плотности вероятности или дифференциальным законом распределения.

Пусть X — непрерывная случайная величина, имеющая функцию распределения $F(x)$. Если эта функция дифференцируема, то можно рассматривать ее производную:

$$F'(x) = f(x) \Rightarrow F(x) = \int_{-\infty}^x f(x) dx. \quad (1.12)$$

Функция $f(x)$, производная от функции распределения, называется **плотностью вероятности** случайной величины X или функцией распределения вероятностей.

Из определения плотности вероятности (1.12), принимая во внимание (1.11), следует, что:

$$P(a \leq x < b) = P(a < x < b) = \int_a^b f(x) dx. \quad (1.13)$$

Следовательно, если известна плотность распределения непрерывной случайной величины, то можно вычислить вероятность попадания этой случайной величины в любой заданный промежуток ее возможных значений.

Для задачи 2 предыдущего пункта

$$f(x) = F'(x) = 0.02e^{-0.02x}.$$

Используя (1.13), можно найти искомую вероятность через функцию плотности вероятности:

$$P(30 < t < 40) = \int_{30}^{40} 0.02e^{-0.02t} dt \approx 0.1.$$

Основные свойства функции плотности вероятности:

1. Функция плотности вероятности — неотрицательная функция, $f(x) \geq 0$;

2. Интеграл в бесконечных пределах от функции плотности вероятности (если она задана на всей числовой оси) равен 1. Действительно, как частный случай формулы (1.13) имеем:

$$F(\infty) = P(-\infty < x < \infty) = \int_{-\infty}^{\infty} f(x) dx = 1. \quad (1.14)$$

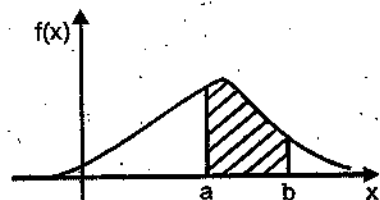


Рис. 1.2. Определение вероятности попадания в заданный интервал по функции плотности вероятности

Замечание. Если случайная величина задана только на отрезке $[x_1, x_2]$, то в (1.14) пределы интегрирования заменяются на x_1 и x_2 .

Свойство (1.14) называют условием нормировки. Общий вид функции плотности вероятности представлен на рис. 1.2. Очевидно, что вероятность попадания в заданный промежуток (a, b) численно равна площади криволинейной трапеции с основанием (a, b) .

1.2.3. Числовые характеристики случайных величин

Законы распределения являются наиболее полными характеристиками случайных величин. Знание этих законов позволяет решать практически все задачи теории вероятностей. Вместе с тем на практике часто затруднительно найти закон распределения. С другой стороны, при решении многих задач достаточно знание лишь основных характеристик случайных величин. К таким характеристикам в первую очередь относятся: *математическое ожидание, дисперсия и среднее квадратичное отклонение*.

1.2.3.1. Математическое ожидание. Рассмотрим пример. Бросаются одновременно три игральные кости и рассматривается случайная величина — сумма выпавших очков. Вполне очевидно,

что выпадение 18 очков будет в среднем происходить реже, чем 15 очков, и значительно реже, чем 12 очков. Если с некоторым «весом» усреднить все возможные значения суммы выпавших очков (от 3 до 18), то получим величину, являющуюся «центром» распределения возможных значений рассматриваемой случайной величины и называемую ее математическим ожиданием.

Для дискретной случайной величины математическое ожидание вычисляется по формуле:

$$m_x \equiv M[X] \equiv \sum_i x_i \cdot p(x_i). \quad (1.15)$$

Для непрерывных случайных величин математическое ожидание есть величина:

$$m_x \equiv M[X] \equiv \int_{-\infty}^{\infty} x \cdot f(x) dx. \quad (1.16)$$

Итак, **математическим ожиданием** (или средним) называют характеристику положения случайной величины X , которая равна средневзвешенному возможных ее значений. Обозначение $m_x \equiv M[X]$ показывает, что к случайной величине X надо применить оператор математического ожидания M по правилам (1.15) или (1.16). К формуле (1.16) относится примечание сделанное к (1.14).

Следует подчеркнуть, что математическое ожидание есть число (неслучайная величина) — центр группирования значений случайной величины или центр рассеивания.

Например. Если плотность вероятности времени безотказной работы прибора есть функция

$$f(x) = 0.02e^{-0.02x},$$

то математическое ожидание, т. е. среднее время безотказной работы, есть величина

$$m_x = \int_0^{\infty} t \cdot 0.02e^{-0.02t} dt = \frac{1}{0.02} = 50 \text{ (час)}.$$

Заметим, что размерность математического ожидания совпадает с размерностью самой случайной величины.

Найдем математическое ожидание для задачи 1 п. 1.2.2.2. Согласно формуле (2.5), имеем:

$$m_x = \sum_0^4 x_i \cdot p(x_i) = \sum_0^4 x_i \cdot C_4^i \cdot p^i \cdot (1-p)^{4-i} = 1.6.$$

Полученное число означает, что, проводя много серий по четыре выстрела в каждой, в среднем будем получать 1.6 попаданий, если вероятность попадания при одном выстреле равна 0.4. В каждой конкретной серии из четырех выстрелов будет *ровно* или 0, или 1, ..., или 4 попадания, но в *среднем* за много серий их будет 1.6. Этот пример показывает, что математическое ожидание не всегда может быть равно одному из возможных значений случайной величины.

1.2.3.2. Дисперсия. Рассмотрим две случайные величины, плотности вероятности которых представлены на рис.1.3. Они имеют одинаковое мате-

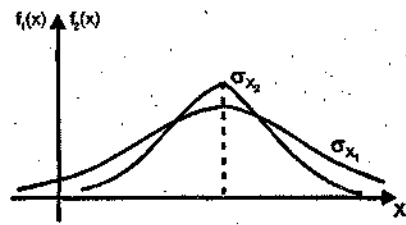


Рис. 1.3. Графики функций распределения различных дисперсий

матическое ожидание, однако значительные отклонения от центра рассеивания у первой случайной величины наблюдаются чаще, чем у второй. В этом случае говорят, что первая случайная величина имеет большее рассеивание, или размытость, чем вторая.

В качестве меры рассеивания значений случайной величины используется дисперсия. **Дисперсией** называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D_x \equiv M[(X - m_x)^2] = M(x_x^2) - m_x^2.$$

Для дискретной случайной величины дисперсия вычисляется по формуле:

$$D_x \equiv M[(X - m_x)^2] \equiv \sum_{i=1}^n (x_i - m_x)^2 \cdot p(x_i) \quad (1.17)$$

Для непрерывной случайной величины дисперсия вычисляется по формуле:

$$D_x \equiv M[(X - m_x)^2] \equiv \int_{-\infty}^{\infty} (x - m_x)^2 \cdot f(x) dx. \quad (1.18)$$

Из приведенных формул следует, что размерность дисперсии есть размерность случайной величины в квадрате. Для практических нужд это не всегда удобно. В этой связи чаще ис-

пользуется так называемое **среднеквадратическое отклонение**, имеющее общепринятое обозначение σ_x . По определению

$$\sigma_x = +\sqrt{D_x}.$$

Очевидно, что для распределений, изображенных на рис.1.3, выполняется неравенство $\sigma_{x_2} > \sigma_{x_1}$.

В качестве примера найдем дисперсию и среднеквадратическое отклонение для первой задачи п.1.2.2.2:

$$D_x = \sum_{i=0}^4 (x_i - 1.6)^2 \cdot C_4^i \cdot (0.4)^i \cdot (0.6)^{4-i} = 0.96;$$

$$\sigma_x = \sqrt{0.96} \approx 0.98.$$

1.2.4. Наиболее распространенные законы распределения вероятностей

В рамках теории вероятностей, на основании обобщения знаний о случайных явлениях в природе и человеческом обществе, построен ряд моделей распределения вероятностей, которые в некоторых случаях удовлетворительно описывают исследуемые закономерности. Для каждой модели (закона распределения) установлены и условия ее применимости.

Если при рассмотрении некоторого явления исследователь считает, что имеющие место условия совпадают с условиями применимости того или иного закона, то можно воспользоваться соответствующим законом распределения и всеми знаниями о нем, накопленными в рамках теории вероятности.

Наиболее распространенные из законов распределения рассматриваются в этом параграфе.

1.2.4.1. Биномиальное распределение. Пусть реализуется схема опытов Бернулли: проводится n одинаковых независимых опытов, в каждом из которых событие A может появиться с постоянной вероятностью p . Число появлений события A в этих n опытах есть дискретная случайная величина X , возможные значения которой: 0; 1; 2; ...; m ; ...; n .

Законом распределения этой случайной величины может явиться формула, связывающая m — число появлений события A в конкретной серии из n опытов с вероятностью того, что $X = m$.

Такая вероятность, а следовательно, закон распределения, задается формулой Бернулли:

$$P(X=m) = P_{m,n} = C_n^m p^m (1-p)^{n-m}.$$

Числовые характеристики случайной величины X , распределенной по биномиальному закону:

$$m_x = np, D_x = np(1-p), \sigma_x = \sqrt{np(1-p)}.$$

Задача. Автомобиль, подъезжая к перекрестку, может продолжить движение по любой из трех дорог А, В или С с одинаковой вероятностью. К перекрестку подъезжают пять автомобилей. Найти среднее число автомашин, которое поедет по дороге А и вероятность того, что по дороге В поедет три автомобиля.

Решение. Число автомашин, проезжающих по каждой из дорог, является случайной величиной. Если предположить, что все подъезжающие к перекрестку автомобили совершают поездку независимо друг от друга, то эта случайная величина распределена по биномиальному закону с $n = 5$ и $p = 1/3$.

Следовательно, среднее число автомашин, которое проследует по дороге А, есть $m_x = 5/3$, а искомая вероятность —

$$P(X=3) = C_5^3 \cdot (1/3)^3 \cdot (2/3)^2 = 0.165.$$

1.2.4.2. Распределение Пуассона. Пусть событие А может появиться в любой момент времени. При этом выполнены следующие условия:

- 1) События происходят независимо друг от друга;
- 2) Появление события А на данном отрезке времени не зависит от расположения временного отрезка на оси времени;
- 3) Вероятность появления события А за бесконечно малый интервал времени Δt более одного раза есть бесконечно малая величина по сравнению с Δt (в этой связи закон Пуассона называют законом редких событий).

Число появлений события А за выбранный промежуток времени t подчиняется закону Пуассона:

$$P(X=m) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}. \quad (1.19)$$

Здесь λ — среднее число событий А, появляющихся за единицу времени.

Этот закон однопараметрический, т. е. для его задания требуется знать только один параметр λ . Можно показать, что математическое ожидание и дисперсия в законе Пуассона численно равны: $m_x = D_x = \lambda$.

Одним из классических примеров применения закона Пуассона является описание числа запросов на соединение, поступающих на телефонную станцию.

Задача. Пусть в середине рабочего дня среднее число запросов равняется 2 в секунду. Какова вероятность того, что: 1) за секунду не поступит ни одной заявки? 2) за две секунды поступит 10 заявок?

Решение. Поскольку правомерность применения закона Пуассона не вызывает сомнения и его параметр задан ($\lambda = 2$), то решение задачи сводится к прямому применению формулы Пуассона (1.19):

$$1) t=1, m=0;$$

$$P(X=0) = \frac{(2)^0}{0!} e^{-2} = 0.135.$$

$$2) t=2, m=10;$$

$$P(X=10) = \frac{(2 \cdot 2)^{10}}{10!} e^{-4} = 0.005.$$

1.2.4.3. Закон равномерной плотности. Пусть непрерывная случайная величина X может принимать любые значения лишь на отрезке $[a, b]$ и нет оснований считать, что появление одних возможных значений вероятней других. При выполнении этих условий говорят, что X распределена с равномерной плотностью. В этой связи целесообразно считать, что плотность вероятности $f(x)$ имеет вид:

$$f(x) = \begin{cases} c, & \text{если } a \leq x \leq b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases}$$

График такой функции представлен на рис.1.4. Поскольку площадь, ограниченная любой кривой распределения, равна 1, легко найти значение константы c из равенства $c(b-a) = 1$:

$$c = \frac{1}{b-a}.$$

Теперь можно сказать, что случайная величина X распределена на отрезке $[a, b]$ равномерно, если:

$$f(x) = \begin{cases} c = \frac{1}{b-a}, & \text{если } a \leq x \leq b, \\ 0, & \text{если } x < a, x > b. \end{cases} \quad (1.20)$$

Основные числовые характеристики равномерно распределенной случайной величины.

$$m_x = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{b+a}{2},$$

$$D_x = \int_{-\infty}^{\infty} (x-m_x)^2 \cdot f(x) dx = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \cdot \frac{1}{b-a} dx = \frac{(b-a)^2}{12},$$

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}.$$

Рассмотрим пример. Минутная стрелка часов делает скачок на соседнее деление, когда реальное время превышает указанное значение на полминуты. При взгляде на часы фиксируется показываемое ими время. Какова средняя ошибка в показаниях таких часов и каков разброс этой ошибки?

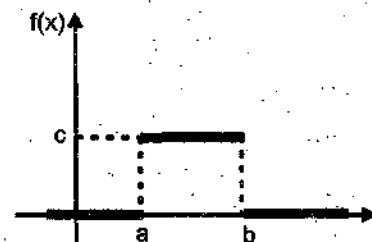


Рис. 1.4. График равномерной плотности распределения

Решение. В каждый момент времени показания часов есть случайная величина, показывающая реальное время с некоторой ошибкой. Взгляд на часы производится в случайно выбранный момент, поэтому целесообразно предположить, что ошибка в показаниях часов имеет равномерную плотность распределения.

Так как рассогласование между реальным временем и показаниями часов находится в пределах от -0.5 до $+0.5$, то следует положить $a = -0.5$, $b = +0.5$. Следовательно:

$$m_x = \frac{0.5 + (-0.5)}{2} = 0$$

Это означает, что систематическая ошибка отсутствует:

$$\sigma_x = \frac{0.5 - (-0.5)}{2\sqrt{3}} = 0.289.$$

1.2.4.4. Нормальный закон распределения. Непрерывная случайная величина X распределена по нормальному закону, если ее плотность вероятности описывается функцией:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где m — математическое ожидание; σ — среднеквадратическое отклонение. График плотности вероятности для нормального закона показан на рис. 1.3.

Нормальный закон возникает в тех случаях, когда случайная величина X есть сумма большого числа случайных величин, распределенных по произвольному закону, но каждая из них не является доминирующей. Наиболее ярким примером является ошибка, возникающая при различных измерениях (длины, объема, массы и т. п.). Действительно, если измерительный прибор хорошо отрегулирован, то он не дает существенных систематических ошибок (иначе его следовало бы отрегулировать). Получаемые же при каждом измерении ошибки складываются из влияния множества факторов, устранить которые практически невозможно. Они зависят от изменений температуры, давления, влажности и т. п. Иногда ошибки складываются, усиливая друг друга, а иногда компенсируя одна другую.

Для вычислений вероятности попадания случайной величины в заданный промежуток возможных значений используются приведенная функция плотности вероятности и приведенная функция распределения (функция распределения для так называемой нормированной случайной величины с $m = 0$ и $\sigma = 1$):

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; \quad \Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Функция $\Phi^*(x)$ обладает следующими свойствами: $\Phi^*(0) = 0.5$, $\Phi^*(x) = 1 - \Phi^*(-x)$.

Значения этой функции табулированы. Задавая x , по таблице легко найти $\Phi(x)$. Если случайная величина распределена по нормальному закону с параметрами m и σ , то:

$$P(\alpha < x < \beta) = \Phi^*\left(\frac{\beta-m}{\sigma}\right) - \Phi^*\left(\frac{\alpha-m}{\sigma}\right). \quad (1.21)$$

При вычислениях, связанных с нормальным законом, часто используют интеграл Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

При этом следует иметь в виду, что $\Phi^*(x) = \Phi(x) + 0.5$, и тогда формула (1.21) остается справедливой.

Рассмотрим задачу. При изготовлении бумаги наблюдается отклонение ее плотности от номинала, равного 100 г на квадратный метр. Если считать, что плотность бумаги есть нормально распределенная величина с $\sigma = 5$ г, найти вероятность того, что плотность бумаги будет отличаться от номинала не более, чем на 2 г, и не менее, чем на 3 г.

Решение задачи дается непосредственным использованием формулы (1.21):

$$\begin{aligned} \Phi^*(97 < x < 102) &= \Phi^*\left(\frac{102-100}{5}\right) - \Phi^*\left(\frac{97-100}{5}\right) = \\ &= \Phi^*(0.4) - \Phi^*(-0.6) = 0.6554 - 0.2743 = 0.3811. \end{aligned}$$

На практике удобно использовать правило «три сигма», которое гласит: «с вероятностью большей, чем 0.997, случайная величина, распределенная по нормальному закону, будет принимать значения в промежутке $(m_x - 3\sigma_x, m_x + 3\sigma_x)$ ».

Пример. Имеется партия изделий, в которой могут попадаться качественные и бракованные. Число бракованных изделий — случайная величина, характеризующаяся так: среднее число бракованных изделий (математическое ожидание) составляет 12% и среднеквадратическое отклонение — 3%. Отобрано 100 изделий. Какое число бракованных изделий окажется среди них?

Решение. Должно быть понятным, что точно ответить на такой вопрос в принципе невозможно. Однако используя правило «три сигма», легко найти следующий ответ: можно быть практически уверенным, что бракованных деталей будет не менее $12 - 3 \cdot 3 = 3$ и не более $12 + 3 \cdot 3 = 21$. Формально это можно записать так: $P(3 < X < 22) > 0.997$.

Глава 1.3. Предельные теоремы теории вероятностей

Математические законы теории вероятностей получены в результате обобщения закономерностей, свойственных массовым явлениям в обществе и в природе. Массовость понимается как большое число повторений опытов в одинаковых или сходных условиях. Было замечено, что при массовых явлениях результаты отдельных опытов практически не влияют на некоторые средние характеристики этих явлений. Этот феномен известен как устойчивость средних: «При очень большом числе испытаний средние характеристики наблюдаемых явлений перестают быть случайными и могут быть предсказаны со сколь угодно высокой точностью».

Еще в глубокой древности люди заметили феномен устойчивости средних. Однако только в девятнадцатом веке ученые нашли общие условия, выполнение которых обязательно влечет за собой статистическую устойчивость средних.

1.3.1. Неравенство Чебышева

Возможные значения случайной величины X группируются вокруг своего математического ожидания m_x . Однако в каждом конкретном опыте X может принять значение и весьма далекое от m_x . Если известен закон распределения X , то оценить вероятность далеких отклонений несложно.

Пусть задана некоторая величина ε . Событие $(|X - m_x| > \varepsilon)$ означает, что значение X отклонилось от своего математического ожидания больше, чем на ε . Такие отклонения будем называть «далекими». Если ограничиться только непрерывным случаем, то вероятность получить далекие отклонения можно вычислить так:

$$P(|X - m_x| > \varepsilon) = \int_{\Omega} f(x) dx,$$

где Ω — область далеких отклонений $\Omega: |X - m_x| > \varepsilon$.

Однако закон распределения известен далеко не всегда. Хотелось бы уметь оценивать вероятность таких событий, опираясь, например, лишь на знание дисперсии D_x . Именно такая задача и решается с помощью неравенства Чебышева.

Дисперсия случайной величины X , распределенной по непрерывному закону $f(x)$, равна:

$$D_x = \int_{-\infty}^{\infty} (x - m_x)^2 \cdot f(x) dx \geq \int_{\Omega} (x - m_x)^2 \cdot f(x) dx. \quad (1.22)$$

Если под интегралом заменить $(x - m_x)^2$ на величину ε^2 , то неравенство в (1.22) лишь усилится. Следовательно:

$$D_x \geq \int_{\Omega} (x - m_x)^2 \cdot f(x) dx \geq \varepsilon^2 \cdot \int_{\Omega} f(x) dx = \varepsilon^2 \cdot P(|X - m_x| \geq \varepsilon).$$

Отсюда имеем неравенство Чебышева:

$$P(|X - m_x| \geq \varepsilon) \leq \frac{D_x}{\varepsilon^2}. \quad (1.22)$$

Задача. Оценить вероятность того, что значения случайной величины будут отклоняться от ее математического ожидания меньше, чем на $3\sigma_x$.

Решение. Нам необходимо оценить вероятность следующего события $|X - m_x| \leq 3\sigma_x$. Это событие будет противоположным к тому, что рассматривается в неравенстве Чебышева, если положить $\varepsilon = 3\sigma_x$. Переход к противоположному событию осуществляется по правилу $P(x > a) = 1 - P(x \leq a)$. Следовательно, используя (1.23) при $\varepsilon = 3\sigma_x$, имеем:

$$P(|X - m_x| \leq 3\sigma_x) > 1 - \frac{D_x}{9\sigma_x^2} = \frac{8}{9}.$$

Мы получили очень важный результат: «Каков бы ни был закон распределения для X , вероятность того, что отклонение этой случайной величины от ее математического ожидания будет в пределах трех среднеквадратических ожиданий, не может быть менее $8/9$ ». Иначе говоря, значения случайной величины весьма редко отклоняются от своего математического ожидания больше, чем на «три сигма». Это правило также получило название **правило «трех σ »** (см. раздел 1.2.4.4).

На практике, получив оценки для m_x и σ_x , легко оценить интервал, в котором будут находиться практически все значения случайной величины. Этот интервал задается условием:

$$m_x - 3\sigma_x < x < m_x + 3\sigma_x. \quad (1.24)$$

Величина $8/9$ является нижней границей оценки. Для конкретных законов распределения попадание в интервал (1.24) значительно больше, чем $8/9$. Например, для нормального закона вероятность попадания в такой интервал оценивается величиной 0.997.

1.3.2. Закон больших чисел

Математическим основанием того факта, что значения случайной величины группируются около некоторых постоянных величин, является закон больших чисел.

Исторически первой формулировкой закона больших чисел стала теорема Бернулли: «При неограниченном увеличении числа одинаковых и независимых опытов n частота появления события A сходится по вероятности к его вероятности», т. е.:

$$\lim_{n \rightarrow \infty} P(|p^* - p| < \varepsilon) = 1, \quad (1.25)$$

где частота появления события A в n опытах $p^* = \frac{m}{n}$.

Содержательно выражение (1.25) означает, что при большом числе опытов частота события A может заменять неизвестную вероятность этого события и, чем больше число проведенных опытов, тем ближе p^* к p . Интересен исторический факт. К. Пирсон бросал монету 12000 раз, и герб у него выпал 6019 раз (частота 0.5016). При бросании этой же монеты 24000 раз он получил 12012 выпадений герба, т. е. частоту 0.5005.

Наиболее важной формой закона больших чисел является **теорема Чебышева**: *при неограниченном возрастании числа независимых, имеющих конечную дисперсию и проводимых в одинаковых условиях опытов среднее арифметическое наблюдаемых значений случайной величины сходится по вероятности к ее математическому ожиданию. В аналитической форме эта теорема может быть записана так:*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n x_i - m_x\right| < \varepsilon\right) = 1. \quad (1.26)$$

Теорема Чебышева кроме фундаментального теоретического значения имеет и важное практическое применение, например, в теории измерений. Проведя n измерений некоторой величины x , получают различные несовпадающие значения $x_1, x_2, \dots, x_n$. За приближенное значение измеряемой величины x принимают среднее арифметическое наблюдаемых значений:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

При этом, чем больше будет проведено опытов, тем точнее будет полученный результат. Дело в том, что дисперсия величины $\bar{x}$ убывает с возрастанием числа проведенных опытов; т. к. $D(x_1) = D(x_2) = \dots = D(x_n) \equiv D(x)$, то

$$D(\bar{x}) = \frac{D(x)}{n}. \quad (1.27)$$

Соотношение (1.27) показывает, что и при неточности приборов измерения (большая величина D_x) за счет увеличения количества измерений можно теоретически получать результат со сколь угодно высокой точностью.

1.3.3. Центральная предельная теорема.

Очень часто значение, которое принимает исследуемая величина Y , является результатом суммарного воздействия ряда случайных величин $x_1, x_2, \dots, x_n$. Если эти величины слабо зависят друг от друга и ни одна из них не является доминирующей, то их сумма имеет закон распределения весьма близкий к нормальному. Такая ситуация наиболее характерна для процессов измерения. Дело в том, что результаты измерения складываются под влиянием многочисленных независимых элементарных причин, причем влияние каждой из причин на суммарный результат незначительно. Например, на величину температуры в конкретном месте комнаты оказывает влияние близость отопительных приборов, наличие и расположение в комнате людей, потоки воздуха, влажность, температура воздуха на улице, наличие источников света и их мощность, расположение окон, дверей и т. д.

Все формы центральной предельной теоремы (ЦПТ) посвящены установлению условий возникновения нормального закона. Наиболее простой формой ЦПТ является теорема **Ляпунова**: «Если $X_1, X_2, \dots, X_n$ — независимые случайные величины, имеющие один и тот же закон распределения, то при неограниченном увеличении числа n закон распределения суммы $Y = X_1 + X_2 + \dots + X_n$ неограниченно приближается к нормальному».

На практике при $n > 10$ закон распределения суммы случайных величин, имеющих какой угодно (возможно и неизвестный) закон распределения, считают нормальным. Рассмотрим пример.

Случайная величина Y есть сумма 24 случайных величин X , распределенных по закону равномерной плотности на отрезке $[0, 1]$. Найти вероятность того, что Y примет значение в пределах от 10 до 13.

Ввиду того, что закон распределения величин X равномерный, легко найти их математические ожидания и дисперсию:

$$m_x = (1 + 0)/2 = 1/2, \quad D_x = (1 - 0)^2/12 = 1/12.$$

Y есть сумма 24 случайных величин. Поэтому можно считать ее закон распределения близким к нормальному с параметрами:

$$m_y = \sum_{i=1}^{24} m_x = 24 \cdot 0.5 = 12; \quad D_y = \sum_{i=1}^{24} D_x = 24 \cdot \frac{1}{12} = 2.$$

Требуемая вероятность может быть найдена при помощи известной формулы (см. (1.21)):

$$P(10 < Y < 13) = \Phi \left(\frac{13 - 12}{\sqrt{2}} \right) - \Phi \left(\frac{10 - 12}{\sqrt{2}} \right) = 0.7$$

Частным случаем ЦПТ является теорема **Лапласа**: «Если производится n опытов, в каждом из которых событие A появляется с вероятностью p и не появляется с вероятностью $q = 1 - p$, то справедливо соотношение:

$$P(a < Y < b) = \Phi \left(\frac{b - np}{\sqrt{npq}} \right) - \Phi \left(\frac{a - np}{\sqrt{npq}} \right) \quad (1.28)$$

Эта формула позволяет вести расчеты, когда реализуется система случаев при большом числе n . Рассмотрим задачу. Производится компьютерный набор текста из 100 страниц. Вероятность сделать ошибку при наборе одной страницы равна 0.2. Найти вероятность того, что при наборе 100 страниц ошибок будет не более 25.

Решение данной задачи сводится к непосредственному применению формулы (1.28) при $p = 0.2$, $q = 0.8$ и $n = 100$, т. е.

$$P(0 < Y < 25) = \Phi \left(\frac{25 - 20}{4} \right) - \Phi \left(\frac{0 - 20}{4} \right) = 0.39.$$

Предельные теоремы образуют своеобразный мостик между теорией вероятностей как математической наукой и одной из наиболее важных и имеющих практическое применение областей современных знаний — прикладной статистикой.

Глава 1.4. Обработка результатов измерений (анализ данных)

Мы постоянно сталкиваемся с событиями, имеющими неопределенный исход. Например, сколько покупателей сегодня посетит магазин, какая завтра будет погода, как изменится к концу месяца курс акций, каков будет состав населения города к началу следующего столетия? Неопределенность в ответах на эти и подобные им вопросы, как правило, заложена в сущности самих этих явлений.

Если в повседневной деятельности, сталкиваясь с неопределенным явлением, человек может сделать выбор на основе своей интуиции и опыта, то решение сложных задач в науке, предпринимательстве и политике требует привлечения серьезных методов анализа информации, обоснованности принимаемых решений и доказательства их правильности.

Нас окружают события, которые мы воспринимаем в виде огромного количества фактов и наблюдений, являющихся, в свою очередь, результатом действий множества других факторов. Часть этих факторов лежит в основе наблюдаемого явления, а факторы другой части второстепенны, несущественны и лишь затеняют рассматриваемое явление. Проблема состоит в том, чтобы, отбросив все лишнее, отразить существенную информацию в такой форме, чтобы можно было принять наиболее правильное решение.

Научная дисциплина, разрабатывающая и систематизирующая понятия, приемы, математические методы и модели, предназначенные для организации сбора, систематизации и обработки данных с целью их интерпретации и получения выводов, называется **анализом данных**. Анализ данных, или прикладная статистика — одна из наиболее важных имеющих практическое применение областей знаний современной математики.

Пусть проводится некоторый опыт (рассматривается некоторое явление). Все возможные результаты опыта образуют так называемую **генеральную совокупность**. Если в данной серии опытов получено несколько результатов, то эта совокупность данных называется **выборкой**. Свойства и характеристики генеральной совокупности обычно неизвестны. Необходимо, исследуя данные выборки, как можно точнее оценить свойства и характеристики генеральной совокупности. Решение такого

рода задач — предмет рассмотрения дисциплины, называемой математической статистикой. Отметим, что она применима только к *массовым явлениям, имеющим вероятностную природу*.

Вместе с тем, встречается множество ситуаций, когда полученные данные, даже если их много, характеризуют не все, а только некоторые объекты. В таких случаях мы не всегда вправе распространять результаты обработки данных на другие (нерассмотренные) объекты. Например, проведено обследование приморских городов страны. Полученные характеристики будут справедливы только для этих городов и не могут быть в полной мере распространены на города, расположенные в глубинных районах.

Другой пример: проведено исследование населения городов с числом жителей более миллиона. Полученные данные нельзя использовать для прогнозирования ситуации в населенных пунктах с другой численностью населения. Или, в 12 часов дня проведен опрос населения по домашним телефонам, хотя в это время основная масса трудоспособного населения не находится дома. Данные опроса нельзя распространить на всех жителей города.

Другой пример: рассматриваются произведения поэтов «серебряного века». Цель рассмотрения — выявление групп поэтов близких по стилю, построению стихов, их размеру, тематической направленности и т. д. Полученные при этом результаты будут справедливы только для рассматриваемой группы поэтов (только для данной выборки). Надо понимать, что данные такого типа имеют другую, не вероятностную природу. Отличительной особенностью таких данных является то, что они уникальны и многократное повторение анализа (опытов) не принесет существенно новой информации. Полученные результаты, вообще говоря, нельзя экстраполировать на другие подобные объекты. Все это надо иметь в виду при интерпретации результатов обработки данных.

Анализу данных, имеющих *невероятностную природу*, посвящены методы кластерного анализа, распознавания образов и т. п. Но, тем не менее, и для анализа данных, имеющих не-вероятностную природу, широко применяются методы статистической обработки наблюдений. Достаточно полное представление о методах обработки данных дано, например, в работах [2,3].

1.4.1. Элементы математической статистики

Основной целью применения аппарата математической статистики является определение некоторых частных характеристик случайной величины, например математического ожидания и дисперсии или ее закона распределения в целом, с тем, чтобы иметь возможность при решении прикладных задач применять аппарат теории вероятностей.

Поскольку рассматривается только выборка (а не генеральная совокупность), при решении задач математической статистики всегда присутствует элемент случайности, и результаты получаются лишь с некоторой степенью уверенности.

На практике, в зависимости от поставленных целей, объемов выборки и имеющихся ресурсов, могут быть поставлены самые разнообразные вопросы. В общем же случае представляется целесообразным рассматривать три так называемые основные задачи математической статистики:

- * Определение вида закона распределения исследуемой случайной величины (задача сглаживания экспериментальных зависимостей);

- * Определение неизвестных параметров распределения;

- * Проверка правдоподобия гипотез.

1.4.1.1. Первичная обработка данных. Решение любой задачи математической статистики предваряет первичная обработка наблюдений. Пусть в результате проведения *в одинаковых условиях* независимых опытов получено n значения исследуемой случайной величины X .

i	1	2	3	...	n
x_i	x_1	x_2	x_3	...	x_n

(1.29)

Расположенные в виде таблицы (1.29) полученные данные, где x_i — результат i -го опыта, называют **простой статистической совокупностью**. Простая статистическая совокупность — первичная форма записи статистического материала. Данные могут быть упорядочены не по номерам опытов, а, например, по мере возрастания полученных значений X или в любом другом порядке.

Если получено большое число данных, а в статистике оперируют, как правило, сотнями и тысячами значений, то их

преобразуют в так называемый статистический ряд. Для этого весь диапазон полученных значений случайной величины X разбивают на разряды (подинтервалы). Для удобства обычно вводят $k = 6 - 20$ разрядов (большее число разрядов сложно анализировать, а при малом их числе происходит излишнее огрубление исследуемых зависимостей). Для каждого разряда подсчитывается число попавших в него значений k_i из совокупности (1.29). Если каждое из этих значений разделить на общее количество имеемых данных n , то получим частоты попадания в соответствующий разряд: $p_i^* = k_i/n$ (здесь и далее знак $*$ говорит о том, что это статистический аналог соответствующего понятия теории вероятностей). Полученные таким образом данные образуют таблицу называемую **статистическим рядом**.

I_i	x_0-x_1	x_1-x_2	...	$x_{k-1}-x_k$
k_i	k_1	k_2	...	k_k
p_i^*	p_1^*	p_2^*	...	p_k^*

(1.30)

Графическое изображение статистического ряда называется **гистограммой** (рис.1.5).

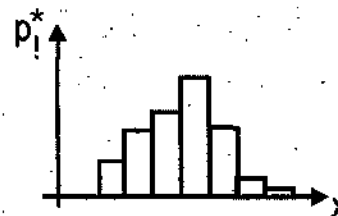


Рис. 1.5. Гистограмма

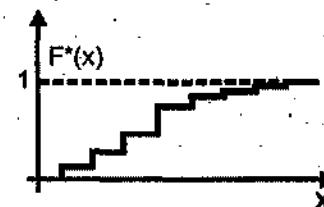


Рис. 1.6. Статистическая функция распределения

Данные статистического ряда можно использовать для построения **статистической функции распределения**:

$$F^*(x) = P^*(X < x).$$

График этой функции, изображенной на рис.1.6, всегда представляет собой характерную ступенчатую функцию.

Полученные в результате первичной обработки данные используются для решения задач математической статистики.

1.4.1.2. Сглаживание экспериментальных зависимостей. Полученные характеристики являются лишь статистическими аналогами исследуемой случайной величины, и им в большей или меньшей мере присущи черты случайности. Так, гистограмма дает наглядное представление о том, как часто случайная величина в проведенных опытах принимала значения из соответствующего разряда.

Очевидно, что при увеличении числа опытов и увеличении числа разрядов, гистограмма для непрерывной случайной величины X будет приближаться к кривой плотности вероятности этой случайной величины. Аналогично, статистическая функция при увеличении числа проведенных опытов и разрядов приближается к универсальному закону — функции распределения.

Задача заключается в подборе математической модели — закона распределения исследуемой случайной величины X . При этом модель должна в некотором смысле слова наилучшим образом описывать свойства этой величины и отражать лишь существенные стороны ее распределения. Сложность задачи заключается в том, чтобы найти сглаживающую кривую только на основании имеющегося экспериментального материала. Формально эта задача является неопределенной, поскольку слова «наилучшим образом» могут пониматься по-разному. Математика разработала различные подходы к решению такого рода задач (см., например, [4]). Ниже мы рассмотрим один из распространенных аналитических методов, который может с успехом применяться для сглаживания зависимостей.

На практике класс сглаживаемых функций получают на основании априорных сведений о характере исследуемой величины X .

Так, если речь идет об измерениях, то сглаживающая кривая для гистограммы, как правило, есть кривая нормального закона распределения; если речь идет о времени телефонного разговора, то это кривая экспоненциального закона распределения и т. п. При большом объеме экспериментальных данных и значительном числе разрядов формы гистограммы и статистической функции распределения могут служить основанием для выбора типа сглаживающей функции.

Задача. Изготавливается тираж книги, один из размеров которой должен быть равен 20 см. Выбрано и измерено 100 готовых экземпляров. Результаты измерений сведены в таблицу, где x_i — отклонение от заданного размера в сотых долях мм i -той книги

i	x_i	i	x_i	i	x_i	i	x_i	i	x_i
1	32	21	101	41	88	61	47	81	27
2	117	22	57	42	57	62	43	82	46
3	63	23	85	43	95	63	128	33	61
4	114	24	43	44	88	64	77	84	76
5	2	25	56	45	70	65	89	85	81
6	54	26	75	46	59	66	50	86	111
7	19	27	109	47	37	67	63	87	21
8	78	28	73	48	90	68	82	88	99
9	95	29	49	49	83	69	66	89	61
10	105	30	57	50	107	70	49	90	79
11	39	31	39	51	77	71	116	91	86
12	87	32	66	52	85	72	67	92	38
13	38	33	84	53	118	73	105	93	79
14	127	34	56	54	79	74	67	94	41
15	68	35	39	55	36	75	101	95	137
16	112	36	55	56	29	76	49	96	103
17	73	37	86	57	84	77	98	97	27
18	90	38	72	58	65	78	106	98	68
19	105	39	78	59	47	79	87	99	95
20	68	40	91	60	30	80	56	100	119

Построить статистический ряд, гистограмму и статистическую функцию распределения.

Решение. Из анализа данных таблицы можно сделать вывод о том, что целесообразно разбить интервал наблюдаемых отклонений на семь разрядов и построить следующий статистический ряд:

I_i	0-20	20-40	40-60	60-80	80-100	100-120	120-140
k_i^*	2	13	19	25	22	16	3
p_i^*	0.02	0.13	0.19	0.25	0.22	0.16	0.03

Гистограмма, построенная на основе этого статистического ряда, представлена на рис. 1.7.

Статистическая функция распределения данной выборки имеет следующий вид (x в см):

$$F^*(x) = \begin{cases} 0; & x < 0.02 \\ 0.02; & 0.02 < x \leq 0.04 \\ 0.15 & 0.04 < x \leq 0.06 \\ 0.34; & 0.06 < x \leq 0.08 \\ 0.59; & 0.08 < x \leq 0.10 \\ 0.81; & 0.10 < x \leq 0.12 \\ 0.97; & 0.12 < x \leq 0.14 \\ 1.00; & x > 0.14 \end{cases}$$

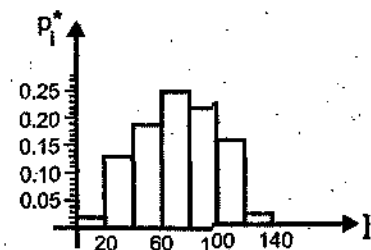


Рис. 1.7. Гистограмма к задаче п. 1.4.1

1.4.1.3. Определение неизвестных параметров закона распределения. Высказанная гипотеза о виде сглаживающей кривой еще не дает возможности применять количественные методы теории вероятностей. Так, например, если установлено, что X имеет нормальный закон распределения —

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

то для нахождения вероятности попадания в интервал (a, b) совершенно необходимо знать и математическое ожидание (m) , и среднеквадратическое отклонение (σ) случайной величины X .

Определение неизвестных параметров — следующая основная задача математической статистики. Решение этой задачи часто осуществляется при помощи **метода моментов**. Согласно этому методу, неизвестные параметры выбираются таким образом, чтобы числовые характеристики X были равны их статистическим аналогам.

Для математического ожидания статистическим аналогом является среднее арифметическое (среднее выборочное):

$$m_x^* = \bar{x} = \sum_{i=1}^n x_i \cdot p_i^*,$$

где x_i — середины соответствующих интервалов в гистограмме.

Для дисперсии — статистическая дисперсия (выборочная дисперсия):

$$D_x^* = \sum_{i=1}^n (x_i - m_x^*)^2 p_i^*.$$

Отсюда находится статистическое среднеквадратическое отклонение:

$$\sigma_x^* = \sqrt{D_x^*}.$$

Для задачи п. 1.4.1.2. можно выдвинуть гипотезу о том, что распределение размера книги подчинено нормальному закону. Параметры этого закона при помощи метода моментов можно определить, используя данные ряда распределения:

$$m_x^* = \sum_{i=1}^n x_i \cdot p_i^* = 10 \cdot 0.02 + 30 \cdot 0.13 + 50 \cdot 0.19 + 70 \cdot 0.25 + 90 \cdot 0.22 + 110 \cdot 0.16 + 130 \cdot 0.03 = 72.4.$$

$$D_x^* = \sum_{i=1}^n (x_i - 72.4)^2 p_i^* = 802.2$$

$$\sigma_x^* = \sqrt{D_x^*} = 28.3$$

Теперь можно сказать, что средний размер выпускаемых книг равен приблизительно 20.72 см и разброс размеров в обе стороны (по правилу «три сигма») практически не превысит 0.85 см.

1.4.1.4. Проверка правдоподобия гипотез. Предположим, что сформулирована гипотеза H о законе распределения исследуемой случайной величины X (ее будем называть теоретическим распределением). Требуется проверить, не противоречит ли эта выдвинутая гипотеза имеющимся выборочным данным.

Чтобы иметь основания принять или опровергнуть гипотезу, необходимо выработать некоторый критерий, позволяющий делать корректные выводы. В математической статистике при проверке гипотез выбрать критерий значит *задать некоторое критическое значение вероятности ошибочного отклонения проверяемой гипотезы*. Эта вероятность называется **уровнем значимости** и обозначается α . Чем весомее для нас потери от ошибочного отклонения гипотезы H , тем меньше α . Обычно эта величина имеет одно из стандартных значений: 0.1, 0.05 или 0.01. Выбрав, например, $\alpha = 0.1$, мы, в среднем, десять раз из ста будем отбрасывать верную гипотезу. Если же выбрать $\alpha = 0.01$, то из ста случаев мы (в среднем) только один раз отбросим верную гипотезу.

Основная идея проверки гипотез заключается в следующем. Выбирается уровень значимости α и задается некоторая мера расхождения U , характеризующая степень расхождения теоретического и статистического распределений. В предположении, что гипотеза верна, вычисляется вероятность того, что

только за счет чисто случайных причин (связанных с ограниченностью числа проведенных опытов) мера расхождения U (критерий согласия) окажется не меньше, чем вычисленная по результатам наблюдений величина этой меры u , т. е. $P(U \geq u)$. Если эта вероятность не меньше α , то говорят, что экспериментальные данные не противоречат выдвинутой гипотезе. По такой схеме проводится проверка практически любых статистических гипотез.

Для проверки гипотезы о законе распределения исследуемой случайной величины применяют различные критерии [4]. Чаще других применяется критерий согласия Пирсона. Здесь в качестве меры рассогласования U выбирается величина (обозначаемая χ^2), которая вычисляется по формуле:

$$U = \chi^2 = n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}, \quad (1.31)$$

где p_i — теоретическая вероятность, т. е. вероятность попадания в i -ый разряд, вычисленная по гипотетическому закону; p_i^* — статистическая (вычисленная по выборке) частота попадания в i -ый разряд.

Вычисленная по такой формуле величина рассогласования (называемая «хи-квадрат») является хорошо изученной случайной величиной, для распределения которой построены таблицы.

Схема применения критерия согласия χ^2 следующая:

1. По формуле (1.31) вычисляется мера рассогласования χ^2 .

2. Вычисляется число степеней свободы $r = k - s - 1$,

где k — число разрядов, на которые разбит диапазон наблюдаемых значений величины X ; s — число параметров теоретического закона распределения, которые определены по выборке. Например, если предполагаемое распределение нормальное, то по выборке определяют среднее значение $\bar{x}$ и среднее квадратическое отклонение σ_x^2 , т. е. $S = 2$, $r = k - 3$. Если предполагаемое распределение — закон Пуассона, то определяют среднее значение $\bar{x} = \lambda$, т. е. $S = 1$, $r = k - 2$.

3. По таблице распределения «хи-квадрат» для найденных χ^2 и r определяется вероятность P . Если эта вероятность не больше заранее выбранного уровня значимости α , то гипотеза отбрасывается как неправдоподобная. Если же $P > \alpha$, то говорят, что гипотеза не противоречит экспериментальным данным.

Для примера п. 1.4.1.2. проверим гипотезу о том, что размеры книги имеют нормальное распределение. Используя формулы приведенной функции Лапласа для интервалов, выбранных при формировании гистограммы, представленной на рис. 1.7, и

найденных ранее $m^* = 72.4$, $\sigma^* = 28.3$, найдем теоретические вероятности $\{p_i\}$. Затем по формуле (1.31) определим значение величины χ^2 , которая равна 6.6. Поскольку выбрано 7 разрядов и из выборки найдены оценки математического ожидания и среднеквадратического отклонения ($s = 2$), то $r = 4$. Теперь по таблице распределения хи-квадрат находим вероятность $P = 0.25$, что существенно больше, чем 0.1 ($\alpha = 0.1$). Следовательно, можно констатировать, что гипотеза о нормальном распределении ошибки размера книги не противоречит экспериментальным данным.

1.4.2. Обработка выборок

Для нахождения законов распределения как исчерпывающих характеристик случайной величины необходим обширный статистический материал, основанный на сотнях и тысячах экспериментальных данных. На практике же часто располагают выборкой из нескольких десятков (а иногда и единиц) наблюдений. В этом случае говорить об определении закона распределения не приходится. Ставится лишь задача приближенного определения наиболее важных характеристик случайной величины.

1.4.2.1. Оценки параметров случайных величин. Ограниченное число опытов приводит к тому, что получаемые на их основе значения параметров содержат элемент случайности. Например, монета подбрасывалась пять раз, и четыре раза выпадала цифра. Отсюда можно сделать вывод о том, что среднее число выпадений цифры в одном опыте равно 0.8. В то же время, очевидное значение этой характеристики — 0.5 (если монета и техника подбрасывания «правильные»). Однако к такому выводу приходят при очень большом количестве подбрасывания монеты. Так, Ж. Бюффон подбрасывал монету 4048 раз, и цифра выпадала в 1956 случаях, т. е. среднее число появлений цифры при одном подбрасывании было получено равным 0.48.

Случайное значение искомого параметра, вычисленного на основании данного статистического материала, называют **оценкой параметра**.

Пусть проведено n опытов и получены значения случайной величины ($x_1, x_2, \dots, x_n$). Требуется построить оценку (называемую иногда точечной оценкой) неизвестного параметра « γ », которую будем обозначать γ . Эта оценка есть функция опытных величин, т. е.

$$\gamma = \gamma(x_1, x_2, \dots, x_n).$$

Задача состоит в том, чтобы построить «наиболее точные оценки». Здесь под термином «наиболее точные» подразумевается предъявление к оценке параметра следующих требований:

1. Оценка параметра должна быть **состоятельной**. Это означает, что при увеличении числа опытов, оценка все более приближается (стремится по вероятности) к точному значению параметра:

$$\gamma \rightarrow \gamma \text{ при } n \rightarrow \infty.$$

2. Оценка параметра должна быть **несмещенной**, т. е. ее математическое ожидание должно равняться оцениваемому параметру:

$$M[\gamma] = \gamma.$$

Несмещенность означает отсутствие постоянной ошибки оценки в какую-то одну сторону.

3. Желательно, чтобы оценка параметра была **эффективной**, т. е. оценка по сравнению с другими должна обладать наименьшей дисперсией.

Сформулированным требованиям обладают следующие точечные оценки основных параметров распределения случайной величины — математического ожидания и дисперсии (см. выборочную среднюю и выборочную дисперсию):

$$\tilde{m} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1.32)$$

$$\tilde{D} \equiv S^2 = \frac{\sum_{i=1}^n (x_i - \tilde{m})^2}{n-1}, \quad \tilde{\sigma} \equiv S = \sqrt{\tilde{D}}. \quad (1.33)$$

Величина S^2 называется исправленной дисперсией, при $n > 30$ считается, что разница между исправленной и выборочной дисперсией незначительна.

1.4.2.2. Качество статистического оценивания. Пусть по результатам опыта построена точечная оценка γ некоторого параметра γ . Возникает вопрос: насколько эта оценка точна и достоверна.

Точность оценивания характеризуется абсолютной ошибкой, которая возникает, если параметр заменить его точечной оценкой, т. е.:

$$\Delta\gamma = |\gamma - \gamma|.$$

Однако, определить истинную погрешность $\Delta\gamma$ невозможно, поскольку γ — случайная величина, а сам параметр γ не известен.



Рис. 1.8. Доверительный интервал

В математической статистике задачу оценки точности решают следующим образом. Откладывают от γ вправо и влево такую величину $\varepsilon_\beta > 0$, что неслучайное (хотя и неизвестное) значение γ , как показано на рис. 4.4, оцениваемого параметра практически навер-

няка оказывается в пределах случайного интервала $I_\beta = (\gamma - \varepsilon_\beta, \gamma + \varepsilon_\beta)$.

Введем понятие доверительной вероятности и доверительного интервала. Вероятность β называется **доверительной**, если интересующее нас событие A , происходящее с этой вероятностью, считается практически гарантированным (достоверным). Как правило, β близка к единице и выбирается равной 0,9, 0,95, 0,99 или 0,999. Максимальной вероятной погрешностью статистической оценки называется ее максимально возможное отклонение $\varepsilon_\beta > 0$ от оцениваемого параметра, гарантируемое с вероятностью β .

Пусть задана доверительная вероятность (надежность) β . Тогда условие

$$P(|\gamma - \gamma| < \varepsilon_\beta) = \beta \quad (1.34)$$

определяет **доверительный интервал** $I_\beta = (\gamma - \varepsilon_\beta, \gamma + \varepsilon_\beta)$, который с вероятностью β накрывает неизвестный параметр γ . Можно сказать, что доверительный интервал — интервал значений параметра γ , совместимых (в вероятностном смысле) с опытными данными.

Очевидно, что выражение (1.34) по заданному β позволяет найти ε_β и наоборот, по заданной величине ε_β можно найти β . Доверительный интервал определяет точность найденной оценки параметра, а доверительная вероятность — ее достоверность.

Изложим методы нахождения доверительных интервалов для математического ожидания m ($\gamma \equiv m$), если известно, что случайная величина имеет нормальный закон распределения.

Малая выборка. На практике малой считается выборка с объемом меньше 30. Пусть по формуле (1.32) найдена оценка математического ожидания, а по формуле (1.33) — оценка дисперсии. Тогда для определения доверительного интервала используют распределение Стьюдента, функция распределения которого $S(t, n)$, в отличие от нормального закона, зависит только от размера выборки n . По распределению Стьюдента значение ε_β есть:

$$\varepsilon_\beta = t_\beta \frac{S}{\sqrt{n}}, \quad (1.35a)$$

где t_β — значение аргумента функции $S(t, n)$:

$$t_\beta = S^{-1}(t, n) \approx \arg S(t, n),$$

при данном n , для которого выполняется равенство $S(t, n) = \beta$.

Дадим таблицу значений t_β при разных объемах выборок n для доверительных вероятностей $\beta \approx 0,95; 0,99; 0,999$.

β	n	5	6	7	8	9	10	11	12	13
	0,95	2,78	2,57	2,45	2,37	2,31	2,26	2,23	2,20	2,18
β	0,99	4,60	4,03	3,71	3,50	3,36	3,25	3,17	3,11	3,06
	0,999	8,61	6,86	5,96	5,41	5,04	4,78	4,59	4,44	4,32
β	n	14	15	16	17	18	19	20	25	30
	0,95	2,16	2,15	2,13	2,12	2,11	2,10	2,09	2,06	2,05
	0,99	3,01	2,98	2,95	2,92	2,90	2,88	2,86	2,80	2,76
	0,999	4,22	4,14	4,07	4,02	3,97	3,92	3,88	3,75	3,66

Таким образом, чтобы найти доверительный интервал для оценки математического ожидания по заданной доверительной вероятности, необходимо:

1. По (1.32) вычислить оценку математического ожидания, по (1.33) вычислить оценку дисперсии S^2 , а по ней величину S ;

2. Для величины β по таблице значений функции распределения Стьюдента определить t_β ;

3. Вычислив по (1.35 а) ε_β , записать доверительный интервал $I_\beta = (m^- - \varepsilon_\beta, m^- + \varepsilon_\beta)$.

Большая выборка, т. е. объем выборки больше 30. Поскольку m^- есть сумма достаточно большого числа слагаемых, то, согласно закону больших чисел, она приближенно распределена по нормальному закону. Запишем условие (1.34), выразив правую часть через функцию Лапласа:

$$P(m^- - m) < \varepsilon_\beta = \beta = 2\Phi\left(\frac{\varepsilon_\beta}{\sigma_{m^-}}\right) \quad (1.36)$$

где среднеквадратическое отклонение случайной величины

можно приближенно положить равным $\sigma_{m^-} = \sqrt{\tilde{D}/n}$.

Решая уравнение (1.36) относительно неизвестного ε_β , получим

$$\varepsilon_\beta = \sigma_{m^-} \cdot \Phi^{-1}\left(\frac{1+\beta}{2}\right) = \sigma_{m^-} \cdot \arg \Phi\left(\frac{1+\beta}{2}\right) \equiv \sigma_{m^-} \cdot t_\beta. \quad (1.35 б)$$

Таким образом, чтобы найти доверительный интервал для оценки математического ожидания по заданной доверительной вероятности, необходимо:

1. По (1.32) вычислить оценку математического ожидания

m^- , а по (1.33) величину $\sigma_{m^-} = \sqrt{\tilde{D}/n}$;

2. Для величины $(1+\beta)/2$ по таблице приведенной функции распределения Лапласа определить t_β ;

3. Вычислив по (1.35 б) ε_β , записать доверительный интервал

$$I_\beta = (m^- - \varepsilon_\beta, m^- + \varepsilon_\beta).$$

Для оценки дисперсии ($\gamma = D$), если проделать аналогичные выкладки, доверительный интервал равен:

$$I_\beta = (D^- - \varepsilon_\beta, D^- + \varepsilon_\beta),$$

где максимальная вероятностная погрешность ε_β вычисляется по формуле:

$$\varepsilon_\beta = t_\beta \cdot \sqrt{2/(n-1)} \cdot \tilde{D}.$$

Приведем сокращенную таблицу значений β и t_β .

β	0,80	0,82	0,84	0,86	0,88	0,90
t_β	1,282	1,340	1,404	1,475	1,554	1,643
β	0,92	0,94	0,95	0,96	0,98	0,99
t_β	1,750	1,880	1,960	2,053	2,325	2,576

В качестве примеров найдем доверительные интервалы для оценок математического ожидания и дисперсии для выборки, приведенной в п. 1.4.1.2, т. е. если $m_x^- = 72,4$ и $D_x^- = 704,4$. Пусть $\beta = 0,9$, тогда доверительный интервал для математического ожидания:

$$I_{0,9} = (25,9, 118,9);$$

доверительный интервал для дисперсии

$$I_{0,9} = (615,8, 998,6);$$

доверительный интервал для среднеквадратического отклонения $I_{0.9} = (24.8, 31.4)$.

1.4.2.3. Оценка вероятности по частоте. Пусть в результате опыта фиксируется наступление (или ненаступление) события А. После проведения n опытов получена частота наступления этого события p^* как точечная оценка неизвестной вероятности $P(A) = p$. Какова достоверность и точность полученной оценки?

Для достаточно большого числа проведенных опытов (если np и nq больше четырех [1], p^* распределена приблизительно по нормальному закону) можно воспользоваться вышеприведенной методикой и записать доверительный интервал:

$$I_B = \left(p^* - t_B \cdot \sqrt{\frac{p^*(1-p^*)}{n}}, \quad p^* + t_B \cdot \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

(здесь использована формула для дисперсии биномиального закона распределения $D = p(1-p)$).

Для малого числа опытов гипотеза о нормальности закона распределения p^* неверна и необходимо пользоваться точно известным законом. Легко видно, что p^* распределена по биномиальному закону (реализуется схема опытов Бернулли). Поскольку биномиальный закон несимметричен, говорят не о доверительном интервале, а о доверительной области. Ее определяют границы $\alpha/2$ и $(1-\alpha)/2$, где $\alpha = 1 - \beta$.

При проведенных n опытах, в которых событие А появилось k раз, и доверительной вероятности β доверительную область определяют путем решения системы уравнений:

$$\sum_{m=k}^n C_n^m p^m (1-p)^{n-m} = \frac{\alpha}{2}$$

$$\sum_{m=0}^k C_n^m p^m (1-p)^{n-m} = \frac{1-\alpha}{2}$$

На практике вместо решения такого рода систем уравнений можно использовать заранее рассчитанные таблицы (см. [5]).

Если в результате n опытов событие А не произошло ни разу, то частота события $p^* = 0$. В этом случае событие А либо невозможно, либо его вероятность весьма мала. Поставим задачу определить доверительные границы для вероятности события

А, совместимые с полученными результатами опыта — это событие ни разу не произошло. Не вызывает сомнения, что нижняя граница p_1 есть 0. Верхняя же граница p_2 может быть оценена на основании следующих соображений. Если в n опытах событие А не произошло ни разу (вероятность $(1-p_2)^n$), то доверительная «совместимость» с этим фактом есть $1 - \beta$, т. е. $(1-p_2)^n = 1 - \beta$. Отсюда:

$$p_2 = 1 - \sqrt[n]{1 - \beta}.$$

Задача. Проводится выборочная проверка изданных книг. Из проверенных 100 единиц не найдено ни одной бракованной. Определить верхнюю границу вероятности издания бракованной книги с доверительной вероятностью 0.9.

Решение. Для получения ответа может быть непосредственно использована полученная выше формула, а именно:

$$p = 1 - \sqrt[100]{1 - 0.9} = 1 - \sqrt[100]{0.1} = 1 - 0.977 = 0.023.$$

1.4.3. Методы анализа статистических зависимостей

Рассмотренные выше методы позволяют определять статистические характеристики результатов опыта. Однако важным является и нахождение и анализ статистических зависимостей между наблюдаемыми результирующими величинами и факторами их вызывающими.

1.4.3.1. Основные задачи анализа статистических зависимостей. Наблюдаемая результирующая величина y может зависеть от нескольких характеризующих ее величин (факторов) $X = (x_1, x_2, \dots, x_k)$. На практике можно встретить самые разнообразные формы причинно-следственных связей между результирующей величиной и величинами ее вызывающими. Рассмотрим наиболее распространенные схемы зависимостей между факторами $(x_1, x_2, \dots, x_k)$ и результатом y .

Схема А — зависимость между случайными параметрами. Здесь связь y с факторами детерминированная, и задача состоит в том, чтобы восстановить по значениям факторов значения результирующих величин, т. е. установить вид функции f :

$$y = f(x_1, x_2, \dots, x_k).$$

Схема В — регрессионная зависимость. Зависимость случайного результирующего показателя y от случайных фак-

торов (η — случайная реализация y):

$$\eta = f(x_1, x_2, \dots, x_k) + \varepsilon(x_1, x_2, \dots, x_k).$$

Данная схема реализуется в том случае, если y связана со случайными обстоятельствами регистрации или с наличием неучтенных случайных факторов. Предполагается, что случайный компонент ε обладает следующими свойствами:

1. Его математическое ожидание равно нулю ($M[\varepsilon] = 0$);
2. Его дисперсия конечна ($D[\varepsilon] < \infty$).

Схема С — корреляционная зависимость. Зависимость случайного результирующего показателя η от случайных факторов $\xi_1, \xi_2, \dots, \xi_k$ (ξ_i — случайная реализация x_i):

$$\eta = f(\xi_1, \xi_2, \dots, \xi_k) + \varepsilon.$$

Существует и целый ряд других зависимостей [2]. Выбор схемы исследуемых зависимостей в первую очередь определяется сущностью решаемой задачи, ее смысловой направленностью и глубиной рассмотрения вопроса. Приведем ряд примеров. Классический пример схемы А — установление зависимости между возрастом дерева (y) и числом колец x на срезе его ствола. Сейчас всем, наверное, известно, что в данном случае $y = x$.

Если рассматривается вопрос о числе гласных букв (y) в словах с различным количеством букв (x_i), являющихся различными частями речи (x_j), то эта ситуация, скорее всего, подходит под схему В. Если искать связь между чистотой хлопка и качеством выпускаемой пряжи, целесообразно принять схему С. Таким образом, выбор схемы искомой зависимости определяется поставленной задачей и является прерогативой исследователя.

Методы исследования зависимостей вида:

$$f(X) = M(\eta | \xi = X),$$

которые описывают поведение условных средних результирующего показателя η (вычисленного при зафиксированном уровне $\xi = X$) в зависимости от изменения X и называются **функциями регрессии** (см., например, [4]).

Можно рассматривать качественные и количественные факторы. Если свойство объекта рассмотрения не выражается числовым показателем, то это качественный показатель. Например, квалификация наблюдателя, социальное положение опрашиваемого, материал, из которого изготовлено изделие, —

качественные факторы, но температура воздуха, скорость автомобиля, этажность зданий — количественные факторы.

Некоторые факторы в зависимости от точки зрения исследователя можно рассматривать и как качественные, и как количественные. Примером может явиться возраст людей. Если результат наблюдения рассматривается как функция от возраста людей, то возраст выступает как количественный фактор. Однако результаты наблюдения можно рассматривать для различных возрастных групп населения (дети, молодежь, пенсионеры и т. д.). Здесь возраст — качественный фактор.

Анализ статистических зависимостей проводится в рамках тех или иных теорий. Так, если все факторы относятся к качественному виду, то анализ зависимостей проводится в рамках дисперсионного и факторного анализа. Когда все факторы количественные, то применимы методы регрессионного анализа. В рамках корреляционного анализа возможно исследование наборов факторов, относящихся к различным типам.

Сформулируем следующую задачу. По результатам опытов получена определенная информация. Исходя из сути рассматриваемых процессов, выбрана схема зависимости. Необходимо проанализировать полученные данные и получить характеристики причинно-следственной связи. Такой анализ сводится, как минимум, к поиску ответов на следующие вопросы:

1. Действительно ли y зависит от факторов ($x_1, x_2, \dots, x_k$)?
2. Если y зависит от x_i , то какова степень этой зависимости?
3. Какой вид имеет причинно-следственная зависимость между y и X , в частности каков между ними вид аппроксимационной зависимости f ?

Рассмотрим кратко наиболее простые и распространенные методы анализа статистической зависимости.

1.4.3.2. Установление зависимости y от качественного фактора x (дисперсионный анализ). Будем считать, что результирующая величина y зависит только от одного фактора x (однофакторный анализ). Требуется, исходя из результатов наблюдений, определить наличие (или отсутствие) влияния x на y .

Предположим, что фактор x может принимать ряд значений, называемых уровнями $x_1, x_2, \dots, x_m$. При каждом заданном уровне i проводится группа из n опытов и в каждом j -м опыте измеряется значение величины y обозначаемое y_{ij} . Общее количество опытов — N . Результаты этих опытов сведены в таблицу следующего вида:

Номер уровня	Номер наблюдения						
	1	2	3	4	j		n
1	y ₁₁	y ₁₂	y ₁₃		y _{1j}		y _{1n}
2	y ₂₁	y ₂₂	y ₂₃				y _{2n}
.....							
i	y _{i1}	y _{i2}	y _{i3}		y _{ij}		y _{in}
.....							
m	y _{m1}	y _{m2}	y _{m3}		y _{mi}		y _{mn}

Из анализа данных обычно следует, что при одном и том же значении фактора имеются различные значения результата ($y_{ij} \neq y_{ik}$). Поскольку значения величин в одной и той же строке получены при одних и тех же уровнях фактора x , то различие между y_{ij} и y_{ik} в одной строке обусловлено чисто случайными обстоятельствами. Но различия в значениях результатов, стоящих в различных строках, могут быть обусловлены или той же причиной (тогда y и x независимы), или они связаны также и с изменением значения фактора, т. е. имеется зависимость y от x . Предположение о независимости значений результатов от значений уровней фактора называется нулевой гипотезой H_0 . Если оказывается, что эта гипотеза не противоречит результатам проведенных опытов, то есть основания считать, что y не зависит от x .

Задача о проверке гипотезы H_0 может быть решена в рамках так называемого дисперсионного анализа. Сущность дисперсионного анализа состоит в разложении по специальным правилам дисперсии результата наблюдения на независимые слагаемые, каждое из которых характеризует влияние того или иного фактора на результат опыта. Более подробно с данной тематикой можно познакомиться, например, в работе [6].

Рассмотрим модель однофакторного дисперсионного анализа. Отклонение случайной величины η от своего математического ожидания

$$\bar{y} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} y_{ij}}{N}$$

может быть представлено следующим образом:

$$\eta - \bar{y} = \alpha + \gamma,$$

где α — отклонение, вызываемое фактором x ; γ — отклонение, вызываемое другими неучтенными факторами.

В однофакторном дисперсионном анализе приняты следующие допущения: ошибка γ распределена по нормальному закону с

$m_x = 0$, и для различных уровней факторов она имеет одинаковую дисперсию. Сравнение дисперсий, характеризующих вклад в рассеивание получаемых результатов, позволяет судить о зависимости или независимости y и x .

Введем следующие виды дисперсий. Общая дисперсия D_0 — характеризует рассеивание результатов наблюдений относительно общего математического ожидания $\bar{y}$.

$$D_0 = M[(y_{ij} - \bar{y})^2].$$

Ее оценка имеет следующий вид:

$$D_0^* = \frac{1}{N-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

где n_i — количество данных в i -той строке. Если все n_i равны, т. е. $n_i = n$ для любого i , то

$$\tilde{D}_0 = \frac{1}{N-1} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y})^2$$

Групповая дисперсия D_i — характеризует рассеивание результатов наблюдений относительно группового математического ожидания

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \quad (\text{оценка группового математического ожидания } \bar{y}_i \text{ есть среднее по результатам серии опытов для фиксированного уровня } i):$$

$$D_i = M[(y_{ij} - \bar{y}_i)^2].$$

Оценка этой дисперсии вычисляется по формуле:

$$\tilde{D}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

Межгрупповая дисперсия $D_m = M(\bar{y}_i - \bar{y})^2$ характеризует рассеивание, обусловленное исключительно фактором x и может быть оценена по формуле:

$$\tilde{D}_m = \frac{1}{m-1} \sum_{i=1}^m (\bar{y}_i - \bar{y})^2$$

Внутригрупповая дисперсия D_γ характеризует различия в результатах наблюдения, обусловленных влиянием неучтенных факторов, определяется как средневзвешенная по объемам групп

$$D_y = \sum_{i=1}^m \frac{n_i}{N} D_i = \sum_{i=1}^m \frac{n_i}{\sum_{i=1}^m n_i} D_i$$

и, если $n_i = n$ для всех i , то может быть оценена по формуле:

$$\tilde{D}_y = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 = \frac{1}{N-m} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

Из сравнения введенных оценок дисперсии следует:

$$\tilde{D}_0 = \tilde{D}_m + \tilde{D}_y$$

Рассмотрим методику проверки существенности влияния фактора x на результаты наблюдения.

1. Выдвигается гипотеза о *несущественности* фактора x (гипотеза H_0). Содержательно понятно, что, если эта гипотеза верна, математические ожидания в различных группах (при различных значениях фактора, т. е. средние по строкам) должны совпадать:

$$\bar{y}_1 = \bar{y}_2 = \dots = \bar{y}_i = \dots = \bar{y}_m$$

При этом условии оценка межгрупповой дисперсии должна быть значительно меньше, чем оценка внутригрупповой дисперсии. В этой связи величина

$$u = \frac{\tilde{D}_m}{\tilde{D}_y} = \frac{\frac{1}{m-1} \sum_{i=1}^m (\bar{y}_i - \bar{y})^2}{\frac{1}{N-m} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2} \quad (1.37)$$

может выступать в качестве меры непротиворечивости выдвинутой гипотезы, характеристикой справедливости гипотезы H_0 .

2. По формуле (1.37), исходя из полученных результатов опытов, вычисляется экспериментальное значение величины u .

Показано [6], что такая величина u распределена по закону распределения Фишера со степенями свободы $k_1 = m - 1$ и $k_2 = N - m$. Распределение Фишера табулировано, и его таблицы приведены во многих учебниках. Заметим, что таблицы этого распределения составлены при условии, что числитель в отношениях типа (4.9) больше знаменателя. Если это не так, то за экспериментальное значение показателя рассогласованности выбирается обратная величина $u_1 = 1/u$.

3. По таблицам распределения Фишера [7] находится теоретическая величина показателя u_α с выбранным уровнем значимости α (обычно 0.1, 0.05 или 0.01) и указанными степенями свободы.

4. Гипотеза не противоречит экспериментальным данным, если выполняется условие

$$u > u_\alpha \quad (1.38)$$

Выполнение этого неравенства свидетельствует о том, что влияние фактора не существенно и все наблюдаемые различия обусловлены не изменением значения фактора x , а какими-то другими обстоятельствами. Следовательно, можно сказать, что гипотеза H_0 не противоречит опытным данным. Невыполнение же условия (1.38) означает, что u , скорее всего, зависит от x и возможна постановка вопроса о нахождении уровня фактора в наибольшей мере влияющего на рассматриваемый результат. Решение этого вопроса здесь не рассматривается, но с ним можно ознакомиться по работе [8].

1.4.3.3. Анализ существенности влияния (тесноты связи) количественных факторов на результаты наблюдений. Рассмотрим наиболее простой случай: результирующая величина y зависит только от одного фактора x . Требуется, исходя из результатов наблюдений, определить степень тесноты связи x и y . Решение этой задачи осуществляется в рамках корреляционного анализа [9].

Пусть проведена первичная обработка результатов опытов и получены оценки необходимых математических ожиданий и дисперсий. Существенность влияний x на y может быть определена, если будет указано, какая доля степени изменчивости интересующего нас результата обусловлена именно изменчивостью описывающей его функции f от контролируемого нами аргумента x . Количественной мерой этой доли может выступать отношение оценки межгрупповой дисперсии к оценке общей дисперсии:

$$\tilde{I}^2 = \frac{\tilde{D}_m}{\tilde{D}_0} = \frac{\frac{1}{m-1} \sum_{i=1}^m (\bar{y}_i - \bar{y})^2}{\frac{1}{N-1} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y})^2} \quad (1.39)$$

Величина $\tilde{I} = +\sqrt{\tilde{I}^2}$ называется **корреляционным отношением** зависимой переменной y от независимой переменной x и является универсальной характеристикой тесноты их связи.

Из (1.39) следует, что $0 \leq \tilde{I} \leq 1$. Если $\tilde{I} = 1$, то наличие однозначная функциональная (жесткая) связь между x и y , т. е. корреляционная зависимость отсутствует. Наоборот, при $\tilde{I} = 0$ приходится констатировать, что усредненные по группе значения результирующего признака (оценки групповых математических ожиданий) не зависят от x и наблюдаемая зависимость — следствие других неконтролируемых факторов. Промежуточные значения $\tilde{I}$ и есть количественная мера тесноты исследуемой связи.

В частном случае, когда связь между y и x является линейной, для оценивания степени зависимости используется оценка коэффициента корреляции, которая находится следующим образом. Пусть в результате n опытов получены следующие значения, представленные таблицей:

x_1	x_2	...	x_i	...	x_n
y_1	y_2	...	y_i	...	y_n

Оценка коэффициента корреляции определяется с помощью соотношения:

$$\tilde{r} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Оценка $\tilde{r}$ определена на основе выборки результатов наблюдения, поэтому является величиной случайной. Следовательно, неравенство полученной оценки нулю ($\tilde{r} \neq 0$) еще не означает, что между x и y есть линейная зависимость, т. к. это может быть следствием чисто случайных обстоятельств, вызванных недостаточно большим числом опытов. В этой связи необходима проверка значимости оценки.

Решение этой задачи проводится по следующей методике.

1. Вычисляется мера согласованности гипотезы об отсутствии линейной зависимости y от x с наблюдаемыми значениями:

$$u = \frac{\tilde{r} \sqrt{n-2}}{\sqrt{1-\tilde{r}^2}}$$

2. По таблицам распределения Стюдента с $(n-2)$ степенями свободы (такое распределение имеет величина u) при выбранном уровне значимости α находится теоретически «допустимая» с уровнем α мера согласованности u_α . (Таблицы распределения Стюдента имеются практически в любом учебнике по математической статистике.)

3. Если окажется, что

$$u_\alpha > |u|,$$

то гипотеза об отсутствии линейной связи между y и x не противоречит опытным данным и на практике обычно принимается. В противном случае полученную оценку можно считать значимой и между x и y скорее всего существует линейная связь.

1.4.3.4. Определение причинно-следственной зависимости между результатом и влияющими на него факторами. Пусть рассматривается некоторая случайная результирующая переменная η (случайные реализации y), зависящая от неслучайных факторов $x_1, x_2, \dots, x_m$. Значения величины η случайным образом распределяются около неизвестного уровня $f(x_1, x_2, \dots, x_m)$. При этом отклонения значений η от $f(x_1, x_2, \dots, x_m)$ представляют собой значения некоторой случайной ошибки $\gamma(x_1, x_2, \dots, x_m)$ с $M[\gamma] = 0$, т. е.:

$$\eta = f(x_1, x_2, \dots, x_m) + \gamma(x_1, x_2, \dots, x_m).$$

Случайный характер величины η требует выбора подходящей характеристики. Такой характеристикой обычно выступает ее условное среднее, которая в виду того, что $M[\gamma] = 0$, имеет вид:

$$M[\eta | x_1, x_2, \dots, x_m] = f(x_1, x_2, \dots, x_m). \quad (1.40)$$

Зависимость (1.40) называется **регрессионной**, что определило название методов определения причинно-следственных зависимостей — **регрессионный анализ**.

Для установления причинно-следственной связи f необходимо:

1. Выбрать класс функций, в рамках которого целесообразно искать общий вид функции с параметрами a_i :

$$f = f(x_1, x_2, \dots, x_m, a_1, a_2, \dots, a_k). \quad (1.41)$$

$\alpha = 0.05$ по таблицам распределения Фишера, находим теоретический уровень рассогласованности $u_{\alpha} = 3.24$.

Поскольку $u_1 = 31.25 > u_{\alpha} = 3.24$, гипотеза о несущественности типа прибора может быть принята.

Задача 2. Произведенные 20 измерений системы двух случайных величин x и y сведены в таблицу.

	y_j		
x_i	1	2	3
2	5	3	2
4	1	3	6

Требуется определить тесноту связи между x и y .

Решение. Вначале определим корреляционное отношение между измеряемыми величинами. Используя формулу (1.39), найдем:

$$\tilde{r} = +\sqrt{\tilde{I}^2} = +\sqrt{\frac{\tilde{D}_m}{\tilde{D}_0}} = +\sqrt{\frac{0.50}{0.725}} = 0.830.$$

Следовательно, между x и y существует достаточно сильная статистическая связь.

Рассмотрим вопрос о степени линейной зависимости между x и y . Для этого вычислим величину u , связанную с оценкой коэффициента корреляции и распределенную по закону Стьюдента:

$$u = \frac{\tilde{r} \sqrt{n-2}}{\sqrt{1-\tilde{r}^2}} = \frac{0.368\sqrt{18}}{\sqrt{1-0.136}} = 1.69$$

Для доверительной вероятности 0.95 и 18 степеней свободы по таблице распределения Стьюдента находим $u_{0.95} = 2.10$. Поскольку

$$|u| = 1.69 < u_{0.95} = 2.10,$$

мы можем принять гипотезу о линейной связи x с y . 1

Задача 3. Проведены измерения величины растворяемой в данном объеме воды некоторой соли при различных температурах.

Определить зависимость объема растворяемой соли как функцию температуры.

Решение. Если нанести полученные значения на плоскость, то можно увидеть, что искомую зависимость можно считать линейной, т. е. $y = at + b$. Найдем неизвестные параметры a и b методом наименьших квадратов.

Система уравнений (1.43) в данном случае имеет вид:

$$\begin{cases} \sum_{i=1}^8 (y_i - at_i - b)t_i = 0 \\ \sum_{i=1}^8 (y_i - at_i - b) = 0. \end{cases}$$

Решение этой системы дает такие результаты: $a = 0.87$, $b = 67.5$. Следовательно, искомая зависимость имеет следующий вид $y = 0.87t + 67.5$.

t^0	4	10	15	21	29	36	51	68
y	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

Раздел 2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ

*Будем же учиться хорошо мыслить —
вот основной принцип морали.*

(Б. Паскаль)

Знание конкретных разделов математики и освоение ряда базовых математических методов является обязательной, но лишь необходимой частью математического образования. Практическое же применение знаний по математике обычно заключается в том, чтобы объекты нашего рассмотрения — объекты реального мира — описать на языке математики. Такой перевод на язык математики называется построением математической модели.

Все развитие современной науки связано с созданием и изучением моделей систем, процессов и явлений. Модель — это условный образ объекта исследования, который создается по аналогии или сходству с объектом. Построенная модель позволяет получить более точное представление о наиболее существенных его свойствах и в определенной степени предсказать будущие события. Процесс построения математической модели и последующее ее применение для решения конкретных задач называется **математическим моделированием**.

Язык науки требует описания моделей на точном уровне, не допускающем принципиальных разноточений. Однако требование точности описания можно трактовать по-разному и рассматривать весьма широкий спектр моделей. На одном конце этого спектра находятся математические (формальные) модели, а на другом — текстовые (вербальные) модели. Между ними находятся самые различные модели смешанного типа. В дальнейшем мы будем рассматривать абстрактные модели, построенные на основе математических символов и их последовательностей, то есть математические модели.

Глава 2.1. Методологические основы моделирования

2.1.1. Постановка задач моделирования

Любой человек (коллектив, учреждение), которого в дальнейшем будем называть «исследователь», в результате своей

целенаправленной деятельности рассматривает некоторый объект (систему, процесс и т. п.), ставит и решает задачи по достижению поставленной цели.

Под задачей в канонической форме обычно понимают логическое высказывание вида:

«Дано V , требуется Z », или $\langle V, Z \rangle$.

Заданные условия V включают: множество возможных состояний рассматриваемой системы S и множество операторов (правил) R , переводящих объект рассмотрения из одного возможного состояния в другое. Решение задачи — достижение цели Z , которая в большинстве случаев понимается как желаемое состояние объекта. Так, например, следовательно, исходя из имеющихся фактов (описания состояния объекта S) должен найти преступника и на основе правовых норм и логических заключений (R) доказать его виновность или невиновность (цель Z).

На стадии постановки задачи исследователь определяет объект исследования и окружающую его среду, возможные факторы, влияющие на объект, способы описания собственно объекта и (или) правил перехода из одного состояния в другое, цели решения задачи и оценки качества решений. Формальная модель поставленной задачи может быть записана следующим образом:

«Дано X, Y, R, W ; требуется Z »,

где X — множество входных управляемых и неуправляемых факторов; Y — множество исходов, т. е. результатов взаимодействия входных факторов с объектом рассмотрения; R — множество операторов, определяющих появление исходов в результате взаимодействия входных факторов с рассматриваемым объектом; Z — множество целей — желаемых состояний системы; W — множество критериев оценки элементов множества Z .

Поставленная цель определяет требования к исходам. Критерии же либо упорядочивают возможные исходы по степени близости их к поставленным целям, либо в явном виде вводят правила оценки этих исходов, т. е. непосредственно позволяют выбрать исход в наибольшей мере соответствующий поставленной цели. Критерии формулируются на основе поставленных целей. В целом ряде случаев критерий непосредственно задает цель.

Рассмотрим пример. Поставлена задача разработки мероприятий по улучшению условий жизни жителей данного города. Цель — улучшение условий жизни — может характеризоваться

следующими критериями: обеспеченность жильем, уровень экологического загрязнения, удобство транспортного обслуживания, качество медицинского обслуживания и т. д. Каждое решение, характеризующееся наиболее высокими значениями этих показателей, можно считать отвечающим поставленной цели. В то же время можно сформулировать задачу, в которой условия жизни будут определяться только объемом финансовых затрат населения на потребление. Тогда цель задачи можно сформулировать как стремление максимизировать средние затраты на потребление, т. е. максимизация показателя затрат на потребление выступает в качестве цели решения задачи.

Процедура решения задачи заключается в построении множества исходов как альтернатив результатов решения задачи, в той или иной мере соответствующих поставленным целям, и в выборе одной из них в качестве собственно решения (ответа задачи).

Реализация этих процедур может быть организована по-разному: от построения и выбора решения одним человеком «в уме» до проведения крупномасштабных исследований, осуществляемых большими коллективами исследователей с привлечением компьютерной техники и сложных математических методов. Озарение, опыт, интуиция и т. п. факторы играли и будут играть огромную роль в процессах принятия решений. Однако, решение сложных и ответственных задач должно осуществляться с применением математических методов, основу которых составляют методы обработки опытных данных и математическое моделирование, которые в настоящее время реализуются почти исключительно с помощью компьютерной техники. Приведем наиболее характерные постановки задач принятия решений [10].

Постановка типа А.

Цель решения задачи — нахождение лишь допустимого варианта решения. Каждое допустимое решение X должно отвечать ограничениям по ресурсам g_{k0} и быть внутренне сбалансированным (непротиворечивым). Цели здесь могут выражаться в виде специальных ограничений типа «быть не менее чем...» или твердых заданий «быть равным ...». Они лишь косвенно выражают приоритет вариантов и направление устремлений. Например, требуется сформировать учебный план неполной средней школы. В рамках этой проблемы может быть

задана следующая задача: за восемь первых лет обучения в школе на математику необходимо выделить не менее 1500 часов учебного времени и непременно изучить методы решения квадратных уравнений за 20 часов.

Математическая запись такого рода моделей может быть следующей: определить $X = (x_1, x_2, \dots, x_n)$ при ограничениях:

$$W_i(X) \geq W_{i0}(X), \quad i = \overline{1, l} \text{ — целевые ограничения,}$$

$$g_k(X) \leq g_{k0}, \quad k = \overline{1, m} \text{ — ресурсные ограничения.}$$

Примерами такого типа постановок являются: модели прямого вычисления (вычисления по формулам, а также решение систем уравнений и неравенств), модели баланса, модели прогнозирования и др. При помощи таких моделей, описывается и анализируется состояние и поведение объектов исследования.

Постановки типа В.

Эти постановки формулируются как требование максимизации (минимизации) некоторых характеристик объектов исследования. Эти характеристики одновременно выступают в качестве критериев оценки и отбора вариантов. Поэтому задачи, имеющие постановку такого типа, называются оптимизационными. Множество допустимых вариантов решения задается системой ограничений на регулируемые и нерегулируемые параметры как требование принадлежности некоторому допустимому множеству D .

Математическая запись может быть следующей:

$$X^* = \operatorname{argmax} W(X) | X \in D \quad \text{или} \quad X^* = \operatorname{argmin} W(X) | X \in D,$$

где $W(X)$ — целевая функция; D — область допустимых решений.

Подобные математические модели предписывают норму поведения объекта и относятся к моделям нормативного типа. Решив подобную математическую задачу, получают вполне определенный и наилучший в смысле принятого критерия план действий.

Если требования экстремизации относятся к нескольким характеристикам, то говорят о векторной (многокритериальной) оптимизации. Например, требуется разработать план подготовки в условиях заочного обучения (D) за самый короткий срок (T) специалистов с наивысшим в данной области образованием V и с наименьшими затратами средств W . Формально это можно было бы записать так:

$$X: T(X) \rightarrow \min, V(X) \rightarrow \max, W(X) \rightarrow \min | X \in D.$$

Однако эти требования, как правило, противоречивы (несовместны), а одновременное их выполнение, с математической точки зрения, абсурдно.

Существуют различные приемы изменения противоречивых постановок с целью их дальнейшей формализации и применения математических методов решения задачи, однако прямое решение задач многокритериальной оптимизации наталкивается на принципиальные трудности.

Постановка типа С.

Это наиболее общий вид задач поиска решения, встречающийся в целенаправленной деятельности людей. В постановках такого типа дают развернутую формулировку целей и не привязываются к структуре выделяемых ресурсов. Цели могут последовательно конкретизироваться вплоть до введения целевых нормативов (как в постановках типа В), однако цели и критерии в этом типе постановок разделены.

Оценка качества решения осуществляется путем вычисления отклонения выбранных показателей от этих нормативов. Если же нормативы просто перечислены, то постановка становится похожей на постановку типа А. Целевые нормативы могут быть как-то ранжированы или даже взвешены. В последнем случае постановка С внешне будет походить на постановку типа В. Нахождение эффективных решений задач имеющих постановки типа С — дело будущего.

Сравнивая типы постановок, можно сказать, что нормативные постановки приводят к жестким моделям. Решение задачи в этом случае получается наилучшим в смысле принятых критериев.

Разделение целей и критериев оценки вариантов решений в постановках типа С позволяет получать более гибкие модели способные адаптироваться к существенным изменениям условий задачи.

Постановки же типа А обычно предназначены только для решения вопросов анализа.

Таким образом, в целенаправленной деятельности могут встречаться самые различные постановки задач, в решении которых важное место занимают математические методы. Какие именно постановки задач применять и какие использовать для решения методы моделирования зависит от исследователя, его целей, знаний и возможностей.

2.1.2. Классификация моделей

Процесс математического моделирования может развиваться по одному из двух сценариев. Наиболее распространен следующий: формулируется задача, затем ее пытаются формализовать в виде известной математической модели, которая, как правило, хорошо известна исследователю и решение которой потенциально доступно. Это путь подгонки задачи под модель. Здесь возникает проблема адекватности полученного решения исходной задаче.

Другой сценарий ориентирован на построение наиболее адекватной математической модели. После построения модели проводится поиск метода решения, который может быть неизвестен исследователю или вообще не существовать (построение модели под задачу). Основная трудность такого подхода, порой непреодолимая, заключается в построении метода решения задачи и оценке точности получаемого результата.

В этой связи специалист должен представлять себе современное состояние науки о математическом моделировании, знать основные модели, их свойства и соответствующие методы решения. Каждый тип математических моделей имеет свои особенности, ориентирован на тот или иной класс задач, связан с определенными требованиями к вычислительной технике и т. п. В этой связи становится важной классификация математических моделей.

2.1.2.1. Классификация моделей может быть проведена с различных точек зрения. Рассмотрим лишь некоторые из них.

1. Классификация по целевому назначению.

Модели структуры описывают связи между средой и компонентами системы. Из них можно выделить: канонические модели, где описана связь с окружающей средой через вход и выход; модели внутренней структуры, описывающие состав компонентов системы и связь между ними; модели иерархической структуры, где целое расчленяется на элементы более низкого уровня (обычно в виде дерева структуры системы) и др.

Модели функционирования — модели жизненного цикла системы в целом; модели операции, представляющие описание процессов функционирования отдельных элементов; информационные модели, описывающие взаимосвязи источников и потребителей информации, характер ее преобразования,

временные и другие количественные характеристики; процедурные модели, отражающие порядок взаимодействия элементов при выполнении отдельных операций; временные модели, описывающие процедуры функционирования во времени.

Стоимостные модели предназначены для комплексной оценки по экономическим критериям.

2. Классификация по типу задач.

Описательные (дескриптивные) модели (к ним часто приводят постановки задач типа А) предназначены для описания изучаемого процесса, объяснения наблюдаемых фактов, а также прогноза поведения системы: модели планирования без оптимизации (балансовые модели); модели для некоторых задач сетевого планирования и управления (расчет по известным формулам); модели для задач учета; модели для задач контроля и анализа (обычно в виде статистических моделей); модели прогнозирования; модели для расчета параметров функционирования случайных систем с неформализованными связями.

Нормативные, или прескриптивные модели, к которым обычно приводят постановки задач типа В. В моделях такого типа отражается то, что должно было бы происходить, если принять некоторые исходные предположения. Построение нормативных моделей преследует цель определения наилучшего эффекта или состояния. С их помощью дается ответ на вопросы о том, как должно быть.

Модели конструирования решений, выступающие в виде формализованных схем построения комплексных решений. Они обычно включают в качестве элементов и дескриптивные, и нормативные модели. К таким моделям обычно приводят постановки задач типа С.

3. Классификация по форме реализации.

Аналитические модели, записывающиеся в виде математических конструкций, не включающих логических условий, приводящих к разветвлению вычислительного процесса.

Алгоритмические модели — это математические модели, в которых присутствуют логические условия, приводящие к разветвлению вычислительного процесса.

2.1.2.2. Основные принципы построения математических моделей. При построении математических моделей целесообразно придерживаться следующих принципов, выработанных практикой.

Достаточность используемой информации. При построении модели целесообразно использовать ту информацию, которая требуется в соответствии с разрабатываемым алгоритмом, что принципиально противоположно подходу: «сначала сбор информации, а затем построение алгоритма по обработке этой информации».

Инвариантность информации. Данный принцип означает, что входная информация должна быть независима от параметров моделируемой системы. Иначе говоря, модель должна работать без коррекции в некотором диапазоне значений входной информации.

Преемственность. Каждая последующая модель не должна нарушать свойств объекта, полученного на предыдущих этапах или при использовании других моделей.

Эффективная реализуемость предполагает соответствие точности исходных данных, точности решения задачи и точности результирующей информации. В этой связи следует заметить, что нахождение оптимальных решений для практики часто иллюзорно.

2.1.2.3. Процедура математического моделирования. Несмотря на произвольный характер процесса математического моделирования, можно выделить наиболее характерные его этапы.

Цель *первого этапа* моделирования — построение концептуальной модели как совокупности качественных зависимостей между существенными факторами. После исследования объекта моделирования обычно строится вербальная (описательная) по форме модель, которая дает содержательное представление о существенных свойствах системы и главных связях между этими свойствами. Она включает: условие функционирования системы; цели исследования; возможности управления системой.

При построении концептуальной модели встает целый ряд проблем: стремление упростить отображение системы с одновременным желанием построить адекватный реальности сценарий; формулировка и формализация целей, часто заменяют критерием, однако при этом модель обычно теряет адекватность и становится «плоской»; формализацию внутренних и внешних ограничений трудно согласовать с желанием учесть большее количество ограничений; неоднозначность появляется при классификации факторов и выделении из них управляемых факторов.

Второй этап — построение математической модели. Здесь главной проблемой является определение количественных математических соотношений, формализующих качественные зависимости. Практика показывает целесообразность введения промежуточного этапа — построения алгоритма. *Под алгоритмом здесь мы понимаем строго определенную последовательность действий, которая удовлетворяет требованиям определенности, массовости и результативности. Если требование результативности, т. е. получения решения за конечное число операций, не выполнено, то говорят об эвристической процедуре. Наличие алгоритма обеспечивает целостное восприятие системы и ее функционирования, уточняет роль и место каждой подсистемы, элемента, функции.*

Необходимость численного значения констант, определения диапазона изменения факторов и переменных, законов распределения и их параметров требует дополнительной кропотливой работы по углубленному изучению системы.

Третий этап моделирования — проведение исследований (собственно решения задачи с помощью модели). Для полученной математической модели, исходя из ее особенностей и целей исследования, выбирается метод решения. Затем осуществляется поиск алгоритма того решения, которое соответствует математической модели, и, в конце концов, целям моделирования. Полученные результаты решения анализируются с целью их версификации. В случае положительного исхода результаты принимаются и модель передается для использования с другими исходными данными и/или при других параметрах модели.

2.1.3. Динамические модели

В 1798 Мальтус опубликовал свою работу, посвященную проблеме выживаемости человечества и содержащую соответствующую математическую модель. Если $N(t)$ обозначает численность популяции на момент времени t , b — мгновенная рождаемость (коэффициент рождаемости), то простейшая математическая модель Мальтуса для оценки темпа роста численности популяции организмов, размножающихся при постоянных условиях выглядит следующим образом [11]:

$$dN(t) = bN(t) \cdot dt.$$

Это обыкновенное дифференциальное уравнение, решение которого при $N(0) = N_0$, имеет вид:

$$N(t) = N_0 \exp(bt).$$

Если ввести показатель мгновенной смертности d (коэффициент смертности), то модель приобретает вид:

$$dN(t) = (b - d) \cdot N(t) \cdot dt.$$

Решение этого уравнения (изменение численности популяции во времени) есть функция:

$$N(t) = N_0 \exp(b - d)t.$$

Величина $r = b - d$ служит мерой, определяющей скорость роста популяции. Если $r < 0$, то популяция в данный момент сокращается, а при $r > 0$ возрастает.

На продолжительном отрезке времени смертность и рождаемость могут значительно изменяться, т. е. быть функцией времени $r = r(t)$. В этом случае:

$$N(t) = N_0 \exp\left(\int_0^t r(t) dt\right).$$

При $r = 0$ популяция устойчива по численности. Экспоненциальный рост величины $N(t)$ показывает, что даже малые отклонения от устойчивого состояния могут привести к неограниченному росту (или исчезновению) популяции. В реальной же действительности существуют некоторые внутренние механизмы, регулирующие численность популяции. Эффективность этих механизмов часто зависит от численности популяции на данный момент, т. е. равна $N(t)$. Учет такого механизма приводит к следующей модели:

$$dN(t) = (b - d - c \cdot N(t)) \cdot N(t) \cdot dt.$$

В частном случае, если K — максимальная численность популяции, которую может обеспечить окружающая среда и изменение скорости роста популяции пропорционально величине $(K - N(t))/K$, то модель примет вид:

$$dN(t) = r(1 - N(t)/K) \cdot N(t) \cdot dt.$$

Это уравнение имеет решение:

$$N(t) = \frac{N_0 \cdot K \cdot \exp(rt)}{K - N_0 \cdot [1 - \exp(rt)]},$$

которое достаточно хорошо описывает изменение численности популяции и может использоваться на практике.

В природе часто встречаются ситуации, когда численность популяций одних видов взаимосвязана с численностью другой популяции. Подобные межвидовые отношения также можно моделировать системой дифференциальных уравнений. Так, для двух взаимосвязанных видов:

$$\begin{cases} dN_1(t) = (r_1 - c_{11}N_1(t) - c_{12}N_2(t)) \cdot N_1(t) \cdot dt \\ dN_2(t) = (r_2 - c_{22}N_2(t) - c_{21}N_1(t)) \cdot N_2(t) \cdot dt, \end{cases}$$

где c_{ij} описывает влияние численности видов на их собственные скорости роста, c_{ij} есть мера подавления вида i видом j ($i, j = 1, 2$).

Данный пример иллюстрирует возможность применения аппарата дифференциальных уравнений для исследования проблем народонаселения, численности популяций животного мира, применяться при описании боевых действий и т. п.

Рассмотрим широко распространенный вариант модели, применяемый в системах массового обслуживания. Пусть в каждый момент времени t система может находиться в одном из следующих состояний: $S_1(t), S_2(t), \dots, S_n(t)$. Предположим, что за малый интервал времени Δt , непосредственно следующий за моментом t , система может остаться в данном состоянии $S_i(t)$ или перейти только в одно из «соседних» состояние, т. е. в $S_{i-1}(t)$ или в $S_{i+1}(t)$. Интенсивность перехода в более высокое состояние есть λ_i , а в более низкое — μ_i . Требуется определить $\Delta S_i = S_i(t+\Delta t) - S_i(t)$ — изменение каждого из возможных состояний время Δt .

Изменение данного состояния $S_i(t)$ за малое время складывается из следующих составляющих:

1. Переход из нижнего состояния в верхнее: $\lambda_{i-1}S_{i-1}(t) \rightarrow S_i(t)$ и $\lambda_i S_i(t) \rightarrow S_{i+1}(t)$;
2. Переход из верхнего состояния в нижнее: $\mu_{i+1}S_{i+1}(t) \rightarrow S_i(t)$ и $\mu_i S_i(t) \rightarrow S_{i-1}(t)$.

Исключения составляют лишь граничные состояния и $S_1(t)$: из $S_0(t)$ можно перейти только в $S_1(t)$, а из $S_n(t)$ только в $S_{n-1}(t)$.

Эти рассуждения позволяют составить следующую систему разностных дифференциальных уравнений:

$$\begin{cases} \frac{\Delta S_0}{\Delta t} = -\lambda_0 S_0(t) + \mu_1 S_1(t) \\ \frac{\Delta S_i}{\Delta t} = \lambda_{i-1} S_{i-1}(t) - \mu_i S_i(t) - \lambda_i S_i(t) + \mu_{i+1} S_{i+1}(t) \\ \frac{\Delta S_n}{\Delta t} = \lambda_{n-1} S_{n-1}(t) - \mu_n S_n(t). \end{cases} \quad (2.1)$$

Система уравнений (2.1) представляет модель «размножения гибели», получившая свое название при изучении проблемы изменения численности популяций в биологии. С точки зрения математики (2.1) это система обыкновенных дифференциальных уравнений, решение которой позволит определить функцию, описывающую динамику состояния системы. Так, $S_i(t)$ может означать на данный момент времени число людей в данном городе, число частиц в атомном реакторе, вероятность занятости i каналов связи на данной телефонной станции и т. п.

2.1.4. Метод статистических испытаний (метод Монте-Карло)

Численный метод решения математических задач при помощи моделирования случайных величин называется методом статистических испытаний или методом Монте-Карло. В основе метода лежит следующий факт: если имеется механизм генерирования (розыгрыша) значений равномерно распределенной на отрезке $[0, 1]$ случайной величины, то легко получить случайные значения другой случайной величины, распределенной по любому заданному закону.

Генерирование значений равномерно распределенной случайной величины осуществляется с помощью так называемых псевдослучайных чисел. Сегодня практически в каждом алгоритмическом языке или пакете прикладных программ имеется стандартная процедура генерирования случайных чисел. Так, например, в программах написанных на языке PASCAL достаточно написать $p := \text{rand}$ и p будет присвоено одно из значений псевдослучайного числа. Рассмотрим механизм метода статистических испытаний.

2.1.4.1. Розыгрывание значений непрерывной случайной величины. Поставим задачу получить значения ξ случайной

величины X , распределенной в отрезке $[a, b]$ с заданной плотностью $f(x)$.

При заданном законе распределения вероятность попадания случайной величины X в $[a, \xi]$ находится по известной формуле:

$$P(a \leq x < \xi) = \int_a^{\xi} f(x) dx. \quad (2.2)$$

Это выражение можно рассматривать в качестве уравнения относительно неизвестной ξ . Покажем, что величина ξ , являющаяся корнем уравнения (2.2), имеет плотность вероятности $f(x)$.

Функция распределения случайной величины X

$$F(x) = \int_a^x f(x) dx \equiv y(x)$$

монотонно возрастает от 0 до 1. Следовательно, прямая $y = \eta$

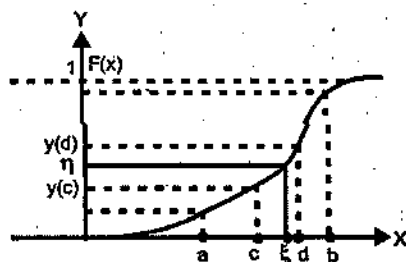


Рис. 2.1. Розыгрыш непрерывной случайной величины

ке $c < x < d$ соответствует ордината кривой $y = y(x)$, удовлетворяющая неравенству $y(c) < y < y(d)$. Поэтому, если ξ принадлежит $c < x < d$, то η принадлежит интервалу $y(c) < y < y(d)$, и наоборот (рис. 2.1). Отсюда следует, что:

$$P\{c < \xi < d\} = P\{y(c) < \eta < y(d)\}. \quad (2.3)$$

Предположим, что случайная величина η на интервале $(0, 1)$ распределена равномерно. В этом случае вероятность

$$P\{y(c) < \eta < y(d)\} = y(d) - y(c) = \int_c^d f(x) dx. \quad (2.4)$$

Сравнивая это выражение с (2.3), получим:

$$P\{c < \xi < d\} = \int_c^d f(x) dx. \quad (2.5)$$

Соотношение (2.5) показывает, что величина ξ имеет плотность вероятности $f(x)$. Этот факт является основанием для построения следующей схемы получения случайного значения непрерывной случайной величины X , имеющей заданный закон распределения $f(x)$:

1. Генерируется η — значение случайной величины, равномерно распределенной в интервале $(0, 1)$;

2. Записывается уравнение:

$$\int_a^{\xi} f(x) dx = \eta; \quad (2.6)$$

3. Решается уравнение (2.6) относительно искомой величины ξ .

Пример 1. Розыгрыш равномерного распределения. Пусть задано распределение

$$f(x) = \frac{1}{b-a} \quad \text{для } a \leq x \leq b.$$

Составляем уравнение:

$$\int_a^{\xi} \frac{1}{b-a} dx = \eta$$

Отсюда легко получить искомый результат:

$$\xi = a + \eta(b - a).$$

Пример 2. Для случайной величины, распределенной по экспоненциальному закону $f(t) = \lambda \cdot \exp(-\lambda t)$ для $t > 0$, окончательная формула имеет следующий вид:

$$\xi = \frac{-1}{\lambda} \ln(1 - \eta). \quad (2.7)$$

Таким образом, чтобы разыграть случайную величину, распределенную по экспоненциальному закону, необходимо разыграть значение равномерно распределенной на отрезке $[0, 1]$ величины ξ , а затем подставить ее в формулу (2.7).

Для более сложных распределений не удастся аналитически решить уравнение типа (2.6). Поэтому используют таблицы функций распределений. Так же разыгрывают равномерно распределенную величину η , а затем по таблице ищут величину ξ , удовлетворяющую условию: $\xi = \arg F(\eta)$.

2.1.4.2. Розыгрыш значений дискретной случайной величины. Пусть заданы значения вероятностей для некоторой дискретной случайной величины X : $p(x_1), p(x_2), \dots, p(x_n)$. Ставится задача случайным образом выбрать одно из возможных значений X , учитывая ее распределение.

Идея решения данной задачи основана на попадании случайной точки на один из интервалов, каждый из которых пропорционален величине, соответствующей вероятности $p(x_i)$, $i=1, n$.

Вначале, как всегда в методе Монте-Карло, генерируется ξ — значение равномерно распределенной в интервале $[0, 1]$ случайной величины. Затем находится искомая величина по правилу:

$$x_k = \left\{ \min(i, n) \mid \left(\sum_{i=1}^n p(x_i) / \sum_{i=1}^n p(x_i) \right) > \eta \right\}. \quad (2.8)$$

Логика этого правила заключается в следующем. Случайная точка попадает в один из возможных интервалов (рис. 2.2). Приведенное правило позволяет последовательно просматривать отношения нарастающей суммы вероятностей к их общей сумме. Считается выбранным то значение x_k , для которого впервые выполнится условие в (2.8).

Замечание. В целом ряде задач практики встает вопрос о случайном выборе одного из возможных вариантов, каждый из которых имеет определенный вес. Его решить можно по аналогии с описанным выше методом, положив в (2.8) вместо вероятности величину соответствующего веса.

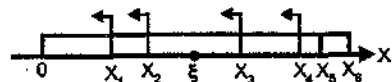


Рис. 2.2. Розыгрыш дискретной случайной величины

Рассмотрим пример. Имеются четыре альтернативы x_1, x_2, x_3, x_4 весами: $q(x_1) = 13$, $q(x_2) = 7$, $q(x_3) = 11$ и $q(x_4) = 3$. Какая альтернатива будет выбрана, если выпала $\eta = 0,737$?

Решение. В соответствии с (1.7) имеем:

- $k = 1$ — « $13/34 = 0.382 > 0.737$ » — нет;
- $k = 2$ — « $20/34 = 0.588 > 0.737$ » — нет;
- $k = 3$ — « $31/34 = 0.912 > 0.737$ » — да!

Таким образом, выбрана третья альтернатива. Если бы выпало $\eta = 0.59$, то был бы выбран тот же вариант, а вот если $\eta = 0.27$, то первый и т. д.

2.1.4.3. Пример применения метода статистических испытаний. Пусть в данный прямоугольник площади S_0 вписана некоторая сложная фигура (рис. 2.3). Требуется определить площадь вписанной фигуры.

Решение этой задачи при помощи метода статистических испытаний может происходить по следующей схеме. Реализуется механизм попадания в прямоугольник: разыгрываются случайные значения равномерно распределенных случайных чисел из интервала (a, b) и (c, d) , которые выступают координатами случайной точки. Всего разыгрываются N точек. Из них N_1 попадает во вписанную фигуру, а N_2 — вне ее ($N = N_1 + N_2$). За отношение площадей фигур принимается отношение числа точек, попавших в фигуру, к общему числу разыгранных точек: $\frac{S}{S_0} = \frac{N_1}{N}$.

Сколько розыгрышей точек N необходимо произвести, чтобы обеспечить заданную точность ϵ ? Обычно поступают следующим образом. Производится серия из достаточно представительного числа точек (k штук), в результате чего получается результат S_k . Затем серия повторяется и, если

$$|S_k - S_{2k}| < \epsilon,$$

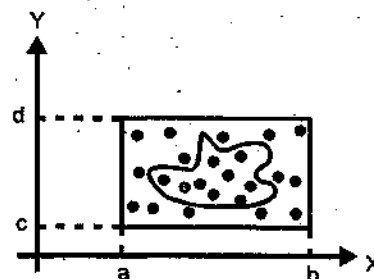


Рис. 2.3. Определение площади методом статистических испытаний

S_{2k} принимается за конечный результат. В противном случае серии повторяются до тех пор, пока два последних результата не дадут отличие меньше, чем ϵ . Формальную оценку числа розыгрышей можно получить на основании следующих рассуждений. Пусть требуется вычислить неизвестную величину ρ . Предположим, что имеется такая случайная величина x , что $M[X] = \rho$ и $D[X] = \sigma^2$. Сгенерируем N значений случайной величины X . Согласно центральной предельной теореме, распределение суммы $\rho_N = x_1 + x_2 + x_3 + \dots + x_N$ приближается к нормальному с параметрами $N\rho$ и $N\sigma^2$. Применяя правило «трех сигм», получаем приближенное равенство:

$$P\{N\rho - 3\sigma\sqrt{N} < \rho_N < N\rho + 3\sigma\sqrt{N}\} = 0.997,$$

или, в более компактной форме,

$$P\left\{\left|\frac{1}{N}\sum_{j=1}^N x_j - p\right| < \frac{3\sigma}{\sqrt{N}}\right\} = 0.997. \quad (2.9)$$

Данное соотношение говорит о том, что среднее значение сгенерированной случайной величины с очень высокой вероятностью равно p . При этом ошибка не превосходит величины $3\sigma/\sqrt{N}$, стремящейся к нулю при возрастании N . Важно подчеркнуть, что (2.9) позволяет оценить число розыгрышей N , которое обеспечивает получение такой точности.

Глава 2.2. Математическое программирование

2.2.1. Основные понятия

В последние десятилетия появилось множество различных по содержанию задач, которые формально можно свести к одной общей постановке: найти решение $X^* = (x_1, x_2, \dots, x_n)$, принадлежащее области допустимых решений $D = D(X)$ и обеспечивающее наибольшее значение скалярной функции $W(X)$, т. е.:

$$X^* = \arg \max W(X) | X \in D(X), \quad (2.10)$$

или, в более традиционной форме записи,

$$W(X) \rightarrow \max \text{ для } X \in D(X).$$

Если область допустимых значений $D(X)$ — n -мерное евклидово пространство, то задача (2.10) является **задачей математического программирования (МП)**. В отличие от термина, используемого в терминологии специалистов по компьютерам, в информатике термин «программирование» следует понимать как «поиск наилучших планов». Функция $W(X)$ называется **целевой (критериальной) функцией**, условие $X \in D(X)$ — **ограничениями задачи математического программирования**.

С математической точки зрения решение задачи на нахождение минимального значения $W(X)$ в принципе не отличается от решения задачи на поиск наибольшего значения целевой функции. Так, переход к поиску минимума $W(X)$ осуществляется заменой в (2.10) знака целевой функции на противоположный: $W(X) \rightarrow -W(X)$. В этой связи обычно говорят о поиске экстремального (наибольшего или наименьшего) значения целевой функции. Знак $\max$ или $\min$ применяется либо для определенности изложения, либо при решении конкретных содержательных задач.

В основу классификации задач МП положены особенности целевой функции и функций, определяющих область допустимых значений, что существенно влияет на сложность решения соответствующих задач. В этой связи условие $X \in D(X)$ удобнее записать так:

$$\begin{cases} g_i(x) \leq 0, i = \overline{1, m} \\ x \in E. \end{cases} \quad (2.11)$$

$$x \in E. \quad (2.12)$$

Решить задачу (2.10) означает найти такой X^* , который максимизирует целевую функцию $W(X)$ на множестве всех допустимых решений. Такое решение X^* называется **оптимальным**. Если $W(X^*) \geq W(X)$ для любого X из $D(X)$, то X^* называется **глобальным максимумом**. Максимум в точке $X^* = X$ называется **локальным**, если условие $W(X^*) \geq W(X)$ выполняется только в некоторой окрестности X^* . Решение X , удовлетворяющее ограничениям, но не обеспечивающее наибольшего (наименьшего) значения целевой функции, называется **допустимым**. Геометрически допустимое решение — это точка в области допустимых решений (ОДР), задаваемой условиями (2.11) и (2.12).

В следующей таблице приведены наиболее известные и хорошо изученные классы задач математического программирования.

	Функции W и g	Множество E	Класс задач МП
1	Непрерывные нелинейные, $m = 0$	Непрерывное R^n	Нелинейная оптимизация без ограничений
2	Непрерывные нелинейные, ограничения - равенства	Непрерывное R^n	Нелинейная оптимизация с ограничениями типа равенства
3	Линейные	Непрерывное R^n	Линейное программирование
4	Линейные	Дискретное Z^n	Целочисленное линейное программирование
5	Выпуклые	Выпуклое R^n	Выпуклое программирование
6	Непрерывные, нелинейные, $m \neq 0$	Непрерывное R^n	Непрерывное (нелинейное), общий случай задачи МП

Для каждого класса задач математического программирования построены математические методы их решения и разработаны соответствующие пакеты компьютерных программ.

Вместе с тем каждая конкретная задача может иметь присущие только ей особенности, которые могут повлиять на применимость тех или иных математических методов. Решение о выборе метода (и соответствующей программы) принимает исследователь, знающий и суть решаемой задачи, и условия применимости того или иного метода.

2.2.2. Классические задачи оптимизации

Задача на нахождение вектора $X = (x_1, x_2, \dots, x_n)$, обеспечивающего максимум целевой функции при ограничениях типа равенства, т. е.

$$W(X) \rightarrow \max(\min) \quad (2.13)$$

при условиях

$$g_i(x) = 0, \quad i = \overline{1, m}, \quad m < n, \quad (2.14)$$

называется **задачей классической оптимизации**. Условие $m < n$ обеспечивает возможность существования более одного допустимого решения, среди которых требуется найти наилучшее (оптимальное). При отсутствии ограничений (2.14) говорят о **задаче на безусловный экстремум**.

2.2.2.1. Задача на безусловный экстремум. Говорят, что функция $W(X)$ имеет в точке X^* локальный максимум (нестрогий), если при любых малых ΔX имеет место условие:

$$W(X^*) \geq W(X \pm \Delta X). \quad (2.15)$$

При выполнении условия

$$W(X^*) \leq W(X \pm \Delta X) \quad (2.16)$$

говорят, что в X^* функция $W(X)$ имеет минимум (нестрогий).

Если неравенства строгие («>» или «<»), то максимум и минимум называют строгими. Когда условия (2.15) и (2.16) имеют место для всей области определения функции, то максимум (минимум) называют глобальным. Точки максимума и минимума функции имеют обобщенное название — **экстремумы**. Функция $W(X)$ может иметь в области своего определения один, несколько или даже бесконечное количество экстремумов. Она может и не иметь экстремума. На рис. 2.4. функция одной переменной $f(x)$, заданная на $[a, b]$, имеет экстремумы в точках x_1, x_2, x_3, x_4, x_5 . Из них: точка x_2 — глобальный минимум, x_3 — глобальный максимум, а в точках x_1 и x_5 функция имеет локальные максимумы, в точке x_4 — локальный минимум. Нахождение точек, в которых функция имеет экстремум, является одной из важных задач математики, имеющих большое практическое значение.

Пусть $f(x)$ — функция одной переменной, заданная на $[a, b]$. Точками экстремума функции $f(x)$ могут быть лишь те точки, в

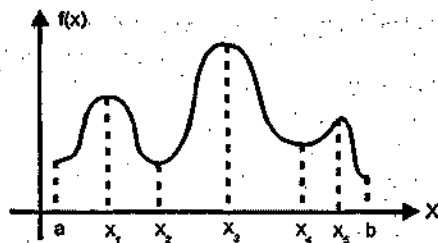


Рис. 2.4. Экстремум функции

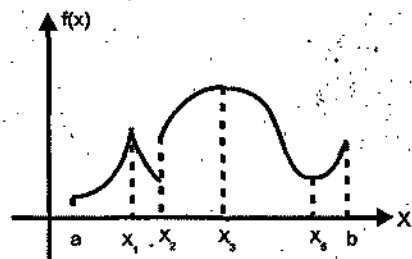


Рис. 2.5. Точки подозрительные на экстремум

Необходимое условие существования экстремума: если дифференцируемая функция $f(x)$ имеет в точке $x = x^*$ экстремум, то ее производная обращается в этой точке в ноль. Для функции нескольких переменных $W(X)$ в точке экстремума требуется равенство нулю всех ее частных производных. Так, для точки максимума имеем:

$$X^*: W(X^*) \geq W(X \pm \Delta X) \rightarrow \frac{\partial W(X)}{\partial x_1} = 0, \frac{\partial W(X)}{\partial x_2} = 0, \dots, \frac{\partial W(X)}{\partial x_n} = 0 \Big|_{X=X^*}.$$

Очевидно, что для точек, где достигается минимум, аналогичное условие также имеет место. В этой связи точки, в которых выполняется условие равенства нулю производных, называются **стационарными точками**.

В окрестностях точек подозрительных на экстремум обычно исследуют знак производной $f'(x)$. Если функция $f(x)$ дифференцируема в окрестности точки x_1 , кроме, может быть, самой этой точки, то, для того, чтобы x_1 была точкой строгого локального максимума, необходимо и достаточно, чтобы производная меняла знак с плюса (слева от максимума функция возрастает) на минус (справа от минимума функция убывает),

которых выполняется одно из следующих условий:

1. $f(x)$ терпит разрыв (точка x_1 на рис. 2.5);
2. $f(x)$ непрерывна, но производная $f'(x)$ не существует (точка x_1 на рис. 2.5);

3. $f'(x) = 0$ (на рис. 2.5 — точки x_3 и x_5);

4. $x = a, x = b$, точки возможных наибольших и наименьших значений.

Такие точки принято называть **подозрительными на экстремум**. Поиск точек экстремума начинается с нахождения всех точек подозрительных на экстремум.

т. е. $f'(x) > 0$ при $x < x_1$ и $f'(x) < 0$ при $x > x_1$. Для строгого локального минимума необходимо и достаточно, чтобы производная меняла знак с минуса на плюс. При поиске глобального экстремума перебирают все точки локального экстремума и среди них выбирают точку с наименьшим и наибольшим значением функции. Процедуры поиска экстремумов аналитическими методами хорошо описаны в литературе (см., например, [13]), однако в компьютерных программах обычно используются численные (приближенные) методы.

Простейшим методом поиска экстремума функции одной переменной, не требующим вычисления производной, является метод деления отрезка пополам. Он был известен еще в древней Греции и назывался **методом дихотомии**.

Пусть функция $f(x)$ на отрезке $[a, b]$ имеет единственный максимум в неизвестной точке x^* . Поскольку x^* находится в неизвестной точке отрезка $[a, b]$, то его называют отрезком неопределенности. Требуется с заданной точностью ε найти точку максимума функции $f(x)$.

Идея метода дихотомии заключается в том, чтобы на каждом шаге сокращать отрезок неопределенности практически вдвое до тех пор, пока его длина не станет меньше заданной точности. Рассмотрим схему алгоритма этого метода.

Пусть известно, что искомая точка x^* находится на отрезке $[a_0, b_0]$. Выберем середину отрезка $c_0 = [a_0 + b_0]/2$. Рассматриваются два отрезка: $[a_0, c_0 + \delta]$ и $[c_0 - \delta, b_0]$, где δ — малое число много меньше ε , т. е. $\delta \ll \varepsilon$. Если $f(c_0 - \delta) > f(c_0 + \delta)$, то максимум находится на отрезке $[a_0, c_0 + \delta]$, иначе на отрезке $[c_0 - \delta, b_0]$. За новый интервал неопределенности принимается отрезок $[a_1, b_1]$, где $a_1 = a_0$, $b_1 = c_0 + \delta$, в противном случае отрезок, где $a_1 = c_0 - \delta$, $b_1 = b_0$. Этот отрезок вновь делится пополам, и процедура продолжается до выполнения условия $|b_k - a_k| < \varepsilon$. При решении задачи на минимум при сравнении значений функций знак меняется на противоположный. На рис. 2.6. иллюстрируется метод дихотомии.

Рассмотрим для примера следующую задачу: на отрезке $[0, 4]$ найти точку минимума функции $f(x) = 2x^2 - 4x + 4$ с точностью до $\varepsilon = 0.5$.

Решение. Выберем параметр $\delta = 0.2$. Исходный отрезок неопределенности $[0, 4]$. Находим точки «середин» этого отрезка:

$$x_1 = \frac{a+b}{2} - \frac{\delta}{2} = 1.9; \quad x_2 = \frac{a+b}{2} + \frac{\delta}{2} = 2.1.$$

Вычислим в этих точках значения функции и сравним их:

$$f(1.9) = 3.62, f(2.1) = 4.42 \rightarrow f(x_1) < f(x_2).$$

В качестве нового отрезка неопределенности выбираем тот, где находится искомая точка минимума, т. е. $[a_1, b_1] = [a, x_1] = [0, 2.1]$.

Вновь определим «средние» точки:

$$x_3 = \frac{a_1 + b_1}{2} - \frac{\delta}{2} = 0.95; \quad x_4 = \frac{a_1 + b_1}{2} + \frac{\delta}{2} = 1.15.$$

Поскольку $f(0.95) < f(1.15)$, то полагаем новый отрезок неопределенности $[a_2, b_2] = [0, 1.15]$. Длина отрезка неопределенности $[a_2, b_2]$ равна $1.15 > \varepsilon$, поэтому продолжаем процесс.

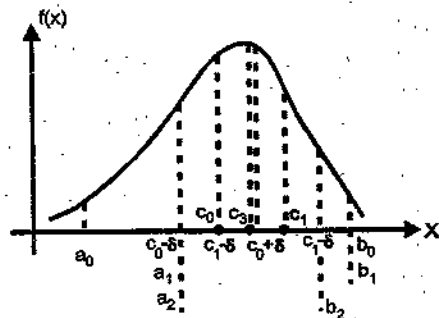


Рис. 2.6. Метод дихотомии

$[a_2, b_2] = [0.475, 1.15]$. Его длина 0.675, поэтому продолжаем процесс дихотомии.

$$x_7 = \frac{a_2 + b_2}{2} - \frac{\delta}{2} = 0.7125;$$

$$x_8 = \frac{a_2 + b_2}{2} + \frac{\delta}{2} = 0.9125.$$

Поскольку $f(x_7) = 2.1653$, $f(x_8) = 2.0153$, то новым отрезком неопределенности будет $[a_3, b_3] = [0.7125, 1.15]$. Для этого отрезка, имеющего длину 0.4375, процесс не продолжаем. Из двух точек x_7 и x_8 в качестве искомого приближения x^* выбираем точку x_8 , т. к. $f(x_8) < f(x_7)$. Можем дать следующий ответ: $x_* = 0.9125$, где $f(x_*) = 2.0153$. Точное же решение этой задачи: $x = 1$, следовательно, погрешность составила величину 0.0875.

Для функции многих переменных обычно применимы поисковые методы, основанные на свойстве градиента указывать направление наибольшего возрастания функции в данной точке.

Напомним, что градиентом скалярной функции $W(x_1, x_2, \dots, x_n)$ называют вектор

$$\text{grad } W(X) \equiv \nabla W(X) \equiv \left(\frac{\partial W(X)}{\partial x_1}, \frac{\partial W(X)}{\partial x_2}, \dots, \frac{\partial W(X)}{\partial x_n} \right)$$

Основное рекуррентное соотношение, определяющее выбор траектории поиска экстремальной точки, имеет следующий вид:

$$X^{k+1} = X^k \pm h \cdot \nabla W(X^k). \quad (2.17)$$

Знак «+» выбирается при поиске максимума, знак «-» — в противоположном случае.

Обобщенная схема градиентных методов такова. Выбирается «нулевое» приближение X^0 и шаг h . Затем по формуле (2.17) вычисляется ряд приближенных значений искомой точки экстремума $X^1, X^2, \dots, X^k, X^{k+1}$ до тех пор, пока не будет удовлетворено установленное условие точности, например $|f(X_{k+1}) - f(X_k)| < \varepsilon$.

Помимо изложенных и целого ряда других детерминированных методов широкое применение находят так называемые методы случайного поиска. Суть этих методов заключается в выборе величины и направления очередного шага случайным образом. Различают направленный и ненаправленный случайный поиск, случайный поиск с обучением и ряд других методов [14, 15].

Схема простейшего метода случайного поиска максимума может быть такой. Выбирается шаг h и из начальной (отправной) точки X^0 в случайном направлении совершается перемещение на расстояние h в точку X^1 . Случайное направление формируется в виде вектора, компоненты которого есть разыгранные значения случайной величины, равномерно распределенной на отрезке $[0, 1]$. Если значение функции $W(X^1)$ не увеличилось по сравнению с $W(X^0)$, то возвращаются в точку X^0 и в новом случайном направлении делают следующую попытку. В том случае, если значение $W(X^1) > W(X^0)$, точка X^1 принимается за отправную и процедура повторяется уже из этой точки. В каждой очередной точке процедура поиска улучшающего направления повторяется заданное число раз, например m раз. Та точка, из которой m раз были совершены неудачные попытки, принимается за точку максимума.

Основное преимущество методов случайного поиска в простоте их реализации и в том, что они не налагают никаких ограничений на функцию W . Недостатки, связанные с медленной

сходимостью этих методов (для получения высоких точностей требуется огромное число шагов), могут быть компенсированы применением высокоскоростных компьютеров.

2.2.2.2. Задачи на условный экстремум. Пусть задана задача следующего вида:

$$W(X) \rightarrow \max, \quad (2.18)$$

$$g_i(X) = 0 \quad i = \overline{1, m}; \quad m < n. \quad (2.19)$$

Классический подход к решению такого рода задач заключается в сведении их к задачам на безусловный экстремум. Наиболее распространенным является метод множителей Лагранжа. Суть его заключается во введении функции Лагранжа $L(X, \Lambda)$ и замены задачи (2.18)–(2.19) на задачу:

$$L(X, \Lambda) = W(X) + \sum_{i=1}^m \lambda_i \times g_i(X) \rightarrow \max, \quad (2.20)$$

где $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ — неизвестный вектор множителей Лагранжа.

Видно, что при выполнении условий (2.19), требование (2.20) эквивалентно требованию (2.18). Поэтому, решив задачу (2.20) и найдя векторы $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ и $\Lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$, получим решение исходной задачи (2.18)–(2.19).

Рассмотрим следующий пример [15]. Требуется найти максимум функции

$$W(x_1, x_2) = (x_1 - 1)^2 + x_2^2$$

при условии $g(x_1, x_2) = x_1^2 - x_2 + 1 = 0$.

В соответствии с методом Лагранжа сведем эту задачу к задаче на безусловный экстремум:

$$L(x_1, x_2, \lambda_1) = (x_1 - 1)^2 + x_2^2 + \lambda_1(x_1^2 - x_2 + 1) \rightarrow \max.$$

Затем вычислим частные производные и приравняем их нулю:

$$\frac{\partial L(x_1, x_2, \lambda_1)}{\partial x_1} = 2(x_1 - 1) + 2\lambda_1 = 0;$$

$$\frac{\partial L(x_1, x_2, \lambda_1)}{\partial x_2} = 2x_2 - \lambda_1 = 0;$$

$$\frac{\partial L(x_1, x_2, \lambda_1)}{\partial \lambda_1} = x_1^2 - x_2 + 1 = 0.$$

Решение этой системы уравнений дает следующий результат: $x_1^* = 0.313$, $x_2^* = 1.098$, $\lambda_1^* = 2.196$. Непосредственной проверкой убеждаемся, что точка $X^* = (x_1^*, x_2^*) = (0.313, 1.098)$ обеспечивает максимум целевой функции $W(x_1^*, x_2^*) = 1.678$.

2.2.3. Линейное программирование

2.2.3.1. Постановка задачи линейного программирования. Рассмотрим так называемую задачу о формировании производственной программы (например, выпуск книг). Типография может выпустить ряд подготовленных к изданию книг. Каждая из этих книг должна быть произведена по определенной технологии. Для издания книги необходимо затратить некоторые материалы, труд различных специалистов, другие ресурсы, которые имеются в ограниченном количестве. Реализация каждого экземпляра книги принесет известный доход. В то же время известны ограничения на число книг, которое может быть куплено. Требуется определить, сколько и каких книг надо произвести, чтобы при соблюдении имеемых ограничений получить наибольшую суммарную прибыль.

Построим математическую модель сформулированной задачи. Введем следующие обозначения. Пусть $X = (x_1, x_2, \dots, x_n)$ — искомый план издательства. Существующие ресурсы $B_1, B_2, \dots, B_m$ в количествах $b_1, b_2, \dots, b_m$ расходуются по определенной технологии: для изготовления одного экземпляра каждой j -ой книги необходим i -ый вид ресурса в количестве a_{ij} . Прибыль от продажи одного экземпляра j -ой книги равна c_j , и на рынке можно реализовать их не более k_j экземпляров.

Предположим, что к изданию подготовлено 26 наименований книг, а при изготовлении книг расходуются 13 видов ресурсов. В этом случае математическая модель задачи может быть записана так:

$$W(X) = c_1 x_1 + c_2 x_2 + \dots + c_{26} x_{26} \rightarrow \max$$

$$a_{1,1} x_1 + a_{1,2} x_2 + \dots + a_{1,26} x_{26} \leq b_1$$

$$a_{2,1} x_1 + a_{2,2} x_2 + \dots + a_{2,26} x_{26} \leq b_2$$

$$\dots \dots \dots$$

$$a_{13,1} x_1 + a_{13,2} x_2 + \dots + a_{13,26} x_{26} \leq b_{13}$$

$$0 \leq x_j \leq k_j \quad j = \overline{1, 26}$$

Это модель математического программирования, в которой и целевая функция, и все ограничения содержат только линейные функции. Модели такого типа называются моделями **линейного программирования**.

В общей постановке задача линейного программирования (ЛП) состоит в отыскании n неотрицательных значений переменных $x_1, x_2, \dots, x_n$, доставляющих максимальное значение линейной целевой функции при линейных ограничениях, а именно:

$$\begin{cases} W(x) = c_1x_1 + c_2x_2 + \dots \rightarrow \max \\ a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \end{cases} \quad (2.21)$$

$$\begin{cases} a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ \dots \dots \dots \end{cases} \quad (2.22)$$

$$\begin{cases} a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \\ \forall x_j \geq 0 \end{cases} \quad (2.23)$$

Можно записать большое число различных задач ЛП. Так, форма (2.21) может стремиться к минимуму, а в (2.22) могут присутствовать знаки и «<», и «>», и «=» в любом сочетании. Однако все задачи ЛП могут быть сведены к так называемой основной задаче линейного программирования (ОЗЛП), имеющей следующий вид:

$$\begin{cases} W(X) = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max \\ a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots \dots \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \\ \forall x_j \geq 0 \end{cases} \quad (2.24)$$

или в матричной форме

$$\begin{cases} W(X) = (C \times X) \rightarrow \max \\ A \cdot X = B \\ \forall x_j \geq 0 \end{cases} \quad (2.25)$$

Нестрогие неравенства в (2.22) можно легко преобразовать в соответствующие равенства (2.25), введя новые неотрицательные (искусственные) переменные. Например, неравенство $a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2$ превращается в строгое равенство $a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+1} = b_2$, т. е. посредством прибавления к левой части неравенства неотрицательной величины x_{n+1} . Если в (2.22) стоит знак «>», то величина x_{n+1} вычитается, т. е.

$$a_{k1}x_1 + \dots + a_{kn}x_n \geq b_k \rightarrow a_{k1}x_1 + \dots + a_{kn}x_n - x_{n+1} = b_k.$$

Таким образом, любую задачу линейного программирования можно привести к стандартной и полностью эквивалентной ей основной задаче линейного программирования (2.25). Для ОЗЛП разработаны специальные стандартизованные алгоритмы решения. Схема решения произвольной задачи ЛП такова: построенная модель линейного программирования приводится к ОЗЛП, находится решение последней, а затем полученное решение интерпретируется в терминах исходной постановки задачи (Например, исключаются из рассмотрения искусственные переменные, меняется знак в значении целевой функции и т. п.).

2.2.3.2. Анализ ОЗЛП. Очевидно, что задача (2.25) относится к задачам выбора, поскольку условие $m < n$ (считая ранг матрицы A равным m) гарантирует в общем случае бесконечное количество допустимых решений.

Так как частные производные от целевой функции есть постоянные числа не все равные нулю

$$\exists x_j: \frac{\partial W(x_1, x_2, \dots, x_n)}{\partial x_j} = c_j \neq 0,$$

то $W(X)$ не может иметь экстремум во внутренних точках ОДР. Это обстоятельство серьезно увеличивает сложность решения задачи, которое в общем виде впервые было получено Л. В. Канторовичем лишь в 1939 году.

Ограничения (2.22) и (2.23) могут определять одну из следующих ситуаций:

- область допустимых решений пустая — нет решения задачи;
- область допустимых решений не ограничена — возможно неограниченное решение задачи (такое решение не реализуемо на практике и обычно считается невозможным);
- область допустимых решений конечна и непуста. Этот вариант и рассматривается в дальнейшем.

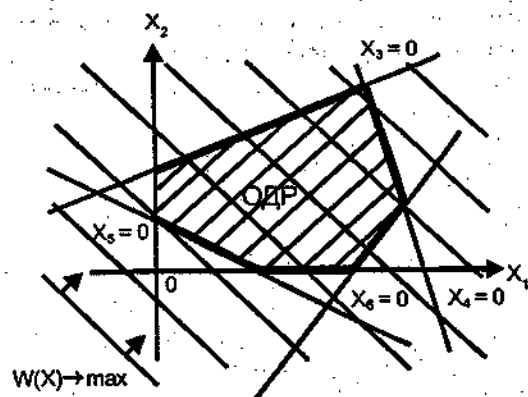


Рис. 2.7. Постановка задачи линейного программирования

Область допустимых решений, задаваемая линейными условиями (2.22) и (2.23), есть выпуклый многогранник. Следовательно, задача имеет глобальный оптимум. Показано [12], что решение ОЗЛП лежит в так называемой угловой точке,

т. е. там, где $n - m$ переменных обращаются в ноль. Отсюда следует основная идея решения задачи: просмотреть допустимые угловые точки и из них выбрать ту, где $W(X)$ достигает максимума. На рис. 2.7 представлена иллюстрация решения ОЗЛП для плоского случая ($n - m = 2$). Для решения задач линейного программирования разработано большое количество методов решения (см., например, [12, 16], для которых имеются доступные пакеты прикладных задач.

2.2.3.3. Двойственность в линейном программировании. Для каждой задачи линейного программирования можно построить так называемую двойственную к ней задачу. Пусть задана исходная задача (2.21)–(2.24). Двойственная к ней строится по следующей схеме:

каждому i -му ограничению ставится в соответствие двойственная переменная y_i , причем неравенству в исходной задаче соответствует неотрицательная переменная ($y_i \geq 0$), равенству в исходной задаче соответствует переменная, от которой не требуется неотрицательности;

матрица ограничений заменяется на транспонированную матрицу $A \rightarrow A^T$;

ограничениями в двойственной задаче выступают коэффициенты целевой функции исходной задачи;

коэффициентами целевой функции двойственной задачи выступают ограничения исходной задачи;

неравенства в ограничениях заменяются на противоположные; смысл экстремума заменяется на противоположный.

В итоге можно рассматривать следующую пару двойственных задач линейного программирования:

$$\begin{cases} W(X) = (C \cdot X) \rightarrow \max \\ A \cdot X \leq B \\ \forall x_j \geq 0 \end{cases} \quad \begin{cases} Z(Y) = (B \cdot Y) \rightarrow \min \\ A^T \cdot Y \geq C \\ \forall y_i \geq 0 \end{cases}$$

Очевидно, что задача двойственная к двойственной является исходной задачей. В этой связи любую из пары двойственных задач можно принять за исходную, тогда другая по отношению к ней будет двойственной.

Введение в рассмотрение пары двойственных задач имело большое теоретическое и практическое значение. Наиболее важный теоретический результат дает основная теорема двойственности: «Если одна из пары двойственных задач имеет решение, то и другая двойственная по отношению к ней также имеет решение, причем

$$\max W(X) = \min Z(Y).$$

Если же одна из задач имеет неограниченный оптимум, то двойственная к ней не имеет решения (система ограничений несовместна)».

Содержательная интерпретация двойственных переменных полезна при решении ряда прикладных задач. Так, например, если прямая задача заключается в определении состава выпускаемой продукции, приносящей наибольший суммарный доход, то двойственная по отношению к ней задача заключается в поиске цен (так называемых теневых цен) на ресурсы, обеспечивающих минимизацию расходов на последние [16].

2.2.3.4. Целочисленное линейное программирование. В целом ряде случаев в задаче линейного программирования присутствует дополнительное условие о том, что все или некоторые переменные должны быть целочисленными. Эти условия возникают тогда, когда дробные значения переменных противоречат их физическому смыслу. С точки зрения математики это условие приводит к принципиально новому классу моделей — целочисленному программированию. Принципиальная особенность таких задач заключается в том, что X принадлежит дискретному

множеству, а это в свою очередь требует пересмотра многих положений теории линейного программирования.

В общем случае задача целочисленного линейного программирования (ЦЛП) имеет вид:

$$\begin{cases} W(X) = (C \cdot X) \rightarrow \max \\ A \cdot X = B \\ \forall x_j \geq 0 \\ \exists x_j - \text{целые.} \end{cases} \quad (2.26)$$

На практике последнее условие в (2.26) часто заменяется условием булевости переменных, т. е. x_j принимает одно из двух возможных значений: 0 или 1.

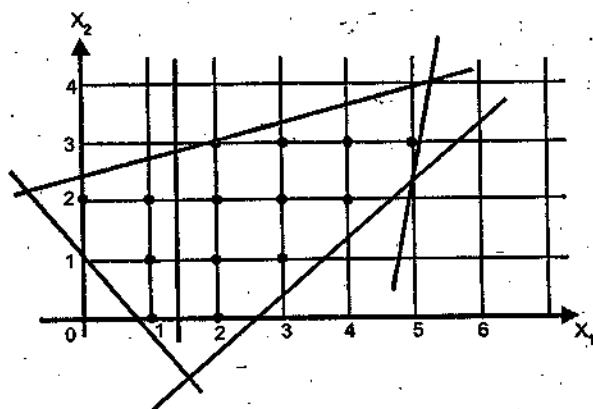


Рис. 2.8. Область допустимых значений в задаче целочисленного программирования

Как показано на рис. 2.8, область возможных значений в ЦЛП — множество целочисленных точек, расположенных внутри ОДР соответствующей нецелочисленной задачи линейного программирования [17]. В самом общем случае решение этой задачи возможно лишь путем перебора всех допустимых решений. При большой размерности задачи такой путь бесперспективен ввиду длительности процедур решения даже при применении современных высокоскоростных компьютеров.

Среди методов решения задач ЦЛП наибольшее распространение на сегодня получил метод ветвей и границ [18]. Суть этого метода в следующем.

Находится решение соответствующей нецелочисленной задачи линейного программирования (в задаче 2.26 отсутствует условие целочисленности переменных). Поскольку ОДР задачи ЛП не меньше ОДР задачи ЦЛП (она больше или равна ей), то найденное решение — верхняя граница задачи ЦЛП — всегда не хуже искомого целочисленного!

Если это решение целочисленное, то оно и является искомым решением. В противном случае область допустимых решений делят на две части (проводят ветвление), причем граница разделения проводится по целому значению одного из неизвестных, которое является ближайшим к полученному нецелому при решении соответствующей нецелочисленной задачи.

Затем решают две задачи, соответствующие полученным областям деления, которые вновь проверяют на целочисленность. Поскольку границы проводятся по целочисленному значению одной из переменных, то с каждым шагом все больше шансов получить целочисленную вершину ОДР для соответствующей нецелочисленной задачи рассматриваемой подобласти. Добавим, что разветвление продолжаем до тех пор, пока целочисленному решению не будет отвечать наибольшее значение целевой функции.

Проиллюстрируем вышесказанное на следующем простом примере [15].

Максимизировать

$$W(x_1, x_2) = 3x_1 + 4x_2 \quad (2.27)$$

при условиях

$$\begin{aligned} 3x_1 + 2x_2 &\leq 8; \\ x_1 + 4x_2 &\leq 10; \end{aligned} \quad (2.28)$$

$$\begin{aligned} x_1, x_2 &\geq 0; \\ x_1, x_2 &\text{ — целые.} \end{aligned} \quad (2.29)$$

Решение. Вначале, решив соответствующую исходной задаче линейного программирования (2.27)–(2.28), получим нецелочисленное решение: $X_0 = (x_1 = 1.2, x_2 = 2.2)$ при значении целевой функции $W(X_0) = 12.4$.

Ветвим задачу на две, введя (рис. 2.9) условия $x_1 \leq 1$ и $x_1 \geq 2$.

$$\left. \begin{aligned} W(x_1, x_2) &= 3x_1 + 4x_2 \\ 3x_1 + 2x_2 &\leq 8 \\ x_1 + 4x_2 &\leq 10 \\ x_1 &\leq 1 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} (2.30)$$

$$\left. \begin{aligned} W(x_1, x_2) &= 3x_1 + 4x_2 \\ 3x_1 + 2x_2 &\leq 8 \\ x_1 + 4x_2 &\leq 10 \\ x_1 &\geq 2 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} (2.31)$$

Решение этих задач дает следующие результаты:

для (2.30) $X_{11} = (x_1 = 1, x_2 = 2.25)$, $W(X_{11}) = 12$;

для (2.31) $X_{12} = (x_1 = 2, x_2 = 1)$, $W(X_{12}) = 10$.

Решение X_{12} — целочисленное. Однако решение X_{11} дает большее значение целевой функции, и там может находиться искомый максимум, поэтому проводим разветвление задачи (2.30). Поскольку x_1 в X_{11} целое, то разветвление проводим по x_2 , вводя условия $x_2 \leq 2$ и $x_2 \geq 3$. Вновь рассматриваем две нецелочисленные задачи:

$$\left. \begin{aligned} W(x_1, x_2) &= 3x_1 + 4x_2 \\ 3x_1 + 2x_2 &\leq 8 \\ x_1 + 4x_2 &\leq 10 \\ x_2 &\leq 2 \\ x_1 &\leq 1 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} (2.32)$$

$$\left. \begin{aligned} W(x_1, x_2) &= 3x_1 + 4x_2 \\ 3x_1 + 2x_2 &\leq 8 \\ x_1 + 4x_2 &\leq 10 \\ x_2 &\geq 3 \\ x_1 &\leq 1 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} (2.33)$$

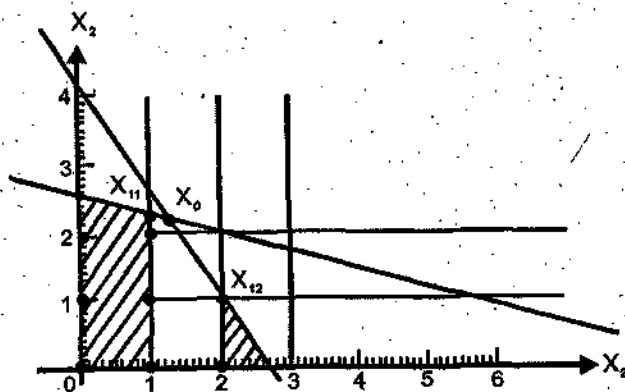


Рис. 2.9. Пример решения задачи ЦЛП

Решение для (2.32): $X_{21} = (x_1 = 1, x_2 = 2)$. Решение для (2.33) не существует. Поскольку полученное решение целочисленное и соответствующее значение целевой функции наибольшее, то X_{21} — решение исходной целочисленной задачи (2.27)–(2.29), т. е. $X^* = (x_1^* = 1, x_2^* = 2)$, причем $W(X^*) = 11$.

2.2.3.5. Примеры построения моделей линейного программирования для некоторых практических задач.

Задача 1. Для поддержания нормальной жизнедеятельности человеку ежедневно необходимо потреблять не менее 118 г белков, 56 г жиров, 500 г углеводов, 8 г минеральных солей. Количество питательных веществ, содержащихся в 1 кг каждого потребляемого продукта, а также цены за 1 кг каждого из этих продуктов приведены в следующей таблице:

Питательные вещества	мясо	рыба	молоко	масло	сыр	крупа	картофель
Белки	180	190	30	10	260	130	21
Жиры	20	3	40	865	310	30	2
Углеводы	-	-	50	6	20	650	200
Мин. соли	9	10	7	12	60	20	10
Цена 1 кг	1.8	1.0	0.28	3.4	2.9	0.5	0.1

Составить дневной рацион, содержащий не менее минимальной суточной нормы потребности человека в необходимых питательных веществах при минимальной общей стоимости потребляемых продуктов.

Решение. Построим соответствующую математическую модель. Пусть x_i — количество (в кг) i -го вида продуктов, приобретаемое человеком в соответствии с дневным рационом, следовательно:

$$x_i \geq 0, \quad i = \overline{0, 7}.$$

Нормальный дневной рацион человека обеспечивают те количества продуктов, которые удовлетворяют условиям:

$$\begin{aligned} 180x_1 + 190x_2 + 30x_3 + 10x_4 + 260x_5 + 130x_6 + 21x_7 &\geq 118; \\ 20x_1 + 3x_2 + 40x_3 + 865x_4 + 310x_5 + 30x_6 + 2x_7 &\geq 56; \\ 50x_3 + 6x_4 + 20x_5 + 650x_6 + 200x_7 &\geq 500; \\ 9x_1 + 10x_2 + 7x_3 + 12x_4 + 60x_5 + 20x_6 + 10x_7 &\geq 8. \end{aligned}$$

Общие расходы на запланированные продукты должны быть минимальны:

$$1.8x_1 + 1.0x_2 + 0.28x_3 + 3.4x_4 + 2.9x_5 + 0.5x_6 + 0.1x_7 \rightarrow \min.$$

Требуется определить состав дневного рациона человека, т. е. найти вектор $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, удовлетворяющий вышеприведенным условиям.

Построенная модель является моделью линейного программирования, т. к. и ограничения, и целевая функция содержат только линейные функции. Решение задачи имеет вид: $X^* = (0, 0, 0, 0, 0.0335, 0.905)$, т. е. необходимо потреблять только 34 г сыра и 905 г картофеля, затрачивая 0.57 денежных единиц. Заметим, что это наилучшее решение поставленной задачи в смысле принятого критерия (расходов на питание). Здесь не ставятся такие условия, как разнообразие питания, ограничения на потребление отдельных видов продуктов и т. п.

Задача 2. Имеется три участка земли различной структуры и размера: 600, 180, и 220 га. С учетом демографического состава семей требуется построить не менее 290, 180, 110 и 420 домов различных проектов. Количество домов, которые можно разместить на одном гектаре площади, задается матрицей, где C_{ij} — количество домов типа j на i -ом участке:

$$C = \begin{pmatrix} 40 & 30 & 18 & 24 \\ 45 & 28 & 22 & 18 \\ 50 & 22 & 14 & 16 \end{pmatrix}.$$

Сколько гектаров каждого участка следует застроить каждым видом домов, чтобы общее число домов было наибольшим?

Решение. Введем обозначения: x_{ij} — количество га застроенного домами j -го проекта на i -м участке. Очевидно, что $x_{ij} \geq 0$.

Построенные дома не могут занимать площадь больше имеемой на каждом участке:

$$\begin{aligned} x_{11} + x_{12} + x_{13} + x_{14} &\leq 600 \\ x_{21} + x_{22} + x_{23} + x_{24} &\leq 180 \\ x_{31} + x_{32} + x_{33} + x_{34} &\leq 220. \end{aligned} \quad (2.34)$$

Построенных домов должно быть не менее заданного количества:

$$\begin{aligned} x_{11} + x_{21} + x_{31} &\geq 290 \\ x_{12} + x_{22} + x_{32} &\geq 180 \\ x_{13} + x_{23} + x_{33} &\geq 110 \\ x_{14} + x_{24} + x_{34} &\geq 420. \end{aligned} \quad (2.35)$$

Количество построенных домов должно быть наибольшим:

$$40x_{11} + 30x_{12} + 18x_{13} + 24x_{14} + 45x_{21} + 28x_{22} + 22x_{23} + 18x_{24} + 50x_{31} + 22x_{32} + 14x_{33} + 16x_{34} \rightarrow \max \quad (2.36)$$

Модель (2.34)–(2.36) является задачей линейного программирования. Отметим, что поскольку в условиях все коэффициенты при неизвестных равны единице, то задача относится к так называемой транспортной задаче, для которой разработаны специализированные методы решения [19]. Ответ в данной задаче следующий: первый участок застраивается домами второго и четвертого проектов на территории $x_{12} = 180$ и $x_{14} = 420$ га, второй участок домами первого и третьего проектов на территории $x_{21} = 70$, $x_{23} = 110$ га, а третий участок только домами первого проекта на территории $x_{31} = 220$ га.

Задача 3. Для проведения психологических тестов каждый из тестируемых должен пройти 5 процедур проверки. При переходе от процедуры к процедуре оборудование проходит переналадку. Время (в мин), необходимое при переналадке с одного теста на другой, задано матрицей переходов, где d_{ij} время перехода с i -той процедуры к j -той:

$$[d_{ij}] = \begin{pmatrix} 0 & 7 & 7 & 2 & 4 \\ 6 & 0 & 2 & 4 & 1 \\ 6 & 1 & 0 & 9 & 1 \\ 2 & 3 & 8 & 0 & 5 \\ 6 & 1 & 2 & 5 & 0 \end{pmatrix}.$$

Требуется определить такой порядок прохождения процедур проверки, чтобы общее время на прохождение всех тестов было наименьшим.

Решение. Построим математическую модель задачи. Пусть величина x_{ij} описывает факт переналадки оборудования с i -ой процедуры на j -ю:

$$x_{ij} = \begin{cases} 1, & \text{если переналадка } (i, j) \text{ входит в решение;} \\ 0, & \text{в противном случае.} \end{cases}$$

Тогда стремление минимизировать общее время на прохождение всех переналадок выразится как стремление минимизировать сумму времен, затрачиваемых на переналадки:

$$W = \sum_{i=1}^5 \sum_{j=1}^5 d_{ij} x_{ij} \rightarrow \min. \quad (2.37)$$

После i -ой процедуры может следовать только одна переналадка, и перед каждой j -ой процедурой была только одна переналадка, т. е.:

$$\forall i: \sum_{j=1}^5 x_{ij} = 1, \quad \forall j: \sum_{i=1}^5 x_{ij} = 1. \quad (2.38)$$

Для того чтобы каждый тестируемый проходил все 5 проверок, необходимо ввести следующие ограничения:

$$\begin{aligned} x_{ii} &= 0 \\ x_{ik} + x_{ki} &\leq 1 \\ x_{ik} + x_{kj} + x_{ji} &\leq 2 \\ x_{ik} + x_{ki} + x_{ij} + x_{ji} &\leq 3 \end{aligned} \quad (2.39)$$

Задача (2.37)–(2.39) есть задача целочисленного линейного программирования, решение которой определяет следующую последовательность переходов $X = (x_{32} = 1, x_{24} = 4, x_{41} = 2, x_{15} = 4, x_{53} = 2)$, обеспечивающую наименьшее общее время, затрачиваемое на переналадки, равное 13.

2.2.4. Нелинейное программирование

В самом общем виде задача нелинейного программирования имеет вид: найти $X = (x_1, x_2, \dots, x_n)$, для которого

$$W(X) \rightarrow \max \quad (2.40)$$

при условиях

$$g_i(X) \leq 0 \quad i = \overline{1, m}, \quad m < n \quad (2.41)$$

$$X \geq 0. \quad (2.42)$$

Основные отличия нелинейной задачи от линейной следующие: область допустимых решений может быть и невыпуклой, следовательно, иметь большое число локальных экстремумов; в нелинейных задачах решение может лежать и во внутренних точках, и на границе ОДР. В этой связи на сегодняшний день не существует общих методов, которые бы позволяли решать любые задачи нелинейного программирования. Общие методы разработаны для задач какого-либо определенного вида, например выпуклого программирования, когда и целевая функция, и функции ограничений являются выпуклыми функциями, т. е.

удовлетворяют условию:

$$f(X) \leq \lambda f(X_1) + (1 - \lambda)f(X_2),$$

где $X = \lambda X_1 + (1 - \lambda)X_2, \quad 0 \leq \lambda \leq 1$.

Одним из эффективных методов решения задач нелинейного программирования является **метод штрафных функций** [15]. Идея этого метода заключается в переводе исходной задачи нелинейного программирования к задаче на безусловный экстремум. Далее задача решается модификацией градиентного метода. При этом целевая функция в области допустимых решений совпадает с целевой функцией исходной задачи, а при нарушении границ ОДР резко меняет значение таким образом, что на следующем шаге выбирается решение вновь в допустимой области.

Итак, пусть задана задача (2.40)–(2.42). Построим соответствующую задачу на безусловный экстремум:

$$F(X) = W(X) - H(X) \rightarrow \max,$$

где $H(X)$ — положительная функция штрафа, строящаяся по схеме:

$$H(X) = \sum_{i=1}^m \alpha_i \cdot g_i(X),$$

$$\text{причем } \alpha_i = \begin{cases} 0, & \text{если } g_i(X) \leq 0 \\ \alpha, & \text{если } g_i(X) > 0. \end{cases}$$

Число α должно быть столь велико, чтобы при нарушении границы ОДР оказывать воздействие, достаточное для возвращения в допустимую область за один шаг. Формально это обстоятельство можно описать условием:

$$\left| \frac{\partial H(X)}{\partial x_j} \right| \gg \left| \frac{\partial W(X)}{\partial x_j} \right|.$$

В общем случае рекуррентное соотношение выбора каждой координаты x_j следующего $(k+1)$ -го приближения имеет вид:

$$x_j^{k+1} = \max \left\{ 0, x_j^k + h \left[\frac{\partial W(X^k)}{\partial x_j} + \sum_{i=1}^m \alpha_i(X^k) \frac{\partial g_i(X^k)}{\partial x_j} \right] \right\} \quad (2.43)$$

Для того чтобы начать решение задачи методом штрафных функций, необходимо задать начальное значение искомой точки экстремума X^0 , а также величины h и α_i . Затем по формуле (2.43) построить ряд уточняющих значений $X^1, X^2, \dots, X^{k+1}$. Величина X^{k+1} считается решением задачи, когда выполнится условие достижения заданного порога точности, например такое:

$$\forall i: |x_i^{k+1} - x_i^k| < \varepsilon \quad (2.44)$$

Рассмотрим пример. Максимизировать целевую функцию:

$$W(x_1, x_2) = x_1 + 2x_2 - x_1^2 - 0.5x_2^2$$

при условиях

$$g_1(x_1, x_2) = -2x_1 - 3x_2 + 6 \geq 0;$$

$$g_2(x_1, x_2) = -x_1 - 4x_2 + 5 \geq 0;$$

$$x_1, x_2 \geq 0.$$

Решение. Зададимся величинами $X^0 = (0, 0)$, $h = 0.2$ и $\alpha_i = 2$.

Затем вычислим производные:

$$\frac{\partial W}{\partial x_1} = 1 - 2x_1 \quad \frac{\partial W}{\partial x_2} = 2 - x_2$$

$$\frac{\partial g_1}{\partial x_1} = -2 \quad \frac{\partial g_1}{\partial x_2} = -3$$

$$\frac{\partial g_2}{\partial x_1} = -1 \quad \frac{\partial g_2}{\partial x_2} = -4$$

Теперь запишем формулу (2.43) для каждой из координат:

$$x_1^{k+1} = \max\{0, x_1^k + 0.2[1 - 2x_1^k + \alpha_1(X^k)(-2) + \alpha_2(X^k)(-1)]\};$$

$$x_2^{k+1} = \max\{0, x_2^k + 0.2[2 - x_2^k + \alpha_1(X^k)(-3) + \alpha_2(X^k)(-4)]\}.$$

Для первого шага координаты X^1 вычисляются так: $\alpha_i = 0$, так как ограничения выполнены:

$$x_1^1 = \max\{0, 0 + 0.2[1 - 0 + 0 + 0]\} = 0.2;$$

$$x_2^1 = \max\{0, 0 + 0.2[2 - 0 + 0 + 0]\} = 0.4.$$

Следовательно, $X^1 = (0.2, 0.4)$, $W(X^1) = 0.88$, ограничения задачи выполняются, поэтому $\alpha_1(X^1) = \alpha_2(X^1) = 0$.

Результаты второго шага;

$$x_1^2 = \max\{0, 0.2 + 0.2[1 - 0.4 + 0 + 0]\} = 0.32;$$

$$x_2^2 = \max\{0, 0.4 + 0.2[2 - 0.4 + 0 + 0]\} = 0.72.$$

Отсюда $X^2 = (0.32, 0.72)$. $W(X^2) = 1.40$. На этом шаге и далее до пятого шага найденные координаты находятся в ОДР, поэтому значения штрафной функции равны нулю. Итак, по аналогии с предыдущими двумя шагами, имеем:

после третьего шага: $X^3 = (0.4, 1.0)$, решение допустимо и $W(X^3) = 1.74$; после четвертого шага $X^4 = (0.44, 1.2)$. Здесь одно из ограничений (второе) не выполняется, поэтому на пятом шаге следует положить:

$$\alpha_1(X^4) = 0 \quad \text{и} \quad \alpha_2(X^4) = 2.$$

Результаты пятого шага:

$$x_1^5 = \max\{0, 0.44 + 0.2[1 - 0.88 + 0 - 2]\} = 0.06$$

$$x_2^5 = \max\{0, 1.2 + 0.2[2 - 1.2 + 0 - 8]\} = 0$$

$$W(X^5) = 0.064.$$

Полученное решение вновь допустимо. Продолжая данный алгоритм, после десятого шага получаем решение $X^{10} = (0.37, 1.07)$ и $W(X^{10}) = 1.8$. Описанный процесс может продолжаться до достижения заданной точности (см. (2.44)).

В ряде пакетов прикладных программ предусмотрено решение задач нелинейного программирования для конкретного типа функций, чаще всего в виде полиномов. В более сложных случаях требуется предварительное нахождение производных и ввод их в компьютер.

Для решения задач нелинейного программирования с успехом могут применяться методы случайного поиска, построенные по схеме аналогичной описанной в п. 2.2.2.1.

2.2.5. Динамическое программирование

Решение задач математического программирования связано с серьезными ограничениями, накладываемыми на целевую функцию и функции ограничений, а также с трудностями организации алгоритма поиска искомого оптимума. Представляет интерес разработать метод, практически не зависящий от характера функций, входящих в задачу математического программирования. И такой метод, получивший название **метод динамического программирования**, был разработан Р. Беллманом. Основная идея метода заключается в том, чтобы развернуть процесс достижения поставленной цели в последовательность

отдельных этапов, на каждом из которых находится так называемое *условное оптимальное решение*.

Рассмотрим схему постановки задачи оптимизации, для которой применим метод динамического программирования.

1. Исследуемая система описывается совокупностью n параметров $x_1, x_2, \dots, x_n$, следовательно, $X = (x_1, x_2, \dots, x_n)$ можно рассматривать как точку в n -мерном фазовом пространстве. Каждое фиксированное значение этих параметров (фазовых координат) задает ее определенное состояние:

$$S^i = S^i(x_1^i, x_2^i, \dots, x_n^i), S \in \Omega,$$

где Ω — пространство возможных состояний системы. В каком бы состоянии S^i ни находилась система, дальнейший процесс ее движения в фазовом пространстве зависит от данного состояния и не зависит от того, каким путем она была переведена в данное состояние (удовлетворяющий таким условиям процесс называют процессом без последствий).

2. Процесс реализуется по этапам (шагам). На каждом k -том этапе принимается некоторое частное решение u_k (управление), под воздействием которого система по закону f переходит в новое состояние, т. е. $S_{k+1}^i = f(S_k^i, u_k)$. Следует подчеркнуть, что на каждом k -том этапе система может находиться в различных состояниях $\{S_k^i\}$, допустимых для этого этапа. Частное управление выбирается только из заданного множества управлений (допустимых управлений):

$$u_k(S) \in U.$$

3. Переход в новое состояние связан с достижением некоторого эффекта w_k , зависящего от достигнутого после предыдущего этапа состояния S_k^i и принятого частного решения u_k :

$$w_k = w(S_k^i, u_k).$$

Общий эффект за N этапов есть сумма эффектов на каждом из этапов:

$$W = \sum_{k=1}^N w_k.$$

4. Требуется найти такую последовательность частных решений $u_1^*, u_2^*, \dots, u_n^*$, называемых оптимальным решением, чтобы, исходя из заданного исходного состояния S_0 , за N этапов получить наибольший суммарный эффект:

$$W_k^*(S_1, U^*) = \max_{u_k} \sum_{k=1}^N w_k.$$

Описанный процесс может быть моделью как естественных динамических многошаговых процессов, так и искусственно разделенных на отдельные шаги процессов, где время вообще не фигурирует.

Приведем некоторые примеры.

1. Задача о рюкзаке. Выбранный для похода рюкзак фиксированного объема требуется наполнить необходимыми в походе вещами, каждая из которых имеет некоторый объем и определенную ценность. Требуется подобрать набор вещей, помещающихся в рюкзак и имеющих наибольшую суммарную ценность. Очевидно, что эта задача может быть формализована моделью целочисленного линейного программирования и решена соответствующим методом. Однако можно представить себе процесс наполнения рюкзака как пошаговый процесс, на каждом шаге которого принимается решение относительно одной вещи (класть ее в рюкзак или не класть) и который укладывается в вышеописанную схему.

2. Задача о поиске кратчайшего пути на сети. Находясь в исходной вершине S_0 требуется построить кратчайший, т. е. имеющий наименьшую сумму весов, путь в конечную вершину S_r .

С математической точки зрения эта задача достаточно элементарна: построить все возможные варианты путей, а затем выбрать из них самый короткий. Проблема же заключается в том, что это путь полного перебора вариантов практически нереализуем для реальных задач. Поэтому требуется найти способ существенного сокращения по отношению с полным перебором числа операций, необходимых для решения задачи.

Если процесс поиска пути представить как последовательный выбор каждого очередного состояния, в которое можно перейти из данного состояния за один шаг (по одной дуге сети) и формализовать его также вышеописанной схемой, то можно применить метод динамического программирования, обеспечивающий, как правило, нахождение оптимального решения задачи за приемлемое число шагов.

Перейдем теперь к изложению метода динамического программирования, опираясь на условия последней задачи. Принципиальным моментом в многошаговых процессах является

то, что эффективность последующего развития процесса существенно зависит от всех решений, принятых на предыдущих шагах. Это приводит к тому, что с линейным ростом числа шагов (этапов) количество вариантов полных траекторий растет в экспоненциальной (степенной) зависимости и приводит к «взрыву» числа альтернатив, требующих рассмотрения.

Однако среди всех этапов только последний обладает тем свойством, что, принимая частное решение, не надо заботиться о последующих шагах. На последнем этапе число альтернатив равно числу вариантов перехода из предпоследнего состояния в последнее. Если же на последнем этапе найдено лучшее частное решение, то уже предпоследний этап обладает вышеописанным свойством, поскольку необходимо выбирать управление лишь на данном (предпоследнем) шаге, прибавляя найденный ранее наилучший эффект на последнем шаге. Продолжая таким образом путь из «конца в начало», можно найти на каждом шаге так называемые условные оптимальные частные решения, совокупность которых и даст искомое решение задачи «из начала в конец».

Опираясь на подобные рассуждения, Р. Беллман сформулировал свой известный принцип, который позволил построить эффективный алгоритм решения задач оптимизации методом динамического программирования. Вот этот принцип: *оптимальная траектория обладает тем свойством, что каково бы ни было данное состояние системы, выбор частного решения на этом этапе должен быть таким, чтобы в совокупности частные решения на всех последующих этапах составляли оптимальную программу.* Иначе говоря, если вся траектория оптимальна, то оптимальна и программа завершающих этапов, начиная с любого этапа.

Этот принцип позволяет построить следующие так называемые рекуррентные соотношения динамического программирования. На каждом k -ом этапе частное управление u_k необходимо выбирать по правилу:

$$\forall i: W_k^*(S_k^i) = \max_{u_k} [w(S_k^i, u_k) + W_{k+1}(f(S_k^i, u_k))], \quad (2.45)$$

где: $k = \overline{1, N}$ — номер этапа; j — номер состояния, в которое попадает система после принятия управления u_k , если перед этим она находилась в состоянии i .

Для последнего же шага ($k = N$) это правило приобретает вид:

$$\forall i: W_N^*(S_N^i) = \max_{u_N} [w(S_N^i, u_N)]. \quad (2.46)$$

Теперь становится ясной схема решения задачи оптимизации методом динамического программирования.

1. Весь процесс разбивается на N этапов (шагов). Определяется исходное состояние системы и функция перехода f , определяющая новое состояние системы, в которое она может перейти под воздействием некоторого управляющего частного решения $S_{k+1}^i = f(S_k^i, u_k)$. Для каждого состояния на всех этапах должен быть определен эффект $w(S_k^i, u_k)$, который получится, если принять управление u_k , когда система была в состоянии S_k^i .

2. Для каждого из возможных состояний системы на предпоследнем этапе решается уравнение (2.46).

3. Последовательно для $k = N-1, N-2, \dots, 3, 2, 1$ решается уравнение (2.45) и для каждого из соответствующих этапов определяются условные оптимальные управления $\{u_k^*\}$.

4. Для заданного исходного состояния системы выписывается последовательность промежуточных состояний и приводящим к ним безусловных оптимальных частных управлений:

$$u_1^* \quad u_2^* \quad u_3^* \quad \dots \quad u_N^* \\ S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow \dots \rightarrow S_N.$$

Решение задачи есть вектор оптимального управления

$$U^* = (u_1^*, u_2^*, u_3^*, \dots, u_N^*) -$$

(оптимальный план), переводящий систему из исходного состояния S_1 с максимальным эффектом W_1 .

Замечание. При решении задачи, когда эффект заключается в минимизации целевой функции, в уравнениях (2.45) и (2.46) требуется выполнять операцию $\min$.

Пример. Сумма в 200 миллионов рублей распределяется среди четырех предприятий. Прибыль, получаемая каждым предприятием при выделении ему определенной суммы, задана следующей таблицей (выделяемая сумма кратна 40 млн. руб.).

Выделенная сумма	Прибыль 1 ^{го}	Прибыль 2 ^{го}	Прибыль 3 ^{го}	Прибыль 4 ^{го}
40	8	6	3	4
80	10	9	6	6
120	11	11	7	8
160	12	13	11	13
200	18	15	18	16

Требуется так распределить выделенную сумму, чтобы суммарная прибыль от всех четырех предприятий была наибольшей.

Решение. Из приведенной таблицы (даже для столь простых условий) трудно быстро подобрать правильный ответ. Так, если все деньги отдать первому предприятию, то получим прибыль в 18 млн руб, если же первому и четвертому дать по 80, а второму — 40, то прибыль составит 22 млн руб. и т. д. Из постановки задачи следует, что выполнены все условия применимости метода динамического программирования. Действительно. В качестве шага естественно принять выделение суммы одному предприятию ($k = 1, 2, 3, 4$). Состояние системы S определяется объемом нераспределенной суммы. На каждом k -ом этапе возможны некоторые из состояний: $S_k^1 = 0$, $S_k^2 = 40$, $S_k^3 = 80$, $S_k^4 = 120$, $S_k^5 = 160$ или $S_k^6 = 200$ млн руб. Управляющее воздействие на k -ом этапе — количество выделенных средств (u_k). При этом получается эффект $w(S_k, u_k)$, значения которого выбираются из заданной таблицы.

На четвертом последнем этапе, независимо от распределений на предыдущих этапах, целесообразно все оставшиеся средства выделить четвертому предприятию. Оставшихся же средств может быть 0, 40, 80, 120, 160 или 200. При этом в соответствии с (2.46) будет получена, соответственно, прибыль 0, 4, 6, 8, 13 или 16 млн руб.

S_k	u_k	3-й этап			2-й этап			1-й этап		
		S_{k+1}^1	W_3	W_4	W_2	W_3	W_4	W_1	W_2	W_3
40	0	40	0	4	0+4=4	0	3	0+4=4	0	6
	40	0	3	0	3+0=3	6	0	6+0=6	8	0
80	0	80	0	6	0+6=6	0	4	0+4=4	0	9
	40	40	3	4	3+4=7	6	3	6+3=9	8	6
	80	0	4	0	4+0=4	9	0	9+0=9	10	0
120	0	120	0	8	0+8=8	0	7	0+7=7	0	13
	40	80	3	6	3+6=9	6	4	6+4=10	8	9
	80	40	4	4	4+4=8	9	3	9+3=12	10	6
	120	0	7	0	7+0=7	11	0	11+0=11	11	0
160	0	160	0	13	0+13=13	0	11	0+11=11	0	16
	40	120	3	8	3+8=11	6	7	6+7=13	8	13
	80	80	4	6	4+6=10	9	4	9+4=13	10	9
	120	40	7	4	7+4=11	11	3	11+3=14	11	6
	160	0	11	0	11+0=11	13	0	13+0=13	12	0
200	0	200	0	16	0+16=16	0	18	0+18=18	0	19
	40	160	3	13	3+13=16	6	11	6+11=17	8	16
	80	120	4	8	4+8=12	9	7	9+7=16	10	13
	120	80	7	6	7+6=13	11	4	11+4=15	11	9
	160	40	11	4	11+4=15	13	3	13+3=16	12	6
	200	0	18	0	18+0=18	15	0	15+0=15	18	0

На третьем этапе, выделяя средства третьему предприятию, необходимо все оставшиеся средства (после выделения первому и второму) разделить между третьим и четвертым предприятиями так, чтобы суммарный эффект был наибольшим. Например, если остались нераспределенными 40 млн рублей, то их можно отдать третьему (прибыль 3 млн руб.) или четвертому (прибыль 4 млн руб.). В соответствии с (2.45) надо выбрать второй вариант распределения. Если же останется 80 млн руб, то возможны три варианта: 80 и 0, 40 и 40 или 0 и 80. Соответствующая прибыль 4, 7 или 6 млн руб. Выгоднее второй вариант, поскольку прибыль будет наибольшей и равна 7 млн руб.

Все варианты дальнейших рассуждений, определяемых алгоритмом динамического программирования, приведены в следующей таблице (в целях экономии места таблицы для всех трех этапов совмещены, и для каждого этапа необходимо рассматривать соответствующие столбцы). Решения, соответствующие условным оптимальным решениям в таблице выделены.

Теперь можно записать все возможные безусловные управления и соответствующие им эффективности.

Состояние	4-й этап		3-й этап		2-й этап		1-й этап	
	W_4	u_4	W_3	u_3	W_2	u_2	W_1	u_1
40	4	40	4	0	6	40	8	40
80	6	80	7	40	10	40	14	40
120	8	120	9	40	13	80	18	40
160	13	16	13	0	16	80	21	40
200	16	200	18	200	19	40	24	40

Из таблицы видно, что для получения наибольшей суммарной прибыли в 24 млн руб. необходимо выделить первому, второму и четвертому предприятиям по 40 млн руб, третьему предприятию — 80 млн рублей. Поэтому ответ задачи: оптимальный план $U^* = (u_1^* = 40, u_2^* = 40, u_3^* = 80, u_4 = 40)$ обеспечивает эффект $W_1^* = 24$ млн руб.

Глава 2.3. Модели принятия решений

В организационных и социальных системах важное место занимают индивидуумы, имеющие право принимать решения. Эти индивидуумы, или лица принимающие решения (ЛПР), реализуют процедуры принятия решений, обеспечивающие преобразование информации о состоянии системы в элементы командной информации. Основные составляющие процесса принятия решения — это формирование множества возможных решений (альтернатив) и выбор одной из них в качестве собственно решения.

Каждый человек, порою не замечая этого, принимает решения. Как одеться при поездке за город, что выбрать себе на обед, какой строить дом, принимать ли участие в конкурсе и т. п.

В зависимости от преследуемых целей, условий внешней среды, степени своей информированности и других факторов ЛПР оказывается в различных ситуациях. В каждой из ситуаций логика процесса принятия решения, а следовательно, и модель, может быть различной. Введем следующую классификацию задач принятия решений: в условиях определенности, в условиях риска, в условиях неопределенности, в конфликтных ситуациях.

2.3.1. Принятие решений в условиях определенности

Принятие решения в условиях определенности характеризуется наличием однозначной или детерминированной связи между принятым решением и его исходом. В наиболее простых случаях каждой альтернативе удается поставить в соответствие определенный исход, оцениваемый некоторым числом — «полезностью» исхода. При этом возможно установить прямую связь: альтернатива — численное значение соответствующего ей исхода, минуя сами исходы. Эта связь описывается целевой скалярной функцией, определенной на множестве альтернатив. Оптимальным решением считают ту альтернативу, которая доставляет целевой функции наибольшее (наименьшее) значение. Иначе говоря, нахождение оптимального решения равносильно нахождению экстремума функции при некоторых условиях. Подобная схема реализуется в рамках методов математического программирования, которые рассматривались в предыдущей главе.

На практике исходы чаще оцениваются по нескольким показателям:

$$W(X) = (W_1(X), W_2(X), \dots, W_m(X)). \quad (2.47)$$

В этих случаях говорят о принятии решений при векторном критерии.

Если все показатели в (2.47) считать показателями типа «выигрыш», то соответствующую задачу принятия решений можно сформулировать в виде следующей системы уравнений —

$$X_j^* = \arg \left(\max_{X \in \Omega} W_j(X) \right), \quad j = \overline{1, m},$$

решение которой есть пересечение отдельных решений, т. е.:

$$X^* = X_1^* \cap X_2^* \cap X_3^* \cap \dots \cap X_m^*$$

К сожалению, показатели, входящие в (2.47), обычно противоречивы. Например, при имеющихся ресурсах хотим соорудить большое, красивое и дешевое здание. Ясно, что удовлетворить одновременно всем этим требованиям невозможно. Формально это означает, что искомое множество часто оказывается пустым: $X^* = \emptyset$. Более того, с точки зрения математики задача принятия оптимального решения при векторном критерии, вообще говоря, не корректна.

В этой связи вводят понятие **эффективного решения** X^* , иногда называемое оптимальным по Парето. Содержательно эффективным является такое допустимое решение, что не существует другого плана, который по всем характеристикам был бы лучше данного. Формально множество эффективных решений при векторном критерии $\{X^*\}$, называемое множеством Парето, определяется следующим образом:

$$\forall X_j \in \Omega, j = \overline{1, m}, \exists X^* \in \Omega: W(X^*) \geq W(X_j),$$

причем, по крайней мере, одно из неравенств является строгим.

Рассмотрим пример. Предположим, что при разработке нового ассортимента нас интересуют: объем производства W_1 и качество моделей W_2 . При этом имеется возможность варьировать в некоторых пределах эти показатели (детали модели, сложность отделки, материал и т. п.).

Каждый возможный вариант можно отобразить в виде точки в двумерном пространстве (W_1, W_2) . На рис. 2.10. показана область возможных вариантов решений (ОДР). Поскольку мы хотим

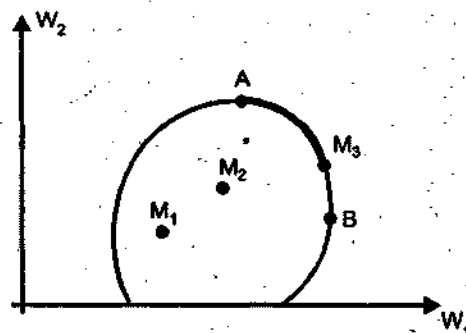


Рис. 2.10. Множество Паретооптимальных решений

максимизировать обе характеристики W_1 и W_2 , ясно, что точка M_1 не является оптимальной по Парето, так как точка M_2 имеет значения W_1 и W_2 большие, чем точка M_1 . Однако и точка M_2 не относится к искомому множеству X^* , поскольку, например, у точки M_3 значения W_1 и значения W_2 выше, чем у

точки M_2 . Из рисунка видно, что множество точек, образующих множество Парето, есть верхняя грань области (AB), образно говоря, ее «северо-восточной граница». Любая точка, лежащая в этом множестве, не хуже произвольной точки из всего множества допустимых решений.

Итак, выбор варианта решения при нескольких критериях целесообразно производить только из множества оптимальных по Парето точек. С точки зрения эффективности все они несравнимы и могут считаться эквивалентными. Для данного примера любое из оптимальных по Парето решений $X^*(W_1^*, W_2^*)$ обладает таким свойством, что увеличение, например, объемов производства может привести к улучшению качества моделей.

Отметим, что поиск оптимальных по Парето множеств представляет весьма сложную задачу, которая в общем виде не решена. *Чтобы принять решение в условиях многокритериальности, всегда необходима дополнительная информация!*

Наиболее известны два формальных способа введения дополнительной информации: упорядочение показателей по важности (ранжирование критериев) и задание весов относительной важности для свертывания критериев.

Пусть все критерии упорядочены по предпочтению: W_1 предпочтительнее критерия W_2 ; он, в свою очередь, предпочтительнее W_3 и т. д. Используя знак предпочтения « $\succ$ », это можно записать так:

$$W_1 \succ W_2 \succ \dots \succ W_m$$

Теперь задача выбора решений может быть записана, например, в следующем виде:

$$W_1(X) \rightarrow \max, \quad (2.48)$$

при этом на другие критерии наложены ограничения:

$$\begin{aligned} W_2(X) &\geq W_2^0(X) \\ W_3(X) &\geq W_3^0(X) \\ &\dots\dots\dots \\ W_m(X) &\geq W_m^0(X), \end{aligned} \quad (2.49)$$

где $W_2^0(X)$, $W_3^0(X)$, ..., $W_m^0(X)$ — назначенные верхние уровни соответствующих показателей.

Очевидно, что задача (2.48)–(2.49) есть задача математического программирования. Решив эту задачу и найдя W_1^* , делают уступку; например, требуют, чтобы $W_1 = \alpha W_1^*$, где α — число «немного» меньше единицы. Формулируют и решают следующую задачу:

$$\begin{aligned} W_2(X) &\rightarrow \max \\ W_1(X) &= \alpha \cdot W_1^*(X) \\ W_3(X) &\geq W_3^0(X) \\ &\dots\dots\dots \\ W_m(X) &\geq W_m^0(X). \end{aligned}$$

Далее может быть сделана уступка для второго показателя и т. д. Слабость такого метода уступок заключается в необходимости ранжирования критериев и определения величины уступок, что и является дополнительной информацией, которой нет в условии исходной задачи. Тем самым решается некоторая «другая» задача.

Другой подход заключается в введении весов относительной важности критериев и свертки их в один интегрированный показатель. Это также позволяет свести задачу к одномерной оптимизации. Наиболее известна следующая форма свертки:

$$W(X) = \sum_{j=1}^m \alpha_j \frac{W_j(X) - q_j^*}{q_j^*}, \quad (2.50)$$

где α_j — вес j -го критерия, причем $q_j^* = \max_{X \in W} W_j(X)$,

$$\sum_{j=1}^m \alpha_j = 1, \forall \alpha_j \geq 0, j = \overline{1, m}.$$

Рассмотрим пример. Для постройки здания имеются средства в объеме 120 единиц. Разработано 6 проектов зданий. Каждое из зданий характеризуется размером жилой площади, качеством планировки и общей стоимостью. Пользуясь значениями характеристик, сведенных в таблицу, требуется выбрать лучший проект.

Номер проекта	Размер площади	Цена	Качество планировки
1	1500	19	1
2	1200	18	2
3	950	20	2
4	700	15	4
5	600	16	5
6	500	14	3

Из таблицы следует, что есть проект с самой качественной планировкой (5), есть самый дорогой (3) и самый дешевый (6) проекты, можно выбрать проекты самого большого (1) и самого маленького здания (6). Но самый «лучший» проект здания из таблицы выбрать невозможно. Для этого нужна дополнительная информация. Во-первых, следует уточнить, что означает в данном случае понятие «лучший проект». Во-вторых, если сделаем попытку выбрать проект с наибольшими (наименьшими или средними и т. д.) значениями показателей, то окажется, что таких в таблице вообще нет. Проект с наибольшим значением одной характеристики не имеет «лучшего значения» другой характеристики. Именно в таких случаях говорят, что критерии противоречивы. При попытке формирования обобщенного критерия мы столкнемся с необходимостью определения в (2.50) численных значений весов α_j , что также есть введение дополнительной информации.

Одним из наиболее продуктивных подходов в разрешении проблемы многокритериальности является введение полезности [21]. Под полезностью данного варианта решения x понимают некоторую скалярную величину $u(x)$, обладающую следующими свойствами:

1. Если вариант решения x_i предпочтительнее, чем вариант x_j , $x_i \succ x_j$, то их полезности соотносятся так: $u(x_i) > u(x_j)$;

2. Если $x_i \succ x_j$, а $x_j \succ x_k$, то $u(x_i) > u(x_k)$ — транзитивность;

3. Если $u(x_i + x_j)$ — полезность варианта x_i и варианта x_j , то $u(x_i + x_j) = u(x_i) + u(x_j)$ — аддитивность.

Введение понятия полезности позволяет формализовать процесс принятия решения, если ЛПР может упорядочить альтернативы по предпочтению. Однако и в этом случае необходимо ввести дополнительную информацию.

Рассмотрим возможную методику введения полезности варианта решения. Пусть имеется два варианта решений x_i и x_j , причем x_i предпочтительнее, чем x_j , т. е. $x_i \succ x_j$. В этом случае, используя понятие полезности, поступают так:

— наименее предпочтительному варианту приписывают полезность, равную единице: $u(x_j) = 1$;

— для более предпочтительного варианта определяют вероятность $\alpha < 1$, при которой достижение результата x_i эквивалентно результату x_j : $\alpha u(x_i) = u(x_j) = 1$. Отсюда $u(x_i) = 1/\alpha$. Содержательно это означает, что полезность предпочтительного варианта, реализующегося с вероятностью α (меньше единицы!), может быть приравнена к менее предпочтительному варианту, который можно реализовать наверняка.

Если $x_1 \succ x_2 \succ x_3 \dots x_n$, то описанная выше методика дает следующие полезности:

$$u(x_1) = 1, u(x_2) = 1/\alpha_2, u(x_3) = 1/(\alpha_2 \cdot \alpha_3), \dots, u(x_n) = 1/(\alpha_2 \cdot \alpha_3 \cdot \dots \alpha_n).$$

2.3.2. Принятие решений в условиях риска

Рассмотрим ситуацию, когда любое решение x_i связано с возможностью наступления одного из целого ряда результатов (исходов) $r_1, r_2, \dots, r_m$, причем каждый из них имеет известную вероятность появления $p(r_j | x_i)$, $j = \overline{1, m}$. Это означает, что в результате принятия i -го решения может появиться любой из m исходов, причем заранее неизвестно, какой именно. Иначе говоря, принимая конкретное решение, мы рискуем получить любой из возможных исходов.

В этих условиях (условиях риска) решение ЛПР должно быть основано на некоторых осредненных оценках, поскольку они могут быть сформированы на некоторой объективной научной основе. Конечно, ЛПР всегда рискует в конкретной ситуации получить не тот результат, на который его ориентируют осредненные оценки. Однако в случае, когда в схожих ситуациях решение принимается многократно, получение рекомендаций,

основанных на экстремальных значениях осредненных показателей эффективности может быть очень полезным.

Пусть для каждого варианта решения x_i и возможного результата r_j найдена полезность $u(r_j, x_i)$. Тогда в качестве меры целесообразности для выбора i -ого решения может быть принята некоторая ожидаемая (по аналогии с математическим ожиданием) полезность:

$$E(x_i) = \sum_{j=1}^m u(r_j, x_i) \cdot p(r_j | x_i)$$

Наилучшим в условиях риска принимается решение, обеспечивающее наибольшую ожидаемую полезность, т. е.:

$$x^* = \arg \max_{x_i} E(x_i)$$

2.3.3. Принятие решений в условиях концептуальной неопределенности

На практике часто встречаются задачи, когда принятие решения x_i ($i = 1, n$) может привести к любому из результатов $r_1, r_2, \dots, r_m$, однако вероятность появления каждого из этих результатов существенно зависит не только от принятого решения, но и от неконтролируемого состояния окружающей среды $S_1, S_2, \dots, S_k$, т. е. $p(r_j | x_i, S_k)$.

Если вероятность нахождения окружающей среды в состоянии S_k известна и равна $p(S_k)$, то решение аналогично случаю схемы принятия решения в условиях риска.

$$x^* = \arg \max_{x_i} E(x_i) = \arg \max_{x_i} \sum_{j=1}^m \sum_{k=1}^l U(r_j, x_i) \cdot p(r_j | x_i, S_k) \cdot p(S_k)$$

На практике же обычно неизвестно, какое из состояний среды на момент принятия решения имеет место и каково распределение вероятностей $p(S_k)$. Такую ситуацию, чтобы отличать от условий риска, иногда называют концептуальной неопределенностью.

Принимая решение в условиях концептуальной неопределенности, необходимо дополнительно сформулировать правило (критерий), обеспечивающее выбор оптимальной стратегии. Выбор такого правила принятия решений зависит от ЛПР и эквивалентен введению им в задачу дополнительной информации. Решение x^* , принятое в соответствии с одним из таких правил называют оптимальным в смысле принятого критерия.

Рассмотрим наиболее распространенные правила принятия решений в условиях концептуальной неопределенности.

1. **Правило Вальда**, позволяющее реализовать схему «осторожного ЛПР», имеет следующий вид:

$$x^* = \arg \max_{x_i} \min_{S_k} \sum_{j=1}^m U(r_j, x_i) \cdot p(r_j | x_i, S_k)$$

Это правило гарантирует наибольший выигрыш при «наихудшем» состоянии среды.

2. **Правило Лапласа**, реализующее схему «безразличного ЛПР». Оно обычно применяется в тех случаях, когда есть основания полагать все возможные состояния среды равновероятными, т. е. $p(S_k) = 1/N$:

$$x^* = \arg \max_{x_i} \frac{1}{N} \sum_{j=1}^m U(r_j, x_i) \cdot p(r_j | x_i, S_k)$$

где N — число возможных состояний среды.

3. **Правило Гурвица** носит «универсальный» характер и записывается следующим образом:

$$x^* = \arg \max_{x_i} \left\{ \alpha \max_{S_k} U(x_i, S_k) + (1 - \alpha) \min_{S_k} U(x_i, S_k) \right\}$$

где $U(x_i, S_k) = \sum_{j=1}^m U(r_j, x_i) \cdot p(r_j | x_i, S_k)$, α — коэффициент доверия, т. е. вероятность, с которой среда может находиться в самом выгодном для ЛПР состоянии.

Очевидно, что при $\alpha = 0$ правило Гурвица превращается в правило Вальда, при $\alpha = 1$ получаем правило «оптимиста», когда ЛПР надеется, что среда будет находиться в самом благоприятном для него состоянии.

4. **Правило Сэвиджа** (минимизация сожалений). Вводят понятие «сожаления» как величину, равную изменению полезности результата при данном состоянии среды относительно ее наилучшего (в смысле ЛПР) состояния. «Сожаление» определяют следующим образом. В каждом столбце матрицы

$\|U(x_i, S_k)\|$ находят максимальный элемент,

$$U_k^* = \max_{x_i} U(x_i, S_k),$$

который вычитают из всех элементов соответствующего столбца. Для полученной матрицы — матрицы «сожаления»

$$\|U_{ik}^*\| = \|U(x_i, S_k) - U_k^*\| -$$

применяют правило Севиджа:

$$x^* = \arg \max_{x_i} \min_{S_k} U_{ik}^*.$$

Это правило можно рассматривать как минимизацию возможных потерь при условии, что состояние среды наихудшим образом отличается от предполагаемого.

Очевидно, что применение при принятии решения того или иного правила является субъективным по существу. Вместе с тем такой подход полезен для анализа ситуации. Например, если рекомендации совпадают с точки зрения нескольких правил, то это является серьезным фактором, говорящим в пользу рекомендованного решения. В противном случае имеется возможность оценить каждое решение с других позиций и принять более обоснованное решение. Следует заметить, что отказ от принятия решения в условиях неопределенности также есть решение и не всегда самое лучшее.

Рассмотрим пример. При подготовке книги к издательству требуется определить ее тираж. Типография может предложить следующие тиражи: 5, 8, 15 или 50 тысяч экземпляров. В зависимости от конъюнктуры рынка возможный спрос может составить 1, 2, 5, 12, 25 или 40 тысяч экземпляров. Анализ возможных расходов и доходов издательства позволил построить следующую таблицу ежегодных прибылей (U_{ik}).

S_k	1	2	5	12	25	40
x_i						
5	-12	6	25	25	25	25
8	-17	1	20	38	38	38
15	-22	-3	15	33	52	52
50	-26	-8	10	28	47	65

Определим целесообразный тираж, используя различные правила принятия решения. Правило Вальда:

$$x^* = \arg \max_{x_i} \min_{S_k} U_{ik} = 5.$$

Ясно, что в данном случае правило Вальда не применимо, так как при наименьшем спросе всегда будет убыток. Можно сказать, что «осторожный» ЛПР от выпуска книги в данной ситуации откажется.

Правило Лапласа:

$$x^* = \arg \max_{x_i} \frac{1}{6} \{94, 118, 127, 116\} = 15.$$

Правило Гурвица, например, при $\alpha = 0.2$:

$$x^* = \arg \max_{x_i} \{-4.6, -6, -7.2, -7.8\} = 5.$$

Правило Севиджа. Построим матрицу «сожалений»:

S_k	1	2	5	12	25	40
x_i						
5	0	0	0	-13	-27	-40
8	-5	-5	-5	0	-14	-27
15	-10	-9	-10	-5	0	-13
50	-14	-14	-15	-10	-5	0

Из нее находим величину

$$\max_{x_i} \min_{S_k} U_{ik}^* = \max(-40, -27, -13, -15) = -13.$$

Следовательно, оптимальным в смысле критерия Севиджа является решение $x^* = 15$.

В итоге можно констатировать, что тираж должен быть: если применять правило Вальда, то 5 тыс. экз. (или же вообще отказаться от тиража);

если применять правило Лапласа, то 15 тыс. экз.;

если применять правило Севиджа, то 15 тыс. экз.;

если применять правило Гурвица, то 5 тыс. экз., когда заказчик — пессимист, и 50, когда оптимист. При степени уверенности в наибольший спрос 0.2 рекомендуется тираж 5 тыс. экз.

Каков ответ в данной задаче? Если Вы скажете, к какому типу ЛПР относитесь, то можно рекомендовать то или иное из полученных выше решений. В целом же заслуживает внимания вариант издания с тиражом в 15 тыс. экз.

2.3.4. Принятие решений в условиях конфликта

Часто приходится принимать решения в ситуациях, когда интересы сторон (например, конкурентов по бизнесу) частично или полностью противоположны. В этих случаях перед каждой стороной возникает задача выбора наилучшего образа действий для достижения своей цели. Подобные задачи решаются обычно в рамках теории игр [22].

Введем некоторые понятия. Стороны, участвующие в конфликте, называются игроками. Возможные действия каждого игрока называются стратегиями. Для разрешения конфликта игрокам приходится принимать какие-то решения путем выбора той или иной стратегии. Выбор одним из игроков своей стратегии называется ходом.

Будем рассматривать простейший случай в теории игр — игру двух игроков с нулевой суммой: то есть когда интересы двух игроков А и В прямо противоположны: выигрыш одного игрока равен проигрышу другого, а их сумма всегда равна нулю. Пусть игрок А имеет возможность принять одно из m решений: $A_1, A_2, \dots, A_m$, а игрок В может принять одно из n решений: $B_1, B_2, \dots, B_n$.

Математической моделью такой игры является платежная матрица C , элемент которой c_{ij} определяет результат игры: выигрыш игрока А, если он примет решение A_i , а игрок В примет решение B_j . При этом игрок В проигрывает сумму c_{ij} (или выигрывает сумму $-c_{ij}$).

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & c_{m3} & \dots & c_{mn} \end{pmatrix}$$

Игрок А стремится максимизировать свой выигрыш, а игрок В стремится минимизировать свой проигрыш. Какие из стратегий выберут игроки, если каждый игрок считает своего противника столь же «умным», как он сам, и не надеется на его промахи (принцип разумности противника)?

Если игрок А использует стратегию A_i , то он с гарантией получит выигрыш, по крайней мере $\min_j c_{ij}$, где минимум берется по всем стратегиям игрока В. Поскольку А свободен в

выборе своей стратегии, то он стремится сделать гарантированный выигрыш v_1 как можно большим:

$$v_1 = \max_i \min_j c_{ij}.$$

Эта стратегия называется оптимальной для А. Она обеспечивает при любом поведении игрока В выигрыш для А не менее v_1 , поэтому величина v_1 называется **нижней ценой** игры или **максимумом**. Игрок В, рассуждая подобным образом, выбирает стратегию B_j , при которой он проигрывает не более чем v_2 :

$$v_2 = \min_j \max_i c_{ij}.$$

Такая стратегия для игрока В оптимальна. При любом своем поведении игрок А получит сумму не более чем v_2 . Величина v_2 называется **верхней ценой** игры или **минимумом**. Поскольку число v_1 минимальное в строке i , то:

$$\max_i \min_j c_{ij} \leq c_{ij}.$$

Число v_2 — максимально в столбце j , следовательно:

$$\min_j \max_i c_{ij} \geq c_{ij}.$$

Объединяя эти два неравенства, получаем, что $v_1 \leq v_2$, или

$$\max_i \min_j c_{ij} \leq c_{ij} \leq \min_j \max_i c_{ij},$$

т. е. нижняя цена игры никогда не превосходит верхней цены.

Если верхняя цена игры равна нижней цене игры, то оптимальные стратегии игроков совпадают. Пара таких стратегий $(i, j)^*$ образует **седловую точку**. Принципиально важным является тот факт, что всякое отклонение от стратегии выбора седловой точки невыгодно обоим игрокам. В этой связи стратегии в играх с седловой точкой называются **чистыми стратегиями**.

Рассмотрим пример. Служба маркетинга, проанализировав возможности двух однопрофильных фабрик по производству и сбыту одного и того же вида товара, построила платежную матрицу:

$$C = \begin{pmatrix} -2 & 2 & -1 & 5 \\ 4 & 3 & 2 & 2 \\ 6 & 1 & 0 & -3 \end{pmatrix}$$

где c_{ij} — доход фабрики А (убыток фабрики В). Какие действия фабрик будут наиболее целесообразными?

Найдем нижнюю цену игры. В первой строке меньший элемент $c_{11} = -2$, во второй $c_{23} = 2$ и в третьей $c_{34} = -3$. Отсюда максимин:

$$v1 = \max_i \min_j c_{ij} = \max (-2, 2, -3) = 2,$$

и оптимальной стратегией для фабрики А будет A_2 .

Найдем верхнюю цену игры. В первом столбце наибольший элемент $c_{31} = 6$, во втором $c_{22} = 3$, в третьем $c_{23} = 2$ и в четвертом $c_{14} = 5$. Получаем минимакс:

$$v2 = \min_j \max_i c_{ij} = \min(6, 3, 2, 5) = 2,$$

и оптимальной стратегией для фабрики В будет B_3 .

Поскольку $v1 = v2 = 2$, что соответствует элементу c_{23} , матрица имеет седловую точку (2,3). Следовательно, фабрике А целесообразно придерживаться стратегии A_2 , а фабрике В — стратегией B_3 .

Если платежная матрица не имеет седловой точки, т. е. $v1 < v2$, то игрок А выиграет не менее $v1$, а игрок В проиграет не более $v2$. Однако между верхней и нижней ценами игры остается некоторая область неопределенности, в которой игроки могут попытаться улучшить свою игру. Как поступать в этом случае? Выход может быть получен, если чередовать стратегии случайным образом.

Предположим, что игрок А для каждой возможной стратегии заранее установил соответствующие вероятности $p_1, p_2, \dots, p_m$. Смешанной стратегией S_A игрока А назовем случайное чередование им своих стратегий с установленными вероятностями. Аналогичным образом введем стратегии S_B игрока В с вероятностями $q_1, q_2, \dots, q_n$. Смешанные стратегии удобно обозначать как элементы линейного векторного пространства:

$$S_A = (p_1, p_2, \dots, p_m), \quad \sum_{i=1}^m p_i = 1;$$

$$S_B = (q_1, q_2, \dots, q_n), \quad \sum_{j=1}^n q_j = 1.$$

Очевидно, что при такой ситуации, выигрыш становится случайной величиной и его целесообразно измерять, например, математическим ожиданием выигрыша:

$$V(S_A, S_B) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} \cdot p_i \cdot q_j.$$

Оптимальной смешанной стратегией игрока будем считать ту стратегию S_A^* , применение которой гарантирует игроку наибольший средний выигрыш по сравнению с другими смешанными стратегиями независимо от действий игрока В. Аналогично вводится оптимальная смешанная стратегия S_B^* для игрока В.

Нижней ценой игры на множестве смешанных стратегий будет выступать величина

$$V1 = V2(S_A^*) = \max_{S_A} \min_{S_B} V(S_A, S_B),$$

а стратегия S_A^* будет называться *максиминной* или *оптимальной* стратегией игрока А.

Для игрока В оптимальной будет являться минимаксная стратегия, обеспечивающая верхнюю цену игры:

$$V2 = V2(S_B^*) = \min_{S_B} \max_{S_A} V(S_A, S_B).$$

Можно показать, что всегда выполняется условие $V1 = V2$.

Проанализируем преимущества, которые получают игроки, придерживающиеся оптимальных смешанных стратегий. Поскольку чистые стратегии являются частными случаями смешанных стратегий, имеем:

$$\max_i \min_j c_{ij} \leq \max_{S_A} \min_{S_B} V(S_A, S_B).$$

Следовательно, применение оптимальной смешанной стратегии позволяет игроку А получить гарантированный выигрыш не меньший (а возможно и больший), чем он имел бы, применяя свою оптимальную чистую стратегию.

Аналогично для игрока В:

$$\min_{S_B} \max_{S_A} V(S_A, S_B) \leq \min_j \max_i c_{ij}.$$

Значит, применяя оптимальную смешанную стратегию, игрок В не увеличит (а возможно и уменьшит) свой проигрыш по сравнению со случаем применения им своей оптимальной чистой стратегии.

Для платежной матрицы с седловой точкой всегда достаточно просто отыскать оптимальные чистые стратегии. Если же седловой точки нет, то отыскание оптимальных смешанных стратегий представляет самостоятельную проблему. Ее можно разрешить, например, применяя методы линейного программирования [23].

Глава 2.4. Некоторые перспективные направления в математическом моделировании

2.4.1. Методологические аспекты имитационного моделирования

Математическая модель может быть моделью либо косвенного, либо прямого описания. Модели косвенного описания — это формальные конструкции аналитических соотношений. Язык такого описания ограничен лишь четырьмя элементами: неизвестное, параметр, математическая операция и ограничения.

При использовании моделей косвенного описания всегда возникает задача разрешения математических соотношений относительно неизвестных и (или) параметров модели, что обычно влечет за собой серьезные вычислительные трудности. Для решения задач, основанных на использовании моделей этого типа, разработано большое количество аналитических и численных методов, используются также методы качественного анализа. Именно к моделям косвенного описания применимы методы математического анализа, математического программирования, регрессионные методы и т. п.

В общем случае между объектом и его косвенным описанием отсутствует структурное подобие. Пусть, к примеру, при решении задачи планирования производственной программы построена модель линейного программирования:

$$W(X) = \sum_{j=1}^n c_j x_j = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \rightarrow \max$$

$$\sum_{j=1}^n a_{1j} x_j = a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \leq b_1$$

$$\sum_{j=1}^n a_{2j} x_j = a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n \leq b_2$$

$$\dots \dots \dots$$
$$\sum_{j=1}^n a_{mj} x_j = a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \leq b_m$$

$$\forall x_j \geq 0.$$

Анализируя эту модель, исследователь интерпретирует получаемые результаты в терминах постановки задачи. Однако если другой исследователь, не знающий предметного содержания исходной задачи, увидит эту модель линейного программирования, то у него не будет оснований утверждать,

что речь идет о производственной программе. Иначе говоря, модели косвенного описания теряют структурное подобие описываемым объектам.

Главное достоинство косвенных моделей заключается в том, что они, как правило, описывают широкий круг объектов и с той или иной степенью адекватности применимы для их исследования. Для их исследования разрабатываются эффективные математические методы, они изучаются и широко известны. Можно говорить об их относительной универсальности.

По мере усложнения объекта получение структурно-подобного отображения становится возможным лишь с помощью так называемых моделей прямого описания. Характерной особенностью этих моделей является их структурное подобие рассматриваемому объекту.

Прямое описание всегда возможно, если объект является конструктивным, т. е. вопрос о его существовании непосредственно связан с доказательством возможности построения этого объекта. В основе процесса построения конструктивных объектов лежат следующие положения [27]:

- * задана совокупность исходных объектов, рассматриваемых в качестве элементарных;
- * заданы правила образования новых объектов из построенных ранее;
- * построения осуществляются в рамках заданных правил отдельными шагами относительно простым способом.

В качестве примера конструктивных объектов могут быть названы слова в конечном алфавите, конечные графы и автоматы, некоторые специальные алгоритмы и т. п. В конечном счете, конструктивный объект может быть описан совокупностью натуральных чисел.

Конструктивные объекты являются предметами рассмотрения конструктивной математики. Одно из центральных мест в этом сравнительно новом направлении математики занимает изучение соответствий, заданных при помощи алгоритмов. Так, например, понятие традиционной математики «функция» связывается с теми соответствиями, которые описаны посредством алгоритма, позволяющего эффективно построить значения функций по значениям ее аргументов. Главное преимущество конструктивного способа построения математики состоит в том, что при этом

сравнительно просто выявить те данные, по которым может быть построено решение данной задачи.

На основе вышеизложенного сформулируем следующее определение: **имитационной** называем модель прямого описания конструктивного объекта. Процесс построения имитационной модели и последующего его приложения назовем имитационным моделированием.

Теперь становятся очевидными и все основные черты имитационного моделирования.

1. *Имитационная модель обычно изображается в виде алгоритма над данным алфавитом.* По этой причине они могут быть также названы алгоритмическими. Это означает, что в некотором алфавите задано общепринятое предписание, определяющее потенциально осуществимый процесс над конструктивными объектами достаточно общего вида. Существует несколько стандартных форм изображения алгоритма, например нормальные алгоритмы А. А. Маркова, машины Тьюринга, операторные алгоритмы Ван Хао, которые являются различными уточнениями интуитивного понимания алгоритма. Очень часто имитационные модели изображаются в виде взаимосвязанной совокупности конечных автоматов, что и приводит некоторых авторов к отождествлению понятий автоматных и имитационных моделей. В то же время перспективность построения имитационных моделей на основе конечных, в первую очередь вероятностных и нечетких, автоматов не вызывает сомнений и подтверждается целым рядом работ [24,25].

2. *Имитационные модели обеспечивают явное представление топологии рассматриваемых объектов*, что является следствием прямого описания последних. Процедура прямого описания объекта заключается в следующем. Вначале объект расчленяется на элементарные составляющие. Между этими составляющими устанавливаются несложные, подчас простейшие взаимосвязи, которые могут быть воспроизведены чисто механическим путем, т. е. конструктивно.

Следует подчеркнуть, что в некоторых моделях элементами могут выступать весьма сложные объекты, которые на данном уровне агрегирования воспринимаются в качестве неделимых и рассматриваются с точки зрения «входа-выхода». Именно это обстоятельство позволяет для одного и того же объекта в зависимости от целей и возможностей исследователя строить имитационные модели с различной степенью детализации. Таким

образом, модульность построения имитационных моделей вытекает из самой сути этих моделей и является их характерной чертой.

Конструктивный характер имитационных моделей, по крайней мере теоретически, позволяет достичь любой степени детализации, а, следовательно, адекватности описания рассматриваемого объекта. Имитационные модели могут быть составлены с таким приближением к реальному объекту, с каким это позволяет наше понимание проблемы и необходимость предварительных требований. Это обстоятельство позволяет разрабатывать модели нескольких модификаций, предназначенных для различных уровней управления, ситуаций и решаемых задач, изменяя лишь степень детализации в описании отдельных сторон объекта моделирования.

Отметим также, что имитационные модели допускают использование практически того же языка, что и язык описания реального объекта.

3. *Имитационная модель — это всегда модель какого-либо дискретного (дискретизированного) процесса.* Очевидность этого тезиса также вытекает из конструктивного характера моделируемого объекта. Любая имитационная модель — это математическая структура $\langle S|R \rangle$, где S — множество возможных состояний объекта; R — множество правил, определяющих возможность перехода объекта из одного состояния в другое. Это, в частности, означает, что объект, находящийся в состоянии S_a переводится, если это возможно, в состояние S_b в результате последовательного (поэтапного) перехода в так называемые промежуточные состояния:

$$\begin{array}{ccccccc} R_a & R_1 & R_2 & R_{m-1} & R_m \\ S_a & \rightarrow S_1 & \rightarrow S_2 & \rightarrow \dots & \rightarrow S_m & \rightarrow S_b \end{array}$$

причем $\forall i: S_i \in S, R_i \in R$.

Формально это есть описание траектории движения объекта в виде алгоритмизированной схемы «состояние — переход — состояние», что позволило некоторым авторам принять это свойство имитационных моделей за их определение. Если моделируемый объект не является дискретным, то для построения имитационной модели его необходимо вначале представить в дискретном виде, что, конечно же, приводит к огрублению описания объекта.

По очевидным соображениям не имеет смысла говорить об имитационных моделях для статических объектов. В этом случае

множество правил возможных переходов пусто, а множество возможных состояний представлено одним элементом S_0 .

4. В имитационных моделях широко используется идея *рандомизированного розыгрыша*. Для каждого конкретно данного состояния объекта возможно указать, вообще говоря, несколько состояний, в каждое из которых объект может быть переведен за один шаг. Выбор «следующего» состояния зависит от различных факторов и может реализоваться несколькими способами, например на основании жестких правил предпочтения, случайным образом либо путем принятия волевого решения индивидуумом.

Исторически первым для выбора «следующего» состояния был использован метод статистических испытаний (Монте-Карло), который позволяет реализовывать процедуры вариантных расчетов и проводить эксперименты на математических моделях с помощью ЭВМ. Успехи этого подхода, связанные в первую очередь с резким расширением круга решаемых задач, и относительная простота его реализации способствовали распространению мнения об эквивалентности имитационного моделирования и метода статистических испытаний, что является неверным по существу.

В имитационных моделях при выборе «следующих» состояний часто используется метод статистических испытаний. Однако (и в этом суть имитационного моделирования) прошлая информация, вводимая в стохастические модели извне, заменяется фактическими данными, генерируемыми в машинном эксперименте самой моделью. Реализация этой идеи с широким использованием приложений теории нечетких множеств позволила по-новому взглянуть на возможность решения целого ряда сложных задач управления, в том числе и задач, связанных с созданием моделей искусственного интеллекта.

5. *Имитационные модели реализуются главным образом на быстродействующих компьютерах*. Эта их особенность часто ведет к отождествлению понятия имитационного моделирования с машинными экспериментами на математических моделях. Необходимость же применения для реализации имитационных моделей компьютеров связана с многоактностью соответствующих алгоритмов. Именно это обстоятельство объясняет тот факт, что практическая реализация имитационного моделирования стала возможной после широкого распространения быстродействующих компьютеров.

Следует подчеркнуть, что в имитационном моделировании используется так называемый структуризированный диалог, характерной чертой которого является непосредственное изменение параметров и структуры модели в ходе эксперимента.

Реализация имитационных моделей на компьютере есть одна из эффективных форм машинной имитации как процесса управляемого эксперимента, проводимого на ЭВМ над любой моделью системы.

6. *Имитационное моделирование — основной инструмент практической реализации системного анализа*. Имитационное моделирование развивается под давлением практической потребности исследования сложных систем и благодаря появлению компьютеров с развитым интерфейсом, позволяющим по-новому организовать ввод, вывод и отображение информации, а также интерактивный режим работы.

Известно, что решение сложных задач не может быть достигнуто простыми способами. В данном случае этот тезис находит свое отражение в том, что имитационные модели сложных объектов, в свою очередь, являются весьма сложными конструкциями. Как правило, это целые имитационные системы, которые включают одну или несколько базовых имитационных моделей, их модификации, информационную базу и специальный управляющий алгоритм, обеспечивающий эффективное и целенаправленное взаимодействие всех элементов системы. В имитационных системах предполагается ведение диалога для оперативного анализа получаемых результатов (в том числе и промежуточных), ввода необходимых корректив для проведения дальнейших расчетов, а также для осуществления не формализованных в модели элементов рассматриваемых процессов.

Обобщая вышеизложенное, можно сказать, что методология имитационного моделирования во всем спектре его возможных приложений на сегодняшний день не имеет альтернатив. Вместе с тем разработка имитационных моделей реальных систем и процессов — удел специалистов в данной прикладной области знаний, достаточно подготовленных в вопросах математики и информатики. За имитационными моделями будущее [25].

2.4.2. Нечеткие множества

2.4.2.1. Понятие нечеткого множества. Под четким множеством или просто множеством, обычно понимают некоторую сово-

купность определенных и различных между собой объектов нашей интуиции и интеллекта, мыслимую как единое целое. В данном высказывании отметим следующий момент: множество A есть совокупность определенных объектов. Это означает, что относительно любого x можно однозначно сказать, принадлежит ли он множеству A или нет.

Условие принадлежности элемента x множеству A можно записать, используя понятие функции принадлежности $\mu(x)$, а именно:

$$\mu(x) = \begin{cases} 1, & \text{если } x \in A \\ 0, & \text{если } x \notin A. \end{cases}$$

Следовательно, множество можно задать в виде совокупности пар: элемента и значения его функции принадлежности:

$$A = \{(x|\mu(x))\}. \quad (2.51)$$

Пример 1. Кафедра предлагает пять элективных курсов x_1, x_2, x_3, x_4 и x_5 . В соответствии с программой необходимо сдать три курса. Студент выбрал для изучения курсы x_2, x_3 и x_5 . Запишем этот факт с помощью функции принадлежности:

$$A = \{(x_1|0), (x_2|1), (x_3|1), (x_4|0), (x_5|1)\},$$

где первый элемент каждой пары означает название курса, а второй — описывает факт принадлежности его к подмножеству, выбранному данным студентом («да» или «нет»).

Примеров четких множеств можно привести бесконечно много: список студентов учебной группы, множество домов на данной улице города, множество молекул в капле воды и т. д.

Между тем, огромный объем человеческих знаний и связей с внешним миром включает такие понятия, которые нельзя назвать множествами в смысле (2.51). Их следует скорее считать классами с нечеткими границами, когда переход от принадлежности одному классу к принадлежности другому происходит постепенно, не резко. Тем самым предполагается, что логика человеческого рассуждения основывается не на классической двузначной логике, а на логике с нечеткими значениями истинности — нечеткими связками и нечеткими правилами вывода. Вот несколько тому примеров: объем статьи примерно 12 страниц, большая часть территории, подавляющее превосходство в игре, группа из нескольких человек.

Остановимся на последнем примере. Ясно, что группа людей из 3, 5 или 9 человек принадлежит к понятию: «группа людей,

состоящая из нескольких человек». Однако для них будет неодинаковой степень уверенности в принадлежности к этому понятию, которая зависит от различных, в том числе и субъективных, обстоятельств. Формализовать эти обстоятельства можно, если предположить, что функция принадлежности может принимать любые значения на отрезке $[0,1]$. Причем крайние значения предписываются в том случае, если элемент безусловно не принадлежит или однозначно принадлежит данному понятию. В частности, множество людей A из нескольких человек может быть описано выражением вида:

$$A = \{(1|0), 2|0.1), 3|0.4), (4|1), (5|1), (6|1), (7|0.8), (8|0.3), (9|0.1), (a|0) \text{ при } a \geq 10\}.$$

Приведем определение нечеткого множества, данное основателем теории нечетких множеств Л. А. Заде. Пусть x есть элемент конкретного универсального (так называемого базового) множества E . Тогда нечетким (размытым) множеством A , заданным на базовом множестве E , называют множество упорядоченных пар

$$A = \{x|\mu_A(x)\}, \quad \forall x \in E,$$

где $\mu_A(x)$ — функция принадлежности, отображающая множество E в единичный интервал $[0,1]$, т. е. $\mu_A(x): E \rightarrow [0,1]$.

Очевидно, что если область значений $\mu_A(x)$ ограничить двумя числами 0 и 1, то данное определение будет совпадать с понятием обычного (четкого) множества.

Функция принадлежности нечеткого множества может задаваться не только перечислением всех ее значений для каждого элемента базового множества, но и в виде аналитического выражения. Например, множество вещественных чисел Z очень близких к числу 2, может быть задано так:

$$Z = \{x|\mu_Z(x)\}, \quad \forall x \in R, \quad \text{где } \mu_Z(x) = e^{\frac{(x-2)^2}{0.3}}.$$

Множество же вещественных чисел Y , тоже достаточно близких к числу 2, есть:

$$Y = \{x|\mu_Y(x)\}, \quad \forall x \in R, \\ \mu_Y(x) = \frac{1}{1+(x-2)^2}.$$

Графическое изображение этих двух функций принадлежности дано на рис. 2.12.

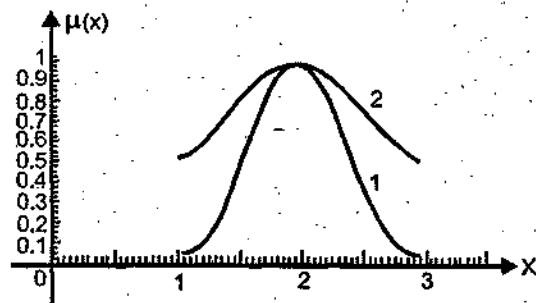


Рис. 2.11. Функции принадлежности нечетких множеств

1. μ_1 — числа очень близкие к 2
2. μ_2 — числа достаточно близкие к 2

Определение. Нечеткое множество A называется нечетким подмножеством B , если A и B заданы на одном и том же базовом множестве E и $\forall x \in E: \mu_A(x) \leq \mu_B(x)$, что обозначают как $A \subset B$.

Условия равенства двух нечетких множеств A и B , заданных на одном и том же базовом множестве E , имеет следующий вид:

$$A = B \text{ или } \forall x \in E: \mu_A(x) = \mu_B(x).$$

Замечание. Между разными по своей сути понятиями «нечеткости» и «вероятности» чувствуется некоторое сходство. Во-первых, эти понятия используются в задачах, где встречается неопределенность либо неточность наших знаний или же принципиальная невозможность точных предсказаний результатов решений. Во-вторых, интервалы изменения и вероятности и функции принадлежности совпадают:

$$P \in [0,1] \text{ и } \mu_A(x) \in [0,1].$$

Вместе с тем вероятность является характеристикой объективной, и выводы, полученные на основе применения теории вероятностей, в принципе могут быть проверены на опыте.

Функция же принадлежности определяется субъективно, хотя обычно она отражает реальные соотношения между рассматриваемыми объектами. Об эффективности применения методов, основанных на теории нечетких множеств, обычно судят после получения конкретных результатов.

Если в теории вероятностей предполагается, что вероятность достоверного события равна единице, т. е.

$$\sum_{i=1}^n P_i = 1,$$

то соответствующая сумма всех значений функции принадлежности

$$\sum_{i=1}^n \mu_A(x_i)$$

может принимать любые значения от 0 до ∞ .

Итак, чтобы задать нечеткое множество A необходимо определить базовое множество элементов E и сформировать функцию принадлежности $\mu_A(x)$, являющуюся субъективной мерой уверенности, с которой каждый элемент x из E принадлежит данному нечеткому множеству A .

2.4.2.2. Операции над нечеткими множествами.

1. Дополнение. Пусть нечеткие множества A и B имеют единое базовое множество E . Нечеткое множество B является дополнением нечеткого множества A (записывается $B = \bar{A}$), если:

$$\forall x \in E: \mu_B(x) = 1 - \mu_A(x).$$

Например, если

$$A = \{(x_1|0.2), (x_2|0.7), (x_3|1), (x_4|0.4), (x_5|0)\},$$

то

$$\bar{A} = \{(x_1|0.8), (x_2|0.3), (x_3|0), (x_4|0.6), (x_5|1)\}.$$

2. Объединение. Объединением двух нечетких множеств A и B , заданных на одном и том же базовом множестве E , называется нечеткое множество $C = A \cup B$, содержащее A и B , причем функция принадлежности определяется по правилу:

$$\forall x \in E: \mu_C(x) = \max(\mu_A(x), \mu_B(x)).$$

Пример 2. Пусть $A = \{(x_1|0.4), (x_2|0.7), (x_3|0.9)\}$ и $B = \{(x_1|0.5), (x_2|0.3), (x_3|1)\}$.

$$\text{Тогда } C = A \cup B = \{(x_1|0.5), (x_2|0.7), (x_3|1)\}.$$

3. Пересечение. Пересечением двух нечетких множеств A и B , заданных на одном и том же базовом множестве E , называется нечеткое множество $D = A \cap B$, содержащее A и B , причем функция принадлежности определяется по правилу:

$$\forall x \in E: \mu_D(x) = \min(\mu_A(x), \mu_B(x)).$$

Для нечетких множеств из примера 2 имеем:

$$D = A \cap B = \{(x_1 | 0.4), (x_2 | 0.3), (x_3 | 0.9)\}.$$

Наглядное представление операций объединения и пересечения двух нечетких множеств, заданных на непрерывном подмножестве, дают схемы, родственные диаграммам Вьенна-Эйлера, которые представлены на рис. 2.13. Границы штрихованных областей изображают графики функций принадлежности соответствующих нечетких множеств.

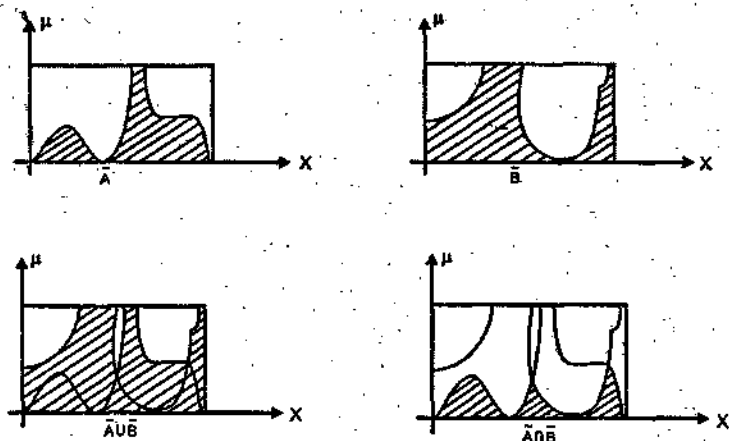


Рис. 2.12. Иллюстрация операций объединения и пересечения нечетких множеств

4. Произведение. Произведением двух нечетких множеств $A \cdot B = A \times B$, заданных на одном и том же базовом множестве E , называется такое нечеткое множество $F = A \cdot B$, функция принадлежности которого определяется по правилу

$$\forall x \in E: \mu_F(x) = \mu_A(x) \cdot \mu_B(x).$$

Для нечетких множеств из примера 2 имеем:

$$F = A \cdot B = \{(x_1 | 0.2), (x_2 | 0.21), (x_3 | 0.90)\}.$$

5. Сумма. Суммой двух нечетких множеств A и B , заданных на одном и том же базовом множестве E , называется нечеткое множество $H = A + B$, содержащее множества A и B , причем функция принадлежности определяется по правилу:

$$\forall x \in E: \mu_H(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x).$$

Для нечетких множеств из примера 2 имеем:

$$H = A + B = \{(x_1 | 0.7), (x_2 | 0.79), (x_3 | 1)\}.$$

Замечание. Операции умножения и суммирования нечетких множеств употребляются значительно реже, чем операции пересечения и объединения, поскольку для них не выполняются некоторые свойства, в том числе и такое, как дистрибутивность.

6. Разность. Разностью двух множеств A и B , заданных на одном и том же базовом множестве E , называется нечеткое множество $G = A - B = A \setminus B$, причем функция принадлежности определяется по правилу:

$$\forall x \in E: \mu_G(x) = \min(\mu_A(x), 1 - \mu_B(x)).$$

Для нечетких множеств из примера 2 имеем:

$$G = A - B = \{(x_1 | 0.4), (x_2 | 0.7), (x_3 | 0.0)\}.$$

7. Возведение в степень. Пусть задано нечеткое множество A , заданное на базовом множестве E . Возведением в неотрицательную степень « α » нечеткого множества A называется нечеткое множество $K = A^\alpha$, функция принадлежности которого $\mu_K(x)$ определяется по правилу:

$$\mu_K(x) = (\mu_A(x))^\alpha.$$

Возведение нечеткого множества в квадрат называется операцией концентрирования:

$$A^2 = \text{CON}(A).$$

Извлечение корня квадратного из нечеткого множества, рассматриваемого как возведение в степень 0.5, называется операцией растяжения:

$$A^{0.5} = \text{DIL}(A).$$

Так, например, если

$$A = \{(x_1 | 0.4), (x_2 | 0.7), (x_3 | 0.9)\},$$

то:

$$\text{CON}(A) = \{(x_1 | 0.16), (x_2 | 0.49), (x_3 | 0.81)\},$$

$$\text{DIL}(A) = \{(x_1 | 0.63), (x_2 | 0.84), (x_3 | 0.95)\}.$$

Нечетким евклидовым расстоянием между двумя нечеткими множествами A и B называется величина

$$e = \sqrt{\sum_{x_i} (\mu_A(x_i) - \mu_B(x_i))^2} \quad \forall x_i \in E.$$

Ближайшим к данному нечеткому множеству A называется четкое множество A , расположенное на наименьшем евклидовом расстоянии от A . Это множество определяется функцией принадлежности, формируемой по правилу:

$$\forall x_i \in E \quad \mu(x_i) = \begin{cases} 0, & \text{если } \mu(x_i) \leq 0.5 \\ 1, & \text{если } \mu(x_i) > 0.5. \end{cases}$$

Пример 3. Пусть заданы два множества:

$$A = \{(x_1 | 0), (x_2 | 0.1), (x_3 | 0.3), (x_4 | 0.7), (x_5 | 0.8), (x_6 | 0.9), (x_7 | 1)\}$$

и

$$B = \{(x_1 | 0), (x_2 | 0), (x_3 | 0), (x_4 | 0.6), (x_5 | 0.8), (x_6 | 1), (x_7 | 1)\}.$$

Тогда $e(AB) = 0.346$, а ближайшим к нечеткому множеству A является четкое множество $A = \{(x_1 | 0), (x_2 | 0), (x_3 | 0), (x_4 | 1), (x_5 | 1), (x_6 | 1), (x_7 | 1)\}$.

В теории нечетких множеств имеет место **принцип обобщения**, который можно записать следующим образом:

$$f(\{x | \mu_A(x)\}) = \{f(x) | \mu_A(x)\},$$

где A — нечеткое множество; f — некоторое отображение $X \rightarrow Y$.

Например, если $y = x^2 + 3$ и $A = \{(1 | 0.4), (2 | 0.6), (3 | 0.9)\}$, то $f(A) = \{(4 | 0.4), (7 | 0.6), (12 | 0.9)\}$.

Этот принцип открывает возможности вводить функциональные описания на нечетких множествах, что имеет важное значение в приложениях теории нечетких множеств.

2.4.2.3. Нечеткие отношения. Пусть P — четкое декартово произведение n множеств $E_1, E_2, \dots, E_n$. Нечеткое подмножество четкого множества называется n -арным нечетким отношением R_n на $P: R_n$, где

Величина μ_{R_n} есть мера того, что совокупность $(x_1, x_2, \dots, x_n)$ принадлежит отношению R_n . Знак $\cup$ в данном случае обозначает объединение соответствующих одноточечных множеств $(x_1, x_2, \dots, x_n)$.

На практике чаще других используется **бинарное нечеткое отношение** R_2 , заданное на двух множествах, например X и Y :

$$R_2 = \bigcup_{x \times y} (x_i, y_j) | \mu_R(x_i, y_j),$$

где $\mu_R: X \times Y \rightarrow [0, 1]$.

Бинарное отношение может рассматриваться в качестве двухместного предиката.

Примеры нечетких отношений: «расстояние в пространстве значительно больше 1 м» — тернарное отношение на множестве точек трехмерного пространства; « X — дальний родственник Y » — бинарное отношение на множестве людей.

Нечеткие бинарные отношения удобно задавать в виде матрицы. Например, R_2 — нечеткое отношение « X значительно больше Y », заданное на множествах $E_x = \{4, 8, 10\}$ и $E_y = \{2, 3, 4\}$ может быть задано следующим образом:

$$R_2 = \begin{matrix} & \begin{matrix} 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 4 \\ 8 \\ 10 \end{matrix} & \begin{pmatrix} 0.7 & 0.1 & 0 \\ 1 & 0.8 & 0.7 \\ 1 & 0.9 & 0.8 \end{pmatrix} \end{matrix}$$

Содержательно это означает, например, следующее: со степенью уверенности лишь 0.7 можно утверждать, что 4 значительно больше 2; но 8 значительно больше 3 со степенью уверенности 0.8.

Нечеткому бинарному отношению можно поставить в соответствие нечеткий граф. **Нечетким графом** G на множествах E_1 и E_2 называется такое нечеткое подмножество, что

$$\forall (x_i, y_j) \in E_1 \times E_2: \mu_G(x_i, y_j) \in M,$$

где M — множество принадлежностей элементов множества $E_1 \times E_2$.

Наглядным изображением нечеткого графа могут служить различные размытые изображения. Над нечеткими отношениями можно определить те же операции, что и над четкими отношениями [26].

2.4.2.4. Нечеткие и лингвистические переменные. Формализация нечетких понятий и отношений естественного языка возможна на основе понятий нечеткой и лингвистической переменных.

Нечеткой переменной называется кортеж $\langle X, U, C \rangle$, где X — название переменной; U — универсальное множество (область определения переменной X); C — нечеткое множество на U , описывающее нечеткое ограничение на значения переменной X .

Множество C описывает семантику нечеткой переменной, и его часто называют функцией совместимости нечеткой переменной. Переменная u является для X базовой переменной. Множество C определяет ту степень, с которой элементу x соответствует значение u . Значения нечеткой переменной есть числа.

Пример. Нечеткая переменная X , именуемая «человек высокого роста». Положим $U = (170-200)$, а C определим следующим образом:

$$C = U_{170}^{200} \left(1 - \frac{1}{1 + [0.2(u - 170)]^3} \right).$$

Лингвистической переменной называется кортеж $\langle X, T(X), U, G, M \rangle$, где X — название переменной; $T(X)$ — термножество, определяющее названия лингвистических значений X из универсального множества U ; G — синтаксические правила, описывающие процесс получения новых значений лингвистической переменной; M — семантическое правило, позволяющее ставить каждой нечеткой переменной X ее смысл $M(X)$.

Лингвистическая переменная — это переменная более высокого порядка, чем нечеткая переменная, поскольку значениями лингвистической переменной являются нечеткие переменные.

Различают числовые и нечисловые лингвистические переменные. Лингвистическая переменная называется числовой, если ее область определения U есть подмножество из R^1 , т. е. из множества вещественных чисел. Значения числовой лингвистической переменной называют нечеткими числами.

Пример. Числовая лингвистическая переменная «НАДЕЖНОСТЬ» может быть описана следующим образом:

$\langle \text{НАДЕЖНОСТЬ}, T, [0, 1], G, M \rangle$,

где $T = \{\text{очень низкая, низкая, средняя, высокая, очень высокая}\}$;

G — процедура перебора элементов из T ; M — ограничения, обусловленные значениями из T и определяющие смысл лингвистических значений. В частности, M могут быть выбраны так:

$$M[\text{очень низкая}] = U_0^{0.3} \exp(-100u^2) u;$$

$$M[\text{низкая}] = U_{0.2}^{0.5} \exp\left[-(u - 0.3)^2 / 0.02\right] u;$$

$$M[\text{средняя}] = U_{0.3}^{0.8} \exp\left[-(u - 0.55)^2 / 0.04\right] u;$$

$$M[\text{высокая}] = U_{0.7}^1 \exp\left[-(u - 0.8)^2 / 0.01\right] u;$$

$$M[\text{очень высокая}] = U_{0.7}^1 \left(1 - \frac{1}{1 + \frac{1}{1 + [10(u - 0.7)]^{1.5}}} \right) u.$$

Примером нечисловой лингвистической переменной может служить переменная КРАСИВЫЙ; формализующая понятие «красивый город» со значениями «не очень красивый», «красивый», «очень красивый», «очень-очень красивый» и т. п.

В дальнейшем будем рассматривать только числовые лингвистические переменные.

Порождение элементов из $T(X)$ возможно двумя способами: процедурой просмотра элементов термножества и путем реализации некоторого алгоритма. Если термножество $T(X)$ и функцию M можно задавать алгоритмически, то такую лингвистическую переменную называют структуризованной.

Рассмотрим один из возможных способов алгоритмического задания синтаксического G и семантического M правил, связанных с данной лингвистической переменной. Для этого отождествим слова: «или», «и», «не», «очень» с отдельными операциями над нечеткими множествами следующим образом:

«или» — операция объединения;

«и» — операция пересечения;

«не» — операция взятия дополнения;

«очень» — операция концентрирования.

Теперь, имея лишь небольшой набор первичных термов, можно аналитически записывать достаточно сложные лингвистические

конструкции. Рассмотрим, например, лингвистическую переменную «ВЕС» на множестве людей. В качестве первичных выберем термы «легкий» T_1 и «тяжелый» T_2 . Тогда терм «не очень легкий и не очень тяжелый» можно записать так: $\neg(T_1^2) \cap \neg(T_2^2)$, а «очень-очень-очень тяжелый» — (T_2^3) и т. д.

Пусть смысл лингвистического значения «легкий» определяется выражением

$$M(\text{легкий}) = \begin{cases} 0 & u < 0 \\ \frac{1}{1 + 0.01(u - 40)^2} & u \geq 0, \end{cases}$$

а смысл значения «тяжелый» — выражением:

$$M(\text{тяжелый}) = \begin{cases} 0 & u \geq 0 \\ \frac{0.01(u - 70)^2}{1 + 0.01(u - 70)^2} & u < 0. \end{cases}$$

Тогда значение «не очень тяжелый» определяется выражением:

$$M(\text{не очень тяжелый}) = 1 - \left[\frac{0.01(u - 70)^2}{1 + 0.01(u - 70)^2} \right]^2.$$

2.4.2.5 О построении функций принадлежности. Одним из основных вопросов, который необходимо решить на пути практического применения теории нечетких множеств, является вопрос о формализации нечетких понятий и отношений. При этом существенное место отводится построению собственно функций принадлежности нечетких множеств. Выше отмечалось, что любая функция принадлежности носит субъективный характер, хотя может отражать мнение не одного лица, а весьма большого числа людей.

Одной из простейших и наиболее ясной по реализации является вероятностная схема, суть которой заключается в следующем. Каждому из n экспертов предлагается ответить на вопрос о принадлежности элемента x к множеству A . Если n_1 экспертов отвечает на вопрос утвердительно, а n_2 — отрицательно, то считают, что

$$\mu(x) = \frac{n_1}{n_1 + n_2}.$$

Существует и целый ряд других, более сложных методик. Однако наиболее перспективной представляется процедура построения функций принадлежности в режиме диалога с компьютером на основе использования стандартного набора графиков.

Стандартный набор графиков функций принадлежности для утверждения «величина x мала» приведена на рис. 2.13 а–е. Из приведенных функций легко, например, составить функции принадлежности утверждений типа «величина $|x|$ большая» (рис. 2.14).

2.4.2.6. Элементы нечетких алгоритмов. Нечетким алгоритмом называют последовательность действий, выполняемых в соответствии с семантикой нечетких операторов и приводящих к нечеткому (размытому) решению задачи. Основными элементами нечетких алгоритмов являются выполнение арифметических операций над нечеткими числами и выполнение условных нечетких операторов типа «если - то - иначе».

Стандартные графики функции принадлежности

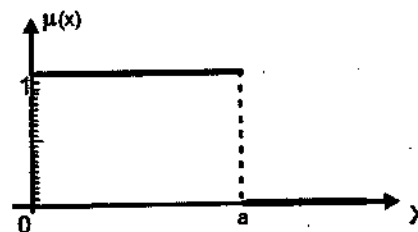


Рис. 2.13 а

$$\mu_1(x) = \begin{cases} 1, & \text{если } (0 \leq x \leq a) \\ 0, & \text{если } (x > a) \end{cases}$$

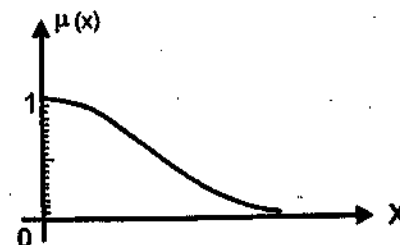


Рис. 2.13 б

$$\mu_2(x) = e^{-kx}, \quad k > 0$$

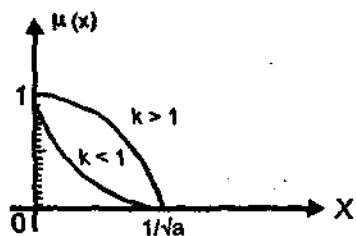


Рис. 2.13 в

$$\mu_A(x) = \begin{cases} 1 - ax^k, & \text{если } 0 \leq x \leq \frac{1}{\sqrt[k]{a}} \\ 0, & \text{если } x > \frac{1}{\sqrt[k]{a}} \end{cases}$$

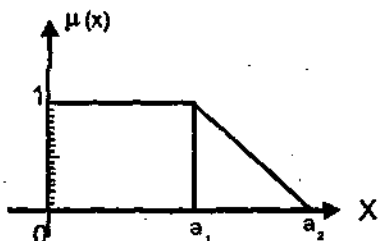


Рис. 2.13 г

$$\mu_A(x) = \begin{cases} 1, & \text{если } 0 \leq x \leq a_1 \\ \frac{a_2 - x}{a_2 - a_1}, & \text{если } a_1 \leq x \leq a_2 \end{cases}$$

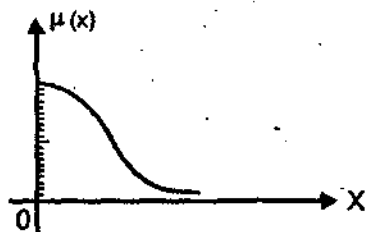


Рис. 2.13 д

$$\mu_A(x) = \frac{1}{1 + kx^2}, k > 1$$

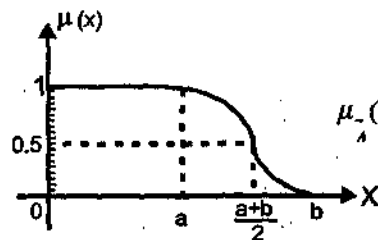
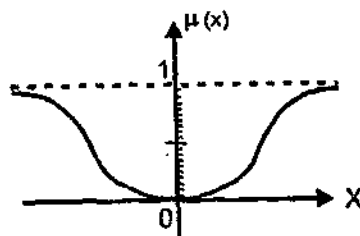


Рис. 2.13 е

$$\mu_A(x) = \begin{cases} 1, & \text{если } 0 \leq x \leq a \\ \frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi}{b-a} \left(x - \frac{a+b}{2}\right)\right), & \text{если } a \leq x \leq b \\ 0, & \text{если } b < x \end{cases}$$



$$\mu(x) = \frac{1}{1 + \frac{1}{kx^2}}$$

Рис. 2.14. Функция принадлежности для понятия «величина |x| большая»

Выполнение арифметических операций над нечеткими числами. В соответствии с принципом обобщения бинарная арифметическая операция $\oplus$ над нечеткими числами A и B есть нечеткое число C , которое определим следующим образом:

$$A \oplus B = C = \bigcup_{\forall c} \min[\mu_A, \mu_B],$$

где $\forall a \in S_A, \forall b \in S_B, c = (a \oplus b), c \in S_C$;

S_A, S_B, S_C — носители нечетких множеств (четкие множества);

$\oplus$ — одна из арифметических операций «+», «-», «*» или «/».

Если результат арифметической операции над элементами носителей нечетких множеств совпадает, т. е. $a_i \oplus b_j = a_k \oplus b_m$, то в S_C из них включается лишь один и ему предписывается значение функции принадлежности, имеющее из двух значений функций принадлежности большее значение. Например, пусть заданы нечеткие числа A и B :

$$A = \{(2|0.3), (6|0.9), (7|0.7)\}$$

$$B = \{(3|0.3), (4|0.8), (5|0.4)\}.$$

Тогда:

$$C_{A+B} = \{(5|0.3), (6|0.3), (7|0.3), (9|0.3), (10|0.8), (11|0.7), (12|0.4)\};$$

$$C_{A-B} = \{(-3|0.3), (-2|0.3), (-1|0.3), (1|0.4), (2|0.8), (3|0.7), (4|0.3)\};$$

$$C_{AB} = \{(6|0.3), (8|0.3), (10|0.3), (18|0.3), (21|0.3), (24|0.8), (28|0.7), (30|0.4), (35|0.4)\};$$

$$C_{A/B} = \{(0.4|0.3), (0.5|0.3), (0.67|0.3), (1.2|0.4), (1.4|0.4), (1.5|0.8), (1.75|0.7), (2|0.3), (2.33|0.3)\}.$$

Выполнение условных нечетких операторов. Будем рассматривать условный нечеткий оператор следующего вида:

$$\text{«если } W, \text{ то } K, \text{ иначе } L\text{»},$$

где W — нечеткое логическое выражение, т. е. любая формула,

в которую входят лингвистические и (или) нечеткие переменные и нечеткие предикаты; K, L — четкие или нечеткие операторы.

Нечетким предикатом называется функция, значениями которой являются нечеткие высказывания:

$$P^n(X) = X \rightarrow \mu_P(X) \in [0, 1], X \subset X_1 \times X_2 \times X_3 \times \dots \times X_n,$$

где $P^n(X)$ — нечеткий предикат; $\mu_P(X^i)$ — степень истинности нечеткого высказывания $P^n(X^i)$, $X^i = X_1^i, X_2^i, X_3^i, \dots, X_n^i$.

$P^n(X^i)$ может рассматриваться как степень уверенности субъекта в том, что он назовет данное высказывание истинным.

Примеры нечетких предикатов:

$P^1(x)$ — «числа x , близкие к нулю»;

$P^2(x, y)$ — «удаленные от районного центра села»;

$P^3(p, v, t)$ — «режим работы двигателя близок к оптимальному»;

p_0, v_0, t_0 .

Результат выполнения нечеткого условного оператора можно задать выражением:

$$R(\text{если } W, \text{ то } K, \text{ иначе } L) = \{R(K); \mu_W, R(L); (1 - \mu_W)\},$$

где $R(I)$ — результат выполнения оператора I ;

μ_W — степень истинности выражения W .

Наиболее целесообразной представляется следующая однозначная схема выполнения операторов. Разыгрывается равномерно распределенная на интервале $[0, 1]$ случайная величина ξ . Результат выполнения условного нечеткого оператора определяется из условия:

$$R(\text{если } W, \text{ то } K, \text{ иначе } L) = \begin{cases} R(K), & \text{если } \mu_W \geq \xi \\ R(L), & \text{если } \mu_W < \xi. \end{cases}$$

Теория нечетких множеств открывает широкие перспективы для построения имитационных моделей в гуманитарной области деятельности человека и ждет активного практического применения.

Для самостоятельной работы по более углубленному изучению материала, изложенного во второй части учебника, рекомендуется литература [27–38].

К введению

1. *Кудрявцев Л.Д.* Современная математика и ее преподавание. — М.: Наука, 1980.
2. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. — М.: Мир, 1976.
3. *Моль А.* Теория информации и эстетическое восприятие. — М.: Мир, 1966.
4. Математическая лингвистика/Под ред. Шрейдера Ю.А. — М.: Мир, 1964.
5. *Миронов Б.Н., Степанов З.В.* Историк и математика. — Л.: Наука, 1976.
6. *Биркгоф Г.* Математика и психология. — М.: ИЛ, 1977.
7. *Яглом И.М.* Математические структуры и математическое моделирование. — М.: Сов. радио, 1980.
8. *Колмогоров А.Н.* Математика в ее историческом развитии. — М.: Наука, 1991.
9. *Стройк Д.Я.* Краткий очерк истории математики. — М.: Наука, 1978.
10. *Бурбаки Н.* Теория множеств. — М.: Мир, 1965.
11. *Гильберт Д.* Основания геометрии. — М.: Наука, 1948.
12. *Лакатос И.* Доказательства и опровержения. — М.: 1967.
13. *Беляев Е.А., Перминов В.Я.* Философские и методологические проблемы математики. — М.: МГУ, 1981.
14. *Курант Р. и Роббинс Г.* Что такое математика? — М.: Наука, 1967.
15. Нормативные и дескриптивные модели принятия решений. — М.: Наука, 1981.
16. Информатика: Энциклопедический словарь для начинающих. — М.: Педагогика-Пресс, 1994.
17. История математики с древнейших времен до начала XIX столетия. В 3-х т. — М.: Наука. Т. 1 — 1970, Т. 2 — 1970, Т. 3 — 1972.
18. *Бурбаки Н.* Элементы математики. — М.: Изд. ИЛ, 1963.
19. *Даан-Дальмедико А., Пейффер Ж.* Пути и лабиринты: очерки по истории математики. — М.: Мир, 1986.
20. *Стюарт Я.* Концепции современной математики. — Мн.: Высш.школа, 1980.

К части 1.

1. *Сигорский В.П.* Математический аппарат инженера. — Киев: Техника, 1975.
2. *Кузнецов О.П., Адельсон-Вельский Г.М.* Дискретная математика для инженера. — М.: Энергия, 1980.
3. *Бурбаки Н.* Теория множеств. — М.: ИЛ, 1965.
4. *Шиханович Ю.А.* Введение в современную математику. — М.: Наука, 1965.
5. *Шрейдер Ю.А.* Равенство, сходство, порядок. — М.: Наука, 1971.
6. *Столл Р.* Множества. Логика. Аксиоматические теории. — М.: Просвещение, 1968.
7. *Калужнин Л.А.* Что такое математическая логика? — М.: Наука, 1964.
8. *Эдельман С.Л.* Математическая логика. — М.: Высшая школа, 1975.
9. *Оре С.* Теория графов. — М.: ИЛ, 1968.
10. *Кофман А.* Введение в прикладную комбинаторику. — М.: Наука, 1975.
11. *Привалов И.И.* Аналитическая геометрия. — М.: Высшая школа, 1962.
12. *Ильин В.А., Позняк Э.Г.* Линейная алгебра. — М.: Наука, 1974.
13. *Никольский С.М.* Курс математического анализа. — М.: Наука, 1975.
14. *Бронштейн И.Н. и Семендяев К.А.* Справочник по математике. — М.: Физ.-мат. лит. 1981.
15. *Хартман Ф.* Обыкновенные дифференциальные уравнения. — М.: Мир, 1970.
16. *Натансон И. П.* Краткий курс высшей математики. — СПб: «Лань», 1997.
17. *Данко П.Е. и Попов А.Г.* Высшая математика в упражнениях и задачах. Ч. 1 и 2. — М.: Высшая школа, 1996.
18. *Горбатов В. А.* Основы дискретной математики. — М.: Высшая школа, 1986.
19. *Яблонский С. В.* Введение в дискретную математику. — М.: Наука, 1986.
20. *Френкель А. А., Бар-Хиллел А.* Основания теории множеств. — М.: Мир, 1966.
21. *Белов В.В. и др.* Теория графов. — М.: Высшая школа, 1982.

22. *Камерон П., Ван Линт Дж.* Теория графов, теория кодирования и блок-схем. — М.: Наука, 1980.

23. *Ежов И. И.* Элементы комбинаторики. — М.: Наука, 1977.

24. *Сачков В. Н.* Введение комбинаторные методы для дискретной математики. — М.: Наука, 1982.

25. *Головина Л. И.* Линейная алгебра и некоторые ее приложения. — М.: Наука, 1979.

26. *Кремер Н. Ш. и др.* Высшая математика для экономистов. — М.: «Юнити», 1999.

27. *Ермаков В. И. и др.* Общий курс математики для экономистов. — М.: ИНФРА — М, 2000.

28. *Гусак А. А.* Высшая математика. В 2-х т. — Минск: Тетра Системс, 1998.

29. *Краснов и др.* Вся высшая математика в двух томах. — М.: Эдиториал УРСС, 2000.

30. *Буров А.В., Вадзинский Н.Н., Воронов М.В. и др.* Высшая математика. — Л.: ВМА, 1988.

31. *Щипачев В. С.* Основы высшей математики. — М.: Высшая школа, 1994.

32. *Щипачев В. С.* Сборник задач по высшей математике. — М.: Высшая школа, 1994.

К части 2.

1. *Вентцель Е.С.* Теория вероятностей. — М.: Физматгиз, 1972.

2. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. — М.: Финансы и статистика, 1983.

3. Статистические методы обработки результатов наблюдений/ Под ред. Юсупова Р.М. — М.: МО СССР, 1984.

4. *Вапник В.Н.* Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.

5. *Ивченко Г.И., Медведев Ю.А.* Математическая статистика. — М.: Высшая школа, 1984.

6. *Кендалл М.Дж., А.Стьюарт.* Многомерный статистический анализ и временные ряды. — М.: Наука, 1976.

7. Справочное пособие по прикладной математике. — М.: Изд-во МО СССР, 1975.

8. *Шеффе Г.* Дисперсионный анализ. — М.: Наука, 1980.

9. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Исследование зависимостей. — М.: Финансы и статистика, 1985.

10. Вилкас Э.И., Майминес Е.З. Решения: теория, информация, модели. — М.: Радио и связь, 1981.
11. Исследование операций Под ред. ДЖ. Моудера, С. Элмаграби. М.: Мир, 1981.
12. Вагнер Г. Основы исследования операций. В трех томах. — М.: Мир, 1972.
13. Акулич И. Л. Математическое программирование в примерах и задачах. — М.: Высшая школа, 1993.
14. Юдин Д.Б. Задачи и методы стохастического программирования. — М., Сов. Радио, 1979.
15. Балашевич В.А. Математические методы в управлении производством. — Минск: Высшая школа, 1976.
16. Канторович Л.В., Горстко А.Б. Оптимальные решения в экономике. — М.: Наука, 1972.
17. Юдин Д.Б., Цой Э.В. Целочисленное линейное программирование. — Изв. АН СССР. Техническая кибернетика, 1974, № 1.
18. Коршунов Ю. М. Математические основы кибернетики. — М.: Энергия, 1990.
19. Вентцель Е.С. Исследование операций. — М.: Сов.радио, 1972.
20. Базара М., Шетти К. Нелинейное программирование. — М.: Мир, 1982.
21. Фишберн П. Теория полезностей для принятия решений. — М.: Наука, 1978.
22. Буров А.В., Воронов М.В. Математическое моделирование процессов принятия решений. — Л.: ВМА, 1990.
23. Оуэн Г. Теория игр. — М.: Мир, 1971.
24. Штойер Р. Многоканальная оптимизация. Теория приложения. — М.: Радио и связь, 1992.
25. Моисеев Н.Н. Математика ставит эксперимент. — М.: Наука, 1980.
26. Модели принятия решений на основе лингвистической переменной. — Рига, Зинатне, 1982.
27. Математическая энциклопедия, т. 2. — М.: Сов. энциклопедия, 1979.
28. Кремер Н. Ш. и др. Теория вероятностей и математическая статистика. — М.: «Юнити», 1999.
29. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1999.
31. Осипов Г. В., Андреев А. П. Методы измерения в социологии. М.: Наука, 1980.

32. Изченко Г. И. и др. Теория массового обслуживания. — М.: Высшая школа, 1982.
33. Острейковский В. А. Теория систем. — М.: Высшая школа, 1997.
34. Грэхем Р., Кнут Д., Паташник О. Конкретная математика. — М.: Мир, 1998.
35. Таняев В.С., Шкурба В.В. Введение в теорию расписания. — М.: Наука, 1975.
36. Юдинов Ю. В. и др. Экспертные системы: инструментальные средства разработки. — СПб.: Политехника, 1996.
37. Блинов А.Н., Воронов М.В. Разработка экспертных систем в инструментальной среде. — С-Пб.: СПГУТД, 1998.
38. Воронов М.В. Нечеткие множества в моделях организационного управления. — Л.: ВМА, 1988.

Предметный указатель

А

алгебраическое дополнение — 123
алгоритм — 288
— нечеткий — 357

Б

базис — 131
булева переменная — 59

В

величина бесконечно большая — 156
— малая — 156
вектора
— ортогональные — 128
— компланарные — 128
вероятность доверительная — 263
вершина — 68
— изолированная — 68
вершины связанные — 72
— смежные — 68
— степень — 69
ветвь — 74
выборка — 81, 264
высказывание — 51

Г

гипербола — 104
гистограмма — 255
граф — 68
— изоморфный — 69
— гамильтонов — 73
— нечеткий — 353
— ориентированный — 70
— планарный — 69
— простой — 69

— связный — 72
— частичный — 71
— эйлеров — 72
график — 32
— соответствия — 33
генеральная совокупность — 252

Д

дерево — 74
— покрывающее — 75
дисперсия — 240
— внутригрупповая — 271
— групповая — 271
— межгрупповая — 271
— общая — 271
дифференциальное уравнение — 205
— линейное — 211
— порядок — 205
— с разделяющимися переменными — 209
— общее решение — 206
— частное решение — 207
длина маршрута — 71
доверительный интервал — 263
доказательство — 61
дуга — 70

З

задача оптимизации — 299

И

инверсия — 40
интегральная сумма — 176
— кривая — 206

К

кардинальное число — 35
квантификация — 66
квантор всеобщности — 64
— существования — 65
классы — 29
композиция функций — 37
— отношений — 44
компоненты кортежа — 31
континуум — 36
контур — 73
— простой — 73
координаты точки — 100, 109
— вектора — 127
кортеж — 31

Л

лес — 75
линейная комбинация векторов — 127

М

максимум глобальный — 298
— локальный — 298
маршрут замкнутый — 72
математическое ожидание — 239
матрица — 115
— вырожденная — 117
— единичная — 116
— инцидентности — 70
— квадратная — 115
— невырожденная — 117
— обратная — 123
— перехода — 133
— подобная — 137
— преобразования — 136
— присоединенная — 123
— прямоугольная — 115
— расширенная — 139
— симметрическая — 121

— смежности графа — 69
— транспонированная — 120
— треугольная — 116
— эквивалентная — 123
метод дихотомии — 301
— замены переменной — 174
— множителей Лагранжа — 304
— наименьших квадратов — 276
— штрафных функций — 317
минор — 121
— базисный — 122
— главный — 122
множества
— равномощные — 36
множество — 23
— дискретное — 36
— совершенно упорядочено — 47
— счетное — 36
— частично упорядочено — 47
модель алгоритмическая — 286
— аналитическая — 286
— имитационная — 342
— конструирования решений — 286
— нормативная — 286
— описательная — 286
— стоимостная — 286
— структуры — 285
— функционирования — 285
мощность множества — 35
мультиграф — 69

Н

направляющая — 111, 112
начало системы координат — 108
норма вектора — 128

О

образ — 33, 135
образующая — 111, 112
окрестность — 154

окружность — 103
 оператор — 41
 определитель — 116
 ось абсцисс — 100, 108
 — аппликата — 108
 — ординат — 100, 108
 — полярная — 107
 отношение бинарное — 42
 — корреляционное — 273
 — π — местное — 42
 — нечеткое — 353
 — обратное — 43
 — полное — 43
 — строгого порядка — 48
 — тождественное — 43
 — унарное — 42
 отрезки соизмеримые — 99
 отображение — 38
 оценка параметра — 261
 — несмещенная — 262
 — состоятельная — 262
 — эффективная — 262

П

парабола — 105
 переменная лингвистическая — 354
 — нечеткая — 354
 переменные свободные (базисные) — 144
 перестановка — 40
 перестановки — 82
 — с повторениями — 90
 первообразная — 171
 период функции — 38
 петля — 68
 плоскости координатные — 108
 плоскость в пространстве — 110
 поверхность вращения — 113
 — цилиндрическая — 111
 — коническая — 112
 подграф — 71

подмножества
 — несобственные — 25
 — собственные — 25
 подстановка — 40
 полезность решения — 330
 полюс — 107
 полуоси эллипса — 104
 — гиперболы — 104
 последовательность возрастающая — 153
 — расходящаяся — 155
 — ограниченная — 153
 — ограниченная сверху — 153
 — ограниченная снизу — 153
 — числовая — 152
 — убывающая — 153
 правило Вальда — 333
 — Гурвица — 333
 — Лапласа — 333
 — Сэвиджа — 333
 — цепочки — 161
 прадевево — 75
 предел последовательности — 154
 предел функции — 155
 предикат — 62
 предикаты равносильные — 64
 предметная область — 62
 предметные переменные — 62
 преобразования матрицы элементарные — 122
 принцип обобщения — 352
 прообраз — 34
 пространство — 97
 — вещественное векторное — 127
 — вероятностное — 224
 — евклидово — 128
 — линейное — 126
 — элементарных событий — 219
 путь — 73
 — простой — 73

Р

радиус
 — полярный — 107
 — сходимости — 200
 радиус-вектор — 100, 109
 разбиение числа — 85
 — множества — 28
 размерность пространства — 132
 размещения — 88
 — с повторениями — 90
 ранг матрицы — 122
 — системы — 139
 расстояние — 109
 ребро — 68
 — инцидентное — 68
 решение
 — допустимое — 298
 — общее — 206
 — оптимальное — 298
 — частное — 207
 ряд гармонический — 195
 положительный — 193
 — расходящийся — 190
 — статистический — 255
 — степенной — 199
 — сходящийся — 190
 — Тейлора — 203
 — функциональный — 199
 — числовой — 189

 С
 связывание переменной — 66
 седловая точка — 337
 система координат — 100, 108
 — левая — 108
 — правая — 108
 — прямоугольная — 108
 — полярная — 107
 — стандартная — 132
 система операций полная — 58

 — представителей — 46
 система уравнений неоднородная — 138
 — неопределенная — 140
 — несовместная — 139
 — однородная — 138
 — определенная — 140
 — совместная — 139
 — эквивалентная — 139
 скалярное произведение — 127
 случайная величина — 233
 — дискретная — 234
 — — непрерывная — 234
 событие — 218
 — достоверное — 219
 — невозможное — 219
 — случайное — 219
 — элементарное — 219
 события зависимые — 222
 — несовместные — 221
 — независимые — 222
 соответствие — 33
 — взаимно-однозначное — 34
 — всюду определенное — 34
 — инъективное — 34
 — сюръективное — 34
 — функциональное — 34
 сочетания — 89
 — с повторениями — 90
 среднеквадратическое отклонение — 241
 стратегия смешанная — 338
 — чистая — 337
 сумма ряда — 190
 сфера — 110
 сходимость последовательности — 153
 — абсолютная — 198

Т

тавтология — 61
теорема — 61
термы — 62
точка разрыва — 157
— стационарная — 300

У

угловой коэффициент — 102
универсум — 26
уравнение прямой линии — 102
— общее — 102
уравнения базисные — 140
уровень значимости — 259

Ф

факториал — 82
фактор множество — 46
формула Байеса — 231
— замены переменной — 174, 181
— интегрирования по частям — 174, 181
— Ньютона-Лейбница — 179
функция — 36
— булева — 59
— возрастающая — 38
— дифференцируемая — 158
— непрерывная — 157
— непрерывного аргумента — 37
— n -местная — 37
— нечетная — 38
— нестрого монотонная — 38
— обратная — 37
— периодическая — 38
— принадлежности — 347
— распределения — 236
— статистическая — 255
— регрессии — 268
— сложная — 161

— строго монотонная — 38
— убывающая — 38
— целевая — 297
— целочисленного аргумента — 37
— четная — 38
формула тождественно истинная — 59
— тождественно ложная — 59
функционал — 41

Ц

цена игры верхняя — 337
— нижняя — 337
цепь — 72
— простая — 72
цикл простой — 72
— эйлеров — 72

Ч

число рациональное — 99
— иррациональное — 100
числовая ось — 98

Э

экстремум — 299
элементы множества — 24
эллипс — 104
эффективное решение — 327

ОГЛАВЛЕНИЕ

Предисловие.....3

Введение.....7

Часть 1. ОСНОВНЫЕ МАТЕМАТИЧЕСКИЕ ПОНЯТИЯ

Раздел 1. ВВЕДЕНИЕ В ТЕОРИЮ МНОЖЕСТВ

Глава 1.1. Множества.....23
1.1.1. Понятие множества.....23
1.1.2. Подмножества.....25
1.1.3. Операции над множествами.....26
1.1.4. Свойства операций над множествами.....29

Глава 1.2. Соответствия.....31
1.2.1. Прямое произведение множеств.....31
1.2.2. Понятие соответствия.....33
1.2.3. Мощность множества.....35
1.2.4. Функции.....36

Глава 1.3. Отношения.....42
1.3.1. Основные определения.....42
1.3.2. Свойства отношений.....44
1.3.3. Отношение эквивалентности.....46
1.3.4. Отношение толерантности.....46
1.3.5. Отношение порядка.....47
1.4. Задачи и упражнения к разделу 1.....48

Раздел 2. ВВЕДЕНИЕ В ДИСКРЕТНУЮ МАТЕМАТИКУ.....50

Глава 2.1. Математическая логика.....50
2.1.1. Алгебра высказываний.....51
2.1.2. Операции над высказываниями.....53
2.1.3. Булевы функции.....58
2.1.4. Методы доказательств.....60
2.1.5. Алгебра предикатов.....62
2.1.6. Логические операции над высказываниями.....33

Глава 2.2. Графы.....	67
2.2.1. Определение графа.....	68
2.2.1 Маршруты на графах.....	71
2.2.1 Деревья.....	74
Глава 2.3. Основы комбинаторики.....	81
2.3.1. Основные теоремы комбинаторики.....	81
2.3.2. Основные правила комбинаторики.....	86
2.3.3. Комбинации объектов.....	88
2.4. Задачи и упражнения к разделу 2.....	91
Раздел 3. ПРОСТРАНСТВА.....	97
Глава 3.1. Понятие пространства.....	97
Глава 3.2. Геометрические пространства.....	98
3.2.1. Числовая ось.....	98
3.2.2. Системы координат на плоскости.....	100
3.2.3. Системы координат в пространстве.....	108
Глава 3.3. Матрицы.....	115
3.3.1. Основные определения.....	115
3.3.2. Действия над матрицами.....	117
3.3.3. Обратная матрица.....	123
Глава 3.4. Линейные пространства.....	126
3.4.1. Определение линейного пространства.....	126
3.4.2. Линейная независимость векторов.....	128
3.4.3. Базис системы векторов.....	131
3.4.4. Преобразование координат.....	132
3.4.5. Линейные преобразования n -мерного пространства.....	135
3.4.6. Системы линейных алгебраических уравнений.....	137
3.4.7. Метод жордановых исключений.....	144
3.5. Задачи и упражнения к разделу 3.....	148
Раздел 4. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОГО АНАЛИЗА.....	152
Глава 4.1. Пределы.....	152
4.1.1. Предел числовой последовательности.....	152
4.1.2 Предел функции и его свойства.....	155

4.1.3. Бесконечно малые и бесконечно большие величины.....	156
Глава 4.2. Производная функции одной переменной.....	158
4.2.1. Определение производной.....	158
4.2.2. Правила дифференцирования.....	159
4.2.3. Дифференциал.....	162
4.2.4. Производные высших порядков.....	163
4.2.5. Формула Тейлора.....	164
Глава 4.3. Функции многих переменных.....	167
4.3.1. Частные производные.....	167
4.3.2. Полный дифференциал.....	170
Глава 4.4. Интегрирование.....	171
4.4.1. Неопределенный интеграл.....	171
4.4.2. Методы интегрирования.....	173
Глава 4.5. Определенный интеграл.....	176
4.5.1. Определение определенного интеграла.....	176
4.5.2. Основные свойства определенного интеграла.....	177
4.5.3. Вычисление определенного интеграла.....	179
4.5.4. Приложения определенного интеграла.....	182
Глава 4.6. Несобственные интегралы.....	186
4.6.1. Несобственные интегралы с бесконечными пределами.....	186
4.6.2. Несобственные интегралы от разрывных функций.....	187
Глава 4.7. Ряды.....	189
4.7.1. Числовые ряды.....	189
4.7.2. Признаки сходимости числовых рядов с положительными членами.....	193
4.7.3. Знакопередающиеся ряды.....	198
4.7.4. Степенные ряды.....	199
Глава 4.8. Дифференциальные уравнения.....	205
4.8.1. Основные понятия теории дифференциальных уравнений.....	205

4.8.2. Дифференциальные уравнения первого порядка.....	207
4.9. Задачи и упражнения к разделу 4.....	214
ЧАСТЬ 2. МАТЕМАТИЧЕСКИЕ МЕТОДЫ.....	217
Раздел 1. МАТЕМАТИКА СЛУЧАЙНОГО.....	217
Глава 1.1. Основные понятия теории вероятностей.....	218
1.1.1. Случайные события.....	218
1.1.2. Понятие вероятности.....	222
1.1.3. Основные теоремы теории вероятностей.....	226
Глава 1.2. Распределения случайных величин.....	233
1.2.1. Случайные величины.....	233
1.2.2. Законы распределения случайных величин.....	234
1.2.3. Числовые характеристики случайных величин.....	238
1.2.4. Наиболее распространенные законы распределения вероятностей.....	241
Глава 1.3. Предельные теоремы теории вероятностей.....	247
1.3.1. Неравенство Чебышева.....	247
1.3.2. Закон больших чисел.....	249
1.3.3. Центральная предельная теорема.....	250
Глава 1.4. Обработка результатов измерений (анализ данных).....	252
1.4.1. Элементы математической статистики.....	254
1.4.2. Обработка выборок.....	261
1.4.3. Методы анализа статистических зависимостей.....	267
Раздел 2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ.....	280
Глава 2.1. Методологические основы моделирования.....	280
2.1.1. Постановка задач моделирования.....	280
2.1.2. Классификация моделей.....	285
2.1.3. Динамические модели.....	288
2.1.4. Метод статистических испытаний (метод Монте-Карло).....	291
Глава 2.2. Математическое программирование.....	297
2.2.1. Основные понятия.....	297
2.2.2. Классические задачи оптимизации.....	299
2.2.3. Линейное программирование.....	305
2.2.4. Нелинейное программирование.....	316
2.2.5. Динамическое программирование.....	319
Глава 2.3. Модели принятия решений.....	326
2.3.1. Принятие решений в условиях определенности.....	326
2.3.2. Принятие решений в условиях риска.....	331
2.3.3. Принятие решений в условиях концептуальной неопределенности.....	332
2.3.4. Принятие решений в условиях конфликта.....	336
Глава 2.4. Некоторые перспективные направления в математическом моделировании.....	340
2.4.1. Методологические аспекты имитационного моделирования.....	340
2.4.2. Нечеткие множества.....	345
Список литературы.....	361
Предметный указатель.....	366
Оглавление.....	371

Серия «Учебники, учебные пособия»

**Михаил Владимирович Воронов,
Галина Пантелеевна Мещерякова**

МАТЕМАТИКА
для студентов гуманитарных факультетов

Ответственные редакторы	<i>Оксана Морозова Жанна Фролова</i>
Выпускающий редактор	<i>Галина Логвинова</i>
Редактор	<i>Наталья Калиничева</i>
Художник	<i>Алексей Пащенко</i>
Компьютерная верстка	<i>Юлия Лубенцова</i>

Лицензия № 065194 от 2 июля 1997 года

Сдано в набор 3.02.2002.

Подписано в печать 10.06.2002.

Формат 84×08¹/₃₂. Бумага типографская № 2.

Гарнитура SchoolBook.

Тираж 10000. Заказ № 475.

Налоговая льгота — общероссийский классификатор продукции ОК-00-93,
том 2: 953000 — книги, брошюры.

Издательство «Феникс»

344007, г. Ростов-на-Дону, пер. Соборный, 17.

Отпечатано с готовых диапозитивов в ЗАО «Книга».

344019, г. Ростов-на-Дону, ул. Советская, 57.

Качество печати соответствует предоставленным диапозитивам.