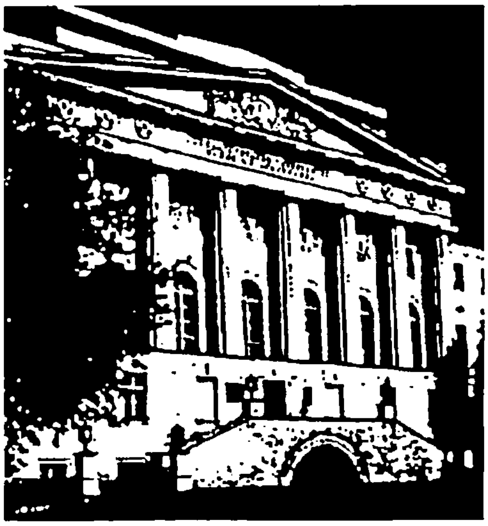
Серия

КЛАССИЧЕСКИЙ УНИВЕРСИТЕТСКИЙ УЧЕБНИК

основана в 2002 году по инициативе ректора
МГУ им. М.В. Ломоносова
академика РАН В.А. Садовничего
и посвящена

250-летию Московского университета



КЛАССИЧЕСКИЙ УНИВЕРСИТЕТСКИЙ УЧЕБНИК

Редакционный совет серии:

Председатель совета
ректор Московского университета
В.А. Садовничий

Члены совета:

Виханский О.С., Голиченков А.К., Гусев М.В.,
Добреньков В.И., Донцов А.И., Засурский Я.Н.,
Зинченко Ю.П. (ответственный секретарь),
Камзолов А.И. (ответственный секретарь),
Карпов С.П., Касимов Н.С., Колесов В.П.,
Лободанов А.П., Лунин В.В., Лупанов О.Б.,
Мейер М.С., Миронов В.В. (заместитель председателя),
Михалев А.В., Моисеев Е.И., Пущаровский Д.Ю.,
Раевская О.В., Ремнева М.Л., Розов Н.Х.,
Салецкий А.М. (заместитель председателя),
Сурин А.В., Тер-Минасова С.Г.,
Ткачук В.А., Третьяков Ю.Д., Трухин В.И.,
Трофимов В.Т. (заместитель председателя), Шоба С.А.

Московский государственный университет имени М.В. Ломоносова

Е.В. Шикин, А.Г. Чхартишвили

МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ В УПРАВЛЕНИИ

РЕКОМЕНДОВАНО УЧЕНЫМ СОВЕТОМ
ФАКУЛЬТЕТА ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ
МГУ им. М.В. ЛОМОНОСОВА
В КАЧЕСТВЕ УЧЕБНОГО ПОСОБИЯ
ДЛЯ СТУДЕНТОВ УПРАВЛЕНЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ ВУЗОВ

Издательство «Дело»
2004

УДК 65.012(075.8)
ББК 65в6я73
Ш-57

Рецензенты:

Самыловский А.И., доктор физико-математических наук, профессор Высшей школы экономики, заведующий кафедрой высшей математики;

Черемных Ю.Н., доктор экономических наук, профессор экономического факультета, заместитель заведующего кафедрой математических методов анализа экономики МГУ им. М.В. Ломоносова.

Шикин Е. В., Чхартишвили А. Г.

Ш-57 Математические методы и модели в управлении: Учеб. пособие. — 3-е изд. — М.: Дело, 2004. — 440 с. — (Сер. "Классический университетский учебник").

ISBN 5-7749-0374-5

Книга содержит изложение основных математических методов и моделей, используемых при выработке управленческих решений.

Рассматриваются сетевая оптимизация, линейное программирование, управление запасами, модель Леонтьева, метод анализа иерархий, методы прогнозирования, вероятностные и статистические методы, методы теории игр, основы теории управления организованными системами и некоторые другие.

Книга рассчитана на студентов и преподавателей вузов, слушателей учебных программ по менеджменту и государственному управлению, руководителей разного уровня, интересующихся современными подходами к проблеме принятия решений в управлении.

УДК 65.012(075.8)
ББК 65в6я73

ISBN 5-7749-0374-5

© Издательство "Дело", 2004
© МГУ им. М.В. Ломоносова, художественное оформление, 2004

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Уважаемый читатель!

Вы открыли одну из замечательных книг, изданных в серии "Классический университетский учебник", посвященной 250-летию Московского университета. Серия включает свыше 150 учебников и учебных пособий, рекомендованных к изданию Учеными советами факультетов, редакционным советом серии и издаваемых к юбилею по решению Ученого совета МГУ.

Московский университет всегда славился своими профессорами и преподавателями, воспитавшими не одно поколение студентов, впоследствии внесших заметный вклад в развитие нашей страны, составивших гордость отечественной и мировой науки, культуры и образования.

Высокий уровень образования, которое дает Московский университет, в первую очередь обеспечивается высоким уровнем написанных выдающимися учеными и педагогами учебников и учебных пособий, в которых сочетаются как глубина, так и доступность излагаемого материала. В этих книгах аккумулируется бесценный опыт методики и методологии преподавания, который становится достоянием не только Московского университета, но и других университетов России и всего мира.

Издание серии "Классический университетский учебник" наглядно демонстрирует тот вклад, который вносит Московский университет в классическое университетское образование в нашей стране и, несомненно, служит его развитию.

Решение этой благородной задачи было бы невозможно без активной помощи со стороны издательств, принявших участие в издании книг серии "Классический университетский учебник". Мы расцениваем это как поддержку ими позиции, которую занимает Московский университет в вопросах науки и образования. Это служит также свидетельством того, что 250-летний юбилей Московского университета — выдающееся событие в жизни всей нашей страны, мирового образовательного сообщества.

Ректор Московского университета
академик РАН, профессор В. Садовничий *В.А. Садовничий*

ОТ АВТОРОВ

Предлагаемая вниманию читателя книга написана на основе курса лекций, который на протяжении последних пяти лет читался студентам факультета государственного управления Московского государственного университета им. М. В. Ломоносова. В этих лекциях и на семинарских занятиях основные математические методы и модели, используемые при выработке управленческих решений, рассматривались с учетом в целом гуманитарной направленности обучения студентов на факультете. Это естественным образом сказалось и на отборе материала, и на характере его изложения.

Авторы стремились к доступному и по возможности наглядному описанию основных идей и количественных подходов к поиску и принятию решений в задачах управления, стараясь на простейших примерах познакомить читателя с возможностями развиваемого аппарата.

Требования к читателю минимальны — предварительного знакомства с какими-либо математическими понятиями и фактами не предполагается. Опыт общения со студентами показывает, что представленный материал вполне хорошо усваивается при относительно небольших усилиях (впрочем, вообще без усилий добиться сколько-нибудь заметных успехов вряд ли возможно). Тем не менее авторы рассчитывают на любознательность читателя, равно как и на его готовность к известному напряжению.

Какие-либо отсылки в книге практически отсутствуют: все необходимые сведения изложены в соответствующих ее разделах.

Авторы неоднократно обсуждали представленные здесь материалы с руководством ФГУ, сотрудниками его кафедр. Неизменно доброжелательное отношение с их стороны вызывает естественное чувство признательности, которое, по нашему мнению, уместно выразить на первых страницах этой книги.

Глава 1
ВСТУПЛЕНИЕ

С незапамятных времен человечество, используя бессмертный метод проб и ошибок, интуицию и опыт, накапливаемый в каждой конкретной ситуации, создавало искусство выработки наилучших решений в самых разных областях своей деятельности.

Принятие решения в реальной задаче управления — проблема многосложная, отягощенная к тому же неохватным разнообразием объективно существующих альтернатив и ограниченными возможностями взявшегося за его поиск. Вовсе не случайно поэтому, что долгое время управление считалось своего рода искусством, а кардинальные решения составили пусть и небольшую, но заметную часть сокровищницы нашей цивилизации (гордиев узел, переход Рубикона, нить Ариадны, колумбово яйцо и др.).

Успехи использования математических методов и стиля мышления в естественных науках с необходимостью, но, разумеется, не сразу привели к мысли о том, чтобы включить в сферу математического влияния и проблему принятия решений и попытаться тем самым превратить древнее искусство в современную науку.

Уровень развития науки и техники, достигнутый к настоящему времени, позволяет задумывать и осуществлять мероприятия, в которые оказываются вовлеченными значительные ресурсы — и материальные, и людские; мероприятия, масштабы, стоимость и последствия которых существенно превышают все, что проводилось когда-либо ранее. Это открывает невиданные ранее возможности, но и таит в себе огромные опасности. Положение усугубляется еще и тем, что на протяжении одной человеческой жизни техника и технологии, а вместе с ними среда, требования и навыки сменяются настолько быстро, что опытные люди, умеющие приводить эту технику в действие и разумно управлять ею, просто не успевают сформироваться — ведь для того, чтобы сложились традиции, нужно время.

Испытанный метод проб и ошибок в наши дни часто теряет свою универсальность: слишком катастрофическими могут оказаться ошибки и слишком мало времени отпущено для проб. Становится все более ясным, что сегодня меньше, чем когда-либо ранее, допустимы произвольные, чисто волевые решения.

На первый план выходит не задача создания все новых и новых образцов техники, а проблема организации и управления, причем управления не только (и не столько) машинами, но и людьми, сложными человеко-машинными системами. А это означает, что ответственные решения должны приниматься на основе предварительных прикидок и расчетов ("семь раз отмерь — один раз отрежь"). Не случайно поэтому в наше время наблюдается бурный рост математических методов во всех областях практики: вместо того чтобы пробовать и ошибаться по отношению к реальным объектам, люди предпочитают делать это на моделях. Формируется *исследование операций* (в англоязычной литературе — OR/MS (operations research/management science)) — наука о предварительном обосновании разумных решений во всех областях целенаправленной человеческой деятельности, широко использующая математический аппарат, но не сводящаяся к нему, наука, занимающая промежуточное положение между науками точными, опытными и гуманитарными.

Отчего же математический аппарат, столь давно используемый в сфере точных и опытных наук, только сравнительно недавно (и то на правах подсобного) стал применяться в науках гуманитарных? Все дело в том, что явления, составляющие предмет гуманитарных наук, неизмеримо сложнее тех, которыми занимаются науки точные. Многие гуманитарные явления гораздо труднее поддаются формализации, если вообще поддаются. Для каждого из этих явлений гораздо шире спектр причин, от которых оно зависит, и в их числе — психология живых людей и коллективов, людские пристрастия и антагонизмы, и потому вербальный способ построения исследования, как это ни парадоксально, часто оказывается здесь точнее формально-логического.

И все же помимо традиционных областей приложения — точных и опытных наук — математика начинает заниматься такими вопросами, которые от века изучались только на гуманитарном уровне: конфликтными ситуациями, иерархическими отношениями в коллективах, согласием, авторитетом, общественным мнением. Строятся и анализируются математические модели, применяются математические методы. Математика не только проникает в ранее чуждые для нее области, но и трансформируется при этом, становится менее “формальной”, меняет свои методологические черты, гибко приближаясь к наукам гуманитарным. Ее методы в области гуманитарных и смежных с ними наук могут служить мощным вспомогательным средством, позволяющим исследователю глубже проникнуть в существо явления, проследить его закономерности, обнаружить скрытые связи, малодоступные наблюдению простым, невооруженным глазом.

Математика не отличается радикально от других форм описания действительности, но, вследствие того что математические объекты более абстрактны, она позволяет отвлечься от большего числа случайных свойств. И потому универсальные закономерности, лишь смутно видимые в других областях, в математическом описании различимы более явно.

Каково же то место, которое, по нашему мнению, следует отвести в совокупном арсенале управленческих приемов математической составляющей, особенно если учитывать в целом гуманитарную ориентированность предполагаемого читателя?

Прежде всего, математические методы можно рассматривать как достаточно эффективное средство структурированного, более компактного и обозримого представления имеющейся информации. Это особенно ясно в тех случаях, когда информация задается в виде числовых массивов, в графической форме и др. Анализ результатов математической обработки данных зачастую позволяет высказать некоторые рекомендации относительно тех или иных способов действия. При принятии решений в больших задачах с их, как правило, огромными объемами информации это играет немаловажную роль.

Кроме того, существует целый ряд типичных управленческих ситуаций, допускающих известную формализацию, где именно математические подходы и соображения обоснованно становятся решающими.

Уже ранние работы (XVIII–XIX вв.) явились важным этапом развития и становления исследования операций. Пионерские попытки разработки научного подхода к организации труда и производства, к учету человеческого фактора в промышленности, предпринятые А. Смитом (A. Smith), Ч. Бэббиджем (Ch. Babbage), Ф. Тейлором (F. Taylor), Г. Гэнтом (H. Gantt) и др., позволили получить эффективные решения целого ряда конкретных задач. Например, введение в Великобритании в 1840 г. почтовой оплаты в 1 пении, существенно упростившей процедуру обработки корреспонденции, явилось результатом анализа операций в почтовом ведомстве, предпринятого

Бэббиджем, который нашел, что бо́льшая часть стоимости письма приходится на его обработку при сортировке, а вовсе не на дальность путешествия от отправителя к получателю, как это считалось ранее.

Начало XX в. отмечено первыми попытками смоделировать математически антагонистический конфликт (модель Ф. Ланчестера (F. Lanchester) исхода артиллерийской дуэли), создать теорию управления инвестициями (Ф. Харрис (F. Harris)), теорию массового обслуживания (теория очередей (А. Эрланг (A. Erlang))) и др. Однако, несмотря на заметные продвижения в разработке математических подходов к решению количественных проблем управления, исследование операций как научное направление (научная дисциплина) было признано лишь в 40–50-е годы XX в. Существенный прорыв обозначился при попытках разрешения целого ряда проблем управления, возникших непосредственно перед и в ходе второй мировой войны, где эффективность междисциплинарного подхода к ним проявилась явно. Наиболее известным примером могут служить результаты работы британской группы экспертов, состоявшей из 11 человек, оказавшие заметное влияние на исход битвы за Англию и сражений в Северной Атлантике. В эту группу, возглавлявшуюся П. М. С. Блэкеттом (P. M. S. Blackett) и ставшую потом известной под названием "Blackett's Circus", входили физиологи, математики, физики, геодезист, астрофизик и военный.

Специфика полученных результатов определенное время была сдерживающим фактором на пути их применения вне военной сферы. Однако заметные теоретические продвижения в теории игр и теории полезности (Дж. фон Нейман (J. von Newmann)) и в линейном программировании (Дж. Данциг (G. Danzig), Л. В. Канторович), а также создание новых мощных средств вычислений обеспечили существенный прорыв в расширении области приложения операционного анализа. Многие задачи управления удалось достаточно хорошо формализовать, и сейчас они уже весьма широко и довольно успешно решаются стандартными методами исследования операций.

Впрочем, зависимость методологии исследования операций от возможностей вычислительных средств не следует преувеличивать. Даже сегодня многие крупномасштабные задачи еще не удается решить при помощи существующих высокоскоростных компьютеров.

Итак, в первой половине XX в. начали разрабатывать (и довольно успешно) элементы научного подхода к поиску решений задач управления, а схемы, хорошо показавшие себя при проведении естественнонаучных и инженерно-технических изысканий, стали пытаться приспосабливать к решению управленческих задач. Сравнительно быстро пришло понимание того, что для поиска перехода от фактически наблюдаемого состояния изучаемой системы к желаемому весьма существенно, насколько хорошо формализована решаемая задача, и что уже имеющихся схем явно недостаточно.

Степень формализации управленческой задачи во многом определяет и методику поиска ее решения. Различают хорошо структурированные, слабоструктурированные и неструктурированные задачи. Резкой грани между ними провести нельзя. К тому же нередко оказывается, что (сначала) слабоструктурированная проблема становится (потом) хорошо структурированной и даже стандартной. Иными словами, для решения последних уже построены хорошо зарекомендовавшие себя схемы. Именно о них по большей части и пойдет речь в этой книге.

Отдельно нужно сказать об информации, перенасыщенный шумами поток которой нарастает с неспадающей стремительностью. Говорят даже об информационном буме. Но информация бывает разная: нужная, полезная и ненужная, загромождающая, утяжеляющая процесс управления. Важно научиться решительно отсекать ненужную, паразитную информацию и оперировать в каждом звене управления только той, которая безусловно необходима. В этом большую пользу могут принести модели, позволяющие сравнивать качество и оперативность управления в более громоздкой системе, перенасыщенной информацией, с тем, что дает более простая система, оперирующая только полезной информацией.

Но не следует забывать, что и в наши дни управление не перестало быть искусством и что некритическое использование для решения управленческих задач методик из иных областей знаний способно привести к неверным выводам. Для того чтобы разобраться в сложном явлении, его нужно рассматривать с различных сторон, под разными углами зрения, сравнивать результаты, обсуждать их, сопоставлять. Следует действовать весьма осторожно: применение математических методов не полезно, а вредно до тех пор, пока явление в достаточной степени не освоено на доматематическом, гуманитарном уровне. Часто бывает полезно вернуться к модели и внести в нее исправления после того, как первый тур расчетов уже проведен. Более того, нередко оказывается плодотворным своеобразный спор моделей, когда одно и то же явление описывается не одной, а несколькими моделями.

Приведем некоторые данные об использовании математических подходов, методов и моделей в задачах управления 125 крупнейшими корпорациями США [из статьи: *Guisseppi A. Forgionne.* Corporate Management Science Activities: An Update, *Interfaces*, 13 (June 1983). P. 20–23].

Метод, модель	**Частота использования, % корпораций**		
	Редко	**Умеренно**	**Постоянно**
Статистический анализ	2	38	60
Имитационное моделирование	13	53	34
Сетевое планирование	26	53	21
Линейное программирование	26	60	14
Теория очередей	40	50	10
Нелинейное программирование	53	39	8
Динамическое программирование	61	34	5
Теория игр	69	27	4

Исследование операций представляет собой применение научного метода к сложным проблемам, возникающим в управлении большими системами людей, машин, материалов и денег в промышленности, деловой сфере, государственном управлении, обороне и др. Его характерной особенностью является построение для соответствующей системы научной модели, включающей факторы вероятности и риска, при помощи которой можно рассчитать и сравнить результаты различных решений, стратегий и методов управления.

Основная задача исследования операций состоит в том, чтобы помочь менеджеру или иному лицу, принимающему решение, научно определить свою политику и действия среди возможных путей достижения поставленных целей. Коротко исследование операций можно назвать научным подходом к проблеме принятия решений. Проблема — это разрыв между желаемым и фактически наблюдаемым состояниями (прежде всего целями) той или иной системы. Решение — это средство преодоления такого рода разрыва, выбор одного из многих объективно существующих курсов действий, который позволил бы перейти от наблюдаемого состояния к желаемому.

В настоящее время под *операцией* понимается система действий, объединенных общим замыслом (*управляемое целенаправленное мероприятие*), а под основной задачей исследования операций — разработка и исследование путей реализации этого замысла. Ясно, что такое весьма широкое понимание операции охватывает значительную часть деятельности людей. Однако наука о принятии решений, о поиске путей достижения цели и особенно ее математическая составляющая еще весьма далеки до завершения даже по основным вопросам.

Совокупность людей, организующих операцию и участвующих в ее проведении, принято называть *оперирующей стороной.* Следует иметь в виду, что на ход операции могут оказывать влияние лица и природные силы, далеко не всегда содействующие достижению цели в данной операции.

Во всякой операции существует лицо (группа лиц), облеченное полнотой власти и наиболее информированное о целях и возможностях оперирующей стороны и называемое *руководителем операции* или *лицом, принимающим решение* (ЛПР). ЛПР несет полную ответственность за результаты проведения операции.

Особое место занимает лицо (группа лиц), владеющее математическими методами и использующее их для анализа операции. Это лицо (*исследователь операции, исследователь-аналитик*) само решений не принимает, а лишь помогает в этом оперирующей стороне. Степень его информированности определяется ЛПР. Так как исследователь-аналитик, с одной стороны, не имеет об операции всей информации, которой обладает ЛПР, а с другой, как правило, более осведомлен в общих вопросах методологии принятия решений, то желательно, чтобы взаимоотношения между исследователем операции и оперирующей стороной имели характер творческого диалога. Результатом этого диалога должен быть выбор (или построение) математической модели операции, на основе которой формируется система объективных оценок конкурирующих способов действий, более четко обозначается окончательная цель операции и появляется понимание оптимальности выбора образа действий. Право оценки альтернативных курсов действий, выбора конкретного варианта проведения операции (принятие решения) принадлежит ЛПР. Это обусловлено еще и тем, что абсолютных критериев рационального выбора не существует — во всяком акте принятия решения неизбежно содержится элемент субъективизма. Единственный объективный критерий — *время* — в конце концов покажет, насколько разумным было принятое решение.

Для того чтобы пояснить, какое место занимает математическая составляющая в исследовании операций, опишем коротко основные этапы разрешения проблемы принятия решения.

1-й шаг — сформулировать проблему (рис. 1).
Это весьма существенный и нетривиальный шаг даже в том случае, когда формулировка проблемы звучит совсем просто. Определение реальной проблемы, а не описание ее симптомов требует понимания и интуиции, некоторого воображения и времени.

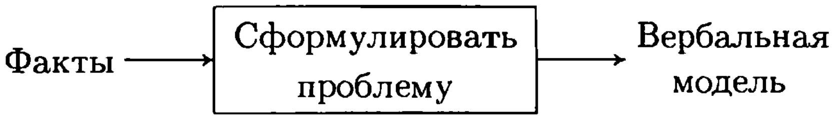


Рис. 1

2-й шаг — выбрать модель (рис. 2).
В случае если проблема сформулирована корректно, появляется возможность выбора готовой модели (из банка моделей, описывающих стандартные ситуации), разработка которой поможет в разрешении рассматриваемой проблемы, либо, если готовой модели нет, возникает необходимость создания такой модели, которая в достаточной степени точно отражала бы существенные стороны данной проблемы.

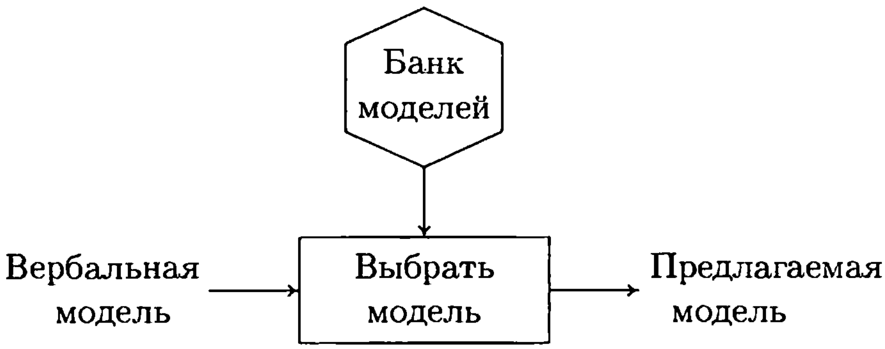


Рис. 2

Модели могут быть очень разными: есть *физические* (iconic) модели, есть *аналоговые* (analog). Мы будем говорить здесь в основном о *математических* моделях.

Существует много разнообразных математических моделей, которые достаточно хорошо описывают различные ситуации, требующие принятия тех или иных управленческих решений. Выделим из них следующие три класса — детерминированные, стохастические и игровые модели.

При разработке детерминированных моделей исходят из той предпосылки, что основные факторы, характеризующие ситуацию, вполне определенны и известны. Здесь обычно ставится задача оптимизации некоторой величины (например, минимизация затрат).

Стохастические модели применяются в тех случаях, когда некоторые факторы носят неопределенный, случайный характер.

Наконец, при учете наличия противников либо союзников с собственными интересами необходимо применение теоретико-игровых моделей.

В ходе дальнейшего изложения мы будем опираться на приведенную классификацию моделей, хотя возможны и другие, например, модели можно делить на статические и динамические, дискретные и непрерывные и т. д.

Как было сказано, в детерминированных моделях обычно имеется некий критерий эффективности, который требуется оптимизировать за счет выбора управленческого решения. (Впрочем, следует иметь в виду, что почти всякая сложная практическая задача является многокритериальной.)

В стохастических и игровых моделях ситуация усложняется еще больше. Зачастую выбор самого критерия зависит здесь от конкретной ситуации, и возможны различные критерии эффективности принимаемых решений.

При выборе и/или создании модели важно суметь найти верный баланс между точностью модели и ее простотой. Привлечение успешно действующих моделей приходит с опытом и практикой в соотнесении конкретных ситуаций с математическим описанием наиболее существенных сторон рассматриваемого явления. Конечно, ни одна математическая модель не может охватить всех особенностей изучаемой проблемы. Поэтому хотя выбор и/или создание модели, дающей математическое описание цели, процесса и результатов проведения операции, является неотъемлемой частью OR/MS, это все еще больше искусство, чем наука.

3-й шаг — найти решение (рис. 3).
Для поиска решения необходимы конкретные данные, сбор и подготовка которых требуют, как правило, значительных совокупных усилий. При этом стоит подчеркнуть, что даже в случае, если необходимые данные уже имеются, их часто приходится преобразовывать к виду, соответствующему выбранной модели.

4-й шаг — тестировать решение (рис. 4).
Полученное решение обязательно должно быть проверено на приемлемость при помощи соответствующих тестов. Неудовлетворительность решения обычно означает, что модель не точно отражает

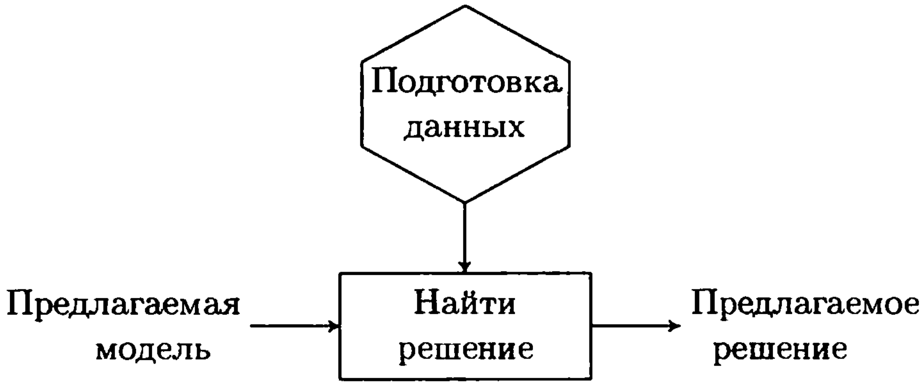


Рис. 3

истинную природу изучаемой проблемы. В этом случае она должна быть либо как-то усовершенствована, либо заменена на другую, более подходящую модель.

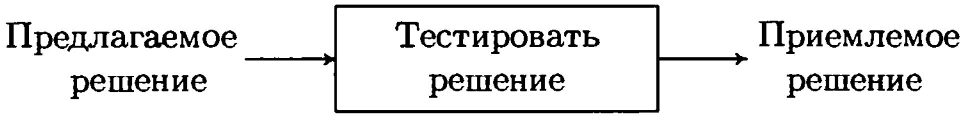


Рис. 4

5-й шаг — организовать контроль (рис. 5).
Если найденное решение оказалось приемлемым, естественно возникает необходимость создания механизма контроля за правильным использованием модели. Основная задача такого контроля состоит в обеспечении соблюдения ограничений, предполагаемых моделью, качества входных данных и получаемого решения. Полезно также иметь в виду, что найденное решение может быть использовано (и

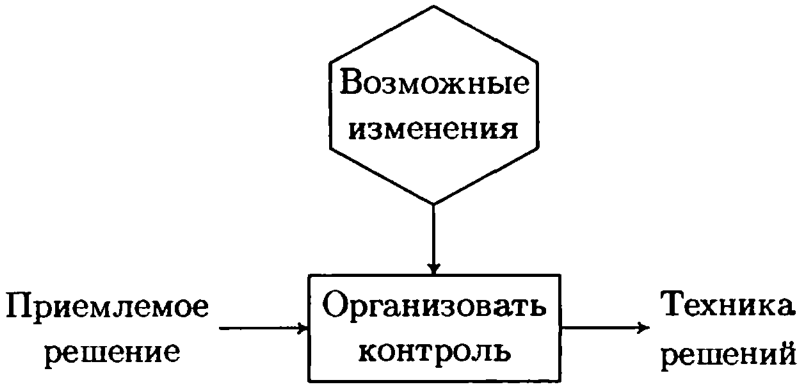


Рис. 5

часто используется) не только для разрешения сиюминутной ситуации, но и при рассмотрении сходных обстоятельств в будущем. Заранее планируемая гибкость выбранной модели дает возможность использовать ее в течение более продолжительного промежутка времени.

6-й шаг — создать режим благоприятствования (рис. 6). Этот шаг часто оказывается самым трудным — внедрение новаций нередко наталкивается на незаинтересованность и даже на сопротивление. Поэтому обучение персонала, реклама, качество подготавливаемой документации и учет разнообразия поведенческих мотивов людей играют здесь решающую роль.

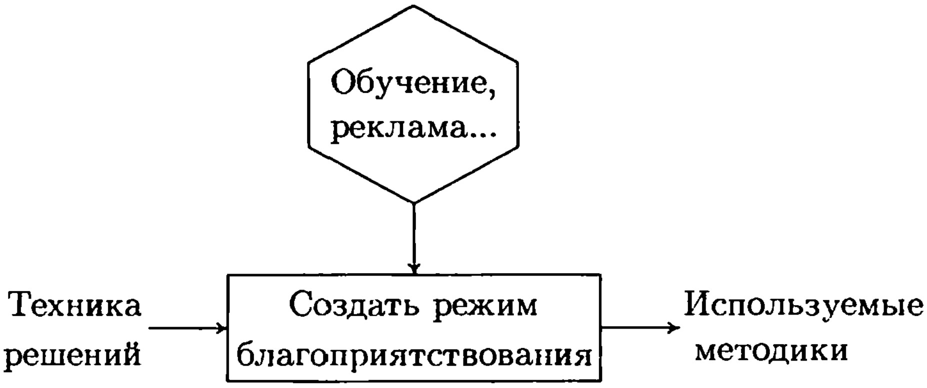


Рис. 6

На схеме (рис. 7) пунктирной линией отмечена та часть процесса принятия решения, где заметную роль играют различные соображения математического характера.

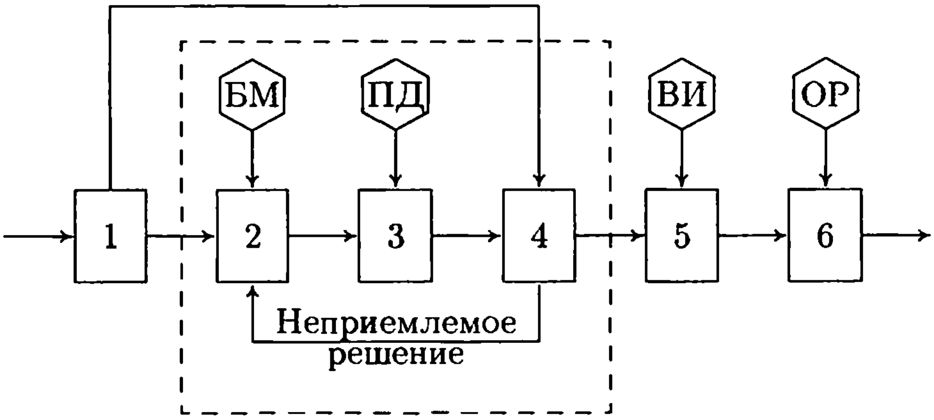


Рис. 7

Отметим, что сам термин "управление" можно понимать по-разному. Это и организация, в том числе и технологическая, той или иной осмысленной деятельности для достижения каких-либо целей (в качестве математического обеспечения здесь используются преимущественно детерминированные и стохастические модели), и изучение моделей поведения взаимодействующих сторон (здесь применяются игровые модели).

В настоящее время к решению сложных управленческих задач, представляющих практический интерес, привлекаются большие коллективы людей (и, добавим, значительные вычислительные средства) с разной профессиональной подготовкой и ориентацией, с разной степенью осведомленности о задаче в целом и, конечно, с разной степенью ответственности — от руководителя (ЛПР) до специалиста-разработчика (исследователя) и рядового исполнителя.

Для того чтобы такое сложное образование могло достаточно плодотворно функционировать, важно подготовить тех, кто был бы способен к действенному связыванию разных его блоков, кто осуществлял бы нетривиальные коммуникационные функции, был посредником как между ЛПР и специалистом-разработчиком, так и между разработчиком и исполнителем. Этому посреднику вовсе не обязательно знать в деталях всю техническую сторону вопроса (это задача для найденных при его посредстве специалистов), а достаточно ориентироваться в основных идеях. Иными словами, если касаться только математической части, у него должны быть определенные представления о возможностях математических методов, об их идейных основаниях и о банке готовых математических моделей и ключевых методов.

Одной из целей книги, которые ставили перед собой авторы, является преодоление математической, методологической и языковой разобщенности исследователей сложной практической управленческой задачи. Только это дает возможность, с одной стороны, как можно точнее отразить в создаваемой (или выбираемой) модели реальные процессы, а с другой — создать (или выбрать) модель, простую настолько, чтобы можно было надеяться решить задачу до конца и получить обозримые и уже этим полезные результаты.

Большинство используемых здесь идей допускает простое и вполне доступное объяснение, разумеется требующее некоторой математической культуры. Это и определило набор математических понятий, формул, подходов и фактов, равно как и методику их изложения.

Не следует также забывать о том, что важно не только ознакомиться с уже имеющимися управленческими технологиями, но и быть готовым к разработке новых. Аппарат, используемый при выработке и принятии нестандартных решений, должен быть настолько прост и нагляден, насколько это возможно. И насколько это возможно, он должен быть согласован с точностью и объемом доступной информации и с вопросами, ответы на которые хочется получить.

Области, где математические методы работают достаточно эффективно, не совпадают с ареалом управленческих задач. Последние слабо формализуемы и часто традиционно консервативны. Отсюда подозрительное отношение к рекомендациям, основанным на точных расчетах, требующих обширных и глубоких математических знаний.

Нужно признать, что определенные основания для этого есть. Методы математики способны решать только те задачи, которые изложены на ее языке. А это предполагает непременные упрощения в реальной сложной ситуации. За разделением определяющих факторов задачи на существенные и второстепенные часто стоят управленческий опыт и интуиция.

Впрочем, даже весьма грубая на вид идеализация может позволить глубже вникнуть в суть проблемы.

Авторы стремятся обходиться минимальным количеством формул и фактов, а привычные в математических курсах теоремы и их доказательства, требующие даже в самых простых случаях сравнительно высокой математической культуры, просто опускаются.

В конце каждой главы приводится небольшой набор задач и упражнений (как правило, с ответами), призванный помочь читателю в его самооценке степени усвоения соответствующей темы.

Все отзывы и замечания по поводу данной книги авторы с благодарностью примут по адресу: alexch@spa.msu.ru.

Часть I

ДЕТЕРМИНИРОВАННЫЕ МЕТОДЫ

Глава 2
ГРАФЫ И СЕТИ

2.1. Графы

“Мне была предложена задача об острове, расположенном в городе Кенигсберге и окруженном рекой, через которую перекинуто 7 мостов. Спрашивается, может ли кто-нибудь непрерывно обойти их, проходя только однажды через каждый мост. И тут же мне было сообщено, что никто еще до сих пор не смог это проделать, но никто и не доказал, что это невозможно. Вопрос этот, хотя и банальный, показался мне достойным внимания тем, что для его решения недостаточны ни геометрия, ни алгебра, ни комбинаторное искусство. После долгих размышлений я нашел легкое правило, основанное на вполне убедительном доказательстве, при помощи которого можно во всех задачах такого рода тотчас же определить, может ли быть совершен такой обход через какое угодно число и как угодно расположенных мостов или не может”.

Из письма Л. Эйлера от 13 марта 1736 г.

Город Кенигсберг (ныне Калининград) располагался на обоих берегах реки Прегель и на двух островах, которые соединялись семью мостами. План расположения мостов приведен на рис. 1. Задача, о которой говорится в письме, состоит в том, чтобы во время прогулки пройти каждый мост по одному разу и вернуться в исходную точку.

Так как нас интересуют только переходы по мостам, то план города можно заменить схемой, представленной на рис. 2. На этой схеме земельные участки, разделенные рукавами реки, как бы сжаты в точки A, B, C, D (*вершины*), а мосты вытянуты в линии a, b, c, d, e,

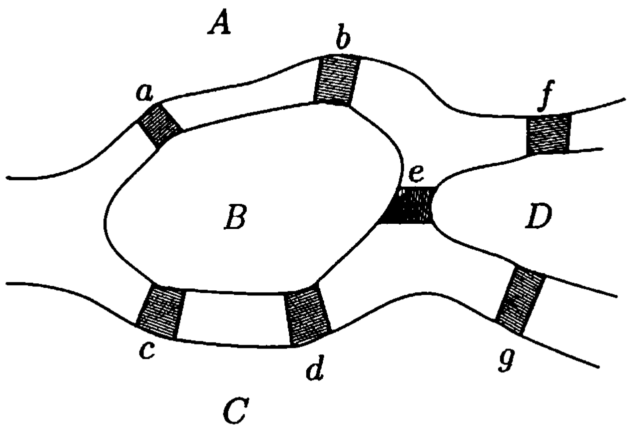


Рис. 1

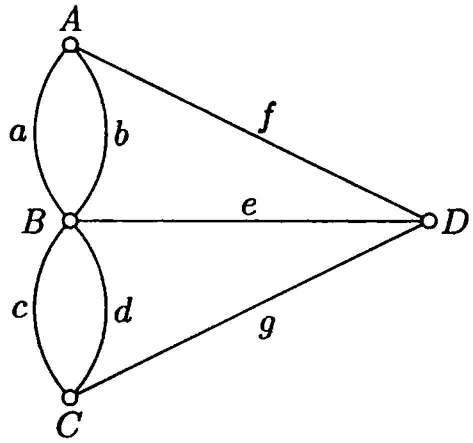


Рис. 2

f, g (*ребра*). Нетрудно проверить (например, перебрав все возможные варианты), что изображенную фигуру нельзя обвести острием пера, не отрывая его от бумаги и проходя по каждой дуге ровно один раз.

Исследуя ситуацию с кенигсбергскими мостами, Эйлер решил значительно более общую задачу. Для того чтобы лучше понять полученный им результат, введем некоторые определения.

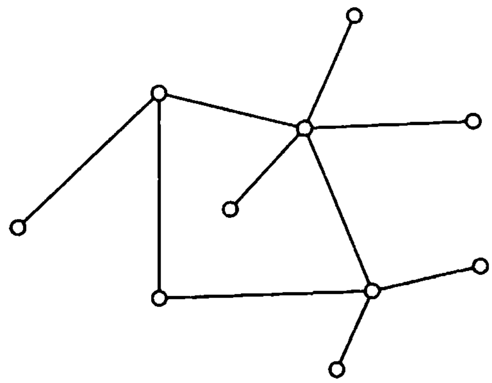


Рис. 3

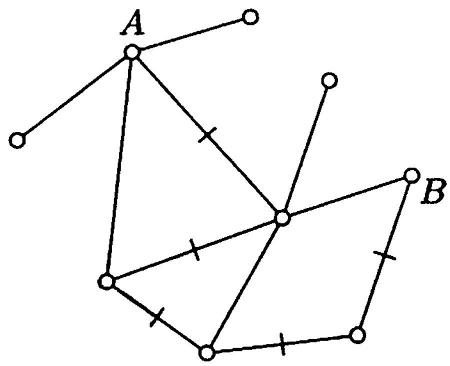


Рис. 4

Фигура, состоящая из точек (*вершин*) и соединяющих их линий (*ребер*), называется *графом* (рис. 3). *Маршрутом*, или *путем*, соединяющим вершины A и B графа, называется такая последовательность его ребер, в которой каждые два ребра имеют общую концевую точку, причем первое ребро выходит из вершины A, а последнее вхо-

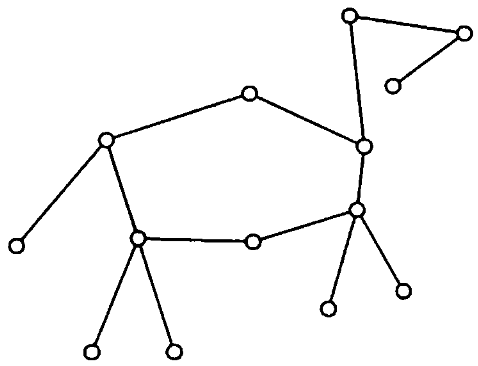


Рис. 5

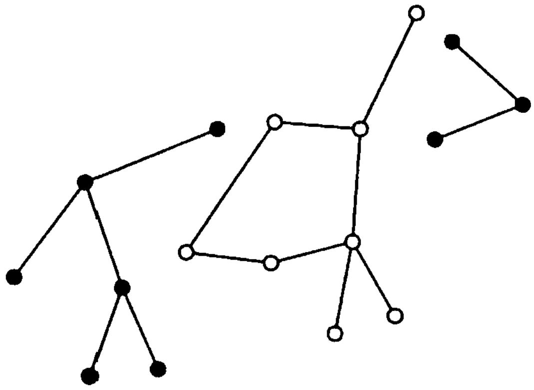


Рис. 6

дит в вершину B (рис. 4). В этом случае вершины A и B называются *связанными*. Граф называется *связным*, если любая пара его вершин связана (рис. 5). Граф, изображенный на рис. 6, несвязен.

Маршрут называется *цепью*, если каждое ребро графа встречается в нем не более одного раза (вершины в цепи могут повторяться и несколько раз) (рис. 7). Цепь, начальная и конечная вершины которой совпадают, называется *циклом* (рис. 8).

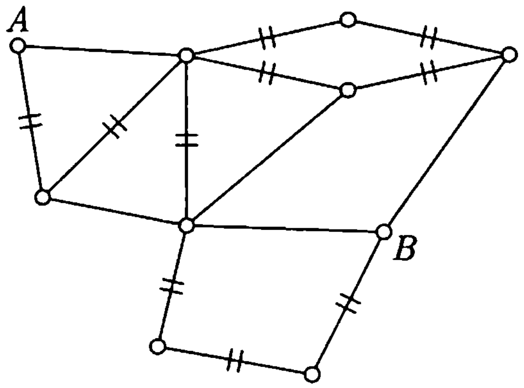


Рис. 7

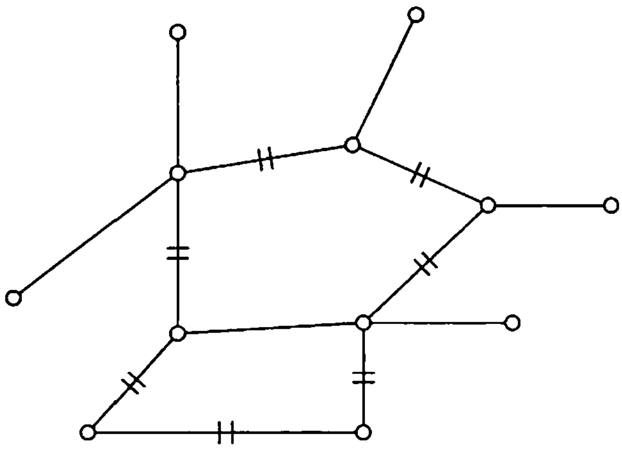


Рис. 8

Вершина называется *четной*, если в ней сходится четное число ребер, и *нечетной*, если число всех сходящихся в ней ребер нечетно. Вершина A на рис. 9 четна — в ней сходятся 6 ребер, а вершина B нечетна — в ней сходятся 5 ребер. Число ребер, сходящихся в вершине графа, называется *степенью (порядком)* этой вершины.

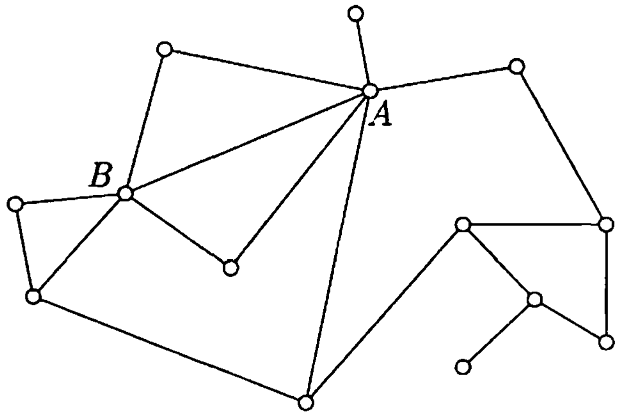


Рис. 9

Наконец, граф называется *конечным*, если множество его ребер конечно. Примером *бесконечного* графа может служить прямоугольная сетка, заданная на всей плоскости.

ТЕОРЕМА **(Эйлер)**. *В любом конечном связном графе, все вершины которого четны, существует цикл, в котором каждое ребро графа участвует ровно один раз.*

Такой цикл называют *эйлеровым циклом*, а граф, все вершины которого четны (и, значит, существует эйлеров цикл), — *эйлеровым графом*.

Обратившись к графу в задаче о кенигсбергских мостах, замечаем, что все четыре его вершины являются нечетными — в каждой из вершин A, C, D сходятся по три ребра, а в вершине B — пять ребер. Значит, этот граф не эйлеров.

Найти эйлеров цикл (разумеется, после того, как вы убедились, что заданный граф эйлеров (все вершины четны)) совсем не трудно: существует универсальный и достаточно простой алгоритм, при помощи которого задача построения эйлерова цикла всегда разрешима.

Покажем эффективность этого алгоритма на конкретном примере (см. рис. 10).

Пример 1. Выйдя из вершины A и не пытаясь еще раз пройти по уже пройденному ребру, мы неизбежно вернемся в вершину A. Это объясняется тем, что, входя в любую вершину графа (кроме, быть может, вершины A), мы всегда имеем возможность выйти из нее (напомним, что в каждой вершине графа сходится четное число ребер). Следовательно, неутомимо продолжая перемещение, мы

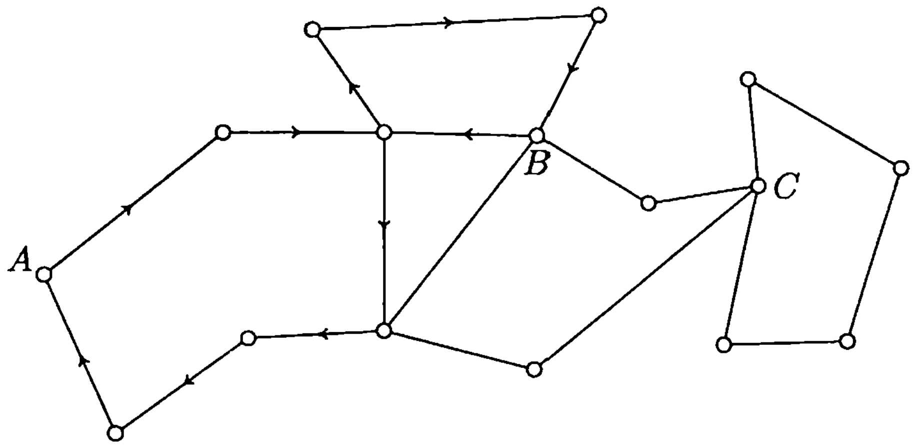


Рис. 10

неизбежно вернемся в вершину A, а вернувшись, окажемся перед двумя возможными ситуациями: 1) в построенный нами цикл входят все ребра графа, 2) остались еще не пройденные ребра.

Первый случай не так интересен: если в построенный цикл входят все ребра, то поставленная задача решена. Что же касается второго случая, то здесь в полученном нами цикле (обозначим его через $\mathcal{A}$) обязательно есть вершина, из которой выходит еще не пройденное нами ребро. Пусть это вершина B. Об этой вершине можно сказать даже больше: число выходящих из нее ребер, не принадлежащих построенному циклу $\mathcal{A}$, обязательно четно. И мы строим новую цепь из вершины B, привлекая только ранее не пройденные ребра. Ясно, что в результате мы вернемся в вершину B и получится новый цикл — $\mathcal{B}$ (рис. 11).

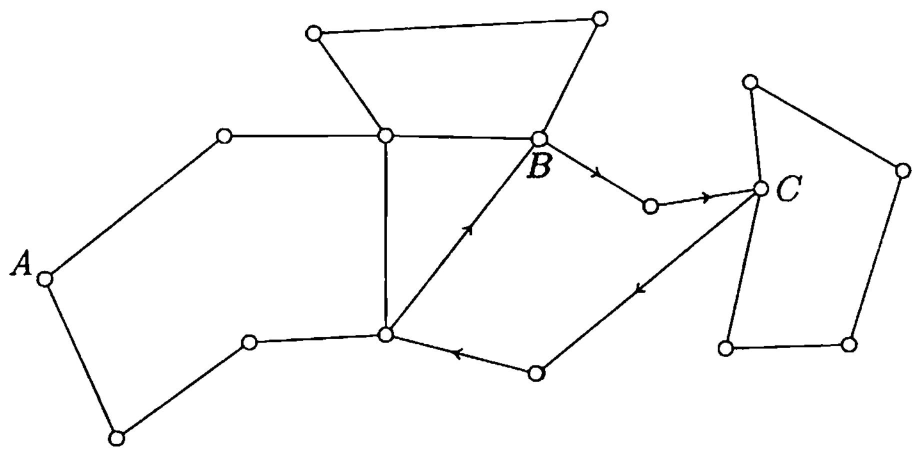


Рис. 11

Теперь легко получить цикл, начинающийся в вершине A и больший построенного ранее цикла $\mathcal{A}$.

Для этого мы сначала перемещаемся по маршруту $\mathcal{A}$ от вершины A до вершины B, затем проходим по циклу $\mathcal{B}$ и, вернувшись в вершину B, завершаем перемещение в вершину A по оставшейся части цикла $\mathcal{A}$ (рис. 12).

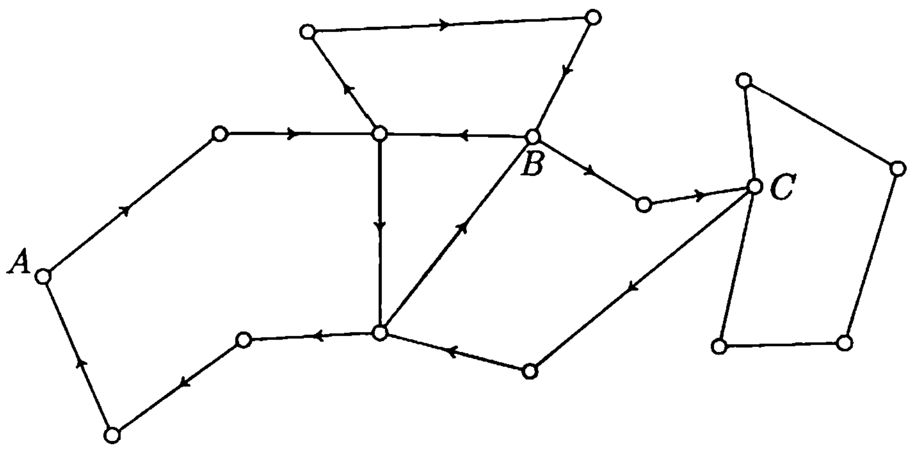


Рис. 12

Если мы и на этот раз не прошли по всем ребрам графа, то, выбрав вершину цикла, построенного по циклам $\mathcal{A}$ и $\mathcal{B}$, из которой исходят ребра, не входящие в этот цикл, расширяем его описанным выше способом.

Повторяя в случае необходимости подобные рассуждения достаточное число раз, мы всегда сможем построить эйлеров цикл за конечное число шагов.

Пример 2. Устроители больших художественных выставок часто вынуждены решать одну и ту же задачу: как организовать осмотр, чтобы дать возможность в отведенное время ознакомиться со всей экспозицией наибольшему числу желающих.

Ясно, что для этого нужно расставить указатели таким образом, чтобы, перемещаясь в соответствии с предложенными в них рекомендациями, любой посетитель мог побывать у каждой картины ровно по одному разу.

Если вход и выход совпадают, то разместить экспонаты следует так, чтобы схема экспозиции была эйлеровым графом. Что же касается указателей, то они должны 1) быть снабжены порядковыми номерами и 2) описывать эйлеров цикл (рис. 13).

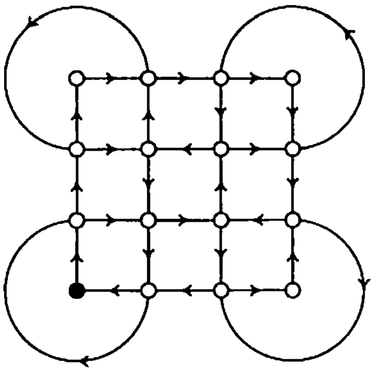


Рис. 13

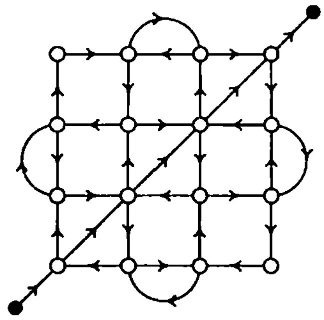


Рис. 14

Если же вход и выход разные, то схема размещения экспонатов должна быть графом, у которого лишь две вершины, соответствующие входу и выходу, являются нечетными. На рис. 14 приведен один из возможных способов расстановки пронумерованных указателей, позволяющий посетителям ознакомиться с каждым экспонатом.

Эйлеровы циклы характеризуются тем свойством, что они проходят через каждое ребро графа в точности по одному разу.

Аналогичным образом, но только по отношению к вершинам определяются для конечных связных графов так называемые *гамильтоновы циклы*:

цикл называется *гамильтоновым*, если он проходит через каждую вершину графа ровно по одному разу.

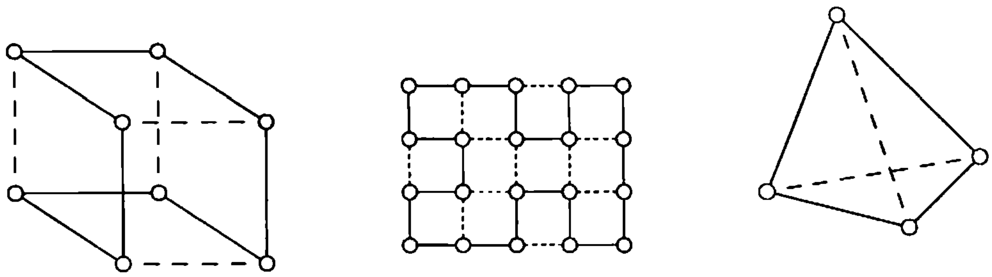


Рис. 15

На рис. 15 приведены гамильтоновы циклы для нескольких простых графов.

Между эйлеровым и гамильтоновым циклами легко просматривается довольно прозрачная аналогия: первый проходит ровно один раз по каждому ребру, второй — ровно один раз через каждую вершину. И на первый взгляд естественно ожидать того, что задача проверки, допускает ли данный граф гамильтонов цикл, должна быть по сложности сравнима с аналогичной задачей для эйлерова цикла (где достаточно подсчитать четность каждой вершины). Однако на деле все обстоит значительно сложнее: несмотря на практическую важность этой проблемы, до сих пор не найдено ни общего критерия, позволяющего устанавливать, является ли заданный граф гамильтоновым, ни универсального эффективного алгоритма построения гамильтонова цикла.

Проблема, восходящая своей постановкой к У. Р. Гамильтону, вообще оказалась очень трудоемкой. Поэтому не стоит удивляться тому, сколько усилий высококвалифицированных математиков и программистов потребовалось и на какие вычислительные затраты пришлось пойти для того, чтобы рассчитать до конца кратчайшую воздушную линию, соединяющую столицы всех североамериканских штатов.

Одной из практических задач, связанных с построением гамильтонова цикла, является *задача о коммивояжере*, в которой нужно найти кратчайший путь, проходящий через заданные пункты (все расстояния известны) и возвращающийся в исходный пункт.

Так как число пунктов конечно, то в принципе задача может быть решена простым перебором.

Пример 3. Торговец, живущий в городе A, намерен посетить города B, C и D, расстояния между которыми ему известны:

$$AB = 11, \quad AC = 13, \quad AD = 17, \quad BC = 6, \quad BD = 9, \quad CD = 10$$

(рис. 16). Требуется указать кратчайший циклический маршрут из города A, проходящий через три других города.

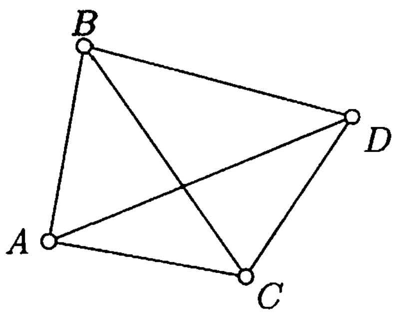


Рис. 16

Возможных циклических маршрутов шесть:

$$ABCDA, \quad ACDBA, \quad ADBCA, \quad ACBDA, \quad ABDCA, \quad ADCBA.$$

Однако для решения задачи достаточно сравнить длины только первых трех:

$$ABCDA, \quad ACDBA, \quad ADBCA.$$

Эти длины равны соответственно

$$11 + 6 + 10 + 17 = 44, \quad 13 + 10 + 9 + 11 = 43, \quad 17 + 9 + 6 + 13 = 45.$$

Тем самым, кратчайшим является любой из маршрутов длиной 43 — $ACDBA$ или $ABDCA$.

Важный класс графов составляют так называемые деревья. *Деревом* называется связный граф, который не имеет циклов (рис. 17).

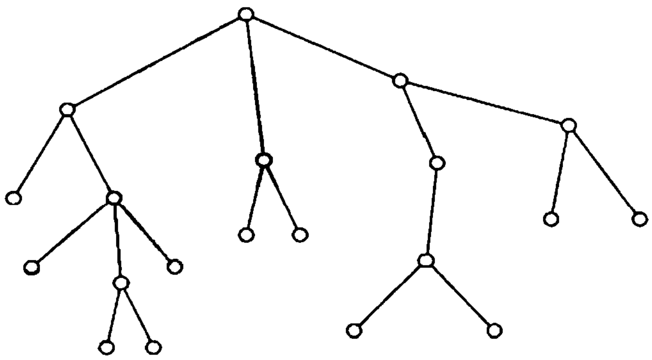


Рис. 17

Число B вершин дерева и число P его ребер различаются на единицу:

$$B = P + 1.$$

2.2. Сети

В приложениях граф обычно интерпретируется как *сеть*, а его вершины называют *узлами*.

Рассмотрим несколько характерных задач.

2.2.1. Дерево решений

При принятии важных решений для выбора наилучшего направления действий из имеющихся вариантов используется так называемое *дерево решений*, представляющее собой схематическое описание проблемы принятия решений.

Применение дерева решений подразумевает понимание (хотя бы интуитивное) таких понятий, как “вероятность” и “ожидаемое значение (математическое ожидание) случайной величины”. Подробно эти понятия обсуждаются в последующих главах.

Пример 4. По настоянию родителей выпускник американской школы должен продолжить учебу. Свободный в своем выборе, он хочет оценить возможности получения диплома в области инжиниринга или в области бизнеса в одном из двух университетов — в университете родного города Йорка и в университете штата, понимая, что вероятность успеха зависит от выбора как университета, так и будущей специальности.

1. Если он останавливает свой выбор на университете штата и бизнесе, то вероятность успеха (получение диплома) считается равной 0,60.
2. Если он останавливает свой выбор на университете штата и инжиниринге, то вероятность успеха считается равной 0,70.
3. Если он останавливает свой выбор на университете г. Йорка и бизнесе, то вероятность успеха считается равной 0,90.
4. Если он останавливает свой выбор на университете г. Йорка и инжиниринге, то вероятность успеха считается равной 0,95.
5. Средний доход за год в течение первых пяти лет у окончившего бизнес-школу университета штата при условии полной занятости равен 35 тыс. долл.
6. Средний доход за год в течение первых пяти лет у окончившего школу инжиниринга университета штата при условии полной занятости равен 30 тыс. долл.
7. Средний доход за год в течение первых пяти лет у окончившего бизнес-школу университета г. Йорка при условии полной занятости равен 24 тыс. долл.
8. Средний доход за год в течение первых пяти лет у окончившего школу инжиниринга университета г. Йорка при условии полной занятости равен 25 тыс. долл.
9. Если по каким-либо причинам выпускник не поступает ни в один из этих университетов, то его средний доход в течение этих пяти лет при условии полной занятости будет равен 18 тыс. долл.

Предположим, что единственным критерием при принятии выпускником окончательного решения является величина ожидаемого среднего дохода в первые пять лет его трудовой деятельности. Сделав это предположение, попробуем решить проблему выпускника, используя дерево решений (рис. 18).

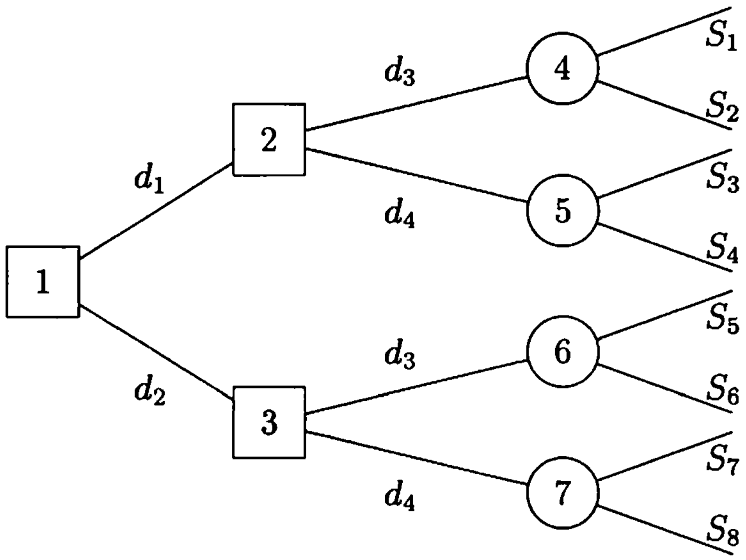


Рис. 18

Поясним некоторые обозначения на рисунке:

S_1 — окончание бизнес-школы университета штата,

S_2 — либо неудача при поступлении в бизнес-школу университета штата, либо невозможность завершения обучения,

S_3 — окончание школы инжиниринга университета штата,

S_4 — либо неудача при поступлении в школу инжиниринга университета штата, либо невозможность завершения обучения,

S_5 — окончание бизнес-школы университета города Йорка,

S_6 — либо неудача при поступлении в бизнес-школу университета города Йорка, либо невозможность завершения обучения,

S_7 — окончание школы инжиниринга университета города Йорка,

S_8 — либо неудача при поступлении в школу инжиниринга университета города Йорка, либо невозможность завершения обучения,

d_1 — выбор университета штата,

d_2 — выбор университета города Йорка,

d_3 — предпочтение отдано бизнесу,

d_4 — предпочтение отдано инжинирингу.

Узлы дерева, в которых делается выбор, обозначены квадратиками; узлы дерева, которые принимающий решение не контролирует, — кружками.

Эти два типа узлов рассчитываются по-разному.

При расчете узлов 4–7 определяются ожидаемые значения: значение в узле 7

$$N_7 = (0{,}95)(25\,000) + (0{,}05)(18\,000) = 24\,650 \text{ долл.},$$

значение в узле 6

$$N_6 = (0{,}90)(24\,000) + (0{,}10)(18\,000) = 23\,400 \text{ долл.},$$

значение в узле 5

$$N_5 = (0{,}70)(30\,000) + (0{,}30)(18\,000) = 26\,400 \text{ долл.},$$

значение в узле 4

$$N_4 = (0{,}60)(35\,000) + (0{,}40)(18\,000) = 28\,200 \text{ долл.}$$

Значение N_3 в узле 3 определяется так:

вследствие того что $N_7 > N_6$, полагаем $N_3 = N_7$ и считаем, что выбор d_4 предпочтительнее выбора d_3.

Тем самым, $N_3 = 24\,650$ долл.

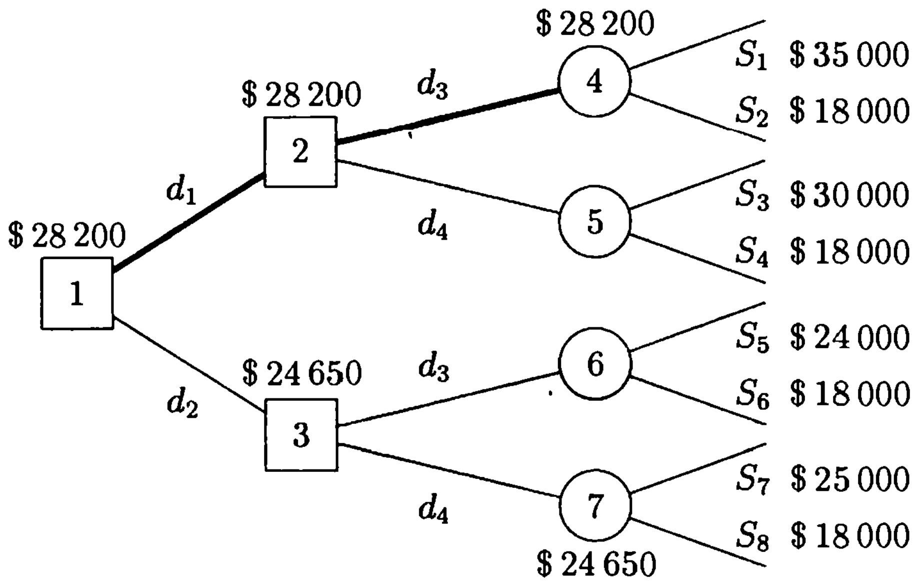


Рис. 19

Значение N_2 в узле 2 определяется так:

вследствие того что $N_4 > N_5$, полагаем $N_2 = N_4$ и считаем, что выбор d_3 предпочтительнее выбора d_4.

Тем самым, $N_2 = 28\,200$ долл.

Значение N_1 в узле 1 определяется так:

вследствие того что $N_2 > N_3$, полагаем $N_1 = N_2$ и считаем, что выбор d_1 предпочтительнее выбора d_2.

Тем самым, $N_1 = 28\,200$ долл.

Наносим результаты расчетов на рис. 18 и принимаем окончательное решение (рис. 19).

2.2.2. Задача о соединении городов

С деревьями связана и одна из проблем минимального соединения, внешне напоминающая задачу о коммивояжере, но значительно проще разрешаемая (для решения этой проблемы построены эффективные алгоритмы).

Имеется n городов — $A_1, A_2, \ldots, A_n$, которые нужно связать между собой сетью дорог. Стоимость сооружения дороги, соединяющей города A_i и A_k, известна и равна $c(A_i, A_k)$.

Какой должна быть сеть дорог, связывающая все города, чтобы стоимость ее сооружения была минимальной?

Граф наиболее дешевой соединяющей сети непременно должен быть деревом. (В противном случае в графе найдется хотя бы один цикл. При удалении любого из звеньев этого цикла стоимость сети уменьшится, а города все еще останутся соединенными.) Тем самым, число ребер искомого графа должно быть равным $n - 1$.

Алгоритм (план реализации)

На 1-м шаге связываем два города с наиболее дешевым соединяющим звеном e_1.

На каждом следующем шаге добавляем самое дешевое из звеньев e_i (если имеется несколько звеньев с одинаковой стоимостью, выбираем любое из них), в результате присоединения которого к уже построенным звеньям найденная сеть удлиняется еще на одно звено, но никакого цикла не образуется. При поиске добавляемого звена надо перебирать все ребра, имеющие общую вершину с уже построенной сетью.

Последний шаг алгоритма имеет номер $n - 1$.

Стоимость строительства полученной сети минимальна и равна

$$c(e_1) + c(e_2) + \cdots + c(e_{n-1}).$$

Пример 5. Пусть, например, нужно соединить города A, B, C и D. Стоимость строительства дорог, соединяющих любые два города, известна и в условных единицах представлена в таблице:

	A	B	C	D
A	–	11	13	12
B	11	–	6	9
C	13	6	–	10
D	12	9	10	–

Сеть дорог минимальной стоимости состоит из 3 $(4-1=3)$ звеньев и строится так: сначала выбирается самый дешевый участок дороги — BC (его цена равна 6), затем он удлиняется на самый дешевый из оставшихся — BD (его цена равна 9). И на последнем, третьем шаге вновь выбирается самый дешевый (но так, чтобы не образовалось никакого цикла) — AB (его цена равна 11) (рис. 20). Таким образом, стоимость строительства равна 26 $(6+9+11)$.

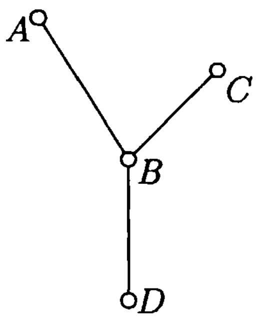


Рис. 20

2.2.3. Максимальный поток

Задана сеть, каждое ребро которой имеет вполне определенную ограниченную пропускную способность. Требуется определить максимально возможный поток в этой сети из заданного узла в другой узел.

Чтобы пояснить основную идею метода решения этой задачи, предположим, что исходный и конечный пункты, пункт A и пункт B, находятся на разных берегах разделяющей их реки (рис. 21). Множество мостов через реку образуют так называемое *разделяющее сечение* (если все мосты по каким-либо причинам выйдут из строя, попасть из пункта A в пункт B будет просто невозможно). Ясно, что пропускная способность разделяющего сечения складывается из пропускных способностей всех мостов.

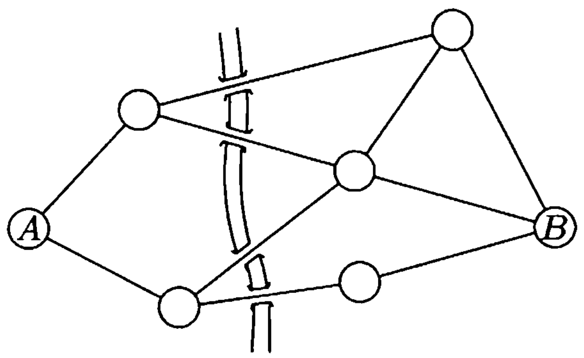


Рис. 21

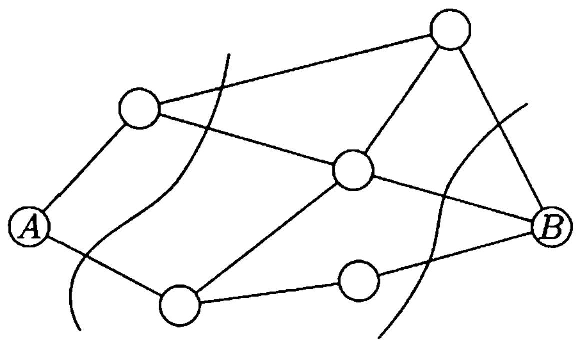


Рис. 22

Подобных сечений, разделяющих пункты A и B, может быть несколько (рис. 22), и каждое из них обладает своей пропускной способностью. Из того, что поток из пункта A в пункт B должен проходить через каждое разделяющее сечение, вытекает, что максимально возможный поток не может превосходить пропускной способности ни одного из этих сечений.

Таким образом, отыскание *макси-потока* (максимально возможного потока) сводится к отысканию *мини-сечения* (разделяющего сечения с наименьшей пропускной способностью).

Пример 6. Рассмотрим сеть, заданную на рис. 23. Требуется найти максимально возможный поток из узла 1 в узел 7.

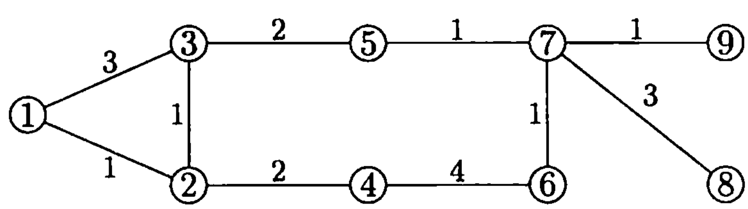


Рис. 23

Вычислим пропускную способность ключевых сечений. Имеем:
пропускная способность сечения $\{(1,2),\ (1,3)\}$ равна 4,
пропускная способность сечения $\{(2,4),\ (3,5)\}$ равна 4,
пропускная способность сечения $\{(1,3),\ (2,3),\ (6,7)\}$ равна 5,
пропускная способность сечения $\{(5,7),\ (6,7)\}$ равна 2.

Сравнивая пропускные способности сечений, получаем, что максимальный поток от вершины 1 к вершине 7 равен 2.

2.2.4. Кратчайший маршрут

Дана сеть, каждое ребро которой помечено числом, равным его длине. Требуется найти кратчайший маршрут, ведущий от выделенного узла к каждому из узлов сети.

Алгоритм решения этой задачи состоит из двух частей.

Покажем, как он работает, на следующем примере.

Пример 7. Рассмотрим сеть, заданную на рис. 24, с выделенным узлом 1.

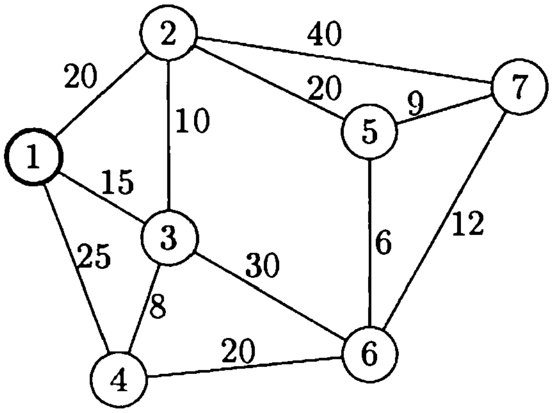


Рис. 24

Прямой ход алгоритма

1-й шаг. Все узлы, которые соединены с выделенным узлом 1 одним ребром, метятся так, как это показано на рис. 25 — первое число в метке равно расстоянию от помеченного узла до узла 1.

Ребро, связывающее узлы 1 и 3, является кратчайшим маршрутом от узла 1 к узлу 3 (любой другой маршрут от узла 1 к узлу 3 длиннее), и поэтому узлу 3 приписывается постоянная метка (15,1).

Таким образом, по окончании 1-го шага узлы 1 и 3 имеют постоянные метки, узлы 2 и 4 — временные метки, а узлы 5, 6 и 7 никаких меток не имеют (рис. 26).

Замечание. При получении постоянной метки узел 3 выделяется так же, как и узел 1.

2-й шаг. Отбираются все узлы, которые соединены с узлом 3 одним ребром и не имеют постоянных меток. Это узлы 2, 4 и 6.

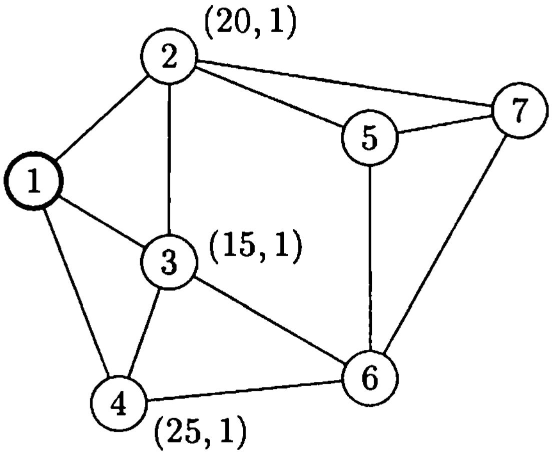


Рис. 25

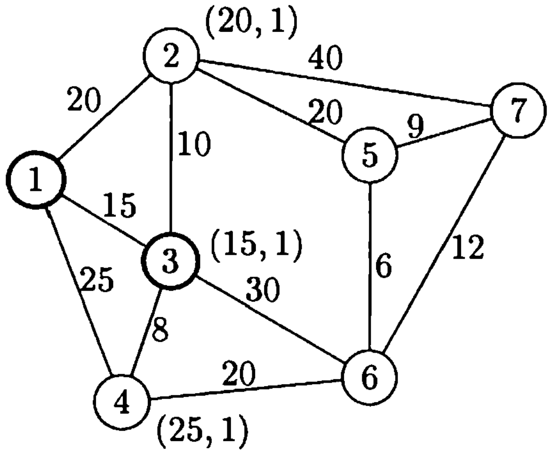


Рис. 26

Сравнивая длины маршрутов 1–2 и 1–3–2, замечаем, что длина первого (20) меньше длины второго ($15 + 10 = 25$). Поэтому метка (20,1) узла 2 остается неизменной.

Сравнивая длины маршрутов 1–4 и 1–3–4, замечаем, что длина первого (25) больше длины второго ($15+8 = 23$). Поэтому временная метка (25,1) узла 4 меняется на метку (23,3).

Узел 6 получает метку (45,3).

Замечание. Первое число в метке указывает длину маршрута от узла 1, а второе — номер предшествующего узла.

Ребро, связывающее узлы 1 и 2, является кратчайшим маршрутом от узла 1 к узлу 2 (любой другой маршрут от узла 1 к узлу 2 длиннее), и поэтому узлу 2 приписывается постоянная метка (20,1).

Таким образом, по окончании 2-го шага узлы 1, 2 и 3 имеют постоянные метки, узлы 4 и 6 — временные метки, а узлы 5 и 7 никаких меток не имеют (рис. 27).

3-й шаг. Отбираются все узлы, которые соединены с узлом 2 одним ребром и не имеют постоянных меток. Это узлы 5 и 7.

Узел 5 получает метку (40,2).

Узел 7 получает метку (60,2).

Маршрут 1–3–4, связывающий узлы 1 и 4, является кратчайшим маршрутом от узла 1 к узлу 4 (любой другой маршрут от узла 1 к узлу 4 длиннее); поэтому узлу 4 приписывается постоянная метка (23,3).

Таким образом, по окончании 3-го шага узлы 1, 2, 3 и 4 имеют постоянные метки, а узлы 5, 6 и 7 — временные метки (рис. 28).

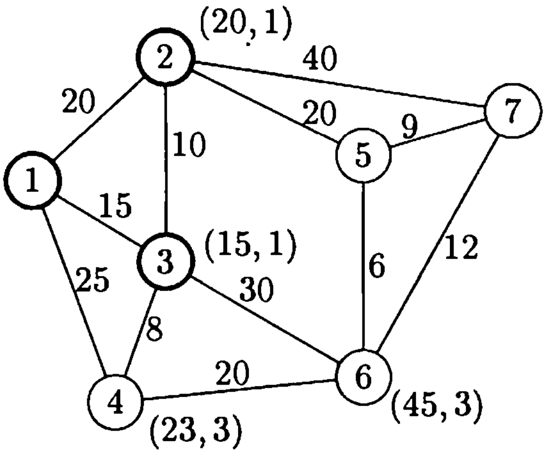


Рис. 27

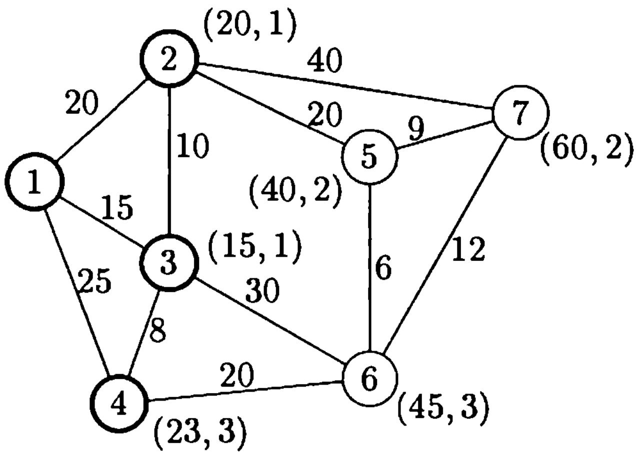


Рис. 28

4-й шаг. Отбираются все узлы, которые соединены с узлом 4 одним ребром и не имеют постоянных меток. Это узел 6.

Сравнивая длины маршрутов 1–3–6 и 1–3–4–6, замечаем, что длины первого (45) и третьего (45) больше длины второго (43). Поэтому временная метка (45,3) узла 6 меняется на метку (43,4).

Маршрут 1–2–5, связывающий узлы 1 и 5, является кратчайшим маршрутом от узла 1 к узлу 5 (любой другой маршрут от узла 1 к узлу 5 длиннее), и поэтому узлу 5 приписывается постоянная метка (40,2).

Таким образом, по окончании 4-го шага узлы 1, 2, 3, 4 и 5 имеют постоянные метки, а узлы 6 и 7 — временные метки (рис. 29).

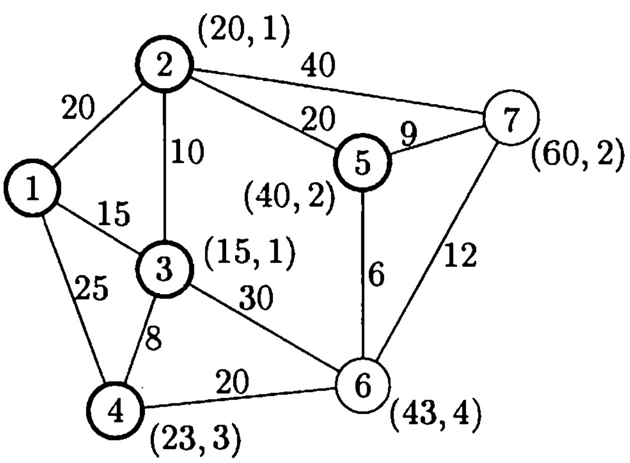


Рис. 29

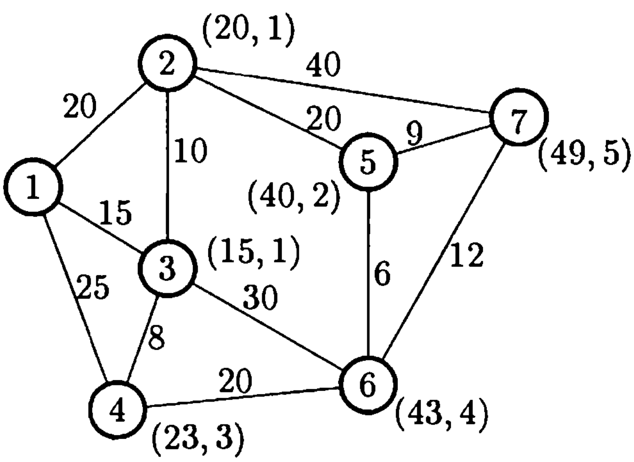


Рис. 30

Следующие два шага позволяют дать постоянные метки узлам 6 и 7 — (43,4) и (49,5) соответственно (рис. 30).

Замечание. На каждом шаге временная метка одного из узлов меняется на постоянную по следующему правилу: рассматриваются все узлы с временными метками и выбирается тот из них, длина маршрута до которого от узла 1 является наименьшей.

Обратный ход алгоритма

Используя вторую компоненту метки, определяем последовательность вершин в каждом кратчайшем маршруте. Например:

метка (49,5) узла 7 указывает на предшествующий узел 5,
метка (40,2) узла 5 указывает на предшествующий узел 2,
метка (20,1) узла 2 указывает на предшествующий узел 1.

В результате обратная последовательность узлов кратчайшего маршрута от узла 1 к узлу 7 имеет вид

$$7 \to 5 \to 2 \to 1.$$

Ответ:

Узел	Маршрут	Длина
2	1 – 2	20
3	1 – 3	15
4	1 – 3 – 4	23
5	1 – 2 – 5	40
6	1 – 3 – 4 – 6	43
7	1 – 2 – 5 – 7	49

2.2.5. Критический путь

Предположим, что требуется проанализировать проект с точки зрения минимальных временны́х затрат на его выполнение. Для этого проект разбивают на отдельные *работы*, или *действия*, оценивают время, необходимое на проведение каждой из них, и записывают последовательность операций, показывающую, какие работы должны быть закончены, прежде чем начнутся другие. Затем вычерчивается *диаграмма работ*, на которой каждая работа изображается направленным ребром, и определяется *критический путь*, имеющий наибольшую общую продолжительность. Он и определяет минимум временны́х затрат на выполнение проекта.

Покажем, как делается временна́я оценка проекта, на примере строительства небольшого загородного дома.

В табл. 1 и 2 указаны работы, их продолжительность и последовательность выполнения.

Таблица 1

Работа	Продолжительность	Работа	Продолжительность
A	2	F	2
B	7	G	6
C	15	H	8
D	8	I	2
E	10	J	3

Здесь A — заливка фундамента, B — изготовление оконных рам и дверей, C — изготовление встроенных шкафов и мебели, D — монтаж водопроводной системы, E — возведение стен, F — оштукатуривание стен, G — возведение крыши, H — благоустройство территории, I — установка встроенных шкафов и мебели, J — покраска.
Продолжительность работ указана в днях.

Таблица 2

D должна следовать за E
E должна следовать за A и B
F должна следовать за D и G
G должна следовать за E
H должна следовать за G
I должна следовать за C, F и H
J должна следовать за I

Таблица 3

a_1	—	2
a_2	—	7
a_3	—	15
a_4	a_1, a_2	10
a_5	a_4	8
a_6	a_4	6
a_7	a_5, a_6	2
a_8	a_6	8
a_9	a_3, a_7, a_8	2
a_{10}	a_9	3

Перенумеруем последовательно все работы, не имеющие предшествующих.

В данном случае это работы A — (a_1), B — (a_2) и C — (a_3).

Затем последовательно нумеруем остальные работы таким образом, чтобы все предшествующие им были уже занумерованы:

E — (a_4) (следует за (a_1) и (a_2)),
D — (a_5) (следует за (a_4)),
G — (a_6) (следует за (a_4)),
F — (a_7) (следует за (a_5) и (a_6)),
H — (a_8) (следует за (a_6)),
I — (a_9) (следует за (a_3), (a_7) и (a_8)),
J — (a_{10}) (следует за (a_9)).

В результате получаем табл. 3.

Составим диаграмму работ.

В верхней части рис. 31 каждая работа представлена на временно́й шкале (точка отсчета совпадает с началом работ) горизонтальным отрезком. Длины этих отрезков пропорциональны продолжительности соответствующих работ, а положения их левых концов определяются возможностью их выполнения (см. табл. 3).

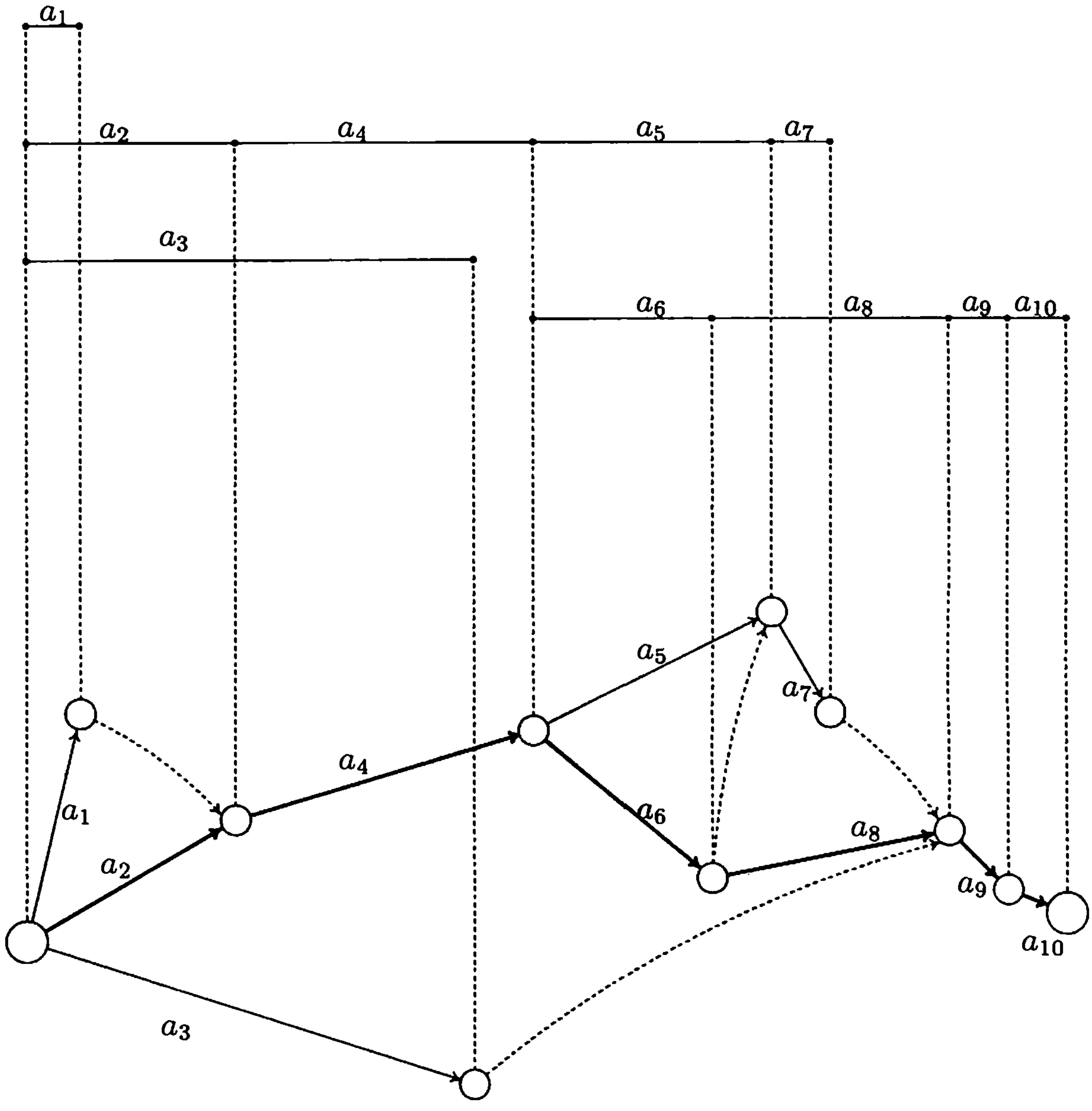


Рис. 31

В нижней части рис. 31 изображена направленная сеть, построенная по данным табл. 3 и в более наглядной форме показывающая, как именно связаны между собой работы по проекту и в какой очередности их следует выполнять. Ребра, обозначенные пунктиром, необходимы для соблюдения правильной последовательности операций.

Жирным выделен *критический путь* (направленный путь из начального события в конечное, имеющий наибольшую общую продолжительность).

Подсчитаем временны́е затраты на критическом пути. Имеем:

$$7 + 10 + 6 + 8 + 2 + 3 = 36.$$

Отсюда следует, что анализируемый проект может быть реализован за 36 дней и ни днем раньше.

Замечание. Всякая работа на критическом пути называется *критической* (малейшая задержка с началом ее выполнения увеличивает общую продолжительность работ). В данном случае критическими являются работы

$$a_2, \quad a_4, \quad a_6, \quad a_8, \quad a_9, \quad a_{10},$$

или, возвращаясь к исходным обозначениям,

B, E, G, H, I, J.

Определение всех таких работ важно для эффективного составления проекта.

В отличие от критической работы момент начала работы, не входящей в критический путь, может быть несколько сдвинут (вперед) без увеличения общей продолжительности. Важно, чтобы сдвинутая некритическая работа была завершена до начала критических работ, которым она предшествует.

2.3. Задания

1. Телефонная компания планирует соединить подземным кабелем шесть городов, расстояния между которыми заданы при помощи таблицы:

	A	B	C	D	E	F
A	–	10	9	30	27	20
B	10	–	15	18	17	20
C	9	15	–	25	21	16
D	30	18	25	–	8	17
E	27	17	21	8	–	13
F	20	20	16	17	13	–

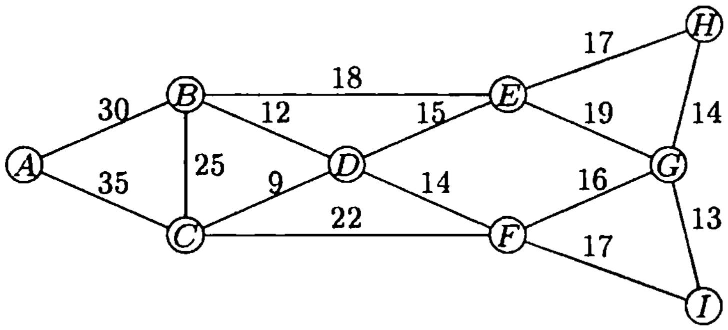


Рис. 32

Найдите минимальную длину кабеля, позволяющего жителям любых двух городов связаться друг с другом по телефону.

2. Найдите кратчайшие маршруты, ведущие из узла A во все другие узлы сети, представленной на рис. 32.

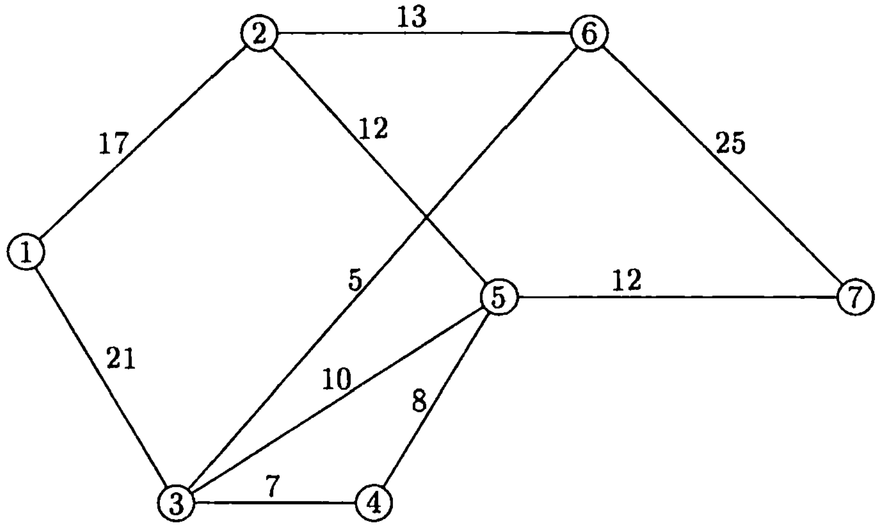


Рис. 33

3. Найдите максимальный поток в сети, представленной на рис. 33 (исходный узел — 1, конечный узел — 7).

4. Найдите критический путь для цикла работ $A(3)$, $B(6)$, $C(12)$, $D(9)$, $E(11)$, $F(3)$, $G(5)$, $H(7)$ и $I(3)$ (в скобках указана продолжительность соответствующего вида работы в днях) и минимальное время, необходимое для выполнения всего цикла, если последовательность операций подчинена следующим требованиям: работа D должна следовать за работой E, E — за A и B, F — за D и G, G — за E, H — за G, I — за C и F.

Глава 3
ЛИНЕЙНЫЕ ЗАДАЧИ

Линейные модели являются одним из наиболее активно используемых классов математических моделей.

Издавна тройное правило (позднее — линейная функция) было важным математическим инструментом в физике, химии, астрономии, экономике и вообще везде, где человек хотел объяснить и упорядочить наблюдаемые явления. И это естественно — для всякого наблюдения линейная функция является самой удобной математической моделью, и ею охотно пользуются.

Конечно, сейчас поле математических приложений значительно расширилось. Но по-прежнему линейные модели привлекают огромное внимание. Они сравнительно просты, хорошо разработаны, допускают полное исследование и достаточно эффективны в целом ряде стандартных ситуаций.

Линейность — это свойство математических выражений и функций. Выражение вида

$$ax + by,$$

где x и y — переменные величины, а a и b — постоянные числа, называется *линейным относительно переменных* x *и* y.

В случае если переменных больше двух — $x_1, x_2, \ldots, x_n$, линейное выражение относительно этих переменных имеет вид

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

где $a_1, a_2, \ldots, a_n$ — постоянные числа.

Заметим, что в линейное выражение все переменные входят в первой степени и никакие переменные не перемножаются.

Линейное программирование является, по-видимому, наиболее известным и одним из наиболее широко используемых инструментов management science. Это математический метод решения задачи оптимального распределения имеющихся ресурсов (или денег, или материалов, или времени) для достижения определенной цели (наибольшего дохода или наименьших издержек).

Пусть, например, отдел маркетинга на основе анализа рынка предлагает фирме выпустить выгодные для реализации новые виды продуктов.

Ясно, что каждый из новых продуктов будет вносить в доход фирмы свой вклад, а его изготовление потребует своей доли в расходе имеющихся в наличии ресурсов. Кроме того, следует учесть, что для одновременного производства всех новых продуктов наличных ресурсов, как правило, оказывается недостаточно.

У руководства фирмы возникает естественный вопрос: какие из этих новых видов продуктов и в каком количестве следует производить?

Для подготовки содержательного ответа на вопрос — производство какого количества новых продуктов способно принести наибольший доход (в рамках имеющихся ресурсов) — во многих случаях может быть использовано линейное программирование.

Теперь о названии *линейное программирование. Программирование* в данном термине имеет смысл *планирования. Линейное* же означает, что ищется экстремум линейной целевой функции при линейных ограничениях (линейных уравнениях и линейных неравенствах).

Тем самым линейное программирование имеет весьма мало общего с программированием, используемым в computer science. Вместе с тем вычислительные средства играют существенную роль в повышении эффективности его приложений. Дело в том, что многие реальные задачи линейного программирования содержат сотни неизвестных, уравнений и неравенств, и их невозможно успешно решать без современных быстродействующих компьютеров.

Укажем несколько общих ситуаций, в которых линейное программирование применяется часто и эффективно:

задачи о составлении смеси, цель которых заключается в выборе наиболее экономичной смеси ингредиентов (руды, нефти, пищевых продуктов и др.) при учете ограничений на физический или химический состав смеси и на наличие необходимых материалов;

задачи производства, целью которых является подбор наиболее выгодной производственной программы выпуска одного или нескольких видов продукции при использовании некоторого числа ограниченных источников сырья;

задачи распределения, цель которых состоит в том, чтобы организовать доставку материалов от некоторого числа источников к некоторому числу потребителей так, чтобы оказались минимальными либо расходы по этой доставке, либо время, затрачиваемое на нее, либо некоторая комбинация того и другого. В простейшем виде это задача о перевозках (*транспортная задача*).

Рассматриваются и комбинированные задачи (например, в случае, когда какой-то товар производится в разных местах, задачи производства и распределения объединяют в единую модель).

Наиболее распространенным методом решения задач линейного программирования является так называемый *симплекс*-метод, а наиболее эффективным из известных — *метод эллипсоидов*.

В простейшем случае, когда число переменных равно двум, удобен простой и наглядный *графический* метод. Он не требует больших предварительных знаний, при аккуратном использовании позволяет получать ответ без привлечения дополнительных средств и, что особенно важно, раскрывает суть идей, лежащих в основе других методов.

Другой важный класс линейных задач образуют задачи, сводимые к системам линейных уравнений, — это линейные задачи, ограничения в которых имеют характер равенств.

Одним из наиболее простых и одновременно эффективных подходов к решению линейных систем является метод последовательного исключения неизвестной.

3.1. Координаты

В этом разделе мы напомним некоторые геометрические факты, обсуждая их и развивая в той мере, в какой они понадобятся нам чуть позже.

Известно, что систематическое изложение метода координат, разнообразные применения которого столь широки и естественны сейчас, было предложено Рене Декартом (Descartes) еще в XVII в. Основной идеей этого метода является налаживание тесных связей между геометрическими объектами и числами. Напомним, как это делается.

3.1.1. Декартовы координаты

Рассмотрим прямую. Отметим на ней точку (точка O) и укажем стрелкой одно из двух возможных направлений перемещения вдоль этой прямой (рис. 1). При помощи выбранного масштабного отрезка (единицы измерения) мы можем поставить в соответствие каждой точке этой прямой число, равное расстоянию от нее до точки O, взятому со знаком плюс или минус в зависимости от того, с какой стороны от точки O эта точка лежит (справа или слева). Например, точке M_1 на рис. 2 соответствует число, равное 2, а точке M_2 — число, равное -3. Эти числа суть *координаты* рассматриваемых точек.

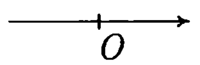


Рис. 1

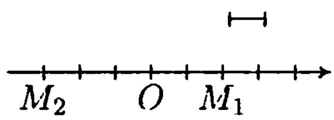


Рис. 2

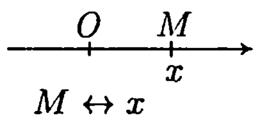


Рис. 3

Так вводится декартова координатная система на прямой (рис. 3). *Обозначение*: $M(x)$.

Предложенный Декартом подход позволяет описывать при помощи чисел точки не только на прямой, но и на плоскости и даже в пространстве.

Проведем на плоскости через фиксированную точку O две взаимно перпендикулярные прямые, укажем на каждой из них стрелкой одно из возможных направлений и выберем единый масштабный отрезок (рис. 4). По правилу, указанному на рис. 5, поставим в соответствие произвольной точке M *упорядоченную пару* чисел. Это *декартовы координаты* точки M — *абсцисса* x и *ордината* y. При

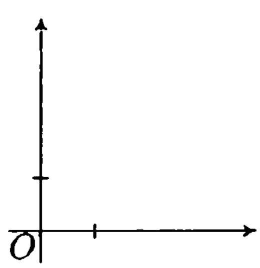


Рис. 4

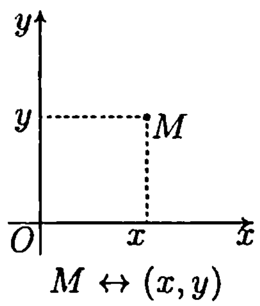


Рис. 5

этом говорят, что *на плоскости введена декартова система координат Oxy*, а оси Ox и Oy называют *координатными осями.*

Обозначение: $M(x, y)$.

Декартова система координат $Oxyz$ в пространстве вводится сходным образом: каждой точке пространства по правилу, указанному на рис. 6, приписывается *упорядоченная тройка* чисел — ее координат; координата z называется *аппликатой*. Координатные оси Ox, Oy и Oz взаимно перпендикулярны.

Обозначение: $M(x, y, z)$.

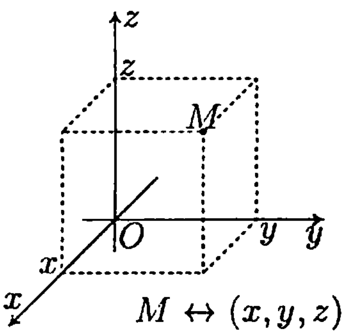


Рис. 6

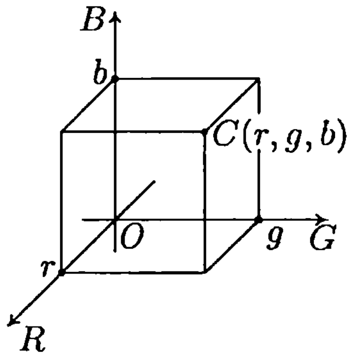


Рис. 7

Пример 1. Основой математического описания цвета является экспериментально установленный факт, что любой цвет при некоторых строго стандартизированных условиях его рассмотрения можно представить в виде смеси (суммы) определенных количеств трех основных цветов — красного (R), зеленого (G) и синего (B). Тогда три числа, описывающие данный цвет, являются количествами основных цветов в смеси, цвет которой зрительно неотличим от данного цвета. Например, цвет (C) создается смесью количества r цвета R, количества g цвета G и количества b цвета B (рис. 7). Числа r, g, b — цветовые координаты данного цвета C — $C(r, g, b)$. Так возникает естественная математическая модель совокупности цветов — трехмерное пространство, элементы которого суть разнообразные цвета, *трехмерное пространство цветов*.

3.1.2. Прямые. Полуплоскости

Задание точки посредством упорядоченного набора чисел позволяет описывать при помощи формул различные множества (это дает возможность добавить к геометрическим рассуждениям действенный аналитический аппарат) и, наоборот, привлекать к исследованию разнообразных формул наглядные геометрические соображения.

Пример 2. Введем на плоскости координатную систему Oxy и рассмотрим множество точек M, координаты x и y которых связаны равенством

$$3x + 4y = 6.$$

Это прямая, пересекающая координатные оси Ox и Oy в точках $P(2, 0)$ и $Q(0, 3/2)$ соответственно (рис. 8).

Вообще прямой на плоскости изображается любое соотношение вида

$$Ax + By = C, \tag{1}$$

где x и y — переменные величины, а A, B и C — постоянные числа, причем хотя бы одно из чисел, A или B, не равно нулю.

В самом деле, если $B \neq 0$, то, перенося слагаемое, содержащее переменную x, в правую часть, после деления на B получаем

$$y = kx + b$$

— уравнение прямой, записанное в более привычной для читателя форме (рис. 9) (здесь $k = -A/B$, $b = C/B$).

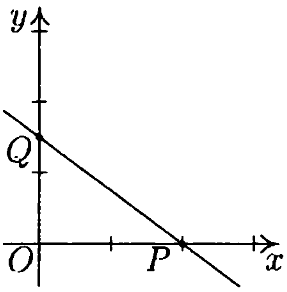


Рис. 8

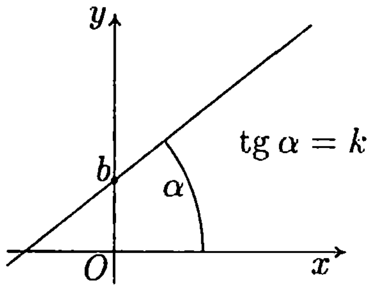


Рис. 9

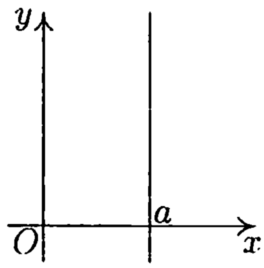


Рис. 10

Однако формула (1) является более общей — она включает в себя еще и вертикальные прямые: при $B = 0$ после деления на A получаем

$$x = a$$

(здесь $a = C/A$) (рис. 10).

Замечание 1. Если обе части равенства умножить на одно и то же не равное нулю число, то геометрический носитель этой связи (прямая) останется прежним.

Например, уравнения

$$150x + 200y = 300$$

и

$$3x + 4y = 6$$

описывают одну и ту же прямую (первое из этих равенств получается из второго после умножения обеих его частей на 50).

Замечание 2. Соотношение (1) принято называть *линейным* в отличие, например, от известного *квадратичного* $y = x^2$.

При помощи формул можно описывать и другие множества.

Пример 3. Множество точек $M(x, y)$ плоскости, у которых неотрицательна абсцисса x, описывается неравенством

$$x \geqslant 0$$

— это правая полуплоскость (рис. 11). Множество точек $M(x, y)$, у которых неотрицательна ордината y,

$$y \geqslant 0,$$

— это верхняя полуплоскость (рис. 12).

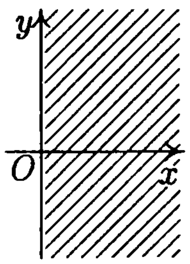


Рис. 11

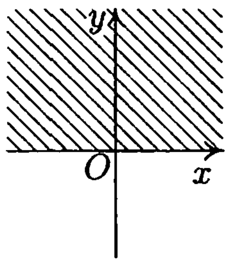


Рис. 12

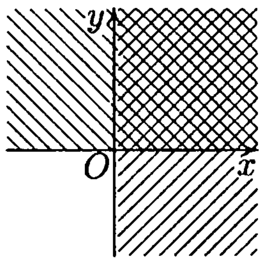


Рис. 13

На рис. 13 множество точек $M(x, y)$, обе координаты которых неотрицательны, помечено двойной штриховкой:

$$x \geqslant 0, \quad y \geqslant 0.$$

Пример 4. Попробуем теперь изобразить все точки плоскости, координаты x и y которых удовлетворяют неравенству

$$15x + 4y \leqslant 14.$$

Начнем с уравнения

$$15x + 4y = 14.$$

Оно описывает прямую $\mathcal{L}$, которую легко можно построить по точкам ее пересечения с осью Ox — точке $A(14/15, 0)$ и с осью Oy — точке $B(0, 7/2)$ (рис. 14).

Ясно, что если точка $M^\star(x^\star, y^\star)$ не лежит на этой прямой, то

$$15x^\star + 4y^\star \neq 14.$$

Это означает, что есть две возможности:
либо

$$15x^\star + 4y^\star < 14,$$

либо

$$15x^\star + 4y^\star > 14.$$

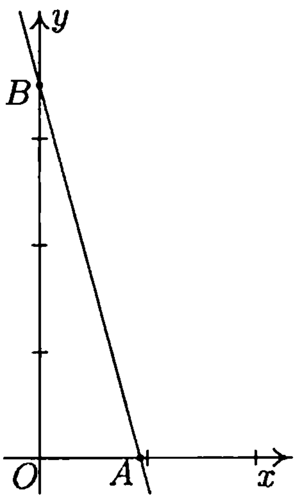


Рис. 14

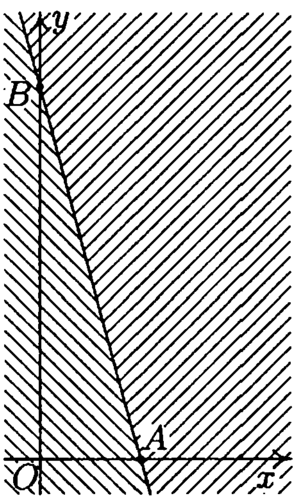


Рис. 15

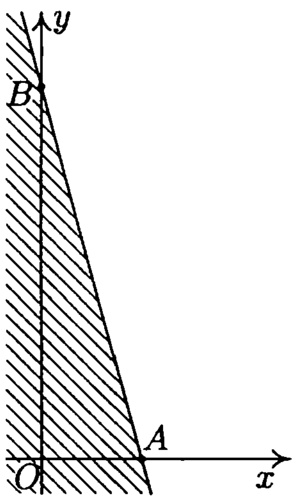


Рис. 16

Но всякая прямая делит плоскость на две полуплоскости. Поэтому если координаты точек одной из полуплоскостей, порожденных прямой $\mathcal{L}$, удовлетворяют неравенству

$$15x + 4y < 14,$$

то для координат точек другой полуплоскости выполняется неравенство

$$15x + 4y > 14.$$

Это означает, что каждому из неравенств

$$15x + 4y < 14$$

и

$$15x + 4y > 14$$

соответствует своя полуплоскость (рис. 15).

Теперь важно определить, какую именно из этих двух полуплоскостей описывает заданное неравенство.

Чтобы найти полуплоскость, описываемую неравенством

$$15x + 4y \leqslant 14,$$

проще всего поступить следующим образом: взять начальную точку $O(0,0)$ и подставить ее (нулевые) координаты в левую часть этого неравенства. Имеем:

$$0 < 14.$$

Это означает, что начальная точка O лежит в искомой полуплоскости (рис. 16).

Таким образом, геометрическое описание неравенства

$$15x + 4y \leqslant 14$$

получено.

Пользуясь аналогичными рассуждениями, можно описать на плоскости множество точек, координаты которых удовлетворяют неравенству

$$150x + 200y \geqslant 300 :$$

1) находим точки пересечения прямой

$$150x + 200y = 300$$

с координатными осями — это $C(2,0)$ и $D(0,3/2)$ (см. пример 2),

2) строим эту прямую (рис. 17),

3) находим, в какой из полуплоскостей, определяемых построенной прямой, лежит начальная точка $O(0,0)$; так как

$$150 \cdot 0 + 200 \cdot 0 = 0 < 300,$$

то искомая полуплоскость точку O не содержит (рис. 18).

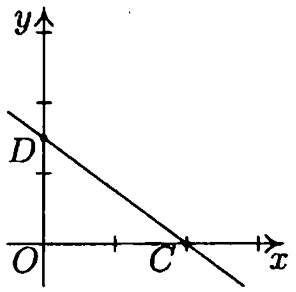


Рис. 17

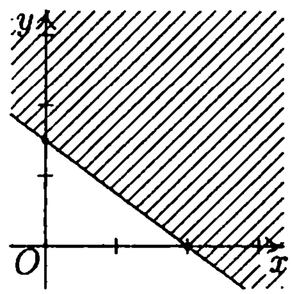


Рис. 18

Подведем некоторые итоги.

На координатной плоскости Oxy линейное уравнение вида

$$Ax + By = C$$

изображается прямой, а линейное неравенство

$$Ax + By \leqslant C,$$

или

$$Ax + By \geqslant C,$$

— полуплоскостью.

Для того чтобы узнать, какую именно полуплоскость описывает заданное неравенство, для определенности

$$Ax + By \geqslant C,$$

необходимо сначала построить соответствующую прямую

$$Ax + By = C$$

(по точкам ее пересечения с координатными осями), а затем определить, по какую сторону от этой прямой лежит начальная точка $O(0,0)$. Если нулевые координаты удовлетворяют заданному неравенству, то полуплоскость, в которой лежит точка $O(0,0)$, является искомой; если же не удовлетворяют, то искомой полуплоскостью будет та, которая точку $O(0,0)$ не содержит.

3.1.3. Пересечения прямых и полуплоскостей

Две произвольные прямые на плоскости могут пересекаться, быть параллельными или совпадать.

Если на плоскости введена координатная система Oxy, то прямые можно задавать уравнениями. Оказывается, что для любой пары прямых, заданных уравнениями, совсем не сложно определить, каково их взаимное расположение, только по виду этих уравнений.

Пример 5. Непосредственным построением легко убедиться в том, что прямые, заданные уравнениями

$$\begin{cases} x + y = 1, \\ x - y = 3, \end{cases}$$

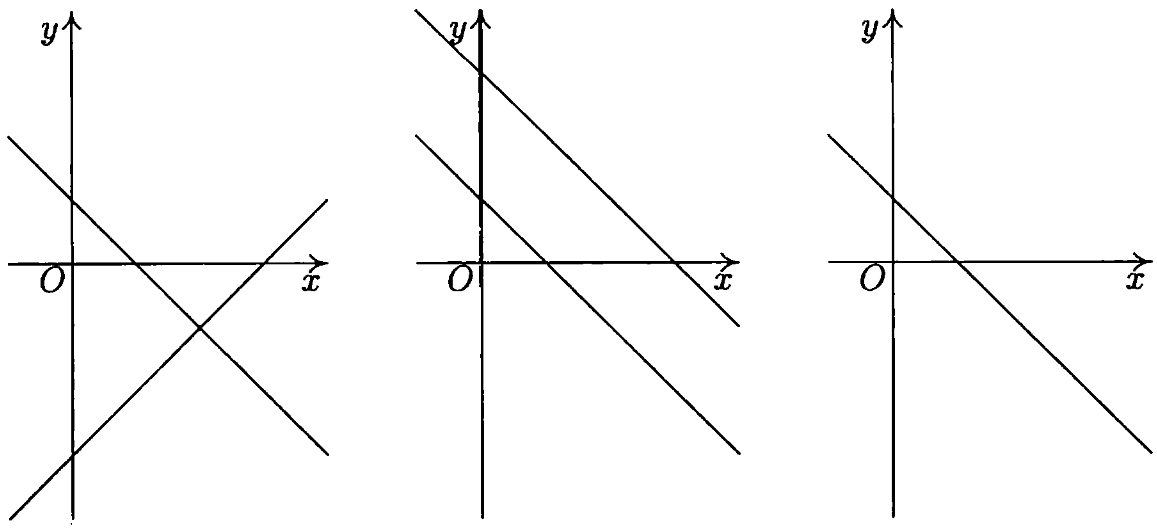
пересекаются (рис. 19*а*), прямые, заданные уравнениями

$$\begin{cases} x + y = 1, \\ x + y = 3, \end{cases}$$

параллельны (рис. 19*б*), а прямые, описываемые уравнениями

$$\begin{cases} x + y = 1, \\ 2x + 2y = 2, \end{cases}$$

совпадают (рис. 19*в*).



а *б* *в*

Рис. 19

Однако вглядитесь в эти уравнения чуть внимательнее, и вы получите те же ответы, не прибегая ни к построениям, ни к вычислениям.

В случае когда прямые заданы уравнениями общего вида

$$\begin{cases} A_1x + B_1y = C_1, \\ A_2x + B_2y = C_2, \end{cases}$$

важно не только уметь определять, пересекаются они или нет, но и находить их общую точку (конечно, в случае, если они пересекаются). Находить точно, а не по графику (примерно).

Для того чтобы отыскать точку пересечения двух прямых, требуется найти ее координаты, т. е. пару чисел $x^\star$ и $y^\star$, обращающую в верное равенство каждое из заданных уравнений.

Опишем один универсальный метод, позволяющий отыскивать решение системы линейных уравнений с двумя неизвестными.

Умножим обе части первого уравнения на B_2, второго — на $-B_1$ и, сложив результаты, получим уравнение, в котором явно присутствует только неизвестная x:

$$(A_1B_2 - A_2B_1)x = C_1B_2 - C_2B_1.$$

Отсюда следует, что

$$x^\star = \frac{C_1B_2 - C_2B_1}{A_1B_2 - A_2B_1}$$

(конечно, если $A_1B_2 - A_2B_1 \neq 0$).

В случае равенства $A_1B_2 - A_2B_1 = 0$ соответствующие прямые либо параллельны, либо совпадают. Отличительным признаком совпадения прямых является соотношение

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}.$$

Значение второй неизвестной находится аналогично: умножив обе части первого уравнения на $-A_2$, второго — на A_1 и сложив результаты, получим уравнение, в котором явно присутствует только неизвестная y:

$$(A_1B_2 - A_2B_1)y = A_1C_2 - A_2C_1.$$

Отсюда следует, что

$$y^\star = \frac{A_1C_2 - A_2C_1}{A_1B_2 - A_2B_1}.$$

Пример 6. Найдем методом исключения неизвестной общую точку прямых

$$\begin{cases} 15x + 4y = 14, \\ 150x + 200y = 300. \end{cases}$$

Умножая обе части первого уравнения на 50, второго — на -1 и складывая результаты, получаем уравнение

$$(750 - 150)x = 700 - 300.$$

Отсюда следует, что

$$x^\star = \frac{400}{600} = \frac{2}{3}.$$

Умножив теперь обе части первого уравнения на -10 и сложив результат со вторым уравнением, получим

$$(-40 + 200)y = -140 + 300,$$

откуда следует, что

$$y^\star = \frac{160}{160} = 1.$$

Пересекаться могут не только прямые, но и полуплоскости.

Пример 7. Рассмотрим пару полуплоскостей, заданных неравенствами

$$15x + 4y \leqslant 14$$

и

$$150x + 200y \geqslant 300.$$

Ранее каждую из них мы уже приводили (см. рис. 16 и 18). Изобразим теперь множество точек, принадлежащих обеим полуплоскостям одновременно. Ясно, что координаты каждой из этих точек должны удовлетворять системе неравенств

$$\begin{cases} 15x + 4y \leqslant 14, \\ 150x + 200y \geqslant 300. \end{cases}$$

На рис. 20 искомое множество отмечено перекрестной штриховкой.

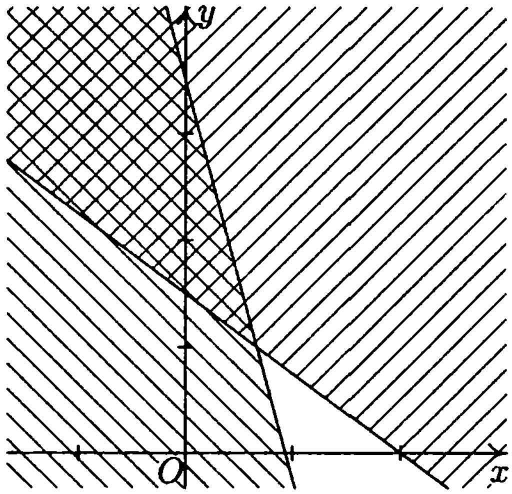


Рис. 20

Замечание. Можно рассматривать пересечение и большего числа полуплоскостей. При этом будут получаться и более сложные фигуры (рис. 21). Однако все они наделены общим свойством — эти фигуры являются *выпуклыми*.

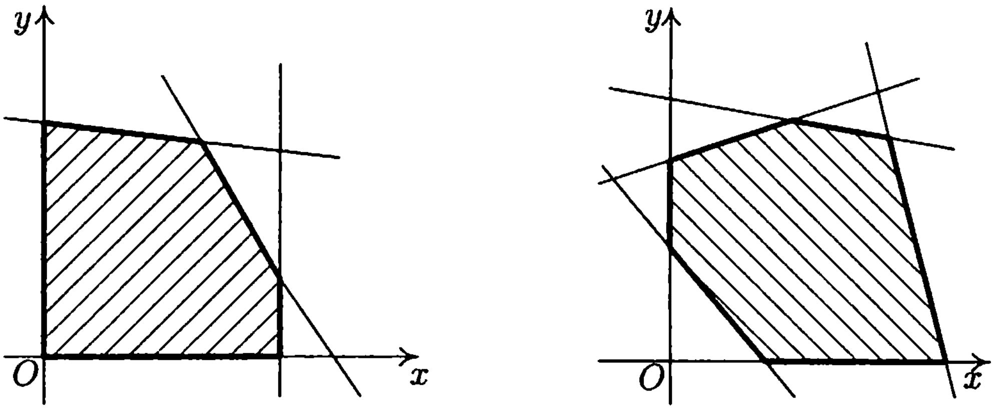


Рис. 21

Выпуклость является естественным и хорошо известным математическим понятием. Даже весьма далекий от математики человек имеет в целом правильное наглядное представление о том, что означают слова "быть выпуклым". Обратившись к такому человеку с просьбой нарисовать что-нибудь выпуклое или распознать, является предлагаемое изображение выпуклым или нет, можно быть уверенным — это будет сделано без особого труда и верно.

Среди выпуклых фигур особое место занимают выпуклые многоугольники. Некоторые из них хорошо известны — это треугольник, прямоугольник, трапеция и т. д.

3.1.4. Экстремальное свойство плоских срезов

В заключение обратимся к линейным соотношениям, связывающим три переменные величины.

Начнем с примера.

Пример 8. Формула

$$z = x + y$$

описывает хорошо известную операцию сложения: x и y — слагаемые, а z — их сумма. Пусть в пространстве введена декартова система координат $Oxyz$. Тогда множество точек, координаты x, y и z которых связаны заданным равенством, является плоскостью. На рис. 22 она показана как плоский срез прямоугольного параллелепипеда, проходящий через точки $O(0,0,0)$, $P(1,0,1)$, $Q(0,1,1)$, $R(1,1,2)$).

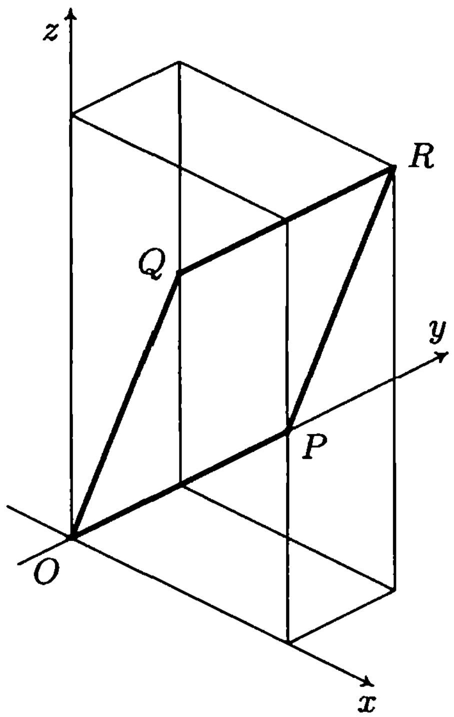


Рис. 22

Приведенное соотношение, как и более общее

$$z = ax + by + c,$$

является *линейным.* В пространстве с введенной координатной системой $Oxyz$ такие линейные соотношения изображаются плоскостями.

Важное замечание. Если внимательно рассмотреть плоский срез на рис. 22, то нетрудно заметить, что и самая нижняя и самая верхняя его точки лежат на ребрах параллелепипеда.

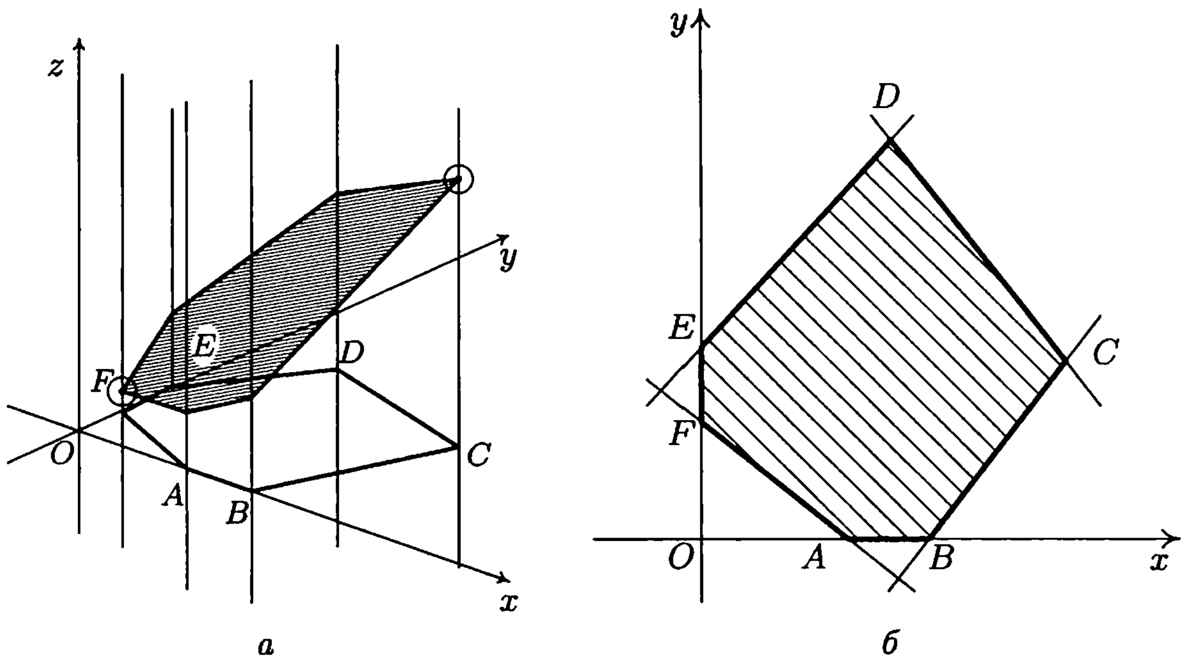


Рис. 23

Оказывается, так будет всегда вне зависимости от того, как именно проведен плоский срез призматической поверхности: самая нижняя точка этого среза и самая верхняя его точка будут лежать на ее ребрах (рис. 23*а*,*б*).

3.2. Линейное программирование

Изложенные положения лежат в основании так называемого *графического метода* решения задач линейного программирования. Покажем теперь, как их можно использовать для решения задач, совершенно "негеометрических" по формулировке.

3.2.1. Задача о диете

Исторически задача о диете является одной из первых задач линейного программирования.

Постановка задачи — первый и наиболее важный этап построения модели, способный обеспечить *правильное* решение проблемы.

Дама просто приятная решила похудеть и, как это нередко случается, обратилась за помощью к подруге.

Построение модели — рассмотрение этого этапа и является главной нашей целью в этом разделе.

Подруга — дама приятная во всех отношениях — посоветовала ей перейти на рациональное питание, состоящее исключительно из двух новомодных продуктов P и Q.

Дневное питание этими новинками должно давать не более 14 единиц жира (чтобы похудеть), но и не менее 300 калорий (чтобы не сойти с дистанции раньше). На банке с продуктом P написано, что в одном килограмме этого продукта содержится 15 единиц жира и 150 калорий, а на банке с продуктом Q — 4 единицы жира и 200 калорий соответственно. При этом цена 1 кг продукта P равна 15 руб., а 1 кг продукта Q — 25 руб.

Так как дама просто приятная в это время была весьма стеснена в средствах, то ее очень интересовал ответ на вопрос: в какой пропорции нужно брать эти удивительные продукты P и Q для того, чтобы выдержать условия диеты и истратить как можно меньше денег?

Переходим к переводу описанных обстоятельств на язык математических символов.

Обозначим через x количество продукта P и через y количество продукта Q, требуемые для выполнения условий диеты.

Количество единиц жира, содержащегося в x кг продукта P и в y кг продукта Q, равно $15x + 4y$ и по условию диеты не должно превосходить 14:

$$15x + 4y \leqslant 14.$$

В свою очередь, количество калорий, содержащихся в x кг продукта P и в y кг продукта Q, равно $150x + 200y$ и по условию диеты должно быть не меньше 300:

$$150x + 200y \geqslant 300.$$

Теперь о стоимости z продуктов. Она равна

$$z = 15x + 25y$$

и в соответствии с высказанными пожеланиями должна быть минимальной.

Последнее записывается так:

$$z = 15x + 25y \to \min.$$

Тем самым мы получили систему формул:

$$z = 15x + 25y \to \min, \tag{2}$$

$$\begin{cases} 15x + 4y \leqslant 14, \\ 150x + 200y \geqslant 300, \\ x \geqslant 0,\ y \geqslant 0, \end{cases} \tag{3}$$

с которой теперь и нужно разобраться.

Для того чтобы применить графический метод, введем на плоскости координатную систему Oxy.

Из условий $x \geqslant 0, y \geqslant 0$ вытекает, что все, что нас может заинтересовать, должно помещаться в первой четверти (рис. 24).

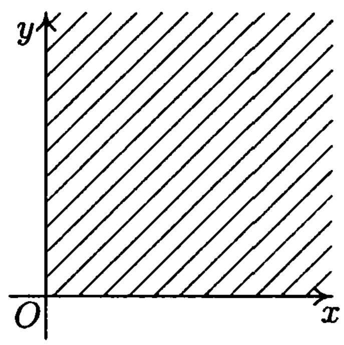


Рис. 24

Геометрический образ первого из неравенств,

$$15x + 4y \leqslant 14,$$

показан на рис. 16, а второго,

$$150x + 200y \geqslant 300,$$

— на рис. 18.

Сводя нарисованные картинки в одну, получим треугольник, изображенный на рис. 25. Две его вершины суть точки B и D, а третья вершина $E(2/3, 1)$ является точкой пересечения прямых

$$15x + 4y = 14 \quad \text{и} \quad 150x + 200y = 300.$$

Графиком линейной функции

$$z = 15x + 25y$$

является плоскость. Нас интересует только та ее часть, которая лежит над треугольником BDE (рис. 26). Соответствующий плоский

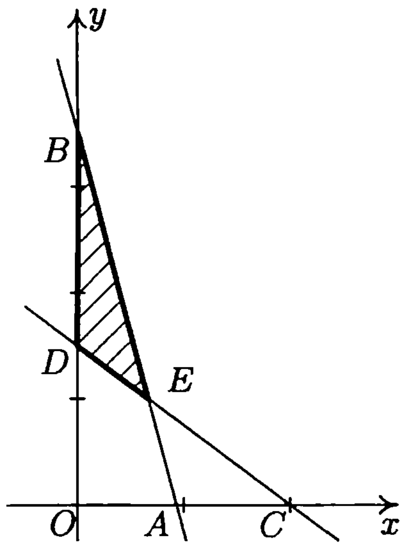


Рис. 25

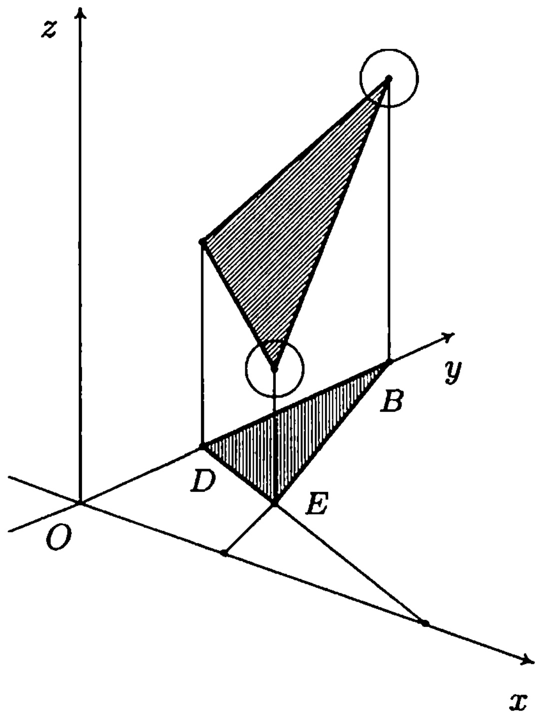


Рис. 26

срез обладает следующим важным свойством: как самая высокая, так и самая низкая его точки располагаются непосредственно над вершинами найденного треугольника.

Вычисляя значения z во всех трех вершинах этого треугольника:

$$z_B = 37{,}5, \quad z_D = 87{,}5, \quad z_E = 35$$

и сравнивая полученные результаты, замечаем, что наименьшее значение (35) достигается в вершине E.

Тем самым,

$$x = \frac{2}{3}, \quad y = 1,$$

и искомая пропорция — 2 : 3.

Этап построения модели завершен.

Ответ на вопрос, волновавший даму просто приятную, получен: для того, чтобы выдержать условия диеты и истратить как можно меньше денег, нужно брать рекомендованные продукты в отношении два к трем.

Воспользовалась ли она этими рекомендациями, неизвестно.

Что же касается остальных этапов, как-то: *проверка модели на достоверность*, *применение модели* (ни одну модель нельзя считать успешно выстроенной, пока она не принята, не понята и не применена на практике) и (возможно) *обновление модели*, то мы оставим их за рамками наших рассмотрений, так как основные трудности, связанные с реализацией этих этапов, носят вовсе не математический характер (и потому особенно сложны).

3.2.2. Задача о выпуске продукции

Фирма выпускает два вида древесно-стружечных плит — обычные и улучшенные. При этом производятся две основные операции — прессование и отделка. Требуется указать, какое количество плит каждого типа можно изготовить в течение месяца так, чтобы обеспечить максимальную прибыль при следующих ограничениях на ресурсы (материал, время, затраты):

Затраты	**Партия из 100 плит**		**Имеющиеся ресурсы на месяц**
	обычных	**улучшенных**	
Материал (фунты)	20	40	4000
Время на прессование (часы)	4	6	900
Время на отделку (часы)	4	4	600
Средства (доллары)	30	50	6000

если за каждые 100 обычных плит фирма получает прибыль, равную 80 долл., а за каждые 100 плит улучшенного вида — 100 долл.

Перейдем к построению математической модели поставленной задачи.

Введем следующие обозначения. Пусть

x — количество партий в 100 плит обычного вида, изготавливаемых в течение месяца, а

y — то же для плит улучшенного качества.

Тогда ожидаемую прибыль можно записать так:

$$z = \$80x + \$100y \to \max .$$

Для изготовления x партий в 100 плит обычного вида и y партий в 100 плит улучшенного вида требуется

$$20x + 40y$$

фунтов дерева. Ясно, что полученное число не может превосходить количество материала, имеющегося в наличии, т. е. 4000 фунтов. Тем самым, ограничения на материал имеют вид:

$$20x + 40y \leqslant 4000.$$

Подобным же образом рассчитываются ограничения на время изготовления и затраты:

$$\begin{aligned} \text{прессование:}\quad & 4x + \ \ 6y \leqslant \ \ 900, \\ \text{отделка:}\quad & 4x + \ \ 4y \leqslant \ \ 600, \\ \text{затраты:}\quad & 30x + 50y \leqslant 6000. \end{aligned}$$

Подведем итог: требуется найти такие значения x и y, подчиненные условиям

$$\begin{cases} 20x + 40y \leqslant 4000, \\ \ \ 4x + \ \ 6y \leqslant \ \ 900, \\ \ \ 4x + \ \ 4y \leqslant \ \ 600, \\ 30x + 50y \leqslant 6000, \\ x \geqslant 0, \\ y \geqslant 0, \end{cases}$$

для которых

$$z = 80x + 100y \to \max.$$

Для того чтобы найти в первой четверти плоскости Oxy множество точек, координаты (x, y) которых удовлетворяют указанным выше неравенствам, необходимо сначала построить прямые (по точкам их пересечения с координатными осями)

$$20x + 40y = 4000, \quad A(200, 0), B(0, 100), \tag{4}$$

$$4x + \ \ 6y = \ \ 900, \quad C(225, 0), D(0, 150), \tag{5}$$

$$4x + \ \ 4y = \ \ 600, \quad E(150, 0), F(0, 150), \tag{6}$$

$$30x + 50y = 6000, \quad G(200, 0), H(0, 120), \tag{7}$$

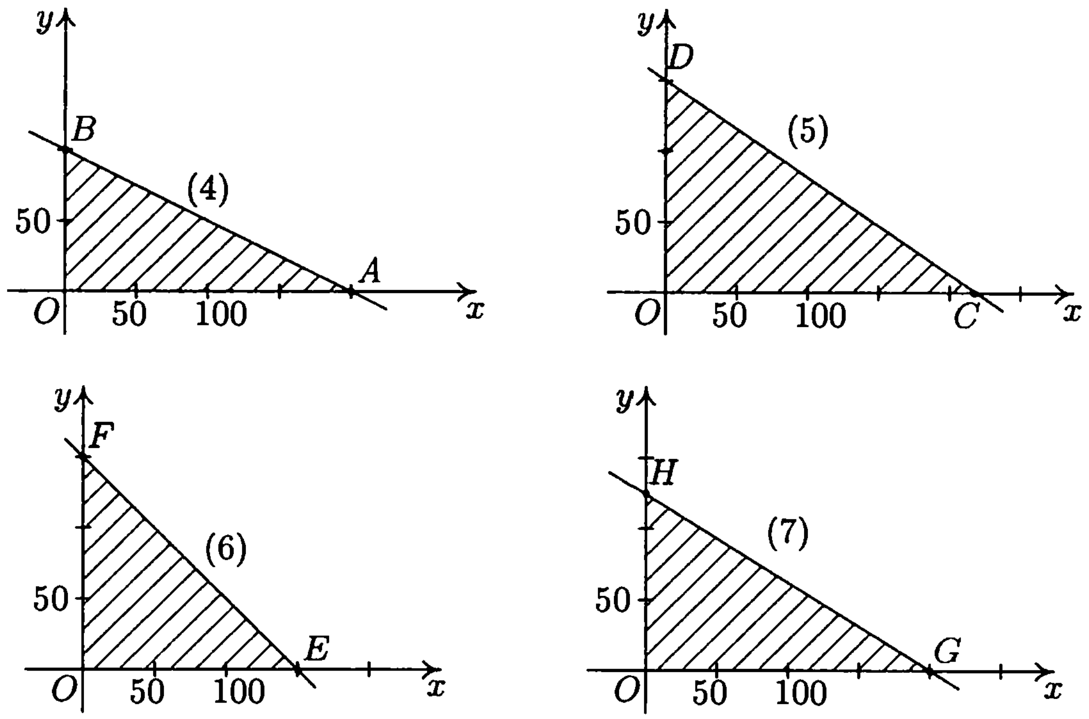


Рис. 27

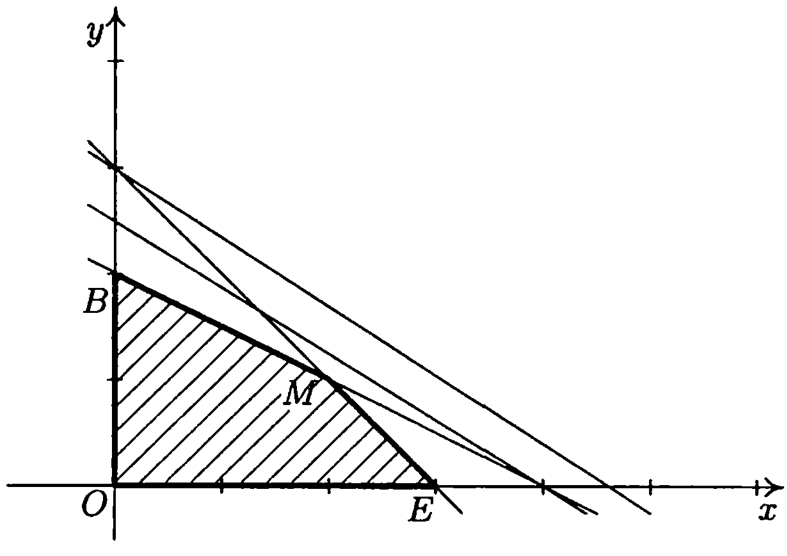


Рис. 28

а затем, используя точку начала отсчета $O(0, 0)$, определить соответствующие полуплоскости (рис. 27).

Пересечением полученных полуплоскостей будет четырехугольник, изображенный на рис. 28.

Напомним, что линейная функция $z = 80x + 100y$ достигает наибольшего значения в одной из вершин этого четырехугольника. Поэтому для ответа на поставленный вопрос необходимо

1) найти координаты точки пересечения прямых (4) и (6) — точки M (координаты вершин B и E известны):

решая систему уравнений

$$\begin{cases} 20x + 40y = 4000, \\ 4x + 4y = 600, \end{cases}$$

находим значения x и y: $x = 100$, $y = 50$;

2) вычислить значения z в точках $B(0, 100)$, $E(150, 0)$, $M(100, 50)$:

$$z_B = 10\,000, \quad z_E = 12\,000, \quad z_M = 13\,000;$$

3) выбрать наибольшее:

$$z_{\max} = 13\,000.$$

Ответ: $x = 100$, $y = 50$, $z_{\max} = 13\,000$.

3.2.3. Общая задача линейного программирования

В общем случае и число неизвестных, и число ограничений могут достигать десятков, сотен, тысяч и т. д. Однако набор соответствующих условий ничем (кроме количества) от рассмотренных выше примеров не отличается. Это нетрудно заметить уже по общей постановке задачи линейного программирования.

Стандартная математическая формулировка общей задачи линейного программирования выглядит так:

требуется найти экстремальное значение показателя эффективности (целевой функции)

$$c_1x_1 + c_2x_2 + \dots + c_nx_n \to \max$$

(линейной функции элементов решения $x_1, x_2, \ldots, x_n$) при линейных ограничительных условиях, накладываемых на элементы решения:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &\leqslant b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &\leqslant b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &\leqslant b_m, \\ x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad \ldots, \quad x_n &\geqslant 0, \end{aligned}$$

где a_{ik}, b_k, c_i — заданные числа.

Что касается существующих методов решения этой задачи с числом переменных, большим двух, то в их основе лежат те же идеи, на которые мы опирались при разработке графического подхода. Конечно, в случае сильного увеличения числа переменных и ограничений техника получения решения заметно усложняется, но она опирается на совершенно стандартные, хорошо разработанные алгоритмы (возникающие трудности связаны лишь с ростом объема необходимых вычислений).

Общую постановку задачи линейного программирования можно записать в более компактной форме, если воспользоваться следующим правилом.

Правило сокращенного суммирования. Для обозначения суммы чисел a_1, a_2, ..., a_n:

$$a_1 + a_2 + \dots + a_n$$

принята такая запись:

$$\sum_{k=1}^{n} a_k,$$

где $\sum$ — знак суммирования, а k — индекс суммирования.

Это обозначение очень удобно:

$$c_1x_1 + c_2x_2 + \dots + c_nx_n = \sum_{k=1}^{n} c_kx_k,$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \sum_{k=1}^{n} a_{1k}x_k.$$

А вот как выглядит запись общей задачи линейного программирования:

$$\sum_{k=1}^{n} c_kx_k \to \max,$$

$$\sum_{k=1}^{n} a_{ik}x_k \leqslant b_i, \quad i = 1, 2, \dots, m,$$

$$x_k \geqslant 0, \quad k = 1, 2, \dots, n.$$

3.2.4. Транспортная задача

Важный тип задач линейного программирования представляет *задача о перевозках.* Называется она так потому, что цель этой задачи заключается в минимизации полной стоимости перевозок известного количества товаров со складов к потребителю.

Мы поговорим здесь о так называемой *сбалансированной транспортной задаче* — задаче о перевозках, в которой общий объем товаров, готовых к отправлению, в точности равен объему товаров, который готовы принять в пунктах назначения.

Пример 9. Компания имеет два товарных склада и трех оптовых покупателей. Известно, что общий объем запасов на складах составляет 300 тыс. единиц продукции и совпадает с общим объемом заказов покупателей.

Конкретные данные о загруженности каждого из складов (в тыс. ед.), потребности каждого покупателя (в тыс. ед.) и стоимости перевозки (млн. руб. за 1 тыс. ед.) приведены в таблице:

		Стоимость перевозок к потребителям (млн. руб. за 1 тыс. ед.)			Наличие (тыс. ед.)
		B_1	B_2	B_3	
Склады	A_1	8	5	6	120
	A_2	4	9	7	180
Запрос (тыс. ед.)		70	140	90	300

Обозначим через x_{ik} количество товара, поставляемого со склада A_i покупателю B_k (рис. 29).

$$A_i \xrightarrow{x_{ik}} B_k$$

Рис. 29

Тогда соответствующая транспортная задача может быть сформулирована следующим образом.

Минимизировать общую стоимость перевозок:

$$z = 8x_{11} + 5x_{12} + 6x_{13} + 4x_{21} + 9x_{22} + 7x_{23} \to \min$$

при условии, что

$$x_{11} + x_{12} + x_{13} = 120,$$
$$x_{21} + x_{22} + x_{23} = 180,$$

$$x_{11} + x_{21} = 70,$$
$$x_{12} + x_{22} = 140,$$
$$x_{13} + x_{23} = 90,$$

$$x_{11} \geqslant 0, \quad x_{12} \geqslant 0, \quad x_{13} \geqslant 0, \quad x_{21} \geqslant 0, \quad x_{22} \geqslant 0, \quad x_{23} \geqslant 0.$$

Получаем задачу линейного программирования, в которой основные ограничения вследствие того, что транспортная задача сбалансирована, являются равенствами.

Замечание. Для лучшего понимания поставленной задачи часто полезно воспользоваться сетью (рис. 30).

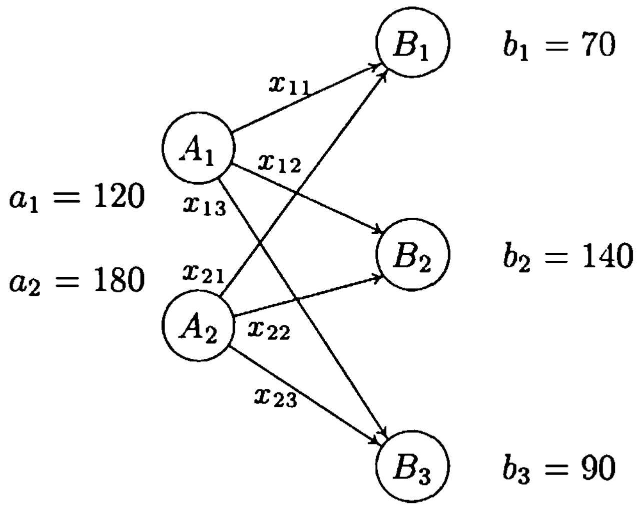


Рис. 30

Перейдем теперь к общей постановке сбалансированной транспортной задачи.

Пусть $A_1, A_2, \ldots, A_m$ — пункты отправления, а $B_1, B_2, \ldots, B_n$ — пункты назначения. Известно, что число единиц товара в пункте A_i равно a_i, в пункте B_k — b_k, причем

$$\sum_{i=1}^{m} a_i = \sum_{k=1}^{n} b_k$$

(задача сбалансирована), и c_{ik} — стоимость перевозки единицы товара из пункта A_i в пункт B_k.

Обозначим через x_{ik} (искомое) число единиц товара, пересылаемого из пункта A_i в пункт B_k.

Тогда общее количество товара, которое можно отправить из пункта A_i в пункты B_1, B_2, ..., B_n, равно

$$x_{i1} + x_{i2} + \cdots + x_{in} = \sum_{k=1}^{n} x_{ik} = a_i, \quad i = 1, 2, \ldots, m,$$

а

$$x_{1k} + x_{2k} + \cdots + x_{mk} = \sum_{i=1}^{m} x_{ik} = b_k, \quad k = 1, 2, \ldots, n,$$

— общее количество товара, которое можно принять в пункте B_k из пунктов A_1, A_2, ..., A_m.

Стоимость перевозки x_{ik} единиц товара из пункта A_i в пункт B_k равна $c_{ik}x_{ik}$, а общая стоимость всех перевозок —

$$z = \sum_{i=1}^{m} \sum_{k=1}^{n} c_{ik}x_{ik}.$$

В результате мы получаем следующую задачу линейного программирования:

$$z = \sum_{i=1}^{m} \sum_{k=1}^{n} c_{ik}x_{ik} \to \min,$$

$$\sum_{k=1}^{n} x_{ik} = a_i, \quad i = 1, 2, \ldots, m; \quad \sum_{i=1}^{m} x_{ik} = b_k, \quad k = 1, 2, \ldots, n,$$

все $x_{ik} \geqslant 0$.

Покажем, как в некоторых простых случаях транспортную задачу можно решать графическим методом.

Пример 10. Рассмотрим транспортную задачу, заданную таблицей

	B_1	B_2	Наличие
A_1	1 тыс. руб.	2 тыс. руб.	20
A_2	2 тыс. руб.	1 тыс. руб.	10
Запрос	16	14	30

Решение. Пусть x_{ik} — искомое число единиц товара, пересылаемого из пункта A_i в пункт B_k. Тогда данные таблицы можно представить в следующем виде:

$$z = x_{11} + 2x_{12} + 2x_{21} + x_{22} \to \min$$

при условии, что

$$\begin{aligned} x_{11} + x_{12} &= 20, \qquad & x_{11} + x_{21} &= 16, \\ x_{21} + x_{22} &= 10, \qquad & x_{12} + x_{22} &= 14, \end{aligned}$$

$$x_{11} \geqslant 0,\ x_{12} \geqslant 0,\ x_{21} \geqslant 0,\ x_{22} \geqslant 0$$

(рис. 31).

Положим $x_{11} = t$ и выразим через t остальные переменные: из первого уравнения

$$x_{12} = 20 - t,$$

из второго уравнения

$$x_{21} = 16 - t,$$

из третьего уравнения

$$x_{22} = 10 - x_{21} = 10 - (16 - t) = t - 6.$$

Тогда

$$z = t + 2(20 - t) + 2(16 - t) + t - 6 = 66 - 2t.$$

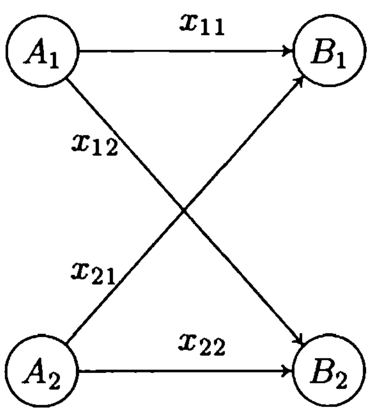


Рис. 31

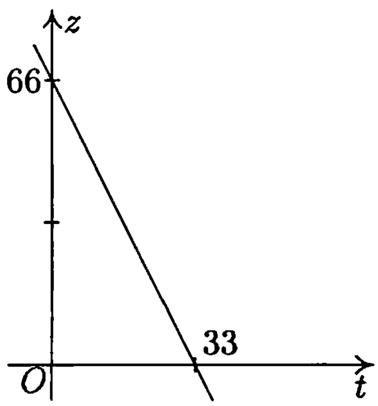


Рис. 32

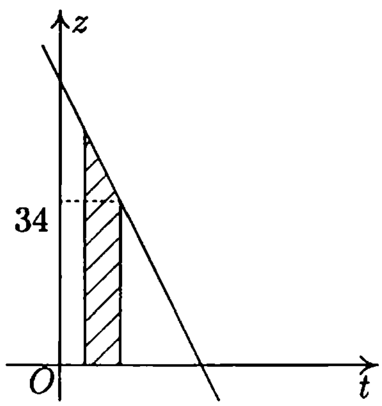


Рис. 33

На плоскости Otz уравнение $z = 66 - 2t$ изображается прямой (рис. 32). Из того, что все x_{ik} неотрицательны, получаем, что переменная t должна удовлетворять одновременно следующим четырем неравенствам:

$$t \geqslant 0, \quad t \leqslant 20, \quad t \leqslant 16, \quad t \geqslant 6.$$

Тем самым, нам нужно найти $z_{\min}$ при условии

$$6 \leqslant t \leqslant 16$$

(см. рис. 33).

Нетрудно видеть, что $z = z_{\min} = 34$ при $t = 16$.

Ответ: $x_{11} = 16$, $x_{12} = 4$, $x_{21} = 0$, $x_{22} = 10$, $z_{\min} = 34$.

Замечание. Подобным же способом можно решать сбалансированную транспортную задачу и с бо́льшим числом неизвестных.

Обратимся к примеру 9:

$$\begin{gathered} x_{11} + x_{12} + x_{13} = 120, \\ x_{21} + x_{22} + x_{23} = 180, \\ x_{11} + x_{21} = 70, \\ x_{12} + x_{22} = 140, \\ x_{13} + x_{23} = 90, \\ z = 8x_{11} + 5x_{12} + 6x_{13} + 4x_{21} + 9x_{22} + 7x_{23} \to \min. \end{gathered}$$

Положим $x_{11} = u$, $x_{12} = v$ и выразим через u и v остальные переменные. Имеем

$$\begin{gathered} x_{13} = 120 - u - v, \quad x_{21} = 70 - u, \quad x_{22} = 140 - v, \\ x_{23} = 90 - x_{13} = 90 - (120 - u - v) = u + v - 30 \end{gathered}$$

и

$$z = 8u + 5v + 6(120 - u - v) + 4(70 - u) + 9(140 - v) + 7(u + v - 30) = \\ = 5u - 3v + 2050.$$

Учитывая, что все x_{ik} неотрицательны, мы приходим к задаче

$$\begin{cases} 120 - u - v \geqslant 0, \\ 70 - u \geqslant 0, \\ 140 - v \geqslant 0, \\ u + v - 30 \geqslant 0, \\ u \geqslant 0, \quad v \geqslant 0, \end{cases}$$

$$z = 5u - 3v + 2050 \to \min.$$

которую можно решить графическим методом.

Выписанные неравенства определяют на плоскости (u, v) пятиугольник с вершинами

$$(30,0),\quad (70,0),\quad (70,50),\quad (0,120),\quad (0,30).$$

Нетрудно убедиться в том, что $z = z_{\min} = 1690$ при $u = 0$, $v = 120$.

Ответ: $x_{11} = 0$, $x_{12} = 120$, $x_{13} = 0$, $x_{21} = 70$, $x_{22} = 20$, $x_{23} = 90$, $z_{\min} = 1690$.

3.2.5. Целочисленное линейное программирование

Если переменные в задаче линейного программирования соответствуют числу машин, станков, людей или каких-либо иных неделимомых объектов, то имеют смысл только целочисленные (целые) значения этих переменных.

Рассмотрим следующий пример.

Пример 11. Найти решение задачи

$$\begin{aligned} z = 11x - y &\to \max,\\ 10x - y &\leqslant 40,\\ x + y &\leqslant 20{,}5,\\ x \geqslant 0, \quad & y \geqslant 0, \end{aligned}$$

ограничиваясь целочисленными значениями переменных x и y.

Решая задачу графическим методом, получаем

$$x = 5{,}5, \quad y = 15, \quad z_{\max} = 45{,}5$$

(рис. 34). Однако это решение недопустимо, так как 5,5 — не целое число. Ближайшие целые значения переменной x — это 5 и 6. Поэтому кажется разумным рассмотреть для (x, y) пары $(5, 15)$ и $(6, 15)$. Первая пара приводит к значению $z = 40$, а вторая пара недопустима: не удовлетворяет первым двум неравенствам задачи.

Однако, исследуя ситуацию графически, нетрудно показать, что, ограничиваясь только целочисленными значениями переменных, можно получить для величины z значение, большее 40. В самом деле, пара (5, 10) приводит к $z = 45$, и это действительно оптимальное решение задачи.

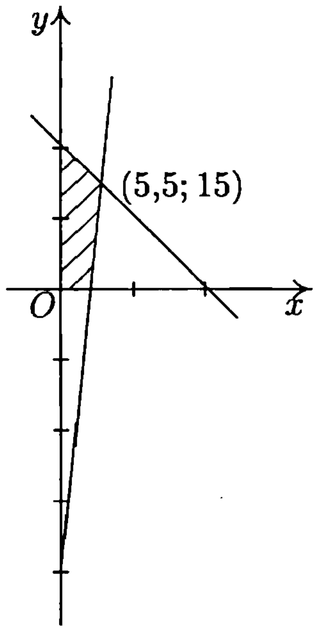


Рис. 34

Покажем, как его можно получить. Для того чтобы найти возможные целые значения переменных x и y, исходная задача разбивается на две подзадачи, выбирается переменная с нецелочисленным значением, в данном случае переменная x, вводятся два новых ограничения: $x \leqslant 5$ и $x \geqslant 6$, где 5 и 6 — целые числа, ближайшие к 5,5, и рассматриваются две новые задачи:

$$
\begin{array}{cc}
(A) & (B) \\
z = 11x - y \to \max, & z = 11x - y \to \max, \\
10x - y \leqslant 40, & 10x - y \leqslant 40, \\
x + y \leqslant 20{,}5, & x + y \leqslant 20{,}5, \\
x \leqslant 5, & x \geqslant 6, \\
x \geqslant 0, & x \geqslant 0, \\
y \geqslant 0, & y \geqslant 0.
\end{array}
$$

Решая обе задачи графическим методом (рис. 35), получаем, что *подзадача* (A): $x = 5$, $y = 10$, $z_{\max} = 45$;

подзадача (B): решения не имеет (противоречивые условия).

Целочисленное решение подзадачи (A) и дает искомый ответ:

$$x = 5, \quad y = 10, \quad z_{\max} = 45.$$

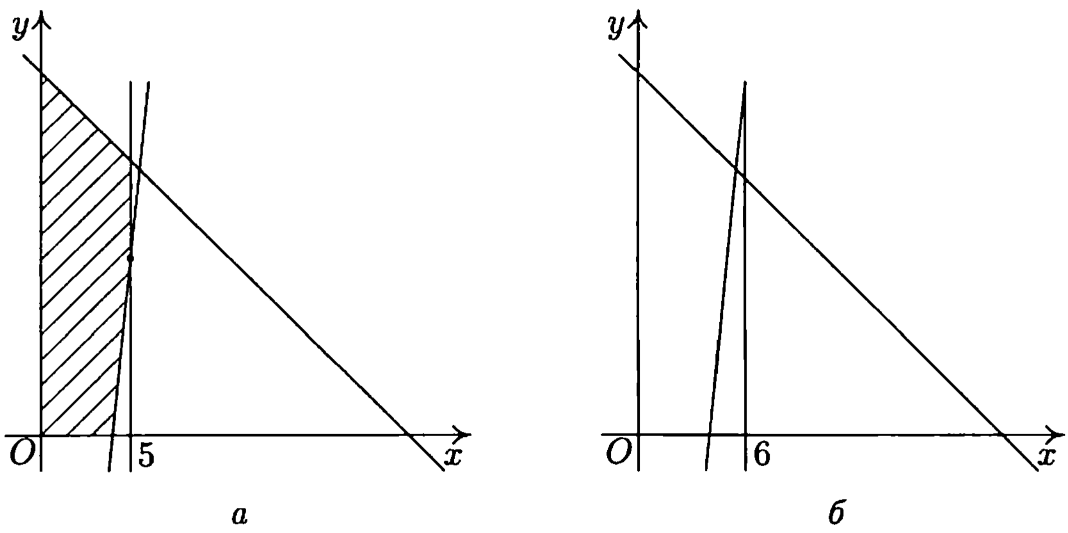


Рис. 35

Этот пример показывает, что к задачам целочисленного линейного программирования следует подходить очень внимательно, а не пользоваться "очевидным" рецептом и округлять нецелое решение до ближайших целых значений.

Замечание. Зачастую описанную операцию разбиения исходной задачи целочисленного программирования приходится повторять несколько раз.

3.3. Линейные системы

Наши рассмотрения мы начнем с несколько условного примера.

Пример 12. У завода есть четыре потребителя, которым еженедельно отгружается готовая продукция. Груз доставляется каждому потребителю на автомашине упакованным в ящики, маркированные в зависимости от вида продукции.

Однажды, когда автомашины были уже отправлены, но еще находились в пути, обнаружилось, что один из четырех видов груза был отправлен по ошибке и его следует возвратить (причем в полной сохранности и без нарушения целостности остальных грузов). Одновременно выяснилось также, что по недосмотру служащего не осталось никаких сведений о том, как именно маркирована та партия ящиков, в которых находился этот подлежащий возврату груз.

А что же известно?

Известно количество маркированных ящиков каждого вида, общий вес груза в каждой автомашине (см. таблицу),

Номер автомашины	Груз (количество ящиков)				
	1-й вид	**2-й вид**	**3-й вид**	**4-й вид**	**Общий вес, ц.**
1	1	4	9	8	51
2	2	9	8	3	45
3	2	6	8	6	48
4	3	5	7	8	51

а также и то, что ящики с возвращаемым грузом должны быть тяжелее остальных.

Возникает вопрос: а нельзя ли дать рекомендации по изъятию этого груза без распаковки и дополнительного взвешивания?

Оказывается, можно.

Приведем расчеты, при помощи которых совсем нетрудно выйти из сложившейся чрезвычайной ситуации.

Обозначим через x_k вес ящика с k-м видом груза. Тогда общий вес груза на автомашине 1 можно подсчитать так:

$$x_1 + 4x_2 + 9x_3 + 8x_4,$$

что равно 51.

Проведем аналогичные подсчеты для трех других автомашин и запишем результаты:

$$\begin{aligned} x_1 + 4x_2 + 9x_3 + 8x_4 &= 51, \\ 2x_1 + 9x_2 + 8x_3 + 3x_4 &= 45, \\ 2x_1 + 6x_2 + 8x_3 + 6x_4 &= 48, \\ 3x_1 + 5x_2 + 7x_3 + 8x_4 &= 51. \end{aligned}$$

Таким образом, мы получили набор (систему) линейных уравнений, в которые входят четыре неизвестные пока величины x_1, x_2, x_3 и x_4, подлежащие определению. Покажем, как их можно найти методом исключения неизвестной.

1-й шаг. Заметим, что если ко второму уравнению прибавить первое, умноженное на -2, то в результате получится соотношение

$$x_2 - 10x_3 - 13x_4 = -57,$$

в котором неизвестной x_1 явно нет.

Поступая аналогичным образом с третьим (первое уравнение нужно умножить на -2 и сложить с ним) и с четвертым (первое уравнение нужно умножить на -3 и сложить с ним) уравнениями и сохраняя неизменным первое, приходим к следующей системе соотношений:

$$\begin{aligned} x_1 + 4x_2 + \ 9x_3 + \ 8x_4 &= \quad 51, \\ x_2 - 10x_3 - 13x_4 &= \ -57, \\ -2x_2 - 10x_3 - 10x_4 &= \ -54, \\ -7x_2 - 20x_3 - 16x_4 &= -102. \end{aligned}$$

В дальнейших преобразованиях первое уравнение, активно работавшее на 1-м шаге, участия не принимает.

2-й шаг. На этом шаге рабочим является преобразованное второе уравнение:

$$x_2 - 10x_3 - 13x_4 = -57.$$

Умножим его на 2 и сложим с третьим уравнением. Полученное соотношение

$$-30x_3 - 36x_4 = -168$$

не содержит ни неизвестной x_1, ни неизвестной x_2.

После сложения четвертого уравнения со вторым, предварительно умноженным на 7, мы приходим к похожему результату:

$$-90x_3 - 107x_4 = -501.$$

В итоге получаем систему уравнений следующего вида:

$$\begin{aligned} x_1 + 4x_2 + \ 9x_3 + \ \ 8x_4 &= \quad 51, \\ x_2 - 10x_3 - \ 13x_4 &= \ -57, \\ -30x_3 - \ 36x_4 &= -168, \\ -90x_3 - 107x_4 &= -501. \end{aligned}$$

3-й шаг. На этом шаге рабочим является преобразованное третье уравнение:

$$-30x_3 - 36x_4 = -168.$$

Прибавляя его к четвертому (предварительно умножив на -3), окончательно получаем:

$$\begin{aligned} x_1 + 4x_2 + \ 9x_3 + \ 8x_4 &= \quad 51, \\ x_2 - 10x_3 - 13x_4 &= \ -57, \\ -30x_3 - 36x_4 &= -168, \\ x_4 &= \quad 3. \end{aligned} \tag{8}$$

Завершающий этап. В результате проделанных преобразований нам удалось найти значение одной из неизвестных, а именно $x_4 = 3$.

Значения остальных неизвестных находятся совсем просто.

Сначала из третьего уравнения системы (8) находим значение неизвестной x_3, предварительно подставив туда уже найденное значение неизвестной x_4:

$$-30x_3 - 36 \cdot 3 = -168.$$

Отсюда $x_3 = 2$.

Подставляя затем найденные значения неизвестных $x_3 = 2$ и $x_4 = 3$ во второе уравнение системы (8), получаем:

$$x_2 - 10 \cdot 2 - 13 \cdot 3 = -57$$

и, далее, $x_2 = 2$.

Наконец, из первого уравнения находим значение неизвестной x_1:

$$x_1 + 4 \cdot 2 + 9 \cdot 2 + 8 \cdot 3 = 51,$$

т. е. $x_1 = 1$.

Запишем полученный ответ: из того, что

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 2, \quad x_4 = 3,$$

вытекает, что нужно вернуть на завод ящики с 4-м видом груза, т. е. $8 + 3 + 6 + 8 = 25$ ящиков.

3.3.1. Что такое — матрица?

Читатель, видимо, уже успел заметить, что довольно часто либо при начальном описании задачи, либо при более жестком ее формулировании конкретный набор числовых данных удобно записывается в виде прямоугольной таблицы. Подобные прямоугольные таблицы естественно возникают при рассмотрении графов и сетей, при описании задач линейного программирования, при рассмотрении систем линейных уравнений и еще во многих других случаях, где они оказываются весьма кстати. Эти обстоятельства были замечены в середине XIX в., и появились матрицы. Сейчас вполне можно сказать, что матрица является одним из основных математических понятий, обладающим массой интересных и полезных свойств. Мы познакомимся лишь с некоторыми из них. Но сначала немного определений.

Матрицей $\mathbf{A}$ размера $m \times n$ называется набор $m \cdot n$ чисел a_{ik} $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$ — *элементов матрицы*, записанных в виде прямоугольной таблицы:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

Набор

$$a_{i1} \quad a_{i2} \quad \ldots \quad a_{in}$$

называется *i-й строкой* матрицы $\mathbf{A}$, а набор

$$\begin{matrix} a_{1k} \\ a_{2k} \\ \ldots \\ a_{mk} \end{matrix}$$

— *k-м столбцом* этой матрицы.

В случае когда $m = n$, матрица $\mathbf{A}$ называется *квадратной*, а число n — ее *порядком*.

Пользуясь понятием матрицы, опишем для примера все три шага преобразований, которые были проведены в примере 12.

Начнем с исходной системы уравнений и аккуратно выпишем все коэффициенты. Имеем

$$\left(\begin{array}{cccc|c} 1 & 4 & 9 & 8 & 51 \\ 2 & 9 & 8 & 3 & 45 \\ 2 & 6 & 8 & 6 & 48 \\ 3 & 5 & 7 & 8 & 51 \end{array}\right).$$

Здесь строки матрицы соответствуют уравнениям системы, а столбцы — неизвестным. Последний столбец отделен от остальных вертикальной чертой — этим подчеркивается его особое место в матрице (члены, расположенные в уравнениях по правую часть от знака равенства, не содержат неизвестных величин (*свободны* от неизвестных)).

Посмотрим теперь, как будут выглядеть матрицы, соответствующие системам, получавшимся в конце каждого шага:

после *1-го шага*:

$$\left(\begin{array}{cccc|c} 1 & 4 & 9 & 8 & 51 \\ 0 & 1 & -10 & -13 & -57 \\ 0 & -2 & -10 & -10 & -54 \\ 0 & -7 & -20 & -16 & -102 \end{array}\right),$$

после *2-го шага*:

$$\left(\begin{array}{rrrr|r} 1 & 4 & 9 & 8 & 51 \\ 0 & 1 & -10 & -13 & -57 \\ 0 & 0 & -30 & -36 & -168 \\ 0 & 0 & -90 & -107 & -501 \end{array}\right),$$

после *3-го шага*:

$$\left(\begin{array}{rrrr|r} 1 & 4 & 9 & 8 & 51 \\ 0 & 1 & -10 & -13 & -57 \\ 0 & 0 & -30 & -36 & -168 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right).$$

Замечание. Обратите внимание на то, как расположены нули в выписанных матрицах.

3.3.2. Линейные системы общего вида

Довольно ясно, что и число линейных уравнений, и число связанных ими неизвестных могут быть любыми.

Основные определения. Совокупность соотношений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned} \qquad (9)$$

где a_{ik}, b_i — заданные числа, а числа x_k рассматриваются как величины, подлежащие определению (*неизвестные*), называется *системой m линейных уравнений с n неизвестными* (*линейной системой*).

Заметим, что линейная система (9) полностью описывается матрицей

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right).$$

Решением линейной системы (9) называется любой упорядоченный набор чисел $\gamma_1, \gamma_2, \ldots, \gamma_n$, который при подстановке в каждое уравнение системы (9) вместо набора неизвестных $x_1, x_2, \ldots, x_n$ обращает это уравнение в верное равенство.

Замечание. Интересно отметить, что и в общем случае (как и для системы двух уравнений с двумя неизвестными) при исследовании системы могут встретиться только три варианта:

система имеет единственное решение,

система имеет бесчисленное множество решений,

система не имеет решений (такая система называется *несовместной*).

3.3.3. Исследование линейных систем

Метод последовательного исключения неизвестной, примененный в ходе решения примера 12, универсален. Пользуясь этим методом, можно исследовать любую систему уравнений и, если она имеет решение, найти его.

Пример 13. Дана система линейных уравнений

$$\begin{aligned} x_1 + \ \ 4x_2 - 3x_3 &= 2, \\ 2x_1 + \ \ 9x_2 + \ \ x_3 &= 0, \\ 7x_1 + 30x_2 - 7x_3 &= c+7 \end{aligned} \qquad \text{или} \qquad \left(\begin{array}{rrr|l} 1 & 4 & -3 & 2 \\ 2 & 9 & 1 & 0 \\ 7 & 30 & -7 & c+7 \end{array}\right).$$

Требуется найти значения параметра c, при которых система:

1) несовместна,

2) совместна.

В случае, когда система совместна, отыскать ее решение.

1-й шаг. При помощи первого уравнения исключаем неизвестную x_1 из второго и третьего уравнений (умножая первое уравнение на -2 и -7 и соответственно складывая со вторым и третьим). В результате получаем

$$\begin{aligned} x_1 + 4x_2 - \ \ 3x_3 &= 2, \\ x_2 + \ \ 7x_3 &= -4, \\ 2x_2 + 14x_3 &= c-7 \end{aligned} \qquad \text{или} \qquad \left(\begin{array}{rrr|l} 1 & 4 & -3 & 2 \\ 0 & 1 & 7 & -4 \\ 0 & 2 & 14 & c-7 \end{array}\right).$$

2-й шаг. Сохраняя первое уравнение неизменным, при помощи второго уравнения исключаем неизвестную x_2 из третьего уравнения (умножая второе уравнение на 2 и складывая с третьим). В результате получаем

$$\begin{aligned} x_1 + 4x_2 - 3x_3 &= 2, \\ x_2 + 7x_3 &= -4, \\ 0 &= c+1 \end{aligned} \qquad \text{или} \qquad \left(\begin{array}{rrr|l} 1 & 4 & -3 & 2 \\ 0 & 1 & 7 & -4 \\ 0 & 0 & 0 & c+1 \end{array}\right).$$

Заданная система
1) несовместна, если $c+1 \neq 0$,
2) совместна, если $c+1=0$.
При $c=-1$ получаем

$$\begin{aligned} x_1+4x_2-3x_3&=2,\\ x_2+7x_3&=-4,\\ 0&=0 \end{aligned} \qquad \text{или} \qquad \left(\begin{array}{ccc|c} 1 & 4 & -3 & 2\\ 0 & 1 & 7 & -4\\ 0 & 0 & 0 & 0 \end{array}\right).$$

В итоге на три неизвестные величины x_1, x_2 и x_3 имеем только два условия.

Полагая $x_3=t$, из второго уравнения находим, что

$$x_2+7t=-4.$$

Отсюда

$$x_2=-4-7t.$$

Подставим теперь полученные выражения в первое уравнение. Имеем

$$x_1+4(-4-7t)-3t=2.$$

Отсюда

$$x_1=18+31t.$$

Ответ:
1) при $c \neq -1$ система несовместна,
2) при $c=-1$ система совместна и

$$x_1=18+31t, \quad x_2=-4-7t, \quad x_3=t$$

— ее решение (здесь t — любое число).

Замечание. С похожей ситуацией мы уже встречались в примере 10. Соответствующая линейная система полностью описывается матрицей

$$\left(\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 20\\ 0 & 0 & 1 & 1 & 10\\ 1 & 0 & 1 & 0 & 16\\ 0 & 1 & 0 & 1 & 14 \end{array}\right),$$

а решение этой системы имеет вид

$$x_{11}=t, \quad x_{12}=20-t, \quad x_{21}=16-t, \quad x_{22}=t-6,$$

где t — любое число.

3.4. Операции над матрицами

Матрицы предоставляют весьма удобный способ записи количественной информации и потому часто используются в самых разных ситуациях.

Приведем лишь два примера.

Пример 14 (расписание). В колледже m студенческих групп: $g_1, \ldots, g_m$ и n преподавателей: $l_1, \ldots, l_n$. Известно, что в i-й группе g_i k-й преподаватель l_k должен провести a_{ik} часов занятий. Таблица (матрица) занятости студентов и преподавателей может быть записана в следующем виде:

	l_1	$\ldots$	l_n
g_1	a_{11}	$\ldots$	a_{1n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$
g_m	a_{m1}	$\ldots$	a_{mn}

(10)

(в i-й строке

$$\begin{pmatrix} a_{i1} & \ldots & a_{in} \end{pmatrix}$$

указано количество часов занятий, которые i-я группа g_i проведет с каждым из преподавателей $l_1, \ldots, l_n$, а в k-м столбце

$$\begin{pmatrix} a_{1k} \\ \ldots \\ a_{mk} \end{pmatrix}$$

— количество часов, которые k-й преподаватель l_k проведет с каждой из групп $g_1, \ldots, g_m$).

Пример 15. Мороженщица, торгующая в кинотеатре, перед утренним сеансом продала 36 порций пломбира: 8 порций в стаканчиках, 10 порций в брикетах, 7 порций в трубочках и 11 порций в рожках; перед дневным сеансом — 62 порции: соответственно 16, 15, 13 и 18. Наибольший спрос пришелся на вечер — 101 порция: 25, 21, 31 и 24 соответственно. Это можно записать компактно в виде таблицы

	У	Д	В
ст.	8	16	25
бр.	10	15	21
тр.	7	13	31
р.	11	18	24

(11)

Нередко матрицы выступают не только в роли хранителей информации. Интересно, что с ними можно проводить различные операции, подобные тем, которые мы привычно совершаем с числами: матрицы можно складывать, вычитать, умножать и пр. Однако действовать здесь приходится осторожно, с оглядкой — не всякие матрицы можно сложить, а с перемножением дело обстоит еще сложнее.

3.4.1. Сложение матриц

Пусть **A** и **B** — матрицы одинаковых размеров (т. е. состоящие из одинакового числа строк и одинакового числа столбцов),

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}.$$

Матрица

$$\mathbf{C} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{m1} & \dots & c_{mn} \end{pmatrix}$$

называется *суммой* матриц **A** и **B**, если ее элементы вычисляются по правилу

$$c_{ik} = a_{ik} + b_{ik}, \qquad i = 1, \dots, m; \quad k = 1, \dots, n.$$

Иными словами, складываются элементы матриц **A** и **B**, стоящие в одинаковых позициях (в i-й строке и в k-м столбце), и полученная сумма записывается в новой матрице **C** в ту же позицию (i, k).

Это можно записать и так:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \dots & \dots & \dots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

Обозначение: $\mathbf{A} + \mathbf{B}$.

Вычитание матриц определяется аналогично.

Замечание. Операция сложения определена лишь для матриц, имеющих одинаковые размеры. Если матрицы имеют разное число строк или разное число столбцов, то складывать их нельзя.

Продолжение примера 14. По заданной матрице занятости (10) требуется составить расписание занятий таким образом, чтобы общая продолжительность h всех проведенных занятий была минимальной, считая, что проблем с аудиторным фондом нет.

Попробуем сначала оценить общую продолжительность занятий снизу.

Преподаватель l_k, $k = 1, \ldots, n$, должен провести всего

$$a_{1k} + \cdots + a_{mk} = \sum_{i=1}^{m} a_{ik}$$

часов занятий, так что число h не может быть меньше, чем

$$l = \max_{1 \leqslant k \leqslant n} \sum_{i=1}^{m} a_{ik},$$

т. е.

$$h \geqslant l.$$

Общее число часов занятий в группе g_i, $i = 1, \ldots, m$, равно

$$a_{i1} + \cdots + a_{in} = \sum_{k=1}^{n} a_{ik},$$

вследствие чего число h не может быть меньше, чем

$$g = \max_{1 \leqslant i \leqslant m} \sum_{k=1}^{n} a_{ik},$$

т. е.

$$h \geqslant g.$$

Тем самым, общая продолжительность занятий h не меньше наибольшего из чисел l и g.

Иными словами, должно выполняться условие

$$h \geqslant s = \max\{l, g\}.$$

На самом деле для выполнения учебного плана требуется *ровно* s часов.

Покажем, как можно составить такое расписание занятий в случае, когда $m = 5$, $n = 3$, а матрица загруженности имеет вид

	l_1	l_2	l_3
g_1	2	1	3
g_2	0	1	0
g_3	2	3	1
g_4	1	0	2
g_5	0	2	0

Складывая элементы матрицы по строкам и по столбцам,

2	1	3	6
0	1	0	1
2	3	1	6
1	0	2	3
0	2	0	2
5	7	6	

легко убедиться в том, что $l = 7$ и $g = 6$, причем в наибольшей степени загружен преподаватель l_2, а g_1 и g_3 — наиболее загруженные группы.

В данном случае $h = s = 7$. Тем самым, минимальная продолжительность занятий равна семи, причем 2-й преподаватель должен быть загружен каждый час.

Покажем, как именно можно составить соответствующее оптимальное расписание.

Матрица (запись)

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

означает, что 1-й час занятий преподаватель l_1 проводит с группой g_4 (и, следовательно, не может провести этот час ни с какой другой группой, да и группа g_4 не может провести этот час ни с каким другим преподавателем), преподаватель l_2 проводит этот час с группой g_3, а преподаватель l_3 — с группой g_1.

По истечении 1-го часа занятий получаем новую матрицу загруженности на оставшиеся шесть часов:

$$\begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 2 \\ 0 & 1 & 0 \\ 2 & 2 & 1 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}.$$

Выбирая матрицу нагрузки на 2-й час занятий в виде

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

по истечении 2-го часа получим

$$\begin{pmatrix} 2 & 1 & 2 \\ 0 & 1 & 0 \\ 2 & 2 & 1 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}.$$

Повторяя подобные действия, т. е. строя каждый раз матрицу, состоящую из нулей и единиц так, что ни в одной ее строке и ни в одном столбце не может быть больше одной единицы, в итоге придем к следующему разложению:

$$\begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} +$$
$$+ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Полученное соотношение означает, что все занятия можно провести за 7 часов, а каждую группу и каждого преподавателя снабдить фактическим расписанием. Так, на 5-м часу преподаватель l_1 проводит занятия с группой g_1, преподаватель l_2 — с группой g_5 и преподаватель l_3 — с группой g_3, в то время как студенты групп g_2 и g_4 могут позаниматься в библиотеке или отдохнуть.

3.4.2. Умножение матрицы на число

Матрица $\mathbf{D}$ называется *произведением матрицы* $\mathbf{A}$ *на число* α, если ее элементы d_{ik} вычисляются по правилу

$$d_{ik} = \alpha a_{ik}, \qquad i = 1, \ldots, m; \quad k = 1, \ldots, n.$$

Иными словами, умножив элемент матрицы $\mathbf{A}$ из i-й строки и k-го столбца на число α, нужно записать полученный результат в i-ю строку и k-й столбец новой матрицы.

Сказанное можно записать и так:

$$\alpha\mathbf{A} = \alpha \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \multicolumn{3}{c}{\dotfill} \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \ldots & \alpha a_{1n} \\ \multicolumn{3}{c}{\dotfill} \\ \alpha a_{m1} & \ldots & \alpha a_{mn} \end{pmatrix}.$$

Обозначение: $\alpha\mathbf{A}$.

В частности, при умножении матрицы $\mathbf{A}$ на число $\alpha = 0$ получается матрица, все элементы которой равны нулю (*нулевая матрица* того же размера, что и матрица $\mathbf{A}$):

$$\mathbf{O} = \begin{pmatrix} 0 & \ldots & 0 \\ \multicolumn{3}{c}{\dotfill} \\ 0 & \ldots & 0 \end{pmatrix}.$$

Пример 16. Матрица выплат

$$\begin{pmatrix} 15000 & 20000 & 15000 \\ 30000 & 10000 & 10000 \end{pmatrix}$$

в результате деноминации с 1 января 1998 г. выглядит так:

$$\begin{pmatrix} 15 & 20 & 15 \\ 30 & 10 & 10 \end{pmatrix} = \frac{1}{1000} \begin{pmatrix} 15000 & 20000 & 15000 \\ 30000 & 10000 & 10000 \end{pmatrix}.$$

3.4.3. Транспонирование матрицы

Результат торговли мороженым можно записать двояко: либо так, как это сделано в примере 15:

	У	Д	В
ст.	8	16	25
бр.	10	15	21
тр.	7	13	31
р.	11	18	24

$$\begin{pmatrix} 8 & 16 & 25 \\ 10 & 15 & 21 \\ 7 & 13 & 31 \\ 11 & 18 & 24 \end{pmatrix},$$

либо так:

	ст.	бр.	тр.	р.
У	8	10	7	11
Д	16	15	13	18
В	25	21	31	24

$$\begin{pmatrix} 8 & 10 & 7 & 11 \\ 16 & 15 & 13 & 18 \\ 25 & 21 & 31 & 24 \end{pmatrix}.$$

Ясно, что обе матрицы содержат одну и ту же информацию. Разница лишь в том, что записанное в одной матрице в столбцы в другой помещено в строки, причем в том же порядке, и наоборот.

Пусть

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \multicolumn{3}{c}{\ldots\ldots\ldots\ldots\ldots} \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$$

— заданная матрица. Построим матрицу $\mathbf{B}$ по следующему правилу: взяв в матрице $\mathbf{A}$ произвольный элемент

$$a_{ik}, \qquad i = 1, \ldots, m; \quad k = 1, \ldots, n,$$

запишем его в новой матрице в k-ю строку и i-й столбец, т. е. положим

$$b_{ki} = a_{ik}.$$

В результате получим матрицу

$$\mathbf{B} = \begin{pmatrix} a_{11} & \ldots & a_{m1} \\ \multicolumn{3}{c}{\ldots\ldots\ldots\ldots\ldots} \\ a_{1n} & \ldots & a_{mn} \end{pmatrix},$$

про которую говорят, что она получена из матрицы $\mathbf{A}$ путем *транспонирования*, а сама операция над матрицей $\mathbf{A}$, которая, не изменяя самих ее элементов, определяет для них новый порядок расположения, называется *транспонированием матрицы* $\mathbf{A}$.

Обозначение: $\mathbf{A}^T$.

3.4.4. Умножение матрицы на столбец

Пусть

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

— матрица размера $m \times n$ и

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

— столбец высоты n.

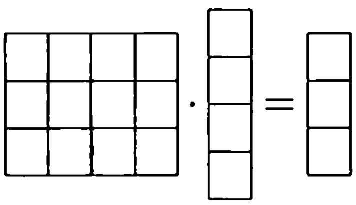
Произведение матрицы $\mathbf{A}$ *на столбец* $\mathbf{x}$ определяется так:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix}$$

или в несколько сокращенном виде:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \dots \\ \sum_{k=1}^{n} a_{mk}x_k \end{pmatrix}.$$

Заметим, что матрицу можно умножать не на любой столбец, а лишь на такой, число элементов которого равно числу столбцов матрицы. Символически это можно описать так:



Обозначение: $\mathbf{A}\mathbf{x}$.

Замечание. Результирующий столбец $\mathbf{A}\mathbf{x}$ содержит ровно m элементов. При $m = n$ оба столбца (и $\mathbf{x}$, и $\mathbf{A}\mathbf{x}$) имеют одинаковую высоту.

Пример 17 (подсчет выручки). Вновь вернемся к мороженщице и подсчитаем ее выручку, зная цену каждого сорта проданного ею мороженого. При цене 3,0 руб. за одну порцию пломбира в стаканчиках, 1,5 руб. за одну порцию пломбира в брикетах, 2,0 руб. за одну порцию пломбира в трубочках и 2,5 руб. за одну порцию пломбира в рожках утренняя выручка оказывается равной:

$$8 \cdot 3{,}0 + 10 \cdot 1{,}5 + 7 \cdot 2{,}0 + 11 \cdot 2{,}5 = 80{,}5,$$

дневная:

$$16 \cdot 3{,}0 + 15 \cdot 1{,}5 + 13 \cdot 2{,}0 + 18 \cdot 2{,}5 = 141{,}5$$

и вечерняя:

$$25 \cdot 3{,}0 + 21 \cdot 1{,}5 + 31 \cdot 2{,}0 + 24 \cdot 2{,}5 = 228{,}5.$$

Те же самые результаты мы получим, умножив матрицу продаж на ценовой столбец:

$$\begin{pmatrix} 8 & 10 & 7 & 11 \\ 16 & 15 & 13 & 18 \\ 25 & 21 & 31 & 24 \end{pmatrix} \cdot \begin{pmatrix} 3{,}0 \\ 1{,}5 \\ 2{,}0 \\ 2{,}5 \end{pmatrix} = \begin{pmatrix} 80{,}5 \\ 141{,}5 \\ 228{,}5 \end{pmatrix}$$

— столбец выручки.

3.4.5. Умножение строки на матрицу

Пусть

$$\mathbf{p} = (p_1 \quad \ldots \quad p_m)$$

— строка из m элементов и

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$$

— матрица размера $m \times n$.

Произведение строки $\mathbf{p}$ *на матрицу* $\mathbf{A}$ определяется формулой

$$(p_1 \quad \ldots \quad p_m) \cdot \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} = \left(\sum_{i=1}^{m} p_i a_{i1} \quad \ldots \quad \sum_{i=1}^{m} p_i a_{in}\right)$$

или символически

$$\begin{array}{|c|c|c|}\hline \ & \ & \ \\ \hline\end{array} \cdot \begin{array}{|c|c|c|c|}\hline \ & \ & \ & \ \\ \hline \ & \ & \ & \ \\ \hline \ & \ & \ & \ \\ \hline\end{array} = \begin{array}{|c|c|c|c|}\hline \ & \ & \ & \ \\ \hline\end{array}.$$

Обозначение: $\mathbf{pA}$.

Пример 18. Цены на билеты в кинотеатр зависят от того, когда начинается сеанс — утром, днем или вечером. Цена билета на утренний сеанс равна 10 руб., на дневной — 15 руб., а на вечерний — 20 руб.

Пусть матрица

$$\mathbf{A} = \begin{pmatrix} 50 & 60 & 45 & 60 \\ 50 & 90 & 80 & 70 \\ 105 & 100 & 95 & 110 \end{pmatrix}$$

описывает число посетителей кинотеатра в первые четыре дня (столбцы) показа нового кинофильма; строки отражают посещения утром, днем и вечером соответственно. Тогда выручку по дням можно рассчитать так:

$$\begin{pmatrix} 10 & 15 & 20 \end{pmatrix} \cdot \begin{pmatrix} 50 & 60 & 45 & 60 \\ 50 & 90 & 80 & 70 \\ 105 & 100 & 95 & 110 \end{pmatrix} = \begin{pmatrix} 3350 & 3950 & 3550 & 3850 \end{pmatrix}.$$

3.4.6. Собственные столбцы и собственные значения матрицы

Начнем с примера.

Пример 19. Умножим матрицу

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

на столбец

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

и на столбец

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Соответственно получим

$$\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}\begin{pmatrix}1\\ 2\end{pmatrix}=\begin{pmatrix}5\\ 4\end{pmatrix} \quad \text{и} \quad \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}=\begin{pmatrix}3\\ 3\end{pmatrix}.$$

Сравнивая результирующие столбцы с исходными, замечаем, что во 2-м случае (в отличие от 1-го) полученный столбец пропорционален заданному (коэффициент пропорциональности равен 3).

Пусть $\mathbf{A}$ — квадратная матрица порядка n. При умножении ее на столбец $\mathbf{x}$ высоты n получаем столбец $\mathbf{y}$ той же высоты:

$$\mathbf{A}\mathbf{x}=\mathbf{y}.$$

Поставим следующий вопрос: для всякой ли квадратной матрицы можно указать столбец, после умножения ее на который мы получим столбец, пропорциональный исходному, т. е. для всякой ли квадратной матрицы $\mathbf{A}$ существует столбец $\mathbf{x}$ и число λ такие, что

$$\mathbf{A}\mathbf{x}=\lambda\mathbf{x}. \tag{12}$$

Тривиальный случай $\mathbf{x}=\mathbf{O}$ отбросим сразу (в этом случае равенство (12) выполняется для любого λ).

Если такие столбец $\mathbf{x}\neq\mathbf{O}$ и число λ существуют, то они называются *собственным столбцом* и *собственным значением* матрицы $\mathbf{A}$.

Покажем, как практически можно ответить на поставленный выше вопрос, для простоты ограничившись подробным рассмотрением случая $n=2$.

Запишем соотношение (12) для матрицы

$$\mathbf{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

и столбца

$$\mathbf{x}=\begin{pmatrix}x_1\\ x_2\end{pmatrix}.$$

Имеем

$$\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\lambda\begin{pmatrix}x_1\\ x_2\end{pmatrix}.$$

Перемножая, получим

$$\begin{pmatrix}a_{11}x_1+a_{12}x_2\\ a_{21}x_1+a_{22}x_2\end{pmatrix}=\begin{pmatrix}\lambda x_1\\ \lambda x_2\end{pmatrix},$$

откуда
$$\begin{cases} a_{11}x_1 + a_{12}x_2 = \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 = \lambda x_2 \end{cases}$$
и, далее,
$$\begin{cases} (a_{11} - \lambda)x_1 + \qquad a_{12}x_2 = 0, \\ \qquad a_{21}x_1 + (a_{22} - \lambda)x_2 = 0. \end{cases}$$

Эта система двух уравнений с двумя неизвестными x_1 и x_2 имеет нулевое решение $x_1 = 0$, $x_2 = 0$ при любом λ. Но этот тривиальный случай нас не интересует. А вот нельзя ли выбрать параметр λ (который пока тоже неизвестен) так, чтобы эта система имела и другие решения?

Замечание. Уравнение вида
$$\alpha x_1 + \beta x_2 = 0$$
описывает прямую, проходящую через точку $O(0,0)$. Для того чтобы уравнение
$$\gamma x_1 + \delta x_2 = 0$$
описывало ту же прямую, должно выполняться условие
$$\frac{\alpha}{\gamma} = \frac{\beta}{\delta},$$
или
$$\alpha\delta = \beta\gamma$$
(в ином случае прямые будут пересекаться (рис. 36)).

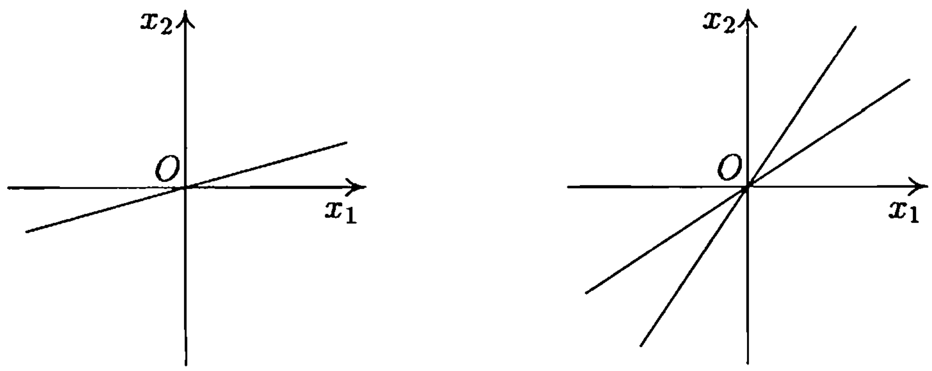


Рис. 36

Последнее соотношение в применении к рассматриваемому случаю будет выглядеть так:

$$(a_{11} - \lambda)(a_{22} - \lambda) = a_{12}a_{21}.$$

Вытекающее из него квадратное уравнение

$$\lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0$$

имеет не более двух корней.

Пример 20. Рассмотрим три конкретные матрицы:

$$1) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \qquad 2) \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad 3) \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

и попытаемся найти их собственные значения.

1) Переходя от матричного равенства

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

к системе уравнений, получим

$$\begin{cases} (1-\lambda)x_1 + \qquad 2x_2 = 0, \\ \qquad 2x_1 + (1-\lambda)x_2 = 0. \end{cases}$$

Отсюда

$$(1-\lambda)^2 = 2 \cdot 2$$

и, далее,

$$\lambda^2 - 2\lambda - 3 = 0.$$

Это уравнение имеет два корня:

$$\lambda_1 = -1 \quad \text{и} \quad \lambda_2 = 3,$$

которые и являются собственными значениями матрицы

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

2) Для матрицы

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

соответствующая система имеет вид

$$\begin{cases} (1-\lambda)x_1 + \quad\quad 2x_2 = 0, \\ \quad\quad\quad\quad (1-\lambda)x_2 = 0, \end{cases}$$

откуда

$$(1-\lambda)^2 = 0$$

и $\lambda = 1$ — единственное собственное значение матрицы

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

3) Система уравнений

$$\begin{cases} (1-\lambda)x_1 + \quad\quad 2x_2 = 0, \\ \quad -2x_1 + (1-\lambda)x_2 = 0 \end{cases}$$

приводит к равенству

$$(1-\lambda)^2 = -4,$$

которое легко пребразуется к квадратному уравнению, не имеющему корней.

Тем самым квадратная матрица второго порядка может иметь не более двух собственных значений.

Замечание. У квадратной матрицы n-го порядка не больше n собственных значений.

Итак, мы описали способ отыскания собственных значений матрицы. А как найти соответствующий собственный столбец в случае, если собственное значение уже найдено?

Продолжим рассмотрение примера 20. Ясно, что рассматривать нужно только первые два случая.

1) Подставив первое из найденных собственных значений в систему уравнений

$$\begin{cases} (1-\lambda)x_1 + \quad\quad 2x_2 = 0, \\ \quad\quad 2x_1 + (1-\lambda)x_2 = 0, \end{cases}$$

получим:

$$\begin{cases} 2x_1 + 2x_2 = 0, \\ 2x_1 + 2x_2 = 0. \end{cases}$$

Из первого уравнения (второе уравнение никакой новой информации не содержит) видно, что
$$x_1 = -x_2.$$
Положим $x_2 = 1$. Тогда $x_1 = -1$ и искомый собственный столбец имеет вид
$$\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Замечание. В качестве x_2 можно взять любое отличное от нуля число. Найдя по нему значение x_1, получим собственный столбец матрицы, пропорциональный предъявленному (отличающийся от предъявленного множителем).

Второе собственное значение $\lambda = 3$ приводит к системе уравнений
$$\begin{cases} -2x_1 + 2x_2 = 0, \\ \ \ 2x_1 - 2x_2 = 0, \end{cases}$$
откуда
$$x_1 = x_2$$
и
$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
— собственный столбец матрицы, отвечающий собственному значению $\lambda = 3$.

2) После подстановки $\lambda = 1$ в систему уравнений
$$\begin{cases} (1-\lambda)x_1 + \qquad 2x_2 = 0, \\ \qquad\qquad (1-\lambda)x_2 = 0 \end{cases}$$
получим
$$\begin{cases} 2x_2 = 0, \\ \ \ \ 0 = 0. \end{cases}$$
Это означает, что $x_2 = 0$, а x_1 может быть любым не равным нулю числом. Тем самым, собственный столбец матрицы имеет вид
$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Замечание. Описанный выше способ отыскания собственных значений матрицы и отвечающих им собственных столбцов состоит из двух основных этапов: сначала ищется собственное значение, а затем по нему строится собственный столбец. Однако в некоторых задачах возникает необходимость проверить, является ли заданный столбец

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

собственным столбцом матрицы

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

и если это так, то найти соответствующее собственное значение.

Конечно, это более простая задача, и она решается следующим образом.

Умножим матрицу $\mathbf{A}$ на столбец $\mathbf{x}$:

$$\mathbf{Ax} = \mathbf{y}$$

и поделим элементы полученного столбца $\mathbf{y}$ на соответствующие элементы столбца $\mathbf{x}$

$$\frac{y_1}{x_1}, \dots, \frac{y_n}{x_n}.$$

Если все эти отношения равны между собой,

$$\frac{y_1}{x_1} = \dots = \frac{y_n}{x_n} = \lambda,$$

то λ — искомое собственное значение. Если же хотя бы одно отношение

$$\frac{y_k}{x_k}$$

отлично от других, то $\mathbf{x}$ собственным столбцом матрицы $\mathbf{A}$ не является.

3.4.7. Неотрицательные и положительные матрицы

Матрица

$$\mathbf{A} = (a_{ik})$$

называется *неотрицательной*, если

$$a_{ik} \geqslant 0,$$
$$i = 1, \ldots, m; \quad k = 1, \ldots, n,$$

и *положительной*, если

$$a_{ik} > 0,$$
$$i = 1, \ldots, m; \quad k = 1, \ldots, n.$$

Пример 21. Матрица занятости в примере о составлении расписания неотрицательна, а матрица реализации порций мороженого положительна.

Обозначения: $\mathbf{A} \geqslant \mathbf{O}$, $\mathbf{A} > \mathbf{O}$.

Пусть $\mathbf{A}$ и $\mathbf{B}$ — матрицы одинаковых размеров.

Будем писать

$$\mathbf{A} > \mathbf{B},$$

если

$$a_{ik} > b_{ik},$$
$$i = 1, \ldots, m; \quad k = 1, \ldots, n,$$

и

$$\mathbf{A} \geqslant \mathbf{B},$$

если

$$a_{ik} \geqslant b_{ik},$$
$$i = 1, \ldots, m; \quad k = 1, \ldots, n.$$

ТЕОРЕМА. *Пусть* $\mathbf{A}$ *— неотрицательная квадратная матрица. Тогда у нее обязательно есть собственные значения, наибольшее из которых*

$$\lambda_{\mathrm{A}} = \lambda_{\max}$$

неотрицательно, и соответствующий ему собственный столбец также неотрицателен.

В случае когда $\mathbf{A}$ *— положительная квадратная матрица, ее наибольшее собственное значение* λ_{A} *положительно и положителен соответствующий ему собственный столбец.*

3.5. Задания и ответы

1. Небольшая фирма производит два вида продукции: столы и стулья. Для изготовления одного стула требуется 3 фута древесины, а для изготовления одного стола — 7 футов. На изготовление одного стула уходит 2 часа рабочего времени, а на изготовление стола — 8 часов. Каждый стул приносит 1 долл. прибыли, а каждый стол — 3 долл. Сколько стульев и сколько столов должна изготовить эта фирма, если она располагает 420 футами древесины и 400 часами рабочего времени и хочет получить максимальную прибыль?

Ответ: $z = x + 3y \to \max$,
$3x + 7y \leqslant 420, \quad 2x + 8y \leqslant 400, \quad x \geqslant 0, y \geqslant 0.$

$x = 56$, $y = 36$, $z_{\max} = 164$ долл.

2. Некоторая фирма выпускает два набора удобрений для газонов: обычный и улучшенный. В обычный набор входит 3 фунта азотных, 4 фунта фосфорных и 1 фунт калийных удобрений, а в улучшенный — 2 фунта азотных, 6 фунтов фосфорных и 3 фунта калийных удобрений. Известно, что для некоторого газона требуется по меньшей мере 10 фунтов азотных, 20 фунтов фосфорных и 7 фунтов калийных удобрений. Обычный набор стоит 3 долл., а улучшенный — 4 долл. Какие и сколько наборов удобрений нужно купить, чтобы обеспечить эффективное питание почвы и минимизировать стоимость?

Ответ: $z = 3x + 4y \to \min$,
$3x + 2y \geqslant 10, \quad 4x + 6y \geqslant 20, \quad x + 3y \geqslant 7, \quad x \geqslant 0, y \geqslant 0.$

$x = 2$, $y = 2$, $z_{\min} = 14$ долл.

3. На имеющихся у фермера 400 акрах земли он планирует посеять кукурузу и сою. Сев и уборка кукурузы требует на каждый акр 200 долл. затрат, а сои – 100 долл. На покрытие расходов, связанных с севом и уборкой, фермер получил ссуду в 60 тыс. долл. Каждый акр, засеянный кукурузой, приносит 30 бушелей, а каждый акр, засеянный соей, – 60 бушелей. Фермер заключил договор на продажу, по которому каждый бушель кукурузы принесет ему 3 долл., а каждый бушель сои — 6 долл. Однако, согласно этому договору, фермер обязан хранить убранное зерно в течение нескольких месяцев на складе, максимальная вместимость которого равна 21 тыс. бушелей.

Фермеру хотелось бы знать, сколько акров нужно засеять каждой из этих культур, с тем чтобы получить максимальную прибыль.

Ответ: $z = 3x + 6y \to \max$,
$x+y \leqslant 400$, $200x+100y \leqslant 60\,000$, $30x+60y \leqslant 21\,000$, $x \geqslant 0, y \geqslant 0$.
$x = 100$, $y = 300$, $z_{\max} = 2\,100$.

4. Решите задачу линейного программирования:
$z = 3x + 6y \to \min$,
$3x + 2y \leqslant 18$, $x + y \geqslant 5$, $x \leqslant 4$, $x \leqslant 7$, $x/y \geqslant 7/8$, $x \geqslant 0, y \geqslant 0$.

Ответ: $x = 4$, $y = 1$, $z_{\min} = 18$.

5. Решите задачу линейного программирования:
$z = 3x + 2y \to \max$,
$2x - 3y \leqslant 12$, $-x + 2y \leqslant 6$, $x \leqslant 6$, $2x + 5y \leqslant 20$, $x \geqslant 0, y \geqslant 0$.

Ответ: $x = 6$, $y = 1{,}6$, $z_{\max} = 21{,}2$.

6. Решите задачу линейного программирования:
$z = 2x + 2y \to \min$,
$x - y \geqslant 0$, $-3x + y \geqslant 3$, $x \geqslant 0, y \geqslant 0$.

Ответ: Решения не имеет (противоречивые условия).

7. Решите транспортную задачу, заданную таблицей

	B_1	B_2	Наличие
A_1	3	4	25
A_2	5	2	15
Запрос	20	20	

Ответ: $x_{11} = 20$, $x_{12} = 5$, $x_{21} = 0$, $x_{22} = 15$, $z_{\min} = 110$.

8. Решите транспортную задачу, заданную таблицей

	B_1	B_2	Наличие
A_1	5	2	20
A_2	4	3	15
Запрос	10	25	

Ответ: $x_{11} = 0$, $x_{12} = 20$, $x_{21} = 10$, $x_{22} = 5$, $z_{\min} = 95$.

9. Найдите решение задачи

$$z = 6x + 6y \to \min,$$

$$3x + 2y \geqslant 12,$$

$$3x + 8y \geqslant 24,$$

$$x \geqslant 2,$$

$$x \geqslant 0,$$

$$y \geqslant 0,$$

ограничиваясь целочисленными значениями переменных x и y.

Ответ: $x = 2$, $y = 3$, $z_{\min} = 30$ и $x = 3$, $y = 2$, $z_{\min} = 30$.

Глава 4
ФУНКЦИИ. ПРОИЗВОДНАЯ. ИНТЕГРАЛ

Материал, изложенный в этой главе, обладает двумя особенностями: во-первых, он существенно пересекается с элементами математического анализа, изучаемыми в средней школе, во-вторых, носит в достаточной степени вспомогательный характер: излагаемые понятия и методы применяются в последующих главах.

В силу этого читатель, знакомый с перечисленными в названии главы понятиями хотя бы в объеме средней школы, вполне может ограничиться беглым просмотром.

4.1. Примеры числовых функций

Понятие функциональной зависимости между величинами — одно из важнейших в математике вообще и в математическом моделировании в частности. Как правило, под функцией понимают отображение, которое каждому вещественному числу (точке) из *области определения* ставит в соответствие некоторое число, называемое *значением функции в точке.* В общем случае функциональную зависимость между величинами x и y записывают в виде соотношения:

$$y = f(x),$$

где x называют *аргументом*, а y — *значением функции.*

Наряду с *формульным* заданием часто применяется представление функциональной зависимости в виде *таблицы* либо *графика.* При табличном способе задания в явном виде выписывают значения аргумента и соответствующие им значения функции. При графическом способе на координатной плоскости (x, y) отмечают точки с координатами (x_0, y_0), образующие график функции (рис. 1).

У каждого из трех перечисленных способов есть свои достоинства и недостатки.

Функция может зависеть и от многих переменных. В главе о линейных моделях приведено линейное выражение, зависящее от переменных $x_1, x_2, \ldots, x_n$.

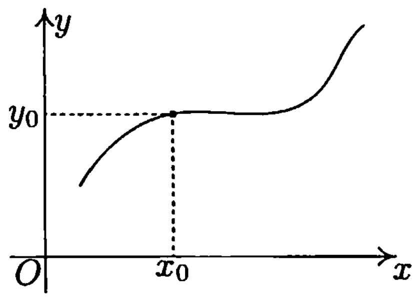


Рис. 1

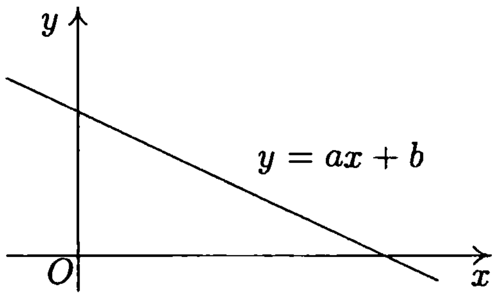


Рис. 2

Рассмотрим ряд элементарных числовых функций.

Линейная функция — это функция вида

$$y = ax + b,$$

где a и b — некоторые числа. Графиком линейной функции является прямая (рис. 2).

Замечание. Линейная функция, как простейший вид связи между величинами, наиболее широко применяется в моделировании. Некоторые примеры будут приведены в последующих главах.

Квадратичная функция — это функция вида

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

где a, b и c — некоторые числа, причем $a \neq 0$ (если $a = 0$, то функция не квадратичная, а линейная). Графиком квадратичной функции является парабола, ветви которой направлены вверх (при $a > 0$) или вниз (при $a < 0$) (рис. 3).

Замечание. Квадратичная функция — простейший вид зависимости между величинами, позволяющей моделировать процессы следующего типа: с ростом объема производства (x) доход (y) сначала увеличивается, затем уменьшается.

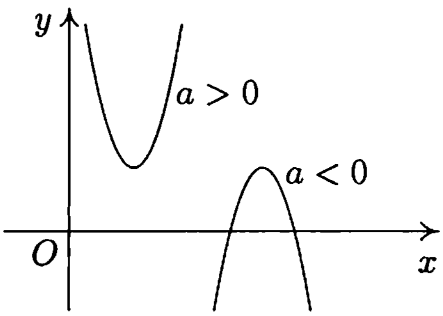


Рис. 3

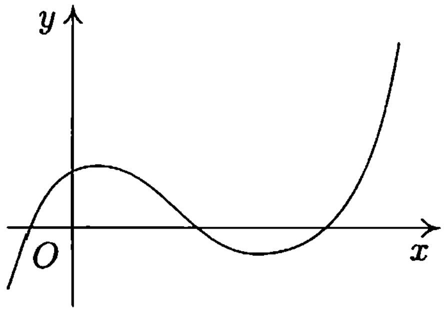


Рис. 4

Кубическая функция — это функция вида

$$y = ax^3 + bx^2 + cx + d,$$

где a, b, c и d — некоторые числа, причем $a \neq 0$. Графиком кубической функции является кубическая парабола (рис. 4).

Замечание. Пользователи графических редакторов (например, редактора Paint) для изображения различных кривых линий отмечают точки на экране, после чего компьютер сам проводит гладкую кривую. Обычно эта кривая подбирается программой как раз среди кубических парабол.

К простейшим элементарным функциям относится и функция

$$y = \frac{k}{x},$$

где $k \neq 0$. Графиком этой функции является гипербола (рис. 5).

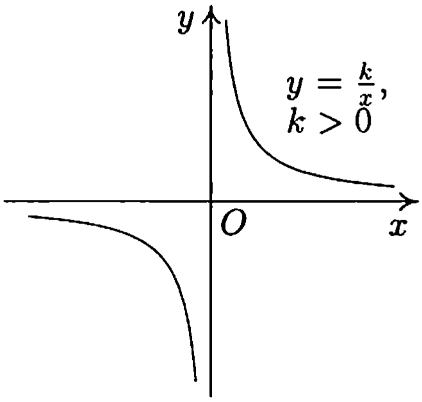


Рис. 5

Замечание. Функции такого вида (с $k > 0$) — простейший способ моделирования, например, зависимости спроса (y) от цены (x).

Функция может задаваться различным образом на отдельных участках области определения. Простейшим примером может служить функция $y = \operatorname{sgn} x$ (читается "сигнум x", т. е. знак числа x), определяемая следующим образом (рис. 6):

$$\operatorname{sgn} x = \begin{cases} -1, & \text{при } x < 0, \\ 0, & \text{при } x = 0, \\ 1, & \text{при } x > 0. \end{cases}$$

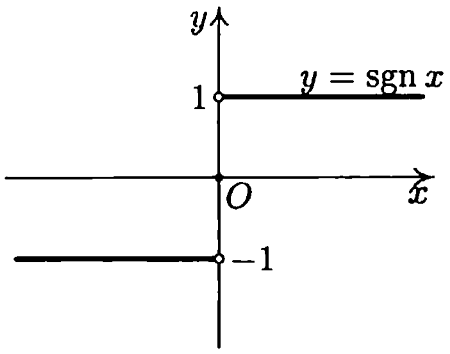


Рис. 6

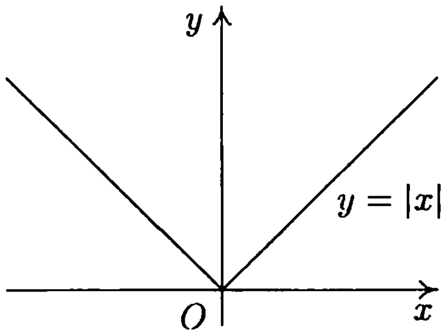


Рис. 7

Еще один пример — функция $y = |x|$ ("модуль x") (рис. 7):

$$|x| = \begin{cases} -x, & \text{при } x < 0; \\ x, & \text{при } x \geqslant 0. \end{cases}$$

Показательная функция задается уравнением

$$y = a^x, \qquad \text{где} \quad a > 0, a \neq 1$$

(см. рис. 8, 9).

Среди показательных функций выделяют *экспоненту* — функцию

$$y = e^x,$$

где $e \approx 2{,}718$.

Замечание. Быстро возрастающая показательная функция моделирует процессы типа роста народонаселения (y) с течением времени (x).

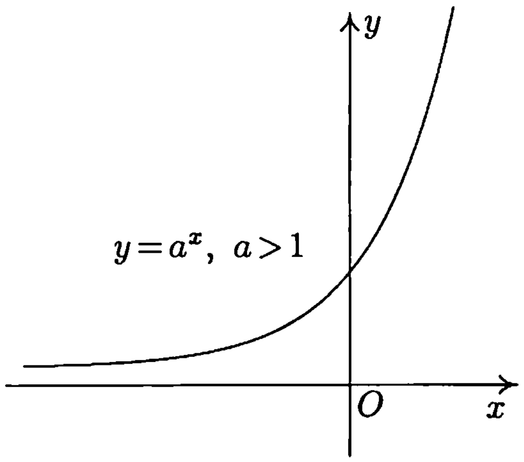


Рис. 8

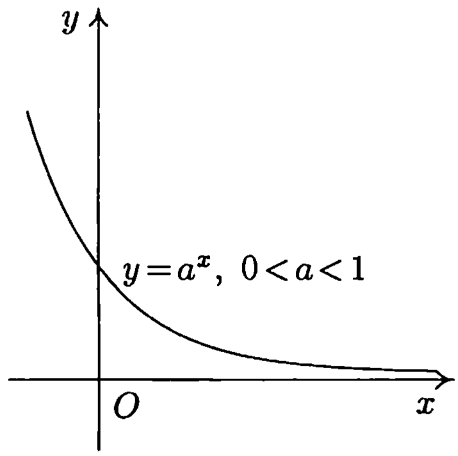


Рис. 9

Используемые в практических задачах функции зачастую имеют и более сложный вид, однако их основные свойства можно проанализировать и на простых примерах. В некоторых случаях простая функциональная зависимость бывает весьма выразительной.

Приведем несколько примеров.

Функции спроса на товары в зависимости от доходов потребителя (*функции Торнквиста* (Tornquist)) и их графики.

А. Спрос на товары первой необходимости:

$$y = \frac{ax}{x+c}, \quad x \geqslant 0$$

(в примерах А, Б, В x — доход, y — спрос, a, b и c — положительные постоянные) (рис. 10).

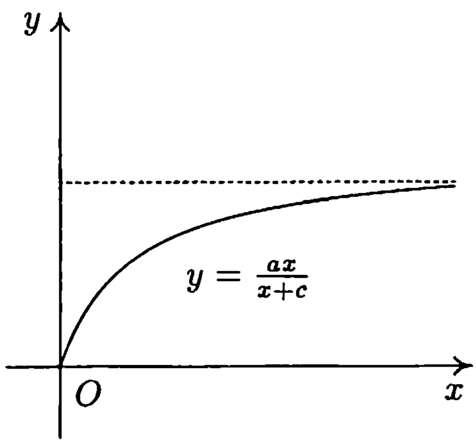


Рис. 10

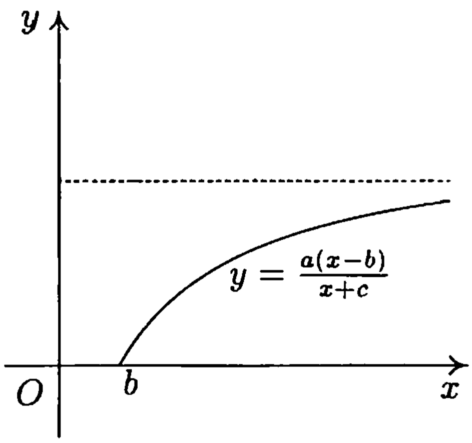


Рис. 11

Б. Спрос на товары относительной роскоши:

$$y = \frac{a(x-b)}{x+c}, \quad x \geqslant b$$

(рис. 11).

В. Спрос на предметы роскоши:

$$y = \frac{ax(x-b)}{x+c}, \quad x \geqslant b$$

(рис. 12).

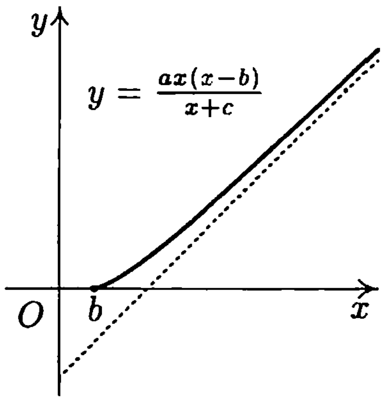


Рис. 12

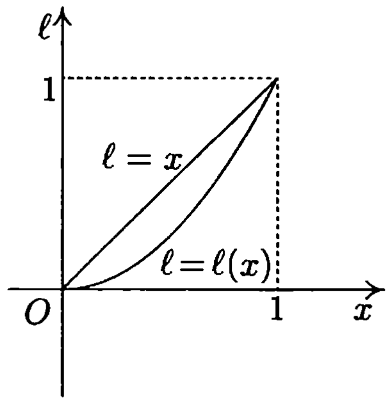


Рис. 13

Г. Функция распределения доходов в обществе (*функция Лоренца* (Lorentz)):

$$l = l(x)$$

(примерный график изображен на рис. 13). Здесь x — доля населения с наименьшим доходом, l — доля дохода. Чем больше отклоняется график функции Лоренца от прямой $l = x$, тем неравномернее доходы.

4.2. Простейшие свойства числовых функций

Важнейшим свойством функции является ее *непрерывность*. График непрерывной на отрезке функции можно начертить, не отрывая ручки от листа бумаги. Линейная и квадратичная функции непрерывны на любом отрезке. Функции $y = (k/x)$ и $\operatorname{sgn} x$ непрерывными на отрезке, содержащем точку $x = 0$, не являются — в этой точке обе функции *разрывны*.

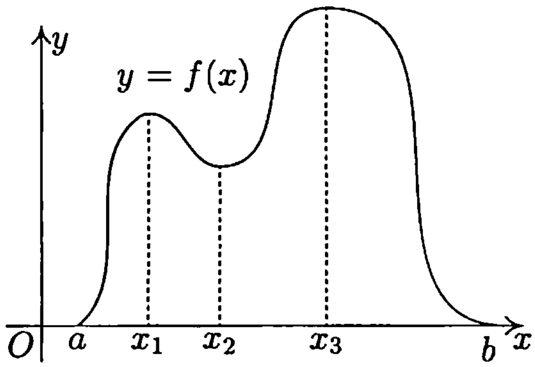


Рис. 14

Для рассмотрения дальнейших свойств обратимся к рис. 14, на котором изображен график непрерывной на отрезке $[a; b]$ функции $y = f(x)$. На отрезке $[a; x_1]$ эта функция *возрастает*: чем больше значение аргумента $\bar{x}$, тем больше и значение функции $f(\bar{x})$ (при $\bar{x}$, лежащем на отрезке $[a; x_1]$).

Можно сказать и так: функция $f(x)$ возрастает в каждой точке интервала $(a; x_1)$. Это означает, что какую бы точку $\bar{x}$ из этого интервала мы ни взяли, сдвинувшись от нее чуть-чуть вправо по числовой оси, мы получим значения функции большие, чем $f(\bar{x})$, а сдвинувшись чуть-чуть влево, — меньшие, чем $f(\bar{x})$.

На отрезке $[x_2; x_3]$ функция $f(x)$ также возрастает.

На отрезке $[x_1; x_2]$ функция $f(x)$ *убывает* — чем больше значение аргумента, тем меньше значение функции. Говоря более строго: для любых двух чисел u и v, $u > v$, принадлежащих отрезку $[x_1; x_2]$, имеет место неравенство

$$f(u) < f(v).$$

Точки x_1 и x_3 являются *точками (локального) максимума* функции $f(x)$. Точка x_2 — *точка (локального) минимума* функции $f(x)$. Точки локального максимума и локального минимума имеют общее название — точки локального *экстремума.*

В точке x_3 функция $f(x)$ принимает свое *наибольшее значение*, в точках a и b — *наименьшее значение.* Отметим следующий интуитивно ясный факт:

непрерывная на отрезке функция может принимать максимальное (минимальное) значение либо в точке локального максимума (минимума), либо на границе отрезка.

Замечание. Понятие экстремального значения функции является очень важным. Ведь при принятии того или иного решения мы заинтересованы в том, чтобы оно было наилучшим. Переводя сказанное на математический язык, мы стремимся найти точку, в которой некоторая функция принимает экстремальное значение. Разумеется, в реальной (а не модельной) ситуации свести задачу принятия решения к нахождению экстремума функции нелегко. Однако в некоторых случаях это удается сделать. Поэтому очень важно уметь находить экстремальные значения функций. С одним из основных способов нахождения экстремума мы познакомимся ниже.

4.3. Производная и экстремум

Наряду с возрастанием (убыванием) функции можно учитывать скорость этого возрастания (убывания). В одной точке эта скорость больше, в другой — меньше. Скорость возрастания функции в точке характеризуется числом, называемым *производной* функции в точке. Производная функции $y = f(x)$ в точке x_0 обозначается $f'(x_0)$ или $y'(x_0)$. Операция нахождения производной называется *дифференцированием.*

Коротко говоря, производная — это скорость возрастания функции. Перечислим некоторые свойства производной в предположении, что она существует.

Итак:

если

$$f'(x_0) > 0,$$

то функция $f(x)$ возрастает в точке x_0;

если

$$f'(x_0) < 0,$$

то функция $f(x)$ убывает в точке x_0;

если функция $f(x)$ имеет экстремум в точке x_0, то

$$f'(x_0) = 0.$$

Из последнего обстоятельства вытекает, что для нахождения экстремума важно знать точки, где производная равна нулю.

Для вычисления производной полезны следующие правила.

1. Производная постоянной равна нулю:

$$f(x) = c \qquad \Longrightarrow \qquad f'(x) = 0,$$

где c — произвольное фиксированное число.

2. Постоянная выносится за знак производной:

$$[cf(x)]' = cf'(x).$$

3. Производная суммы равна сумме производных:

$$[f(x) + g(x)]' = f'(x) + g'(x).$$

4. Для любого фиксированного n

$$(x^n)' = nx^{n-1}.$$

Примеры:

$$f(x) = 5 \quad \Longrightarrow \quad f'(x) = 0;$$

$$(x)' = 1 \qquad (\text{по правилу } 4,\ n = 1);$$

$$(3x^2)' = 6x \qquad (\text{по правилам } 2 \text{ и } 4);$$

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2} \qquad (\text{по правилу } 4,\ n = -1);$$

$$(x^3 - 3x^2 + 3)' = (x^3)' - 3(x^2)' = 3x^2 - 6x.$$

Производная в точке существует не всегда. Опишем три случая, когда она не существует.

Первый случай — когда функция не определена в данной точке. Например, функция $y = 1/x$ не имеет производной в точке $x_0 = 0$.

Второй случай — когда функция определена, но разрывна в точке. Например, функция $y = \operatorname{sgn} x$ не имеет производной в точке $x_0 = 0$.

Наконец, третий случай — когда график имеет в точке излом. Например, функция $y = |x|$ не имеет производной в точке $x_0 = 0$.

Опишем способ нахождения максимального и минимального значений непрерывной на отрезке функции. *Эти значения могут достигаться либо в точке, где производной не существует, либо в точке, где производная равна нулю, либо на границе отрезка.* При исследовании функции $f(x)$ на отрезке $[a; b]$ надо решить уравнение

$$f'(x) = 0,$$

а также найти точки, где производной не существует. Значения функции во всех этих точках и на границах отрезка включают в себя и максимальное, и минимальное значения.

Пример 1. Найти максимальное и минимальное значения функции $y = x^3 - x$ на отрезке $[-1; 2]$.

Решение. Эта функция дифференцируема (т. е. имеет производную) во всех точках. Найдем производную. Имеем:

$$y' = 3x^2 - 1.$$

Решая уравнение

$$y' = 0,$$

получаем два корня:

$$x = \sqrt{\frac{1}{3}} \approx 0,58, \quad x = -\sqrt{\frac{1}{3}} \approx -0,58.$$

Таким образом, экстремальные значения могут достигаться в одной из четырех точек: $x = 0{,}58$, $x = -0{,}58$, $x = -1$, $x = 2$. Вычисляя значения функции в этих точках, получаем

$$y(0,58) \approx -0{,}38; \quad y(-0{,}58) \approx 0{,}38; \quad y(-1) = 0; \quad y(2) = 6.$$

Таким образом, максимальное значение равно 6 (в точке $x = 2$), минимальное равно $-0{,}38$ (в точке $x = 0{,}58$) (рис. 15).

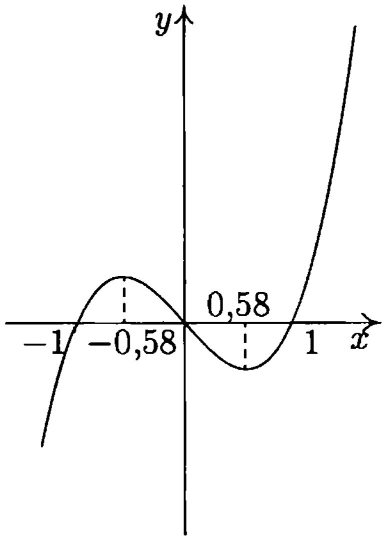


Рис. 15

Пример 2. Зависимость дохода I и издержек C от объема производства x задается функциями следующего вида:

$$I(x) = -2x^2 + 20x;$$
$$C(x) = x^3 - 35x^2 + 150x.$$

Производственные мощности позволяют производить до 25 единиц продукции. При каком объеме производства прибыль максимальна?

Решение. Прибыль, равная разности дохода и издержек, задается следующей функцией объема производства x:

$$f(x) = I(x) - C(x) = (-2x^2 + 20x) - (x^3 - 35x^2 + 150x) = \\ = -x^3 + 33x^2 - 120x.$$

Производственные мощности диктуют следующее ограничение на объем производства:

$$0 \leqslant x \leqslant 25.$$

Таким образом, задача свелась к нахождению максимального значения функции

$$f(x) = -x^3 + 33x^2 - 120x$$

на отрезке $[0; 25]$.

Найдем производную и приравняем ее к нулю. Имеем

$$f'(x) = -3x^2 + 66x - 120 = 0,$$

откуда получаем уравнение

$$x^2 - 22x + 40 = 0.$$

Корнями этого квадратного уравнения являются числа

$$x = 2 \quad \text{и} \quad x = 20.$$

Оба этих значения принадлежат отрезку $[0; 25]$. Поэтому для нахождения максимального значения функции $f(x)$ надо посчитать ее значения в четырех точках. Нетрудно убедиться, что

$$\begin{aligned} f(0) &= 0, \\ f(2) &= -116, \\ f(20) &= 2800, \\ f(25) &= 2000. \end{aligned}$$

Таким образом, прибыль максимальна (и равна 2800) при объеме производства 20 единиц.

Пример 3. Найти максимальное и минимальное значения функции

$$f(x) = \begin{cases} x + \dfrac{2}{x}, & \text{при } 1 \leqslant x < 2; \\ x + \dfrac{1}{x}, & \text{при } 2 \leqslant x \leqslant 3 \end{cases}$$

на отрезке [1;3].

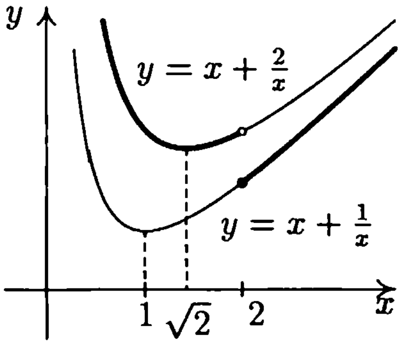


Рис. 16

Найдем производную функции $f(x)$ (ее график см. на рис. 16). В точке $x = 2$ производной не существует (функция разрывна). На интервале $(1; 2)$ имеем:

$$f'(x) = \left(x + \frac{2}{x}\right)' = 1 - \frac{2}{x^2}.$$

Решая уравнение

$$1 - \frac{2}{x^2} = 0,$$

получаем два корня: $x = \pm\sqrt{2} \approx \pm 1,41$. Интервалу (1;2) принадлежит лишь корень $x = 1,41$.

На интервале (2;3) имеем

$$f'(x) = \left(x + \frac{1}{x}\right)' = 1 - \frac{1}{x^2}.$$

Решая уравнение

$$1 - \frac{1}{x^2} = 0,$$

получаем $x = \pm 1$. Интервалу (2;3) не принадлежит ни один из найденных корней. Это означает, что на интервале (2;3) у функции нет локального экстремума.

Таким образом, максимальное и минимальное значения надо искать среди значений в точках $x = 1$, $x = 3$ (концы отрезка), $x = 2$ (точка, где не существует производной), $x = 1,41$ (точка нуля производной). Получаем

$$f(1) = 1 + \frac{2}{1} = 3;$$

$$f(3) = 3 + \frac{1}{3} \approx 3{,}33;$$

$$f(2) = 2 + \frac{1}{2} = 2{,}5;$$

$$f(1{,}41) = 1{,}41 + \frac{2}{1{,}41} \approx 2{,}83.$$

Максимальное значение равно 3,33 и достигается в точке $x = 3$, минимальное значение равно 2,5 и достигается в точке $x = 2$.

Замечание. У разрывной функции максимального либо минимального значения может и не существовать.

Пример 4. У функции

$$f(x) = \begin{cases} 1, & \text{при } x \leqslant 0, \\ x, & \text{при } x > 0 \end{cases}$$

(рис. 17) нет минимального значения на отрезке $[-1; 1]$.

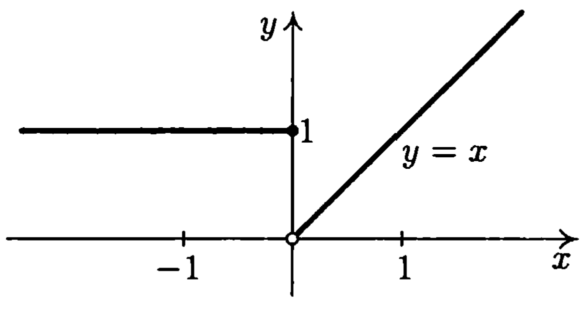


Рис. 17

4.4. Интеграл

Выше мы рассмотрели операцию дифференцирования, т. е. нахождения производной:

$$f(x) \quad \Longrightarrow \quad g(x) = f'(x).$$

Обратная ей операция называется *интегрированием* или нахождением *первообразной*. Первообразная функции $f(x)$ — это функция $F(x)$, производная которой равна $f(x)$:

$$F'(x) = f(x).$$

Если функция $F(x)$ является первообразной функции $f(x)$, то функция $F(x) + C$ (C — произвольное число) также является первообразной, поскольку

$$(F(x) + C)' = F'(x) = f(x).$$

Рассмотрим функцию $f(x)$, принимающую неотрицательные значения на отрезке $[a; b]$ (рис. 18).

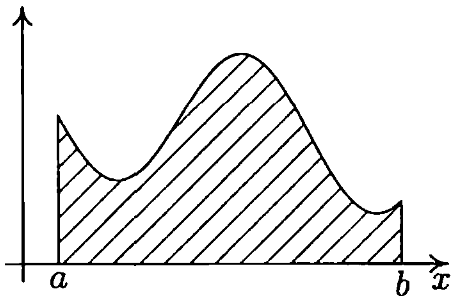


Рис. 18

Интеграл от функции $f(x)$ на отрезке $[a; b]$ обозначается так:

$$\int\limits_a^b f(x)\,dx.$$

Этот интеграл равен площади фигуры, ограниченной сверху графиком функции $y = f(x)$, снизу — прямой $y = 0$, слева и справа — прямыми $x = a$ и $x = b$ соответственно (на рис. 18 эта фигура заштрихована).

Справедлива следующая *формула Ньютона–Лейбница*:

$$\int_a^b f(x)\,dx = F(b) - F(a),$$

где $F(x)$ — любая первообразная функции $f(x)$.

Отметим два свойства интегралов:

$$\int_a^b [f(x) + g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx,$$

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Второе свойство позволяет вычислять интегралы от функций, задаваемых на разных отрезках различными формулами. Отметим, что эти функции не обязательно должны быть непрерывными.

В дальнейшем мы будем пользоваться следующим удобным обозначением:

$$F(b) - F(a) = F(x)|_a^b.$$

Замечание. Для функций, принимающих отрицательные значения, геометрическая интерпретация интеграла несколько сложнее. Однако формула Ньютона–Лейбница верна и в этом случае.

Пример 5. Вычислить интеграл

$$\int_1^2 x^2\,dx.$$

Решение. Поскольку

$$\left(\frac{1}{3}x^3\right)' = x^2,$$

то функция

$$y = \frac{1}{3}x^3$$

является первообразной функции $y = x^2$. По формуле Ньютона–Лейбница получаем:

$$\int\limits_1^2 x^2\,dx = \frac{1}{3}x^3\Big|_1^2 = \frac{1}{3}(2)^3 - \frac{1}{3}(1)^3 = \frac{1}{3}\cdot 8 - \frac{1}{3} = \frac{7}{3}.$$

Пример 6. Вычислить интеграл

$$\int\limits_0^2 f(x)\,dx,$$

где

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1, \\ -x^2+5, & 1 < x \leqslant 2. \end{cases}$$

Решение. Имеем

$$\int\limits_0^2 f(x)\,dx = \int\limits_0^1 f(x)\,dx + \int\limits_1^2 f(x)\,dx = \int\limits_0^1 1\,dx + \int\limits_1^2 (-x^2+5)\,dx =$$

$$= x|_0^1 + \left(-\frac{1}{3}x^3 + 5x\right)\Big|_1^2 = \frac{11}{3}.$$

Выше в качестве примера функциональной зависимости упоминалась функция Лоренца $l(x)$, характеризующая неравномерность распределения доходов в обществе. Степень этой неравномерности выражает *индекс Джини* (Gini) (рис. 19):

$$G.I. = \int\limits_0^1 (x - l(x))\,dx.$$

Пример 7. Пусть функция Лоренца имеет вид

$$l(x) = x^2.$$

Найти индекс Джини.

Решение. Имеем

$$G.I. = \int\limits_0^1 (x - x^2)\,dx = \left(\frac{1}{2}x^2 - \frac{1}{3}x^3\right)\Big|_0^1 = \frac{1}{6}.$$

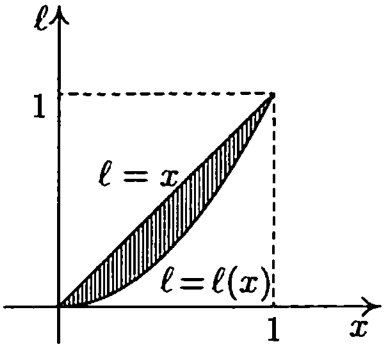


Рис. 19

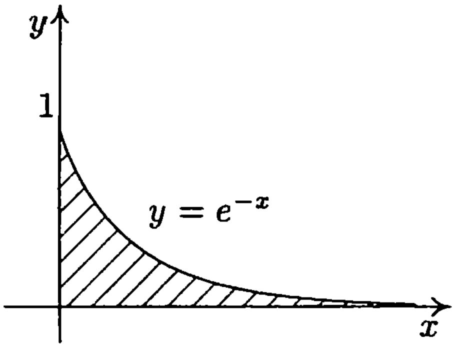


Рис. 20

Часто возникает необходимость вычисления так называемого *несобственного интеграла*, равного площади неограниченной области. Рассмотрим, например, функцию

$$y = e^{-x}$$

на полупрямой $[0; +\infty)$ (рис. 20). Напомним, что e — это специальная постоянная, равная примерно 2,718.

Площадь неограниченной фигуры, образованной осями координат и графиком функции $y = e^{-x}$ (на рис. 20 эта фигура заштрихована), можно вычислить, воспользовавшись несобственным интегралом

$$\int_0^{+\infty} e^{-x}\,dx = 1.$$

В дальнейшем мы познакомимся и с другими несобственными интегралами.

4.5. Задания и ответы

1. Найдите максимальное и минимальное значения функции $f(x)$ на отрезке $[a, b]$, где

а) $f(x) = 2x + 1, \quad a = 4,\ b = 6;$

б) $f(x) = x^2 - x, \quad a = 0,\ b = 2;$

в) $f(x) = 3x + \dfrac{2}{x}, \quad a = 1,\ b = 3;$

г) $f(x) = x^3 - 4x^2 + 2x, \quad a = -2, \ b = 0;$

д) $f(x) = 3x^3 - 8x, \quad a = -3, \ b = 3;$

е) $f(x) = \begin{cases} x^2 - 2, & \text{при } x < 0, \\ -x^2 + 2, & \text{при } x \geqslant 0, \end{cases} \quad a = -1, \ b = 3.$

Ответ: а) 13 и 9; б) 2 и $-0{,}25$; в) $9\frac{2}{3}$ и 5; г) 0 и -28; д) 57 и -57; е) 2 и -7.

2. Как изменится оптимальный объем производства в примере 2, если максимальная загрузка производственных мощностей позволяет производить 18 единиц продукции?

Ответ: оптимальный объем производства составит 18 единиц продукции.

3. Вычислите следующие интегралы:

а) $\displaystyle\int_1^3 (3x^2 + 6)\,dx;$

б) $\displaystyle\int_0^1 (2x^2 - 9x + 7)\,dx;$

в) $\displaystyle\int_{-6}^2 (x^3 + 2)\,dx;$

г) $\displaystyle\int_1^2 \left(x - \frac{1}{x^2}\right)dx;$

д) $\displaystyle\int_0^3 f(x)\,dx, \quad \text{где } f(x) = \begin{cases} x^2 + x + 1, & \text{при } x < 1, \\ 2x - 1, & \text{при } x \geqslant 1. \end{cases}$

Ответ: а) 38; б) $\frac{19}{6}$; в) -304; г) 1; д) $7\frac{5}{6}$.

4. Может ли индекс Джини принимать значение 0,6?

Ответ: нет.

Глава 5
БАЛАНСОВОЕ УРАВНЕНИЕ

В этой главе мы расскажем о том, как можно оценить экономическую эффективность планируемых капитальных вложений.

5.1. Сложные проценты

Проценты на капитал можно рассматривать как награду, которую получает *кредитор* от *заемщика* за пользование *капиталом*, принадлежащим кредитору.

Предположим, что заемщик кладет в банк, выплачивающий $p\%$ годовых (*процентную ставку*), некоторую сумму денег, которую мы для определенности обозначим через x_0.

Это означает, что ровно через год у заемщика на счету будет сумма, равная

$$x_1 = x_0 + x_0 \cdot \frac{p}{100} = x_0 \left(1 + \frac{p}{100}\right),$$

а еще через год —

$$x_2 = x_1 + x_1 \cdot \frac{p}{100} = x_1 \left(1 + \frac{p}{100}\right) = x_0 \left(1 + \frac{p}{100}\right)^2.$$

Терпеливый заемщик через k лет станет обладателем суммы, равной

$$x_k = x_0 \left(1 + \frac{p}{100}\right)^k.$$

Пример 1. При рождении внука дедушка положил в банк на 18 лет под 3% годовых некоторую сумму. К моменту совершеннолетия на счете оказалось 100 тыс. руб. Каким был первоначальный вклад?

По условию имеем:

$$x_{18} = x_0(1 + 0{,}03)^{18} = x_0 \cdot 1{,}03^{18},$$

где

$$x_{18} = 100\,000 \text{ руб.}$$

Отсюда получаем, что

$$x_0 = \frac{100\,000}{1{,}03^{18}} = 58\,739 \text{ руб.}$$

Тем самым, первоначальный вклад равен 58 тыс. 739 руб.

Из формулы

$$x_k = x_0 \left(1 + \frac{p}{100}\right)^k,$$

зная любые три из входящих в нее величин: x_0, x_k, p и k, легко найти четвертую величину. В частности,

$$x_0 = \left(1 + \frac{p}{100}\right)^{-k} x_k.$$

При этом говорят, что величина x_0 получена *дисконтированием* x_k (от английского *discount* — скидка, вычет), и называют ее *дисконтированным* значением x_k.

В примере 1 мы узнали размеры первоначального вклада дисконтированием $x_{18} = 100$ тыс. руб.

5.2. Погашение кредита

Пример 2. Предположим, что две стороны, кредитор и заемщик, договариваются о плане погашения кредитов:

кредит в 10 млн. руб. берется на 6 лет при годовой ставке 9% с условием, что через 3 года в счет погашения кредита будет внесено 4 млн. руб., через год — 2 млн. руб. и еще через год — 3 млн. руб.

Какая сумма должна быть внесена через 6 лет для полного погашения кредита?

Плату за кредит принято называть *процентом.*

На рис. 1 приведена схема выплаты процентов.

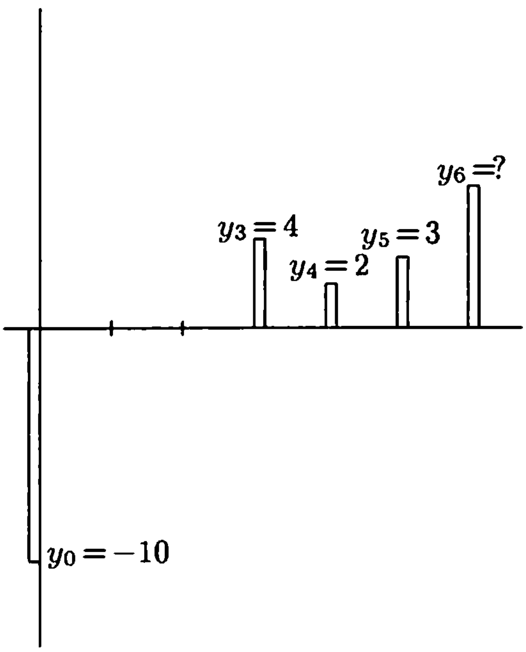


Рис. 1

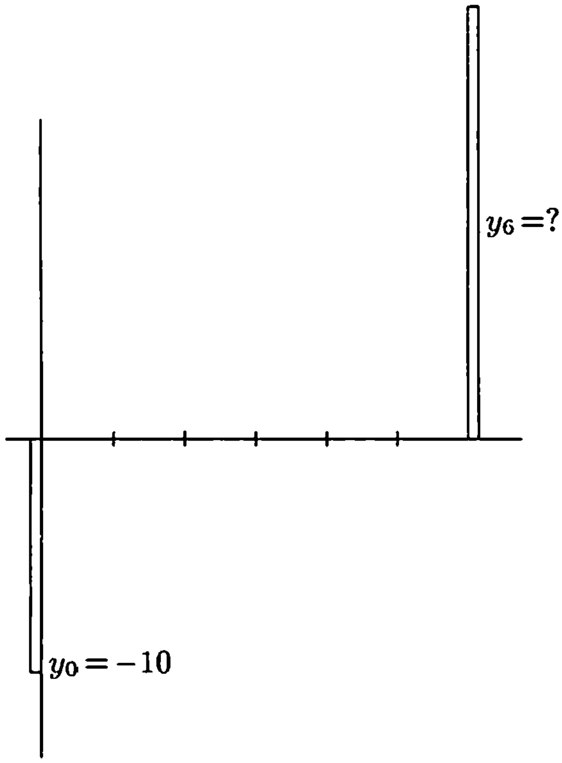


Рис. 2

Для того чтобы ответить на поставленный вопрос, нужно пересчитать все суммы, о которых идет речь, на какой-то определенный момент. Обычно это момент получения кредита.

Пересчитать суммы

$$y_0 = 10{,}0, \quad y_3 = 4{,}0, \quad y_4 = 2{,}0, \quad y_5 = 3{,}0, \quad y_6$$

означает дисконтировать их. В результате получим

$$10{,}0; \quad 4{,}0 \cdot 1{,}09^{-3}; \quad 2{,}0 \cdot 1{,}09^{-4}; \quad 3{,}0 \cdot 1{,}09^{-5}; \quad y_6 \cdot 1{,}09^{-6}.$$

Величина y_6 находится из условия погашения кредита:

$$10{,}0 = 4{,}0 \cdot 1{,}09^{-3} + 2{,}0 \cdot 1{,}09^{-4} + 3{,}0 \cdot 1{,}09^{-5} + y_6 \cdot 1{,}09^{-6},$$

откуда после несложных вычислений получаем, что

$$y_6 = 5\,944\,684 \text{ руб.}$$

Тем самым, при договоренном порядке погашения процент по кредиту равен 4 млн. 944 тыс. 684 руб., т. е. почти половине взятой суммы.

А вот другая схема погашения (рис. 2).

Здесь

$$y_1 = y_2 = y_3 = y_4 = y_5 = 0.$$

В этом случае

$$y_6 = 10{,}0 \cdot 1{,}09^6 = 16\,771\,000 \text{ руб.}$$

При таком порядке погашения процент по кредиту еще больше — 6 млн. 771 тыс. руб.

Весьма распространена практика выплаты в счет погашения кредита каждый год одной и той же суммы (рис. 3).

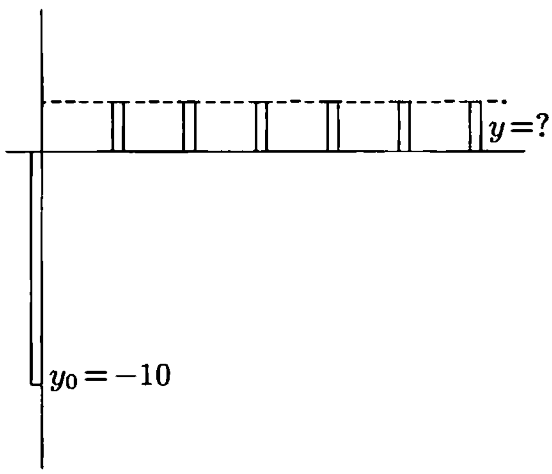


Рис. 3

Здесь

$$y_1 = y_2 = y_3 = y_4 = y_5 = y_6 = y.$$

Как ее вычислить при той же взятой сумме и той же процентной ставке?

Имеем:

$$10{,}0 = y(1{,}09^{-1} + 1{,}09^{-2} + 1{,}09^{-3} + 1{,}09^{-4} + 1{,}09^{-5} + 1{,}09^{-6}),$$

откуда

$$y = 2\,229\,200 \text{ руб.}$$

5.3. Балансовое равенство

Рассмотрим задачу посложнее.

Пример 3. Предположим, что заемщик берет кредит по частям у одного и того же кредитора под 9% годовых:

сразу — 10 млн. руб. ($y_0 = -10$),

через год — еще 8 млн. руб. ($y_1 = -8$)

и еще через два года — 5 млн. руб. ($y_3 = -5$),

а схема погашения кредита выглядит так:

$$y_5 = 4, \quad y_6 = 6, \quad y_7 = 9, \quad y_8 = 7, \quad y_9 = 5, \quad y_{10} = 7, \quad y_{11} = ?$$

(рис. 4).

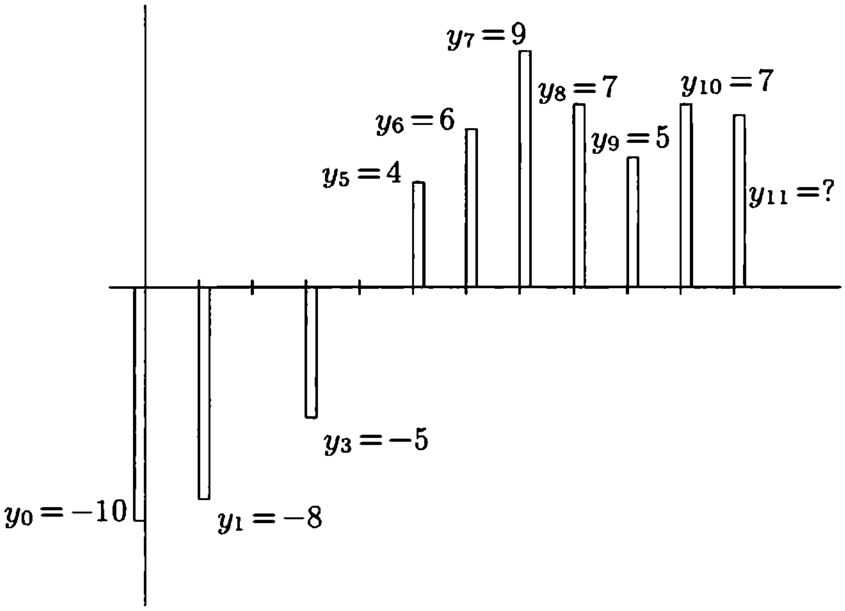


Рис. 4

Дисконтируя все суммы на момент выдачи первой части кредита и приравнивая суммы, соответствующие кредитам и погашениям, получаем:

$$10 + 8q^{-1} + 5q^{-3} = 4q^{-5} + 6q^{-6} + 9q^{-7} + 7q^{-8} + 5q^{-9} + 7q^{-10} + y_{11}q^{-11};$$

здесь

$$q = 1 + 0{,}01p = 1{,}09.$$

Перепишем последнее соотношение формально более подробно, имея в виду, что

$$y_2 = y_4 = 0.$$

В результате получим равенство

$$(-10)q^0 + (-8)q^{-1} + 0q^{-2} + (-5)q^{-3} + 0q^{-4} + 4q^{-5} + \\ + 6q^{-6} + 9q^{-7} + 7q^{-8} + 5q^{-9} + 7q^{-10} + y_{11}q^{-11} = 0.$$

После необходимых вычислений находим y_{11}:

$$y_{11} = 6{,}7 \text{ млн. руб.}$$

Подведем некоторые итоги.

Пусть y_k — величина взноса в конце k-го года, $k \geqslant 0$ (отрицательное значение y_k трактуется как кредит).

Все кредиты погашены за n лет, если имеет место *балансовое равенство*

$$y_0 + q^{-1}y_1 + \cdots + q^{-n}y_n = 0, \tag{1}$$

где

$$q = 1 + 0{,}01p,$$

т. е. сумма всех дисконтированных кредитов и взносов равна нулю.

При этом $p\%$ называют *нормой процента*, или *ставкой дисконта*, или *нормой дисконта*.

5.4. Балансовое уравнение

Метод дисконта можно использовать для оценки экономической эффективности вариантов капитальных вложений.

Пусть y_k — известные нам доходы предприятия за k-й год (отрицательное значение y_k трактуется как капитальное вложение) в проекте, рассчитанном на n лет.

Используем равенство (1)

$$y_0 + q^{-1}y_1 + \cdots + q^{-n}y_n = 0, \tag{2}$$

считая на этот раз, что величины $y_0, y_1, \ldots, y_n$ известны, а величина q (а значит, и p) подлежит определению.

Соотношение (2) при этих условиях называется *балансовым уравнением*.

Индексом прибыльности, или *внутренней нормой процента по капвложениям*, называется ставка дисконта, при которой сумма всех дисконтированных капитальных затрат и дисконтированных доходов равна нулю.

Обозначение: P.I. (сокращение от *profitability index*).

Тем самым, находя значение $q = q^*$, удовлетворяющее уравнению (2), мы определяем индекс прибыльности.

Обычная рыночная процентная ставка составляет примерно 8%.

Вложение считается выгодным, если

$$\text{P.I.} \geqslant 15\%.$$

Итак, пусть

$$y_0, \; y_1, \; \ldots, \; y_n$$

— обсуждаемый вариант капитальных затрат и ожидаемых доходов. Для того чтобы найти P.I., составляем балансовое уравнение

$$y_0 + q^{-1}y_1 + \cdots + q^{-n}y_n = 0.$$

Пусть $q = q^*$ — его решение. Тогда

$$\text{P.I.} = 100(q^* - 1)\%.$$

Пример 4. Обсуждаемый вариант капитальных затрат и ожидаемых доходов показан на рис. 5, т. е.

$$y_0 = -6, \quad y_1 = -8, \quad y_2 = -9, \quad y_3 = -4, \quad y_4 = 0, \quad y_5 = 6,$$
$$y_6 = 12, \quad y_7 = 15, \quad y_8 = 16, \quad y_9 = 13, \quad y_{10} = 4{,}79$$

и

$$-6 - 8r - 9r^2 - 4r^3 + 6r^5 + 12r^6 + 15r^7 + 16r^8 + 13r^9 + 4{,}79r^{10} = 0,$$

где

$$r = q^{-1} = (1 + 0{,}01p)^{-1}.$$

Из последнего соотношения получаем, что

$$r^* = 0{,}862 \quad \text{и} \quad q^* = \frac{1}{r^*} = 1{,}16.$$

Тем самым,

$$\text{P.I.} = 16\%.$$

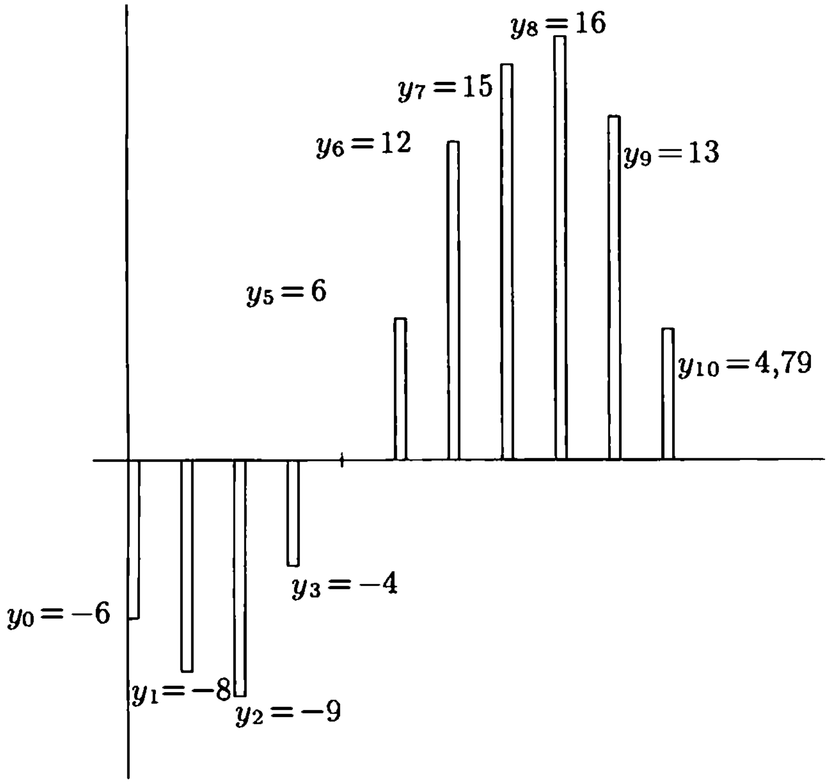


Рис. 5

5.5. Задания и ответы

1. В банк под $a\%$ годовых была положена некоторая сумма. Через n лет на счете оказалось b млн. руб. Каков размер положенной суммы?

Ответ: $\dfrac{b}{(1+0{,}01a)^n}$.

2. Условие задачи изображено на рис. 6:

$$y_0 = -(a+b), \quad y_3 = a, \quad y_5 = b, \quad p = 2\%, \quad n = 7; \quad y_7 = ?$$

Ответ: $y_7 = 1{,}02^4\,(1{,}02^3 - 1)\,a + 1{,}02^2\,(1{,}02^5 - 1)\,b$.

3. Условие задачи изображено на рис. 7:

$$y_0 = -(a+b), \quad p = 2\%, \quad n = 7; \quad y = ?$$

Ответ: $y = \dfrac{0{,}02 \cdot 1{,}02^7}{1{,}02^7 - 1}(a+b)$.

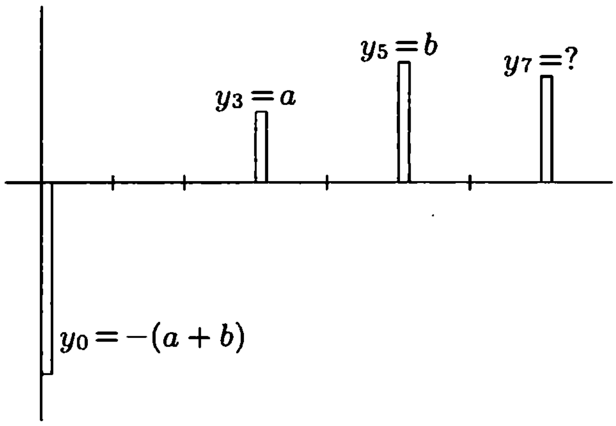


Рис. 6

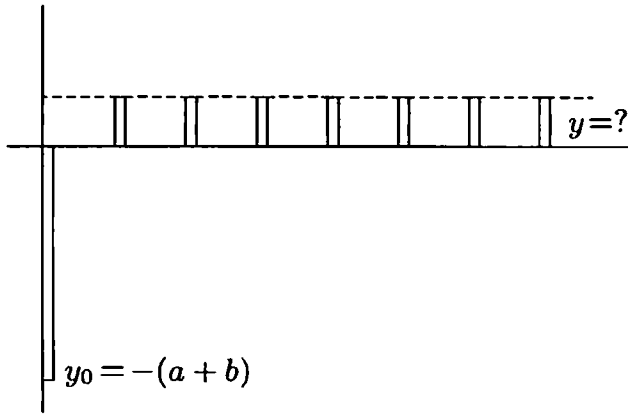


Рис. 7

4. Условие задачи изображено на рис. 8:

$$y_0 = -(a+b), \quad y_2 = 2a, \quad y_4 = 4b; \quad \text{P.I.} = ?$$

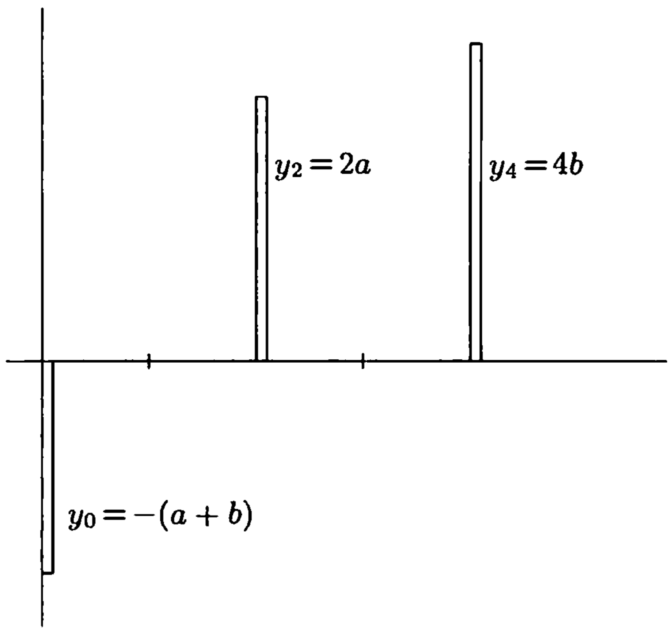


Рис. 8

Ответ: P.I. = 41%.

Глава 6
УПРАВЛЕНИЕ ЗАПАСАМИ

6.1. Вводные замечания

Фирмы часто делают различные запасы. Хранятся сырье, заготовки, готовая продукция, предназначенная для продажи.

Запасов не должно быть ни слишком много, ни слишком мало. В первом случае возникает необходимость неоправданных затрат на хранение, на амортизацию товара. Во втором случае может оказаться так, что на складе не будет нужного товара. Кроме того, малое количество запасов подразумевает их частое пополнение, что также требует затрат.

Задача управления запасами состоит в том, чтобы избежать обеих крайностей и сделать общие затраты по возможности меньше. Отметим, что в целом эта область науки управления развита довольно хорошо, разработаны многочисленные модели с применением различных математических методов. Мы рассмотрим несколько простейших детерминированных моделей управления запасами.

6.2. Основная модель

Важнейшую роль в наших рассмотрениях будет играть *функция изменения запаса*. Это связь между количеством единиц товара на складе (обозначим его через Q) и временем t. Будем считать, что имеется один вид товара.

Если на товар имеется спрос, то функция изменения запаса $Q = Q(t)$ убывает. Если товар, наоборот, завозят на склад, то эта функция возрастает. Мы будем считать возможным мгновенное пополнение запаса.

Затраты, связанные с запасами, можно разделить на три части.

А. *Стоимость товара.*

Б. *Организационные издержки.* Это расходы, связанные с оформлением товара, его доставкой, разгрузкой и т. д.

В. *Издержки на хранение товара.* Это затраты на аренду склада, амортизацию в процессе хранения и т. д.

Рассмотрим основные величины и предположения относительно них, принятые в рамках основной модели. Мы будем в основном использовать в качестве единицы измерения денежных средств условные единицы (УЕ), это могут быть рубли, доллары и т. п.; в качестве единицы измерения времени — год, хотя можно было бы взять месяц, квартал и т. п.

1. *Цена единицы товара* — c УЕ. Цена постоянна, рассматривается один вид товара.

2. *Интенсивность спроса* — d единиц товара в год. Будем считать, что спрос постоянный и непрерывный.

3. *Организационные издержки* — s УЕ за одну партию товара. Будем считать, что организационные издержки не зависят от размера поставки, т. е. от количества единиц товара в одной партии.

4. *Издержки на хранение запаса* — h УЕ на единицу товара в год. Будем считать эти издержки постоянными.

5. *Размер одной партии товара* постоянен — q единиц. Партия поступает мгновенно в тот момент, когда возникает дефицит, т. е. когда запас на складе становится равным нулю.

При сделанных предположениях график функции изменения запаса будет таким, как показано на рис. 1: он состоит из повторяющихся циклов пополнения запаса между двумя соседними дефицитами. Вертикальные отрезки отвечают мгновенному пополнению запаса.

Параметры c, d, s, h считаются заданными. Задача управления запасами состоит в выборе параметра q таким образом, чтобы минимизировать годовые затраты.

Для решения сформулированной задачи надо прежде всего выразить эти затраты через параметры c, d, s, h, q.

А. Поскольку годовая интенсивность спроса равна d, а цена единицы товара — c, то общая стоимость товара в год равна

$$cd.$$

Б. Поскольку в одной партии q единиц товара, а годовой спрос равен d, то число поставок равно d/q. В течение года организационные издержки равны

$$\frac{d}{q} \cdot s.$$

В. Средний уровень запаса равен отношению площади под графиком за цикл к продолжительности цикла. Этот средний уровень равен $q/2$ (на рис. 1 обозначен пунктиром). Поскольку годовые издержки на хранение единицы товара равны h, то общие издержки на хранение составляют

$$\frac{q}{2} \cdot h.$$

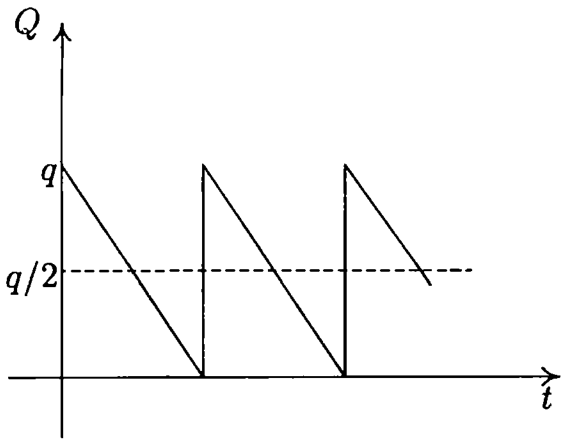


Рис. 1

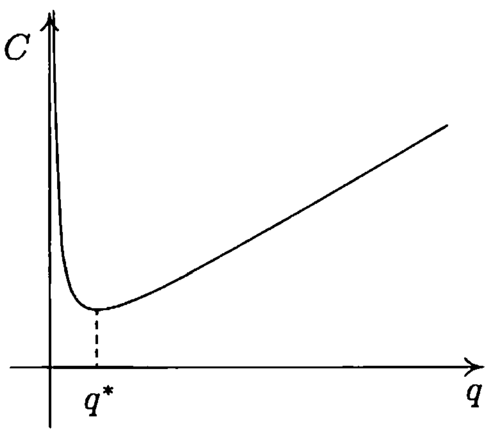


Рис. 2

Таким образом, общие издержки C вычисляются по формуле

$$C = cd + \frac{sd}{q} + \frac{qh}{2}.$$

Еще раз напомним, что в рамках модели параметры c, d, s, h считаются заданными и требуется найти такое число q^*, чтобы функция $C = C(q)$ принимала наименьшее значение на множестве $q > 0$ именно в точке q^*.

График функции $C = C(q)$ показан на рис. 2.

Для нахождения точки q^* минимума функции $C = C(q)$ найдем ее производную (c, d, s, h — фиксированные числа):

$$C'(q) = (cd)' + \left(\frac{sd}{q}\right)' + \left(\frac{qh}{2}\right)' = -\frac{sd}{q^2} + \frac{h}{2}.$$

Приравнивая $C'(q)$ к нулю, получаем

$$-\frac{sd}{q^2} + \frac{h}{2} = 0.$$

Отсюда можно найти q^*. Имеем:

$$q^* = \sqrt{\frac{2sd}{h}}.$$

Полученная формула называется *формулой оптимального запаса* или *формулой Харриса* (Harris).

Пример 1. Пусть интенсивность равномерного спроса составляет 1000 единиц товара в год. Организационные издержки равны 10 УЕ, издержки на хранение — 4 УЕ на единицу товара в год, цена товара — 5 УЕ.

Определить оптимальный размер партии в предположении, что система подчиняется основной модели.

Решение. Имеем:

$$d = 1000, \quad s = 10, \quad h = 4, \quad c = 5.$$

Общие затраты равны:

$$C(q) = cd + \frac{sd}{q} + \frac{qh}{2} = 5000 + \frac{10\,000}{q} + 2q.$$

Тогда

$$C'(q) = -\frac{10\,000}{q^2} + 2,$$

а оптимальный размер поставки q^* является решением уравнения

$$-\frac{10\,000}{q^2} + 2 = 0,$$

т. е. $q^* = \sqrt{5000} \approx 71$.

Замечание. Найдя оптимальный размер заказа, можно определить оптимальное число поставок за год n^* и соответствующую продолжительность цикла изменения запаса t^*:

$$n^* = \frac{d}{q^*} = \frac{1000}{71} \approx 14,$$

$$t^* = \frac{365}{n^*} = \frac{365}{14} \approx 26 \text{ дней}.$$

6.3. Модель производственных поставок

В основной модели предполагалось, что поступление товаров на склад происходит мгновенно. Это предположение достаточно хорошо отражает ситуацию, когда товар поставляется в течение одного дня (или ночи). Если товары поставляются с работающей производственной линии, необходимо модифицировать основную модель. В этом случае к параметрам c, d, s и h добавляется еще один — производительность производственной линии p (единиц товара в год). Будем считать ее заданной и постоянной.

Эта новая модель называется *моделью производственных поставок*. Величина q по-прежнему обозначает размер партии. В начале каждого цикла происходит "подключение" к производственной линии, которое продолжается до накопления q единиц товара. После этого пополнения запасов не происходит до тех пор, пока не возник дефицит.

График функции изменения запаса имеет вид, изображенный на рис. 3.

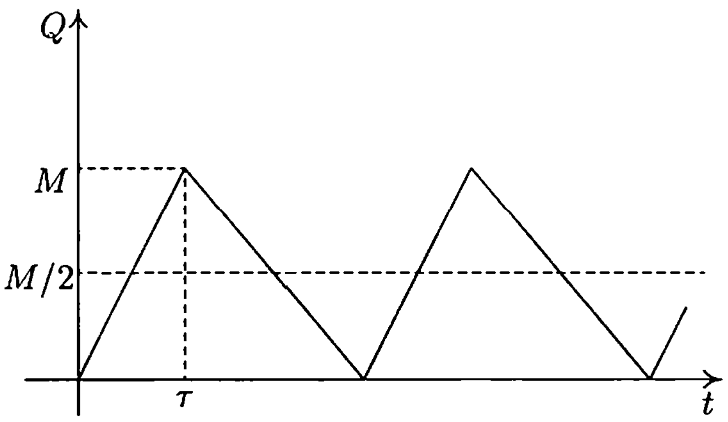


Рис. 3

Общие издержки $C(q)$, как и в основной модели, состоят из трех частей.

А. Общая стоимость товара в год равна

$$cd.$$

В. Годовые организационные издержки равны

$$\frac{sd}{q}.$$

В. Издержки на хранение вычисляются следующим образом. Пусть τ — время поставки (рис. 3). В течение этого времени происходит как пополнение (с интенсивностью p), так и расходование (с интенсивностью d) запаса. Увеличение запаса происходит со скоростью $p-d$. Поэтому достигнутый к концу периода пополнения запаса максимальный его уровень M вычисляется по формуле

$$M = (p-d)\tau$$

(заметим, что $M < q$). Однако,

$$p\tau = q$$

(за время τ при интенсивности производства p произведено q единиц товара). Из последних двух равенств следует, что

$$M = (p-d)\frac{q}{p}.$$

Средний уровень запаса, как и в основной модели, равен половине максимального, т. е. $M/2$. Таким образом, издержки на хранение запаса равны

$$\frac{(p-d)qh}{2p}.$$

Общие издержки вычисляются по формуле

$$C = cd + \frac{sd}{q} + \frac{(p-d)qh}{2p}.$$

Оптимальный размер поставок q^* получаем из уравнения

$$C'(q) = -\frac{sd}{q^2} + \frac{(p-d)h}{2p} = 0.$$

Имеем:

$$q^* = \sqrt{\frac{2psd}{(p-d)h}}.$$

Пример 2. Интенсивность равномерного спроса составляет 1 тыс. единиц товара в год. Товар поставляется с конвейера, производительность которого составляет 5 тыс. единиц в год. Организационные издержки равны 10 УЕ, издержки на хранение — 2 УЕ, цена единицы товара — 5 УЕ.

Чему равен оптимальный размер партии?

Решение. Имеем:

$$d = 1000, \quad p = 5000, \quad s = 10, \quad h = 2, \quad c = 5.$$

Далее,

$$C(q) = cd + \frac{sd}{q} + \frac{(p-d)qh}{2p} = 5000 + \frac{10\,000}{q} + \frac{4}{5}q,$$

$$C'(q) = -\frac{10\,000}{q^2} + \frac{4}{5}.$$

В итоге получаем

$$q^* = \sqrt{10\,000 \cdot (5/4)} \approx 112.$$

Замечание. Найдя оптимальный размер заказа, можно определить оптимальное число поставок за год n^* и соответствующие продолжительность поставки τ^* и продолжительность цикла пополнения запаса t^*:

$$n^* = \frac{d}{q^*} = \frac{1000}{112} \approx 9,$$

$$\tau^* = \frac{q^*}{p} = \frac{112}{5000} \cdot 365 \approx 8 \text{ дней},$$

$$t^* = \frac{365}{n^*} = \frac{365}{9} \approx 41 \text{ день}.$$

6.4. Модель поставок со скидкой

Рассмотрим ситуацию, описываемую в целом основной моделью, но с одной особенностью, которая состоит в том, что товар можно поставлять по льготной цене (со скидкой), если размер партии достаточно велик. Иными словами, если размер партии q не менее заданного числа q_0, товар поставляется по цене c_0, где $c_0 < c$.

Функция общих издержек $C(q)$ задается в таком случае следующим образом:

$$C(q) = \begin{cases} cd + \dfrac{sd}{q} + \dfrac{qh}{2}, & \text{если } q < q_0, \\[2ex] c_0 d + \dfrac{sd}{q} + \dfrac{qh}{2}, & \text{если } q \geqslant q_0. \end{cases}$$

Нетрудно видеть, что функция $C(q)$ в точке $q = q_0$ разрывна.

Обе функции

$$f(q) = cd + \frac{sd}{q} + \frac{qh}{2}$$

и

$$f_0(q) = c_0 d + \frac{sd}{q} + \frac{qh}{2}$$

имеют минимум в точке, где

$$f'(q) = f_0'(q) = 0,$$

т. е. в точке

$$\bar{q} = \sqrt{\frac{2sd}{h}}.$$

Для выяснения вопроса о том, какой размер партии оптимален, следует сравнить значения функции $C(q)$ в точках $\bar{q}$ и q_0, и та точка, где функция $C(q)$ принимает меньшее значение, будет оптимальным размером партии q^* в модели поставок со скидкой (см. рис. 4, 5).

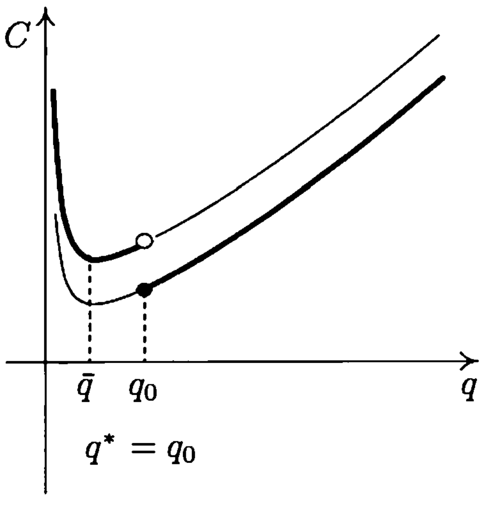


Рис. 4

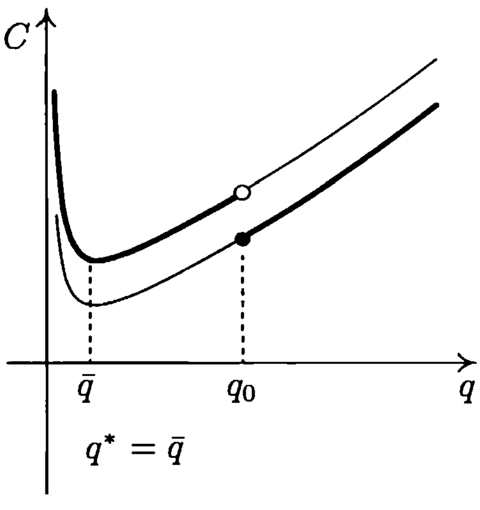


Рис. 5

Замечание. Может случиться так, что $C(\bar{q}) = C(q_0)$. Тогда в качестве q^* можно взять любое из чисел $\bar{q}$ и q_0.

Пример 3. Предположим, что интенсивность равномерного спроса составляет 1000 единиц товара в год. Организационные издержки равны 10 УЕ, издержки на хранение — 4 УЕ. Цена единицы товара равна 5 УЕ, однако, если размер партии не менее 500 единиц, цена снижается до 4 УЕ. Найти оптимальный размер партии.

Решение. Здесь

$$d = 1000, \quad s = 10, \quad h = 4, \quad c = 5, \quad q_0 = 500, \quad c_0 = 4.$$

Общие издержки определяются функцией $C(q)$:

$$C(q) = \begin{cases} f(q) = 5000 + \dfrac{10\,000}{q} + 2q, & \text{при } q < 500, \\ f_0(q) = 4000 + \dfrac{10\,000}{q} + 2q, & \text{при } q \geqslant 500. \end{cases}$$

Найдем точку локального минимума. Имеем:

$$f'(q) = f_0'(q) = -\frac{10\,000}{q^2} + 2 = 0,$$

откуда

$$\bar{q} = \sqrt{5000} \approx 71.$$

Поскольку $\bar{q} < 500$, то

$$C(\bar{q}) = f(\bar{q}) = f(71) \approx 5000 + \frac{10\,000}{71} + 2 \cdot 71 \approx 5283.$$

В точке $q = q_0$ получаем

$$C(q_0) = f_0(q_0) = f_0(500) \approx 4000 + \frac{10\,000}{500} + 2 \cdot 500 = 5020.$$

Таким образом,

$$q^* = 500.$$

6.5. Задания и ответы

1. Система управления запасами некоторого вида товара подчиняется условиям основной модели. Каждый год с постоянной интенсивностью поступает спрос на 15 тыс. единиц товара, издержки на организацию поставки составляют 10 долл. за одну партию, цена единицы товара — 3 долл., а издержки на ее хранение — 0,75 долл. в год. Найдите оптимальный размер партии.

Ответ: 632.

2. Каковы будут а) продолжительность цикла и б) число поставок за год, если стратегия управления запасами в предыдущей задаче является оптимальной?

Ответ: а) 15 дней; б) 23,7.

3. Система управления запасами описывается моделью производственных поставок и имеет следующие значения параметров. Спрос равен 1,5 тыс. единиц в год, цена 2 долл., издержки хранения единицы товара в течение года — 0,2 долл., организационные издержки — 10 долл. В течение года может быть произведено 4,5 тыс. единиц товара при полной загрузке производственной линии.

Нарисуйте график изменения запасов, вычислите оптимальный размер партии, продолжительность поставки, продолжительность цикла и средний уровень запасов.

Ответ: 474, 38 дней, 116 дней, 158.

4. Интенсивность спроса в модели производственных поставок составляет четверть скорости производства, которая равна 20 тыс. единиц товара в год. Организационные издержки для одной партии равны 150 долл., а издержки хранения единицы товара в течение года — 0,3 долл. Определите оптимальный размер партии.

Ответ: 2582.

5. Мебельной фирме требуется 1000 штук дверных ручек в год, расходуемых с постоянной интенсивностью. Организационные издержки составляют 30 долл. за партию, издержки на хранение одной ручки оценены в 1 долл. Цена дверной ручки составляет 2 долл., а при закупке партиями объемом не менее 750 штук — 1,9 долл. за штуку. Найдите оптимальный размер партии поставок.

Ответ: 245.

6. Торговец имеет стабильный спрос на некоторый товар в количестве 500 единиц в год. Товар он покупает у поставщика по цене 6 долл. за штуку, причем издержки на оформление поставки и другие подготовительные операции составляют в каждом случае 10 долл. Если торговец покупает сразу партию в количестве 150 единиц товара или более, цена сбавляется до 5 долл. за штуку. Каков оптимальный размер партии, если годовые затраты на хранение единицы товара равны 1 долл.?

Ответ: 150.

Глава 7
МОДЕЛЬ ЛЕОНТЬЕВА

В этой главе на ряде простых примеров мы покажем, как можно определить эффективность производства экономической системы по имеющейся количественной информации об объеме необходимых затрат, неизбежно сопровождающих всякое производство.

7.1. Продуктивные матрицы

Пусть имеется экономическая система, сфера производства которой состоит из n отраслей, выпускающих n видов продукта, причем каждая отрасль выпускает ровно один вид.

Предположим, что для производства k-й отраслью единицы k-го продукта требуется $a_{ik} \geqslant 0$ единиц i-го продукта, производимого i-й отраслью. Соответствующая таблица затрат выглядит так:

	1-й продукт	... ...	k-й продукт	... ...	n-й продукт
1-я отрасль	a_{11}	...	a_{1k}	...	a_{1n}
... ...	...	...	...	...	...
i-я отрасль	a_{i1}	...	a_{ik}	...	a_{in}
... ...	...	...	...	...	...
n-я отрасль	a_{n1}	...	a_{nk}	...	a_{nn}

или короче:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{1k} & \ldots & a_{1n} \\ \multicolumn{5}{c}{\dotfill} \\ a_{i1} & \ldots & a_{ik} & \ldots & a_{in} \\ \multicolumn{5}{c}{\dotfill} \\ a_{n1} & \ldots & a_{nk} & \ldots & a_{nn} \end{pmatrix}$$

Полученная неотрицательная матрица **A** называется *матрицей материальных затрат* или *технологической матрицей.*

Замечание. Матрица **A** дает информацию о сложившейся структуре межотраслевых связей, о существующей технологии общественного производства и используется в текущем и долгосрочном планировании.

Будем считать дополнительно, что сложившаяся технология неизменна (стационарна) и что производство линейно. Последнее означает, что если для выпуска единицы k-го продукта требуется a_{ik} единиц i-го продукта, то для выпуска x_k единиц k-го продукта необходимо $a_{ik}x_k$ единиц i-го продукта.

Предположим, что за некоторый отрезок времени, фиксированный во всех дальнейших рассмотрениях (неделя, месяц, квартал или год), выпущено x_1 единиц 1-го продукта, x_2 единиц 2-го продукта, ..., x_n единиц n-го продукта.

Тем самым, задан столбец

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix},$$

называемый *столбцом выпуска* или *режимом работы отраслей.*

При заданном столбце выпуска **x** совокупные затраты i-го продукта в рассматриваемой производственной сфере равны

$$\sum_{k=1}^{n} a_{ik}x_k, \qquad i = 1, \dots, n.$$

Из этих величин составляется *столбец совокупных материальных затрат* в сфере производства:

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} \sum\limits_{k=1}^{n} a_{1k}x_k \\ \dots \\ \sum\limits_{k=1}^{n} a_{nk}x_k \end{pmatrix}.$$

Матрица материальных затрат $\mathbf{A} \geqslant \mathbf{O}$ называется *продуктивной,* если найдется такой столбец выпуска $\mathbf{x} > \mathbf{O}$, для которого выполняется неравенство

$$\mathbf{A}\mathbf{x} < \mathbf{x}.$$

Это неравенство означает, что существует хотя бы один режим работы отраслей данной экономической системы, при котором каждого продукта выпускается больше, чем затрачивается на его производство. Другими словами, при этом режиме сфера производства создает положительный столбец прибавочного (конечного) продукта:

$$\mathbf{x} - \mathbf{A}\mathbf{x} > \mathbf{O}.$$

Возникает естественный вопрос: как следует поступить, чтобы сравнительно несложным путем и как можно раньше выяснить, является ли предъявленная матрица материальных затрат исследуемой сферы производства продуктивной или, напротив, производство убыточно и совокупные материальные затраты превышают объем выпуска?

Справедлив следующий общий факт.

ТЕОРЕМА. *Для любой неотрицательной квадратной матрицы* $\mathbf{A} \geqslant \mathbf{O}$ *формулируемые ниже условия равносильны.*

(1) *Матрица* $\mathbf{A}$ *продуктивна.*

(2) *Для любого столбца* $\mathbf{c} > \mathbf{O}$ *существует, и притом ровно один, столбец выпуска* $\mathbf{x} > \mathbf{O}$ *такой, что*

$$\mathbf{x} - \mathbf{A}\mathbf{x} = \mathbf{c}.$$

(3) *Столбца выпуска* $\mathbf{x} > \mathbf{O}$, *совокупные затраты на создание которого удовлетворяют условию*

$$\mathbf{A}\mathbf{x} \geqslant \mathbf{x},$$

не существует.

(4) *Наибольшее собственное значение матрицы* $\mathbf{A}$ *удовлетворяет неравенству*

$$\lambda_{\mathrm{A}} = \lambda_{\max} < 1.$$

Сказанное выше означает, что при выполнении хотя бы одного из этих условий выполняются и три остальных. В частности, выполнение неравенства

$$\lambda_{\mathrm{A}} < 1.$$

позволяет утверждать, что матрица продуктивна.

В приводимых ниже примерах мы ограничимся рассмотрением случая, когда $n = 2$, т. е. сфера производства экономической системы состоит из двух отраслей.

Пример 1. Для ответа на вопрос, является ли матрица

$$\mathbf{A} = \begin{pmatrix} 1/3 & 1/2 \\ 1/12 & 1/4 \end{pmatrix}$$

продуктивной, найдем ее собственные значения.

Имеем:

$$\left(\frac{1}{3} - \lambda\right)\left(\frac{1}{4} - \lambda\right) = \frac{1}{24},$$

откуда

$$\lambda^2 - \frac{7}{12}\lambda + \frac{1}{24} = 0.$$

Корни этого уравнения легко вычисляются по формуле

$$\lambda_{1,2} = \frac{7}{24} \pm \sqrt{\left(\frac{7}{24}\right)^2 - \frac{1}{24}} = \frac{7}{24} \pm \frac{5}{24}$$

и окончательно

$$\lambda_1 = \frac{1}{12}, \qquad \lambda_2 = \frac{1}{2}.$$

Тем самым,

$$\lambda_{\max} = \frac{1}{2} < 1.$$

Ответ: матрица $\mathbf{A}$ продуктивна.

Из той же теоремы вытекает, что если матрица материальных затрат $\mathbf{A}$ продуктивна, то любой столбец прибавочного продукта может быть произведен при соответствующем режиме работы отраслей.

Итак, пусть матрица

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

продуктивна и

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

— столбец конечного продукта. Покажем, как найти режим работы отраслей, обеспечивающий этот продукт.

Запишем матричное равенство

$$\mathbf{x} - \mathbf{A}\mathbf{x} = \mathbf{c}$$

более подробно:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

После перемножения:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

и вычитания:

$$\begin{pmatrix} x_1 - a_{11}x_1 - a_{12}x_2 \\ x_2 - a_{21}x_1 - a_{22}x_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

окончательно получим

$$\begin{cases} (1 - a_{11})x_1 - \qquad a_{12}x_2 = c_1, \\ \qquad -a_{21}x_1 + (1 - a_{22})x_2 = c_2. \end{cases}$$

Для продуктивной матрицы построенная система имеет решение при любых c_1 и c_2.

Рассмотрим конкретный пример.

Пример 2. Пусть

$$\mathbf{A} = \begin{pmatrix} 1/3 & 1/2 \\ 1/12 & 1/4 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}.$$

Как было установлено в примере 1, матрица $\mathbf{A}$ продуктивна, $\lambda_{\max} = \dfrac{1}{2} < 1$, и потому система

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1/3 & 1/2 \\ 1/12 & 1/4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

имеет решение (всегда совместна).

После простых преобразований получаем:

$$\begin{cases} \quad \dfrac{2}{3}x_1 - \dfrac{1}{2}x_2 = 4, \\ -\dfrac{1}{12}x_1 + \dfrac{3}{4}x_2 = 5. \end{cases}$$

Найдем решение этой системы методом исключения неизвестной.

Умножая первое уравнение на 3/2 и складывая со вторым, получим

$$\left(1 - \frac{1}{12}\right) x_1 = 11, \qquad \frac{11}{12} x_1 = 11$$

и далее

$$x_1 = 12.$$

Подобным же образом, умножая первое уравнение на $\frac{1}{8}$ и складывая со вторым, находим значение второй неизвестной. Имеем:

$$\left(-\frac{1}{16} + \frac{3}{4}\right) x_2 = \frac{11}{2}, \qquad \frac{11}{16} x_2 = \frac{11}{2}.$$

Отсюда

$$x_2 = 8.$$

Таким образом, для того чтобы обеспечить прибавочный продукт

$$\begin{pmatrix} 4 \\ 5 \end{pmatrix},$$

необходимо, чтобы столбец выпуска был равен

$$\begin{pmatrix} 12 \\ 8 \end{pmatrix}.$$

7.2. Ограничения на ресурсы

Модель Леонтьева отражает те потенциальные возможности, которые заложены в технологии производственного сектора. В этой модели предполагается, что все промежуточные продукты к тому моменту, когда они оказываются необходимыми, уже произведены. Однако в реальной ситуации нужно принимать в расчет наличие таких ограничительных факторов производства, как мощность каждой отрасли (материальные ресурсы) и общее количество рабочей силы в системе (трудовые ресурсы).

Пусть L — общее число рабочих и

$$\mathbf{l} = \begin{pmatrix} l_1 & l_2 & \ldots & l_n \end{pmatrix}$$

— *матрица-строка затрат рабочей силы*: каждый ее элемент $l_k > 0$ показывает количество рабочих, необходимое для производства единицы k-го продукта.

В предположении линейности производства произведение

$$\mathbf{lx} = (l_1 \quad l_2 \quad \dots \quad l_n) \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \sum_{k=1}^{n} l_k x_k$$

показывает количество рабочей силы, необходимое в сфере производства при режиме работы $\mathbf{x}$.

Ясно, что оно не может превосходить общего числа рабочих

$$\mathbf{lx} \leqslant L.$$

Ограничения на мощности отраслей можно описать при помощи столбца

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \\ \dots \\ m_n \end{pmatrix},$$

превзойти который столбец выпуска не может,

$$\mathbf{x} \leqslant \mathbf{m}.$$

При ограниченных ресурсах уже нельзя ставить вопрос об удовлетворении любого конечного спроса $\mathbf{c} > \mathbf{O}$. Тем не менее продуктивная система может обеспечить любую структуру прибавочного продукта, т. е. соотношение между количеством прибавочных продуктов первой и второй отраслей.

ТЕОРЕМА. *Пусть дана продуктивная матрица* $\mathbf{A} > \mathbf{O}$, *столбцы* $\mathbf{c} > \mathbf{O}$ *и* $\mathbf{m} > \mathbf{O}$, *строка* $\mathbf{l} > \mathbf{O}$ *и число* $L > 0$. *Тогда задача*

$$\begin{cases} \mathbf{x} - \mathbf{Ax} & \leqslant \alpha\mathbf{c}, \\ \mathbf{lx} & \leqslant L, \\ \mathbf{x} & \leqslant \mathbf{m}, \\ \mathbf{x} & \geqslant \mathbf{O}, \end{cases} \qquad (1)$$

$$\alpha \to \max$$

имеет, и притом ровно одно, решение.

Рассмотрим на конкретном примере, как можно решать такую задачу.

Пример 3. Итак, даны

$$\mathbf{A} = \begin{pmatrix} 1/3 & 1/2 \\ 1/12 & 1/4 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \quad \mathbf{m} = \begin{pmatrix} 4 \\ 3 \end{pmatrix},$$

$$\mathbf{l} = (4 \quad 4), \quad L = 40.$$

Начнем с решения системы

$$\mathbf{x} - \mathbf{A}\mathbf{x} = \alpha\mathbf{c}$$

или подробнее:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1/3 & 1/2 \\ 1/12 & 1/4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \alpha \begin{pmatrix} 4 \\ 5 \end{pmatrix}.$$

Это можно записать в равносильной форме:

$$\begin{cases} \quad \dfrac{2}{3}x_1 - \dfrac{1}{2}x_2 = 4\alpha, \\ -\dfrac{1}{12}x_1 + \dfrac{3}{4}x_2 = 5\alpha, \end{cases}$$

откуда

$$x_1 = 12\alpha, \qquad x_2 = 8\alpha,$$

или

$$\mathbf{x} = \begin{pmatrix} 12\alpha \\ 8\alpha \end{pmatrix}.$$

Полученный столбец должен подчиняться условиям

$$\mathbf{l}\mathbf{x} \leqslant L \quad \text{и} \quad \mathbf{x} \leqslant \mathbf{m},$$

которые в данном случае принимают вид:

$$(4 \quad 4) \begin{pmatrix} 12\alpha \\ 8\alpha \end{pmatrix} \leqslant 40, \qquad \begin{pmatrix} 12\alpha \\ 8\alpha \end{pmatrix} \leqslant \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

Отсюда имеем:

$$\begin{cases} 80\alpha \leqslant 40, \\ 12\alpha \leqslant 4, \\ \ \ 8\alpha \leqslant 3 \end{cases} \qquad \text{или} \qquad \begin{cases} \alpha \leqslant 1/2, \\ \alpha \leqslant 1/3, \\ \alpha \leqslant 3/8. \end{cases}$$

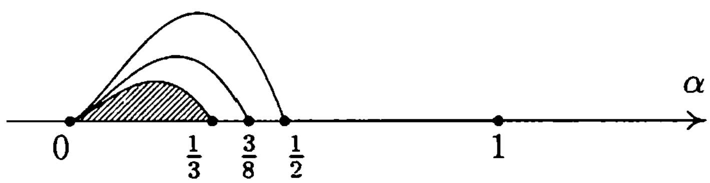


Рис. 1

Наибольшее значение α, удовлетворяющее всем трем условиям, равно $1/3$ (рис. 1).

Ответ: $\alpha_{\max} = 1/3$, столбец выпуска

$$\mathbf{x} = \begin{pmatrix} 4 \\ 8/3 \end{pmatrix},$$

конечный продукт

$$\alpha_{\max}\mathbf{c} = \begin{pmatrix} 4/3 \\ 5/3 \end{pmatrix}.$$

Замечание 1. Соотношение между количеством первого и количеством второго прибавочного продукта $4:5$ — то же, что и в случае отсутствия каких-либо ограничений на материальные и трудовые ресурсы.

Замечание 2. При $n = 2$ соотношения (1) принимают вид:

$$\begin{cases} (1 - a_{11})x_1 - \qquad a_{12}x_2 = \alpha c_1, \\ \quad - a_{21}x_1 + (1 - a_{22})x_2 = \alpha c_2, \end{cases}$$

$$\begin{cases} l_1x_1 + l_2x_2 \leqslant L, \\ \qquad x_1 \leqslant m_1, \\ \qquad x_2 \leqslant m_2, \end{cases} \tag{2}$$

$$\alpha \to \max$$

Решение системы уравнений можно записать так:

$$x_1 = \alpha b_1, \qquad x_2 = \alpha b_2,$$

где b_1 и b_2 выражаются через элементы матрицы $\mathbf{A}$ и столбца $\mathbf{c}$.

Отсюда получаем

$$\frac{x_1}{b_1} = \frac{x_2}{b_2} = \alpha$$

или, исключая α,

$$b_2 x_1 = b_1 x_2.$$

Полученное равенство на плоскости (x_1, x_2) описывает прямую, проходящую через начальную точку $O(0, 0)$.

В свою очередь, неравенства (2) можно проиллюстрировать так, как показано на рис. 2.

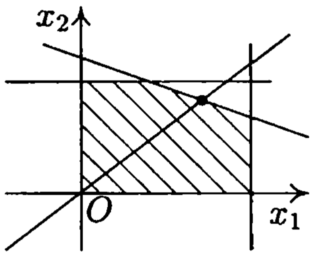


Рис. 2

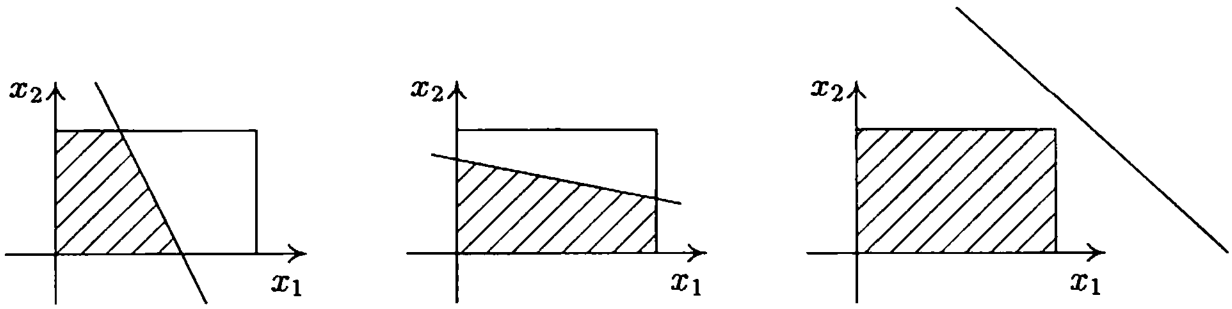


Рис. 3

На рис. 3 представлены все возможные случаи.

Замечание. На рис. 2 жирная точка отвечает $\alpha_{\max}$.

7.3. Прибыльные матрицы

Предположим теперь, что отрасли закупают на внутреннем рынке системы (друг у друга) продукты, которые необходимы им как средства производства.

Пусть $p_i > 0$ — цена единицы i-го продукта. Строка

$$\mathrm{p} = \begin{pmatrix} p_1 & p_2 & \ldots & p_n \end{pmatrix},$$

каждый элемент которой является ценой единицы соответствующего продукта, производимого системой, называется *строкой цен* на продукты или *ценовой строкой*.

Пусть

$$\mathbf{A} = (a_{ik})$$

— матрица материальных затрат системы. Тогда денежные издержки производства единицы k-го продукта будут равны

$$\sum_{i=1}^{n} p_i a_{ik}, \qquad k = 1, \ldots, n.$$

Из этих величин и складывается матрица-строка $\mathbf{pA}$ издержек производства:

$$\mathbf{pA} = \left(\sum_{i=1}^{n} p_i a_{i1} \quad \ldots \quad \sum_{i=1}^{n} p_i a_{in}\right).$$

Квадратная матрица $\mathbf{A} \geqslant \mathbf{O}$ называется *прибыльной*, если существует такая строка $\mathbf{p} > \mathbf{O}$, что

$$\mathbf{pA} < \mathbf{p}.$$

Это означает, что существует хотя бы одна система цен $\mathbf{p}$, при которой цена каждого продукта больше денежных издержек его производства и, следовательно, во всех отраслях обеспечивается положительная прибыль, выражаемая (в расчете на единицу продукции) разностью

$$\mathbf{p} - \mathbf{pA}.$$

Ясно, что возможность получения прибыли неразрывно связана с возможностью получения прибавочного продукта. Более того, условия продуктивности и прибыльности матрицы (материальных затрат) равносильны и всегда справедливо соотношение

$$\mathbf{p}(\mathbf{x} - \mathbf{Ax}) = (\mathbf{p} - \mathbf{pA})\mathbf{x},$$

означающее, что прибыль есть лишь денежное выражение прибавочного продукта, а прибавочный продукт есть материальное выражение прибыли.

7.4. Задания и ответы

Сфера производства некоторой экономической системы состоит из двух отраслей. Найти оптимальный режим работы этих отраслей, обеспечивающих структуру прибавочного продукта, заданного столбцом $\mathbf{c}$, при условии, что матрица материальных затрат $\mathbf{A}$ и строка рабочей силы $\mathbf{l}$ имеют следующий вид:

1. $\mathbf{A} = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/3 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}, \qquad \mathbf{l} = (3 \quad 4)$

и известно, что мощность первой отрасли не превосходит 24, мощность второй отрасли не превосходит 12, а общее число рабочих L равно 120;

2. $\mathbf{A} = \begin{pmatrix} 1/2 & 1/6 \\ 1/6 & 1/2 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 8 \\ 8 \end{pmatrix}, \qquad \mathbf{l} = (3 \quad 2)$

и известно, что мощность первой отрасли не превосходит 8, мощность второй отрасли не превосходит 12, а общее число рабочих L равно 90;

3. $\mathbf{A} = \begin{pmatrix} 1/2 & 1/6 \\ 1/6 & 1/4 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 20 \\ 10 \end{pmatrix}, \qquad \mathbf{l} = (2 \quad 3)$

и известно, что мощность первой отрасли не превосходит 20, мощность второй отрасли не превосходит 11, а общее число рабочих L равно 72;

4. $\mathbf{A} = \begin{pmatrix} 1/3 & 1/6 \\ 1/6 & 1/4 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}, \qquad \mathbf{l} = (4 \quad 3)$

и известно, что мощность первой отрасли не превосходит 12, мощность второй отрасли не превосходит 8, а общее число рабочих L равно 96.

Ответы:

1) $x_1 = 14$, $x_2 = 12$, $\alpha_{\max} = 1/2$;
2) $x_1 = 8$, $x_2 = 8$, $\alpha_{\max} = 1/3$;
3) $x_1 = 20$, $x_2 = 10$, $\alpha_{\max} = 5/12$;
4) $x_1 = 12$, $x_2 = 8$, $\alpha_{\max} = 2/3$.

Глава 8
МНОГОКРИТЕРИАЛЬНЫЕ ЗАДАЧИ

В рассматриваемой выше задаче линейного программирования мы сталкивались с ситуацией, когда при известных условиях на числовые параметры требовалось найти наибольшее (или наименьшее) значение некоторой заданной целевой функции, которую часто называют также *критерием*. И тогда среди всех возможных наборов неизвестных мы выбирали тот, который обеспечивал бы это искомое экстремальное значение критерия. Однако при решении практических задач нередко приходится иметь дело с ситуациями, когда необходимо одновременное выполнение нескольких условий (критериев), зачастую противоречивых. Например, в стремлении приобрести товар получше и подешевле покупатель сталкивается с тем, что товар, вполне подходящий по цене, не удовлетворяет по качеству, а высокое качество товара поднимает и цену.

Пример 1. Из железного листа, имеющего форму квадрата со стороной l, требуется скроить коробку максимально возможного объема при минимальном расходе материала.

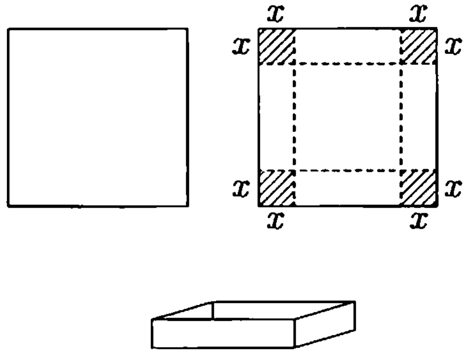


Рис. 1

Вырезав из листа четыре квадрата с неизвестной пока стороной x (рис. 1) и согнув по линиям, отмеченным пунктиром, мы получим коробку, объем которой равен

$$x(l-2x)^2.$$

При этом теряется железо общей площадью

$$4x^2.$$

По условию задачи требуется одновременное выполнение двух условий (критериев):

$$x(l-2x)^2 \to \max,$$
$$4x^2 \to \min$$

(графики функций $y = x(l-2x)^2$ и $y = 4x^2$ представлены на рис. 2).

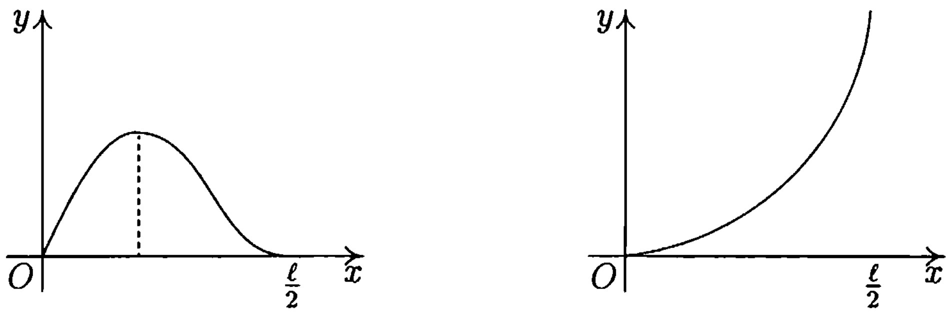


Рис. 2

Ясно, что одновременно удовлетворить обоим критериям невозможно: минимальные отходы получаются при $x = 0$ (не резать вообще), а максимального объема коробка достигает при $x = (1/6)l$.

В этой главе мы рассмотрим несколько задач с двумя критериями, считая их равноправными (вопроса о приоритетах мы коснемся в другой главе).

8.1. Множество Парето

Рассмотрим на плоскости (U, V) множество Ω (рис. 3). Каждая его точка обладает одним из следующих свойств: либо все точки, ближайшие к ней, принадлежат множеству Ω (такая точка называется *внутренней* точкой множества Ω), либо сколь угодно близко от нее расположены как точки множества Ω, так и точки, множеству Ω не принадлежащие (такие точки называются *граничными* точками множества Ω). Граничная точка может как принадлежать, так и не принадлежать множеству Ω. В дальнейшем мы будем рассматривать только такие множества, которым принадлежат все точки границы. Множество всех граничных точек множества называется его *границей*.

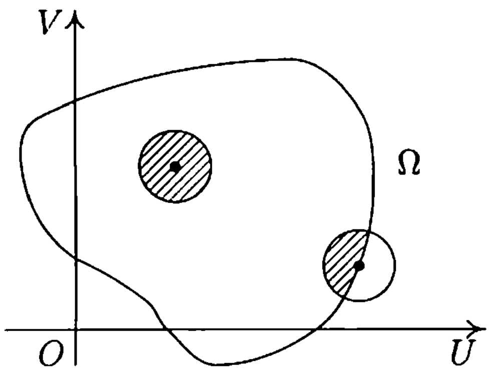


Рис. 3

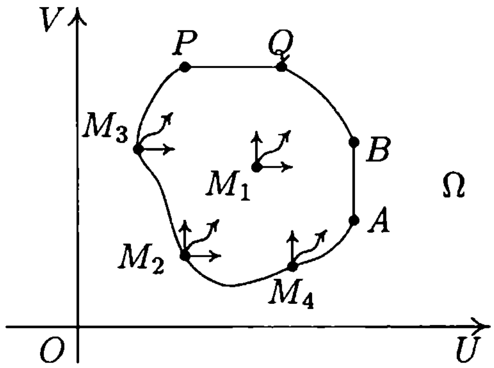


Рис. 4

Обозначение: $\partial\Omega$.

Пусть M — произвольная точка множества Ω, внутренняя или граничная, и (U, V) — ее координаты. Поставим следующий вопрос: можно ли, оставаясь в множестве Ω, переместиться из точки M в близкую точку так, чтобы при этом увеличились обе ее координаты? Если M — внутренняя точка, то это, бесспорно, возможно. Если же M — граничная точка, то такое возможно не всегда (рис. 4). Из точек M_1, M_2 и M_3 это сделать можно, но уже из точек вертикального отрезка AB можно переместиться, увеличивая лишь координату V (координата U при этом остается неизменной). Перемещая точку горизонтального отрезка PQ вправо, мы увеличиваем координату U (при этом координата V сохраняет свое значение). Что же касается дуги BQ, то перемещение вдоль нее способно лишь увеличить одну из координат при одновременном уменьшении другой.

Тем самым точки множества Ω можно разбить на три класса:

к первому классу относятся точки, которые можно сдвинуть так, чтобы одновременно увеличились обе координаты и при этом точки остались в множестве Ω (в этот класс попадают все внутренние точки множества Ω и часть его граничных точек);

второй класс образуют точки, перемещением которых по множеству Ω можно увеличить только одну из координат при сохранении значения второй (вертикальный отрезок AB и горизонтальный отрезок PQ на границе множества Ω);

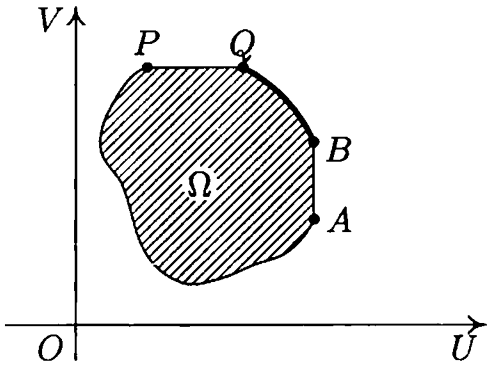


Рис. 5

в третий класс попадут точки, перемещение которых по множеству Ω способно лишь уменьшить хотя бы одну из координат (дуга BQ границы $\partial\Omega$) (рис. 5).

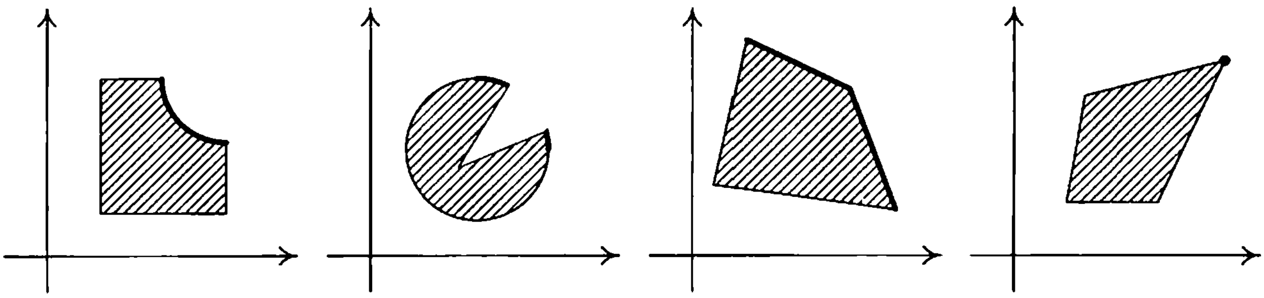


Рис. 6

Множество точек третьего класса называется *границей (множеством) Парето* данного множества Ω (выделено на рис. 5).

Говоря нестрого, граница Парето множества Ω — это точки, из которых нельзя сдвинуться на "север", "восток" либо "северо-восток", оставаясь в том же множестве Ω.

На рис. 6 указаны границы Парето некоторых множеств.

8.2. Постановка задачи

Пусть на плоскости (x, y) задано множество ω (рис. 7) и в каждой точке этого множества определены две непрерывные функции:

$$U = \Phi(x, y) \quad \text{и} \quad V = \Psi(x, y).$$

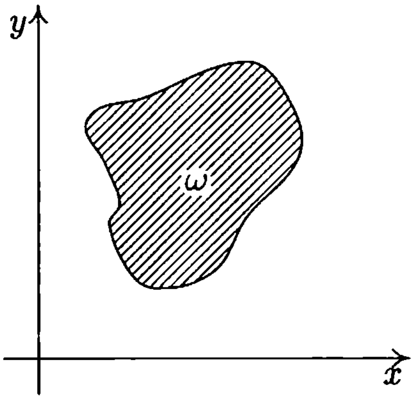


Рис. 7

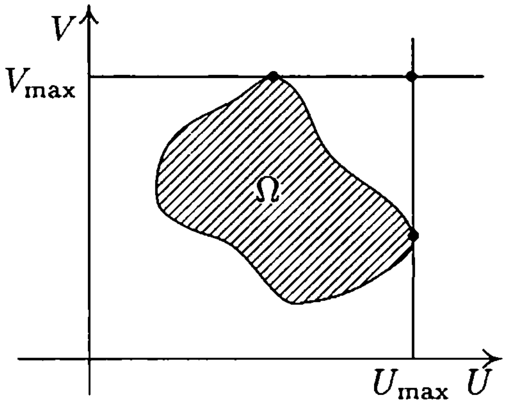


Рис. 8

Рассмотрим следующую задачу.

На множестве ω найти точку (x^0, y^0), в которой

$$\Phi(x^0, y^0) = \max \quad \text{и} \quad \Psi(x^0, y^0) = \max .$$

Обычно это записывается так:

$$\Phi(x, y) \to \max \quad \text{и} \quad \Psi(x, y) \to \max,$$
$$(x, y) \in \omega.$$

Сразу же отметим, что в общем случае поставленная задача решения не имеет.

В самом деле, изобразим на плоскости (U, V) все точки, координаты которых вычисляются по формулам

$$U = \Phi(x, y), \quad V = \Psi(x, y), \qquad (x, y) \in \omega.$$

Обозначим полученное множество через Ω. Из рис. 8 видно, что наибольшее значение U (U_{max}) и наибольшее значение V (V_{max}) достигаются в разных точках, а точка с координатами

$$(U_{max}, V_{max})$$

лежит вне множества Ω.

Тем самым в исходной постановке задача, вообще говоря, неразрешима — удовлетворить обоим требованиями одновременно невозможно. И следовательно, нужно искать какое-то компромиссное решение.

Среди известных стоит отметить два:

1) метод уступок

и

2) метод идеальной точки.

Оба метода используют множество Парето, составленное в данном случае из допустимых точек задачи, которые не могут быть “сдвинуты” в пределах допустимого множества с улучшением сразу по обоим критериям. Иными словами, улучшая значения одного из критериев, мы неизбежно ухудшаем значения другого.

Метод (последовательных) уступок заключается в том, что лицо, принимающее решение (ЛПР), работая в режиме диалога со специалистом, анализирует точки на границе Парето и в конце концов соглашается остановиться на некоторой компромиссной.

Метод идеальной точки состоит в отыскании на границе Парето точки, ближайшей к *точке утопии*, задаваемой ЛПР. Обычно ЛПР формулирует цель в виде желаемых значений показателей, и часто в качестве координат целевой точки выбирается сочетание наилучших значений всех критериев (обычно эта точка не реализуется при заданных ограничениях, поэтому ее и называют точкой утопии).

8.3. Метод идеальной точки. Конкретные примеры

Пусть на множестве ω плоскости (x, y), определяемом системой неравенств

$$\begin{cases} 0 \leqslant x \leqslant 4, \\ 0 \leqslant y \leqslant 2, \\ x + 2y \leqslant 6, \end{cases}$$

заданы две линейные функции:

$$\begin{aligned} U &= x + y + 2, \\ V &= x - y + 6. \end{aligned} \tag{1}$$

Требуется найти решение задачи

$$U \to \max, \qquad V \to \max.$$

Множество ω представляет собой пятиугольник (рис. 9), вершины которого имеют следующие координаты:

$$A(0,0), \quad B(0,2), \quad C(2,2), \quad D(4,1), \quad E(4,0).$$

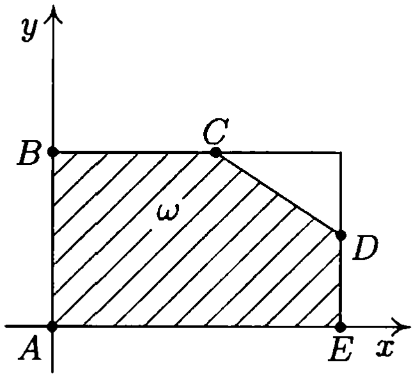


Рис. 9

В силу линейности критериев U и V пятиугольник $ABCDE$ переходит в пятиугольник $A^*B^*C^*D^*E^*$ (рис. 10), координаты вершин которого вычисляются по формулам (1):

$$A^*(2,6), \quad B^*(4,4), \quad C^*(6,6), \quad D^*(7,9), \quad E^*(6,10).$$

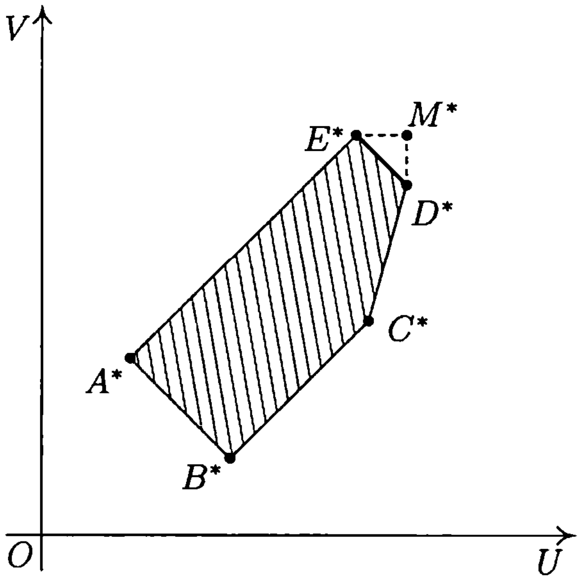


Рис. 10

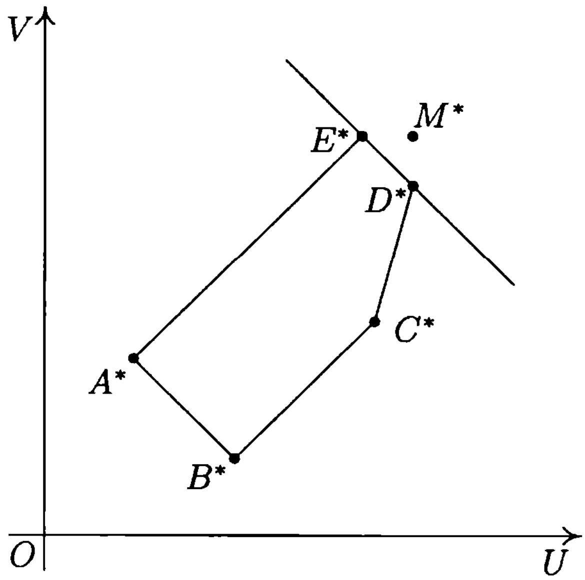


Рис. 11

Находим границу Парето. Это отрезок D^*E^*. Точка утопии $M^*(7,10)$ считается заданной (ее координаты суть наибольшие значения U и V).

Требуется найти на множестве Парето точку, ближайшую к точке утопии M^*. Из рисунка видно, что искомая точка должна лежать на отрезке D^*E^*. Проведем через точки D^* и E^* прямую. Пусть

$$\alpha U + \beta V = \gamma$$

— ее уравнение (рис. 11). Чтобы отыскать конкретные значения параметров α, β и γ, подставим в него координаты обеих точек — и D^*, и E^*. Получим

$$6\alpha + 10\beta = \gamma,$$
$$7\alpha + 9\beta = \gamma.$$

Вычитая из первого равенства второе, после простых преобразований придем к соотношению

$$-\alpha + \beta = 0,$$

откуда

$$\alpha = \beta.$$

Положим $\alpha = \beta = 1$. Тогда $\gamma = 16$ и

$$U + V = 16$$

— искомое уравнение прямой.

Теперь стоит напомнить, как ищется расстояние между точками, заданными своими координатами.

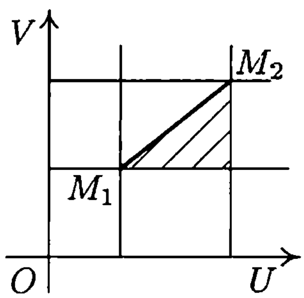


Рис. 12

Пусть $M_1(U_1, V_1)$ и $M_2(U_2, V_2)$ — точки на плоскости (U, V) (рис. 12). Для того чтобы найти расстояние между ними, достаточно вычислить длину гипотенузы построенного прямоугольного треугольника.

Воспользовавшись теоремой Пифагора, получим, что гипотенуза равна

$$\sqrt{(U_1 - U_2)^2 + (V_1 - V_2)^2}.$$

По условию задачи нам нужно определить на прямой

$$U + V = 16$$

точку $M^0(U^0, V^0)$, расстояние которой от точки $M^*(7, 10)$ минимально, т. е. решить экстремальную задачу:

$$z = (U-7)^2 + (V-10)^2 \to \min.$$

Так как $U = 16 - V$, то последнее соотношение можно переписать в виде

$$z = (9-V)^2 + (V-10)^2 \to \min.$$

Возводя в квадрат и приводя подобные, получаем, что

$$z = 2V^2 - 38V + 181.$$

Это уравнение описывает параболу (рис. 13) с вершиной

$$V^0 = \frac{19}{2}, \quad z^0 = \frac{1}{2}$$

(координата V^0 находится из условия равенства нулю производной $z' = 4V - 38$). Тогда

$$U^0 = 16 - \frac{19}{2} = \frac{13}{2}.$$

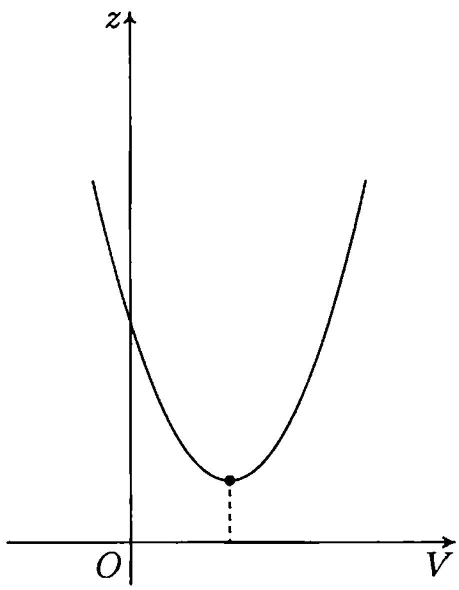


Рис. 13

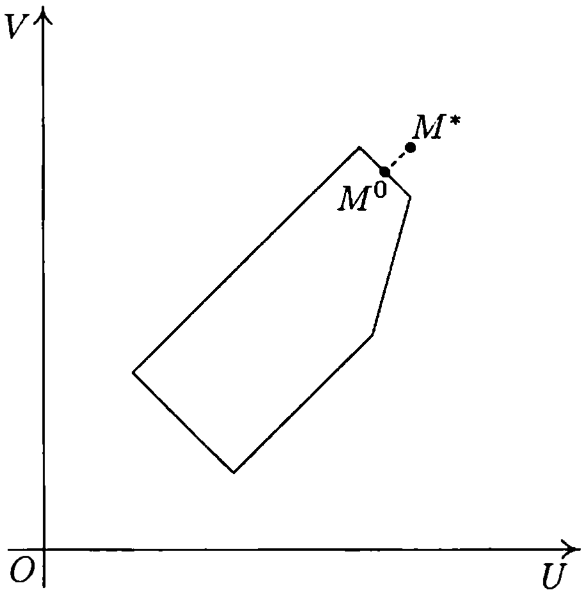


Рис. 14

Идеальная точка

$$M^0\left(\frac{13}{2}, \frac{19}{2}\right)$$

находится на расстоянии $1/\sqrt{2}$ от точки утопии $M^*(7,10)$ (рис. 14). Соответствующие значения x и y легко находятся из системы линейных уравнений

$$x + y + 2 = \frac{13}{2},$$

$$x - y + 6 = \frac{19}{2}.$$

Имеем:

$$x = 4, \qquad y = \frac{1}{2}.$$

Замечание. Мы рассмотрели задачу, в которой

$$\Phi(x,y) \to \max, \qquad \Psi(x,y) \to \max.$$

На практике часто встречаются случаи, когда требования выглядят по-иному —

так:

$$\Phi(x,y) \to \max, \qquad \Psi(x,y) \to \min$$

или даже так:

$$\Phi(x,y) \to \min, \qquad \Psi(x,y) \to \min.$$

Можно, конечно, решать такие задачи и непосредственно. Но значительно удобнее поступить по-другому, если учесть, что функция

$$\Theta = -\Psi$$

обладает следующим свойством: она достигает наибольшего значения в точности в тех точках, где функция Ψ принимает наименьшее значение, и наоборот. Иными словами, условия

$$\Psi \to \min \quad \text{и} \quad \Theta \to \max$$

равносильны. Поэтому, поменяв в случае необходимости знак у критерия на противоположный, мы можем свести любую двухкритериальную задачу к уже рассмотренной:

$$\Phi(x,y) \to \max, \qquad \Psi(x,y) \to \max.$$

Рассмотрим соответствующий пример.

Пример 2. Пусть на множестве

$$\begin{cases} 0 \leqslant x \leqslant 1, \\ 0 \leqslant y \leqslant 1 \end{cases}$$

(рис. 15) заданы две линейные функции:

$$U = 2x \quad \text{и} \quad V = x - y - 1. \tag{2}$$

Требуется найти решение задачи

$$U \to \max, \qquad V \to \min \tag{3}$$

при условии, что точка утопии M^* имеет координаты $(2, -2)$.

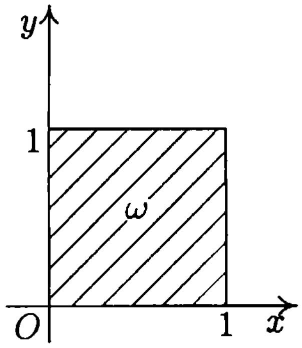


Рис. 15

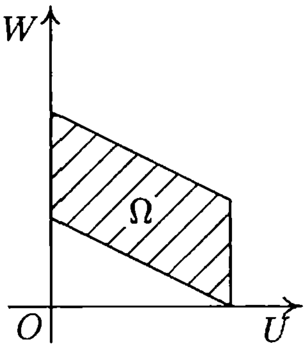


Рис. 16

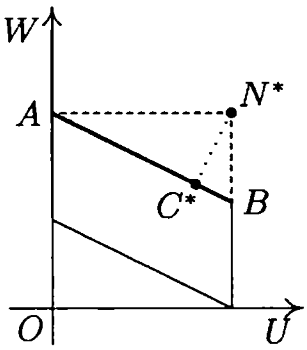


Рис. 17

Введем новую функцию

$$W = -V = -x + y + 1. \tag{4}$$

Тогда требование (3) можно записать так:

$$U \to \max, \qquad W \to \max.$$

Соответственно изменится и точка утопии — $N^*(2, 2)$.

Функции

$$U = 2x \quad \text{и} \quad W = -x + y + 1$$

линейны и преобразуют единичный квадрат ω в параллелограмм Ω (рис. 16), при этом вершины квадрата

$$(0, 0), \quad (1, 0), \quad (1, 1), \quad (0, 1)$$

переходят в вершины параллелограмма

$$(0, 1), \quad (2, 0), \quad (2, 1), \quad (0, 2).$$

Множество Парето образуют точки отрезка с концами $A(0,2)$ и $B(2,1)$ (рис. 17). Проведем через эти точки прямую и найдем коэффициенты ее уравнения

$$\alpha U + \beta W = \gamma.$$

Подставляя в него координаты точек A и B, получаем, что

$$2\beta = \gamma, \qquad 2\alpha + \beta = \gamma.$$

Положим $\beta = 2$. Тогда $\gamma = 4$ и $\alpha = 1$. Тем самым уравнение искомой прямой имеет вид

$$U + 2W = 4.$$

Пусть $C^*(U, W)$ — точка этой прямой, ближайшая к точке $N^*(2,2)$. Это означает, что должно выполняться условие

$$z = (U-2)^2 + (W-2)^2 \to \min.$$

Так как $U = 4 - 2W$, то оно принимает следующий вид:

$$z = 5W^2 - 12W + 8 \to \min.$$

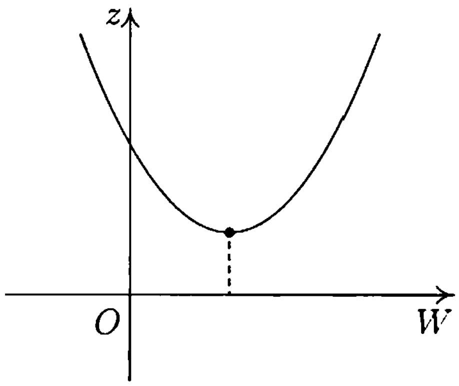


Рис. 18

Чтобы найти минимальное значение функции

$$z = 5W^2 - 12W + 8$$

(рис. 18), приравняем к нулю ее производную. Имеем:

$$z' = 10W - 12 = 0.$$

Отсюда

$$W = \frac{6}{5}, \quad U = 4 - \frac{12}{5} = \frac{8}{5}.$$

Соответствующие значения x и y находятся из уравнений (см. (2))

$$2x = \frac{8}{5}, \quad -x + y + 1 = \frac{6}{5},$$

откуда

$$x = \frac{4}{5}, \quad y = 1.$$

Расстояние от найденной точки

$$C^* \left(\frac{8}{5}, \frac{6}{5}\right)$$

до точки утопии $N^*(2, 2)$ равно $2/\sqrt{5}$.

Ответ: идеальная точка

$$M^0 \left(\frac{8}{5}, \frac{6}{5}\right)$$

находится от заданной точки утопии $M^*(2, -2)$ на расстоянии $2/\sqrt{5}$.

8.4. Задания и ответы

1. На множестве

$$\begin{cases} 0 \leqslant x \leqslant 1, \\ 0 \leqslant y \leqslant 1 \end{cases}$$

заданы две линейные функции:

$$U = 2x + 1 \quad \text{и} \quad V = 2y + 3.$$

Требуется найти решение задачи

$$U \to \max, \qquad V \to \max,$$

при условии, что точка утопии M^* имеет координаты $(5, 7)$.

Ответ: идеальная точка $M^0(3,5)$ находится от заданной точки утопии $M^*(5,7)$ на расстоянии $2\sqrt{2}$; $(1,1)$ — соответствующая точка плоскости (x,y).

2. На множестве, определяемом системой неравенств

$$\begin{cases} 0 \leqslant x \leqslant 2, \\ 0 \leqslant y \leqslant 4, \\ 2x + y \leqslant 6, \end{cases}$$

заданы линейные функции:

$$U = x + y + 2,$$
$$V = x - y + 6.$$

Требуется найти решение задачи

$$U \to \max, \qquad V \to \max,$$

при условии, что точка утопии M^* имеет координаты $(7,8)$.

Ответ: идеальная точка

$$M^0\left(\frac{11}{2}, \frac{13}{2}\right)$$

находится от заданной точки утопии $M^*(7,8)$ на расстоянии $3/\sqrt{2}$; $(2, 3/2)$ — соответствующая точка плоскости (x,y).

Глава 9
ИЕРАРХИИ И ПРИОРИТЕТЫ

...Старинное изречение о том, что нельзя сравнивать яблоки и апельсины, неверно
Т. Саати

9.1. Приоритеты

9.1.1. Измерения и согласованность

Предположим, что имеется некоторое семейство предметов (например, камней)

$$S_1, \ldots, S_n,$$

каждый из которых легок настолько, что его нетрудно удержать в руке, и требуется оценить их относительные веса в отсутствие взвешивающего прибора.

Среди возможных способов разрешения этой проблемы укажем два.

Первый состоит в том, чтобы определить (угадать) вес каждого предмета, взяв за единицу измерения (эталон) самый легкий, сравнить таким образом все предметы и, разделив затем найденный вес каждого S_i на сумму весов всех n предметов, получить его относительный вес.

Это потребует $(n - 1)$ сравнений.

Второй способ состоит в сравнении весов всевозможных пар предметов: сначала мы сравниваем вес предмета S_1 с весами предметов $S_2, \ldots, S_n$, затем вес предмета S_2 с весами предметов $S_3, \ldots, S_n$ и т. д. до тех пор, пока у нас не сформируется суждение об относительном весе (отношении весов) для каждой пары предметов.

В этом случае общее число необходимых сравнений оказывается равным

$$\frac{n(n-1)}{2}.$$

При этом каждый предмет методично сравнивается со всеми остальными.

Конечно, второй способ требует большего времени, чем первый, но оказывается точнее.

Любым измерениям (в том числе и с использованием приборов) присущи ошибки (погрешности), серьезным следствием которых является то обстоятельство, что они могут привести (и нередко приводят) к несогласованным выводам.

Приведем совсем простой пример ошибочного сравнения: предмет A в 1,5 раза тяжелее предмета B, который в свою очередь в 1,5 раза тяжелее предмета C, последний же по весу почти не отличается от предмета A.

Согласованность измерений является весьма важной их характеристикой.

При этом под *согласованностью* при сравнении предметов по весу подразумевается не просто результат типа:

если A тяжелее B и B тяжелее C, то A тяжелее C,

а количественно более точный:

если A в 2 раза тяжелее B, а B в 3 раза тяжелее C, то A в $2 \cdot 3 = 6$ раз тяжелее C.

Замечание 1. Как правило, чем лучше человек знаком с ситуацией, тем более он последователен в своих суждениях. Хотя обратное и необязательно верно — отличная согласованность в суждениях вовсе не означает, что человек хорошо разбирается в ситуации.

Замечание 2. Попарные сравнения позволяют повысить согласованность оценок.

Проблема сравнения возникает повсюду — и при измерении физических величин, и при оценке совершенных поступков.

Для получения хороших результатов в сравнениях требуется уметь:

1) находить подходящую численную шкалу сравнений,

2) определять степень несогласованности наших суждений.

Начнем с обсуждения вопроса о том, как можно оценить согласованность наших суждений практически. А затем поговорим и о шкалировании.

9.1.2. Идеальные измерения

Пусть нам предложено сравнить веса камешков

$$S_1, \dots, S_n.$$

Рассмотрим идеальную ситуацию, предположив, что в нашем распоряжении их идеально точные веса. Обозначим эти веса через

$$w_1, \dots, w_n$$

соответственно.

Отношение

$$a_{ik} = \frac{w_i}{w_k}, \quad i, k = 1, \dots, n, \tag{1}$$

показывает, во сколько раз вес i-го камешка S_i больше веса k-го камешка S_k.

Например, если $w_1 = 305$ г и $w_2 = 244$ г, то отношение

$$a_{12} = \frac{w_1}{w_2} = \frac{305}{244} = 1{,}25$$

говорит о том, что камешек S_1 в 1,25 раза тяжелее камешка S_2.

Запишем отношения (1) в виде квадратной матрицы

$$\mathbf{A} = \begin{pmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \frac{w_2}{w_n} \\ \dots & \dots & \dots & \dots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \frac{w_n}{w_n} \end{pmatrix}$$

и проанализируем некоторые свойства этой *идеальной матрицы сравнений*.

1. Для любого i справедливо равенство $a_{ii} = 1$ (элемент матрицы $\mathbf{A}$, расположенный на пересечении i-й строки и i-го столбца, равен единице).

В самом деле,

$$a_{ii} = \frac{w_i}{w_i} = 1.$$

2. Для любых i и k справедливо равенство $a_{ki} = \dfrac{1}{a_{ik}}$ (произведение элемента матрицы $\mathbf{A}$, расположенного на пересечении i-й строки и k-го столбца, на элемент матрицы $\mathbf{A}$, расположенный на пересечении k-й строки и i-го столбца, равно единице).

В самом деле, из того, что

$$a_{ki} = \frac{w_k}{w_i} \qquad \text{и} \qquad a_{ik} = \frac{w_i}{w_k},$$

следует равенство

$$a_{ki}a_{ik} = \frac{w_k}{w_i} \cdot \frac{w_i}{w_k} = 1.$$

3. Для любых i, k и l справедливо равенство $a_{ik}a_{kl} = a_{il}$ (произведение элемента матрицы $\mathbf{A}$, расположенного в i-й строке и k-м столбце, на элемент матрицы $\mathbf{A}$, расположенный в k-й строке и l-м столбце, равно элементу матрицы $\mathbf{A}$, расположенному в i-й строке и l-м столбце).

В самом деле,

$$a_{ik}a_{kl} = \frac{w_i}{w_k} \cdot \frac{w_k}{w_l} = \frac{w_i}{w_l} = a_{il}.$$

4. Столбец

$$\mathbf{w} = \begin{pmatrix} w_1 \\ \dots \\ w_n \end{pmatrix}$$

является собственным столбцом матрицы $\mathbf{A}$ с собственным значением $\lambda = n$.

В самом деле,

$$\mathbf{A}\mathbf{w} = \begin{pmatrix} \dfrac{w_1}{w_1} & \dfrac{w_1}{w_2} & \dots & \dfrac{w_1}{w_n} \\ \dfrac{w_2}{w_1} & \dfrac{w_2}{w_2} & \dots & \dfrac{w_2}{w_n} \\ \dots & \dots & \dots & \dots \\ \dfrac{w_n}{w_1} & \dfrac{w_n}{w_2} & \dots & \dfrac{w_n}{w_n} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix} = \begin{pmatrix} nw_1 \\ nw_2 \\ \dots \\ nw_n \end{pmatrix} = n \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix} = n\mathbf{w}.$$

9.1.3. Обратно-симметричные и согласованные матрицы

Рассмотрим теперь квадратную положительную матрицу порядка n

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nk} & \dots & a_{nn} \end{pmatrix}.$$

Матрица $\mathbf{A}$ называется *обратно-симметричной*, если для любых i и k выполняется соотношение

$$a_{ki} = \frac{1}{a_{ik}}.$$

Из этого, в частности, следует, что

$$a_{ii} = 1.$$

Матрица $\mathbf{A}$ называется *согласованной*, если для любых i, k и l имеет место равенство

$$a_{ik}a_{kl} = a_{il}.$$

Тем самым, идеальная матрица сравнений — обратно-симметричная и согласованная.

Справедливо следующее утверждение.

ТЕОРЕМА. *Положительная обратно-симметричная матрица является согласованной тогда и только тогда, когда порядок матрицы n ее наибольшее собственное значение совпадают:*

$$\lambda_{\max} = n.$$

9.1.4. Индекс согласованности

Если элементы положительной обратно-симметричной согласованной матрицы $\mathbf{A}$ изменить незначительно («пошевелить»), то максимальное собственное значение $\lambda_{\max}$ также изменится незначительно.

Пусть $\mathbf{A}$ — произвольная положительная обратно-симметричная матрица и $\lambda_{\max}$ — ее наибольшее собственное значение.

Если

$$\lambda_{\max} = n,$$

то матрица $\mathbf{A}$ — согласованная.

Если

$$\lambda_{\max} \neq n$$

(всегда $\lambda_{\max} \geqslant n$), то в качестве степени отклонения положительной обратно-симметричной матрицы $\mathbf{A}$ от согласованной можно взять отношение

$$\frac{\lambda_{\max} - n}{n - 1},$$

которое называется *индексом согласованности* (ИС) матрицы $\mathbf{A}$ и является показателем близости этой матрицы к согласованной.

Замечание. Считается, что если ИС не превышает 0,10, то можно быть удовлетворенным степенью согласованности суждений.

9.1.5. Вычисление собственных характеристик обратно-симметричной матрицы

Довольно естественно встает вопрос о том, как находить наибольшее собственное значение $\lambda_{\max}$ положительной обратно-симметричной матрицы.

Для $n = 2$ такую задачу решать мы умеем. Правда, это не так интересно: обратно-симметричная матрица 2-го порядка всегда согласованная.

В самом деле, пусть

$$\begin{pmatrix} 1 & a \\ 1/a & 1 \end{pmatrix}$$

— обратно-симметричная матрица. Найдем ее собственные значения. Имеем

$$(1 - \lambda)^2 = 1.$$

Отсюда $\lambda = 2$ и $\lambda = 0$.

Однако в общем случае эта задача хотя и разрешима, но технически достаточно сложна. Поэтому, желая содержательно, но относительно просто ответить на поставленный вопрос, мы вынуждены чем-то поступиться. Проще всего поступиться точностью вычислений, т. е. искать приближенное значение наибольшего собственного числа.

Для этого поступают так: сначала приближенно строится собственный столбец, а затем по нему ищется приближенное собственное значение.

Опишем несколько способов приближенного вычисления собственного столбца.

1-й способ:

1) суммируем элементы каждой строки и записываем полученные результаты в столбец,

2) складываем все элементы найденного столбца,

3) делим каждый из элементов этого столбца на полученную сумму.

2-й способ:

1) суммируем элементы каждого столбца и записываем полученные результаты в столбец,

2) заменяем каждый элемент построенного столбца на обратный ему,

3) складываем элементы столбца из обратных величин,

4) делим каждый из этих элементов на полученную сумму.

3-й способ:

1) суммируем элементы каждого столбца,

2) делим элементы каждого столбца на их сумму,

3) складываем элементы каждой строки полученной матрицы,

4) записываем результаты в столбец,

5) делим каждый из элементов последнего столбца на порядок исходной матрицы n.

4-й способ:

1) перемножаем элементы каждой строки и записываем полученные результаты в столбец,

2) извлекаем корень n-й степени из каждого элемента найденного столбца,

3) складываем элементы этого столбца,

4) делим каждый из этих элементов на полученную сумму.

Каждый из этих четырех способов, будучи примененным к идеальной матрице, приводит к одному и тому же точному результату.

Покажем это для 1-го случая (для остальных трех проверка проводится столь же просто).

Пусть

$$\begin{pmatrix} \dfrac{w_1}{w_1} & \dots & \dfrac{w_1}{w_n} \\ \dots & \dots & \dots \\ \dfrac{w_n}{w_1} & \dots & \dfrac{w_n}{w_n} \end{pmatrix} \tag{2}$$

— идеальная матрица.

Просуммируем элементы каждой строки матрицы (2) и, записав полученные результаты в столбец

$$\begin{pmatrix} \frac{w_1}{w_1} + \cdots + \frac{w_1}{w_n} \\ \cdots\cdots\cdots\cdots\cdots \\ \frac{w_n}{w_1} + \cdots + \frac{w_n}{w_n} \end{pmatrix} = \begin{pmatrix} w_1 \sum_{k=1}^{n} \frac{1}{w_k} \\ \cdots \\ w_n \sum_{k=1}^{n} \frac{1}{w_k} \end{pmatrix}, \tag{3}$$

сложим элементы этого столбца. Имеем:

$$w_1 \sum_{k=1}^{n} \frac{1}{w_k} + \cdots + w_n \sum_{k=1}^{n} \frac{1}{w_k} = (w_1 + \cdots + w_n) \sum_{k=1}^{n} \frac{1}{w_k} = \sum_{i=1}^{n} w_i \sum_{k=1}^{n} \frac{1}{w_k}.$$

Поделим каждый из элементов столбца (3) на найденную сумму. В результате получим

$$\begin{pmatrix} \dfrac{w_1}{w_1 + \cdots + w_n} \\ \cdots \\ \dfrac{w_n}{w_1 + \cdots + w_n} \end{pmatrix} = \frac{1}{w_1 + \cdots + w_n} \begin{pmatrix} w_1 \\ \cdots \\ w_n \end{pmatrix}.$$

Нетрудно заметить, что итогом этих операций будет собственный столбец матрицы (2), сумма элементов которого равна единице. Легко убедиться и в том, что соответствующее собственное значение равно n.

В применении к обратно-симметричной, но не согласованной матрице ни один из предложенных способов уже не дает собственного столбца. Тем не менее при вычислении собственных столбцов таких матриц мы будем пользоваться именно этими способами, получая в результате приближенные собственные столбцы.

Сложность вычислений возрастает с увеличением номера способа, но увеличивается и точность.

Замечание. Описанные способы приближенного вычисления собственного столбца матрицы эффективны лишь для обратно-симметричных матриц, достаточно близких к согласованным.

Пример 1. Рассмотрим обратно-симметричную матрицу 4-го порядка

$$\begin{pmatrix} 1 & 5 & 6 & 7 \\ 1/5 & 1 & 4 & 6 \\ 1/6 & 1/4 & 1 & 4 \\ 1/7 & 1/6 & 1/4 & 1 \end{pmatrix}$$

и вычислим приближенно ее собственный столбец всеми четырьмя способами.

1-й способ (указаны результаты каждого шага):

$$1)\ \begin{pmatrix} 19{,}00 \\ 11{,}20 \\ 5{,}42 \\ 1{,}56 \end{pmatrix}, \quad 2)\ 37{,}18, \quad 3)\ \begin{pmatrix} 0{,}51 \\ 0{,}30 \\ 0{,}15 \\ 0{,}04 \end{pmatrix}.$$

2-й способ (указаны результаты каждого шага):

$$1)\ \begin{pmatrix} 1{,}51 \\ 6{,}43 \\ 11{,}25 \\ 18{,}00 \end{pmatrix}, \quad 2)\ \begin{pmatrix} 0{,}66 \\ 0{,}16 \\ 0{,}09 \\ 0{,}06 \end{pmatrix}, \quad 3)\ 0{,}97, \quad 4)\ \begin{pmatrix} 0{,}68 \\ 0{,}16 \\ 0{,}09 \\ 0{,}06 \end{pmatrix}.$$

3-й способ:

в результате 1-го и 2-го шагов получим матрицу

$$\begin{pmatrix} 0{,}66 & 0{,}78 & 0{,}53 & 0{,}39 \\ 0{,}13 & 0{,}16 & 0{,}36 & 0{,}33 \\ 0{,}11 & 0{,}04 & 0{,}09 & 0{,}22 \\ 0{,}09 & 0{,}03 & 0{,}02 & 0{,}06 \end{pmatrix},$$

в результате 3-го и 4-го шагов получим столбец

$$\begin{pmatrix} 2{,}36 \\ 0{,}98 \\ 0{,}46 \\ 0{,}20 \end{pmatrix}$$

и окончательно

$$\begin{pmatrix} 0{,}590 \\ 0{,}245 \\ 0{,}115 \\ 0{,}050 \end{pmatrix}.$$

4-й способ (окончательный результат):

$$\begin{pmatrix} 0{,}61 \\ 0{,}24 \\ 0{,}10 \\ 0{,}04 \end{pmatrix}.$$

Точный метод построения собственного столбца заданной матрицы дает следующий результат:

$$\begin{pmatrix} 0{,}61 \\ 0{,}24 \\ 0{,}10 \\ 0{,}05 \end{pmatrix}.$$

Итак, собственный столбец найден. Теперь остается найти соответствующее собственное значение.

Покажем, как это делается в данном случае.

Как уже отмечалось в гл. 3, если мы хотим проверить, является ли предъявленный столбец $\mathbf{x}$ собственным столбцом матрицы $\mathbf{A}$, следует поступать так:

1) умножить матрицу $\mathbf{A}$ на этот столбец:

$$\mathbf{Ax} = \mathbf{y},$$

или подробнее:

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \ldots \\ y_n \end{pmatrix},$$

2) разделить элементы полученного столбца $\mathbf{y}$ на соответствующие элементы столбца $\mathbf{x}$:

$$\frac{y_1}{x_1}, \ldots, \frac{y_n}{x_n},$$

и если

$$\frac{y_1}{x_1} = \cdots = \frac{y_n}{x_n}, \tag{4}$$

то это отношение и есть собственное значение матрицы $\mathbf{A}$, отвечающее данному столбцу $\mathbf{x}$.

Если же хотя бы одно из равенств (4) нарушается, то столбец **x** не является собственным столбцом матрицы **A**.

В данном случае столбец, получаемый любым из описанных выше четырех способов, мы заранее рассматриваем как приближение собственного столбца, и ожидать выполнения даже одного из равенств (4) нельзя.

Поэтому здесь мы поступим по-иному — считая каждое из отношений

$$\frac{y_1}{x_1}, \ldots, \frac{y_n}{x_n}$$

приближением к искомому собственному значению, выберем в качестве собственного значения их среднее арифметическое:

$$\tilde{\lambda}_{\max} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}.$$

Продолжение примера 1. Для отыскания приближенного значения наибольшего собственного числа заданной матрицы используем приближение собственного столбца, вычисленное по 2-му способу.

Умножив матрицу на соответствующий столбец, получаем

$$\begin{pmatrix} 1 & 5 & 6 & 7 \\ 1/5 & 1 & 4 & 6 \\ 1/6 & 1/4 & 1 & 4 \\ 1/7 & 1/6 & 1/4 & 1 \end{pmatrix}\begin{pmatrix} 0{,}68 \\ 0{,}16 \\ 0{,}09 \\ 0{,}06 \end{pmatrix} = \begin{pmatrix} 2{,}48 \\ 1{,}01 \\ 0{,}25 \\ 0{,}21 \end{pmatrix}.$$

Поделив элементы найденного столбца-произведения на соответствующие элементы исходного столбца-сомножителя, получим следующие числа:

$$3{,}59, \quad 6{,}31, \quad 2{,}78, \quad 3{,}50.$$

Найдем их среднее арифметическое. Имеем:

$$\frac{1}{4}(3{,}59 + 6{,}31 + 2{,}78 + 3{,}50) = \frac{16{,}18}{4} = 4{,}05.$$

Тем самым,

$$\tilde{\lambda}_{\max} = 4{,}05.$$

Теперь уже совсем легко найти:

$$\text{ИС} = \frac{4{,}05 - 4}{4 - 1} = \frac{0{,}05}{3} = 0{,}017.$$

Для точного собственного столбца

$$\lambda_{\max} = 4{,}5.$$

3-й способ дает значение

$$\tilde{\lambda}_{\max} = 4{,}39,$$

более близкое к точному. Правда, в этом случае ИС = 0,13.

9.1.6. Шкалирование

Количественные оценки, вводимые при парных сравнениях, строят исходя из некоторых эмпирических правил, опирающихся на шаткое основание опыта. Тем не менее приобретенное опытным путем удивительным образом оказывается полезным во многих, часто совершенно непохожих ситуациях.

При любом подходе к разрешению задачи сравнения важное значение имеет выбор шкалы сравнений. Главное требование — шкала сравнений должна быть проста и естественна.

Вот некоторые соображения, обосновывающие наш выбор.

Начнем с диапазона. Использование шкалы парных сравнений в пределах от 0 до ∞ может оказаться бесполезным. Дело в том, что наша способность различать находится в весьма ограниченном диапазоне и, когда есть значительная несоразмерность между сравниваемыми объектами, действиями или обстоятельствами, наши предположения тяготеют к тому, чтобы быть произвольными, и обычно оказываются далекими от действительности.

Так как единица является стандартом измерения, то верхняя граница должна быть не слишком далека от нее, хотя и достаточно отдалена для того, чтобы более или менее выразительно представить наш диапазон способности различать.

Поэтому и число сравниваемых объектов должно быть достаточно мало. Обычные пределы — это 7 ± 2.

Почему же выбираются числа от 1 до 9?

Вот только некоторые из возможных объяснений.

1. Способность человека производить качественные разграничения хорошо представлена пятью определениями: *слабый, равный, сильный, очень сильный, абсолютный.* Для большей точности можно пользоваться промежуточными определениями.

2. Классификация по трем основным зонам — *неприятие, безразличие, приятие*, каждая из которых делится на низкую, умеренную и высокую степени.

3. Психологический предел 7 ± 2 предметов при одновременном сравнении подтверждает, что если взять 7 ± 2 отдельных предметов, близких относительно свойства, используемого для сравнения, то требуется 9 точек, чтобы их различить.

Замечание. Здесь уместно упомянуть и о принятой в отечественном образовании системе оценок 3, 4 и 5 с ее градациями $3\pm$, $4\pm$ и $5\pm$.

Опишем один из способов того, как практически придать количественное наполнение сравнению объектов, действий или обстоятельств и построить соответствующую таблицу сравнений.

Пусть даны элементы A, B, C, D и т. д.

Таблица сравнений, имеющая вид

	A	B	C	D	$\dots$
A					
B					
C					
D					
$\vdots$					

строится по следующим правилам:

если A и B одинаково важны, заносим в позицию (A, B) таблицы сравнений число 1,

если A незначительно важнее B — число 3,

если A значительно важнее B — число 5,

если A явно важнее B — число 7,

если A по своей значимости абсолютно превосходит B — число 9.

Числа 2, 4, 6 и 8 используются для облегчения компромиссов между оценками, слегка отличающимися от основных чисел.

Рациональные дроби используются в случае, когда желательно увеличить согласованность всей матрицы при малом числе суждений.

Пример 2. Предположим, что, сравнивая объекты A, B, C и D, мы получили таблицу сравнений

	A	B	C	D
A	1	5	6	7
B	1/5	1	4	6
C	1/6	1/4	1	4
D	1/7	1/6	1/4	1

которая приводит к обратно-симметричной матрице, рассмотренной выше.

Пользуясь одним из способов приближенного вычисления собственных элеметов этой матрицы (для определенности вторым), мы нашли и собственный столбец, и собственное значение, и ИС:

$$\begin{pmatrix} 0{,}68 \\ 0{,}16 \\ 0{,}09 \\ 0{,}06 \end{pmatrix}, \quad \tilde{\lambda}_{\max} = 4{,}05, \quad \text{ИС} = \frac{4{,}05 - 4}{4 - 1} = \frac{0{,}05}{3} = 0{,}017.$$

Сумма всех элементов полученного собственного столбца (его называют *столбцом приоритетов*) равна 1. Он позволяет подвести итог проведенному анализу таблицы сравнений:

среди сравниваемых элементов A, B, C и D наивысший приоритет имеет A (68%), затем идут B (16%), C (9%) и D (6%) соответственно.

9.2. Иерархии

Очень часто при анализе интересующей нас структуры число элементов и их взаимосвязей настолько велико, что превышает способность исследователя воспринимать информацию в полном объеме. В таких случаях система делится на подсистемы. Одним из таких делений является иерархическое.

Иерархии представляют собой определенный вид системы, основанный на предположении, что ее элементы могут группироваться в *не связанные* множества. При этом элементы каждой группы находятся под влиянием элементов некоторой другой вполне определенной группы и в свою очередь оказывают влияние на элементы третьей группы. Мы считаем, что элементы в каждой группе иерархии, называемой *уровнем*, независимы.

Первым требованием при анализе функционирования системы является построение иерархии, воспроизводящей функциональные отношения. Для этого сначала перечисляются все элементы, относящиеся к иерархии. Затем они распределяются по группам в соответствии с влиянием между группами. Так возникают уровни иерархии. Определяются цели, ради которых изучается задача, и строится иерархия.

После того как уровни иерархии заданы, составляются матрицы попарных сравнений между этими элементами относительно каждого элемента следующего, более высокого уровня, который служит критерием при сравнении.

Приведем пример типичной иерархии (рис. 1). Первый уровень иерархии имеет одну цель: общее благосостояние страны. Второй уровень иерархии имеет три цели: сильную экономику, здравоохранение, национальную оборону. Приоритеты этих целей получаются из матрицы попарных сравнений относительно цели первого уровня. Целями третьего уровня являются отрасли промышленности.

Задача заключается в определении влияния отраслей промышленности на общее благосостояние страны через промежуточный второй уровень. Поэтому приоритеты отраслей промышленности относительно каждой цели второго уровня получаются из матриц попарных сравнений относительно этих целей, а полученные столбцы приоритетов взвешиваются затем при помощи столбца приоритетов второго уровня, что позволяет получить в итоге искомый составной столбец приоритетов отраслей промышленности.

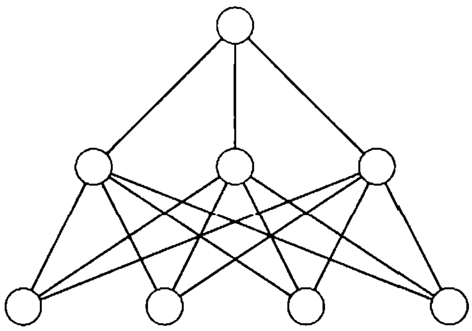


Рис. 1

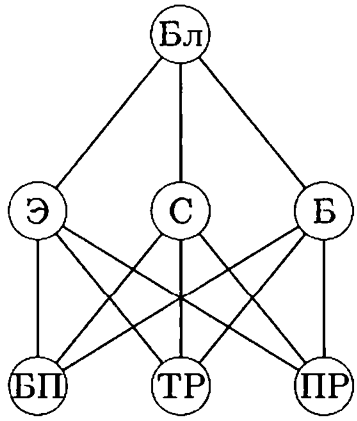


Рис. 2

Пример 3 (распределение энергии). Предположим, что нам необходимо разрешить проблему распределения энергии в некоторой развитой стране между тремя ее крупнейшими пользователями: бытовым потреблением (БП), транспортом (ТР) и промышленностью (ПР). Они составляют третий, или низший, уровень иерархии. Целями, по отношению к которым оцениваются эти потребители, являются вклад в развитие экономики (Э), вклад в качество окружающей среды (С) и вклад в национальную безопасность (Б). Цели составляют второй уровень иерархии. Общая цель — благоприятное социальное и политическое положение (Бл) — первый уровень иерархии (рис. 2).

Построим матрицу попарных сравнений трех целей: Э, С и Б в соответствии с их воздействием на общую цель — Бл. Умышленно навязывая согласованность создаваемой матрице, мы по первой строке находим все остальные ее элементы. Имеем:

Бл	Э	С	Б
Э	1	5	3
С	1/5	1	3/5
Б	1/3	5/3	1

$\lambda_{\max} = 3{,}00, \qquad \text{ИС} = 0{,}0.$

Необходимые пояснения к таблице. Экономика имеет сильное превосходство перед окружающей средой (5) и слабое перед национальной безопасностью (3). Числа во 2-й и 3-й строках выбраны так, чтобы полученная матрица сравнений была обратно-симметричной и согласованной.

Столбец приоритетов, вычисленный любым из описанных выше четырех способов, имеет вид

$$\mathbf{w} = \begin{pmatrix} 0{,}65 \\ 0{,}13 \\ 0{,}22 \end{pmatrix}.$$

Следовательно, в соответствии со сравнением по социально-политическому влиянию экономика получает приоритет 0,65, окружающая среда — 0,13 и национальная безопасность — 0,22 (рис. 3).

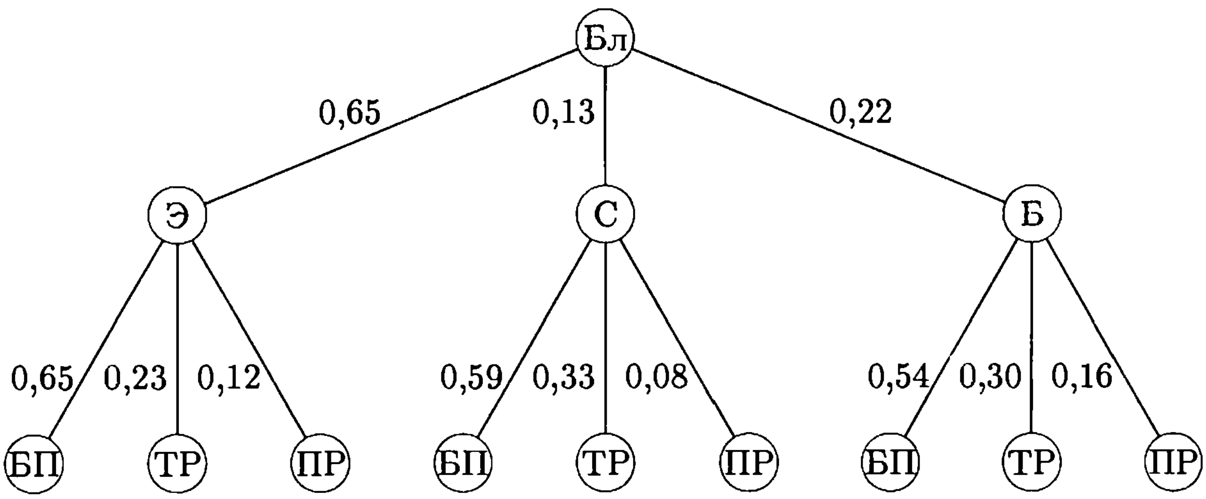


Рис. 3

Проведем теперь оценку относительной важности каждого потребителя с точки зрения экономики, окружающей среды и национальной безопасности (составляющих второй уровень иерархии).

Соответствующие матрицы попарных сравнений, индексы согласованности и столбцы приоритетов имеют следующий вид:

Э	БП	ТР	ПР
БП	1	3	5
ТР	1/3	1	2
ПР	1/5	1/2	1

$$\lambda_{\max} = 3{,}00, \qquad ИС = 0{,}0, \qquad \begin{pmatrix} 0{,}65 \\ 0{,}23 \\ 0{,}12 \end{pmatrix};$$

С	БП	ТР	ПР
БП	1	2	7
ТР	1/2	1	5
ПР	1/7	1/5	1

$$\lambda_{\max} = 3{,}01, \qquad ИС = 0{,}01, \qquad \begin{pmatrix} 0{,}59 \\ 0{,}33 \\ 0{,}08 \end{pmatrix};$$

Б	БП	ТР	ПР
БП	1	2	3
ТР	1/2	1	2
ПР	1/3	1/2	1

$$\lambda_{\max} = 3{,}01, \qquad ИС = 0{,}01, \qquad \begin{pmatrix} 0{,}54 \\ 0{,}30 \\ 0{,}16 \end{pmatrix}.$$

Запишем полученные столбцы в виде матрицы. Имеем

$$\begin{pmatrix} 0{,}65 & 0{,}59 & 0{,}54 \\ 0{,}23 & 0{,}33 & 0{,}30 \\ 0{,}12 & 0{,}08 & 0{,}16 \end{pmatrix}.$$

Умножая эту матрицу на столбец w, находим искомый столбец приоритетов третьего уровня иерархии, представляющего потребителей энергии БП, ТР и ПР (взвешенный согласно их общему влиянию):

$$\begin{pmatrix} 0{,}62 \\ 0{,}26 \\ 0{,}12 \end{pmatrix}.$$

Итак, в соответствии с нашими вычислениями на бытовое потребление следует выделить 62% энергии, на транспорт — 26 и на промышленность — 12%.

9.3. Задание

Попробуйте рассчитать веса распределения времени между учебой, досугом и подработкой в соответствии с их общим вкладом в ваше личное благополучие через 7–10 лет, на которое влияют интересная работа, материальная обеспеченность и здоровье (семья).

Здесь 1-й уровень иерархии — благополучие, 2-й уровень — интересная работа, материальная обеспеченность, здоровье (семья), 3-й уровень — учеба, подработка, досуг.

Глава 10
МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Прогнозирование (от греч. πρόγνωσις — предсказание, предвидение) по своему характеру неразрывно связано со временем — посредством прогноза мы как бы пытаемся разглядеть будущее в настоящем. Способы такого "заглядывания в будущее" весьма разнообразны — от внутреннего голоса ("печенкой чувствую") и исторических аналогий до экспертных оценок и сложных эконометрических моделей. Поэтому необходимость прогноза развития той или иной ситуации, будущих изменений тех или иных обстоятельств ставит нас перед непростой проблемой выбора вполне конкретного метода прогнозирования.

Этот выбор зависит от множества факторов. Отметим некоторые из них: наличие данных (количественное выражение накопленного в прошлом опыта), планируемый момент исполнения и желаемая точность прогноза, а также временны́е и стоимостные затраты на его составление.

По тому, на какой момент или период времени он составляется, прогноз может быть
краткосрочным (до года, но обычно на квартал),
среднесрочным (от года до трех лет) и
долгосрочным (на три года и больше).

Интуитивно ясно, что чем меньше промежуток времени, отделяющий настоящий момент от прогнозируемого, тем большим будет объем хорошо предсказываемых событий — для того, что может произойти завтра, прогноз значительно проще и достовернее, нежели для того, что произойдет через год или через пять лет (рис. 1). Хотя, конечно, реальное развитие событий может оказаться и весьма далеким от прогнозируемого.

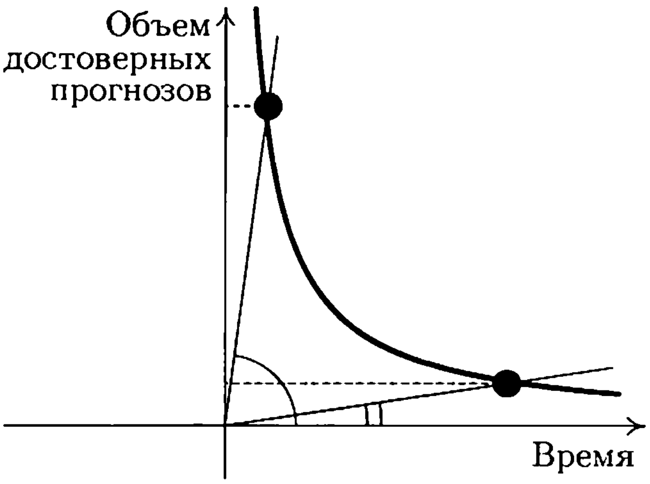


Рис. 1

Многие методы прогнозирования требуют наличия значительного количества начальных данных и при их отсутствии просто не работают. Другие, напротив, разрабатываются при условии отсутствия достоверной количественной информации. Тем самым существующие методы составления прогнозов можно условно разбить на две группы — количественные и качественные (рис. 2).

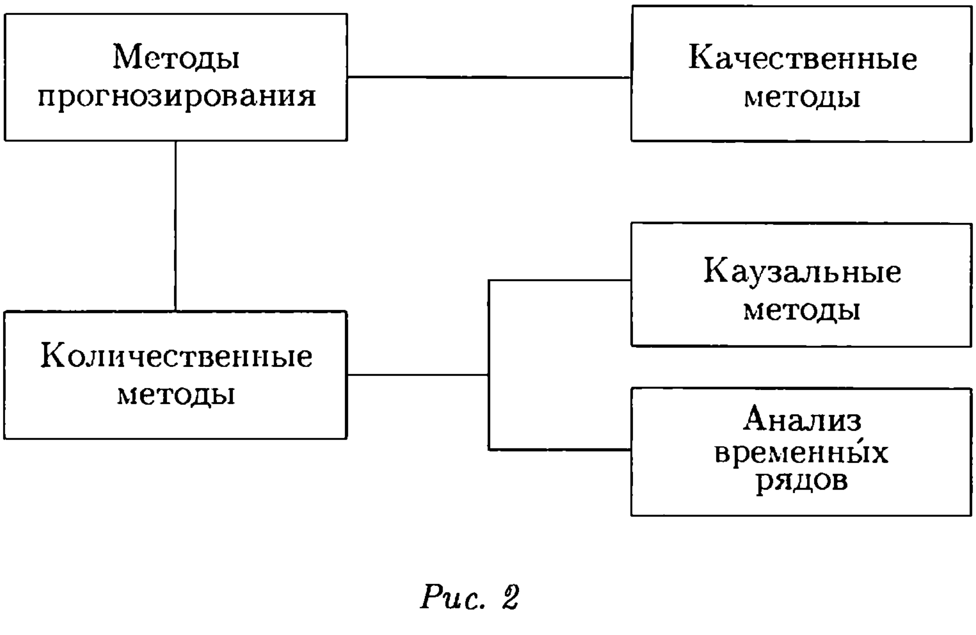


Рис. 2

Качественные, или *экспертные*, методы прогнозирования (qualitative methods) строятся на использовании мнений специалистов в соответствующих областях (экспертов).

Количественные методы прогнозирования (quantitative methods) основываются на обработке числовых массивов данных (как значительных по объему, так и сравнительно небольших) и в свою очередь разделяются на *каузальные*, или *причинно-следственные*, методы (causal metods) и методы *анализа временны́х рядов* (time series methods).

Основанный на допущении, в соответствии с которым происшедшее в прошлом дает хорошее приближение в оценке будущего, анализ временны́х рядов является способом выявления тенденций прошлого и продления их в будущее.

Каузальные методы применяются в тех случаях, когда искомое состояние зависит не только от времени, но и от нескольких, и даже многих переменных. Отыскание математических связей (уравнений и/или неравенств) между всеми этими переменными и составляет суть каузального метода прогнозирования.

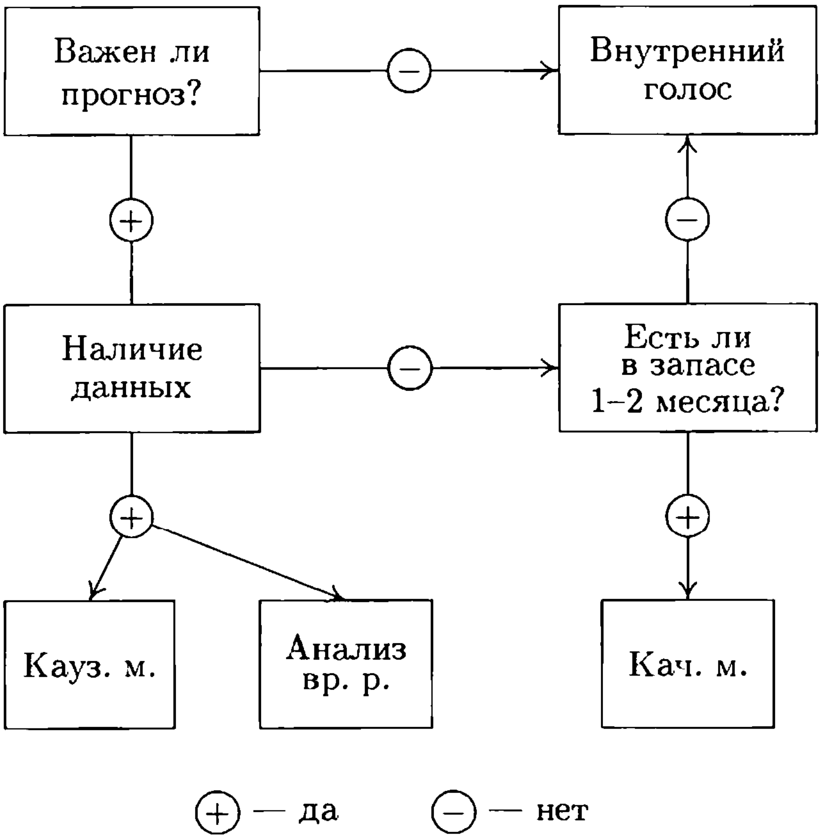


Рис. 3

На рис. 3 приведена логическая схема, пользуясь которой можно выбрать тот тип прогнозирования, который в наибольшей степени отвечает поставленной задаче и ее начальным условиям.

Далее мы остановимся на описании особенностей каждого из трех перечисленных выше типов прогнозирования более подробно, а также расскажем о некоторых конкретных методах составления прогнозов.

Одним из существенных критериев, которым часто руководствуются при выборе того или иного метода прогнозирования, является *полная стоимость прогноза*, слагающаяся из затрат на его составление и цены ошибки прогноза. Поэтому стремление заказчика сделать эту стоимость как можно меньшей нужно воспринимать совершенно естественно.

10.1. Анализ временны́х рядов

Временны́м (*динамическим*, или *хронологическим*) *рядом* называется последовательность значений некоторого показателя во времени (например, объемов продаж).

Различают два вида временны́х рядов — *моментные*, когда значения рассматриваемого показателя

$$x_1, x_2, \ldots, x_n$$

отнесены к определенным моментам времени (например, дням)

$$t_1, t_2, \ldots, t_n;$$

при этом обычно считается, что

$$t_1 < t_2 < \cdots < t_n,$$

и *интервальные*, когда указаны соответствующие промежутки времени, интервалы:

$$[t_0, t_1], (t_1, t_2], \ldots, (t_{n-1}, t_n].$$

Временны́е ряды чаще всего задаются при помощи таблицы:

моментный ряд

Момент времени	t_1	t_2	...	t_n
Значение показателя	x_1	x_2	...	x_n

интервальный ряд

Интервал времени	$[t_0, t_1]$	$(t_1, t_2]$	...	$(t_{n-1}, t_n]$
Значение показателя	x_1	x_2	...	x_n

или графически (рис. 4).

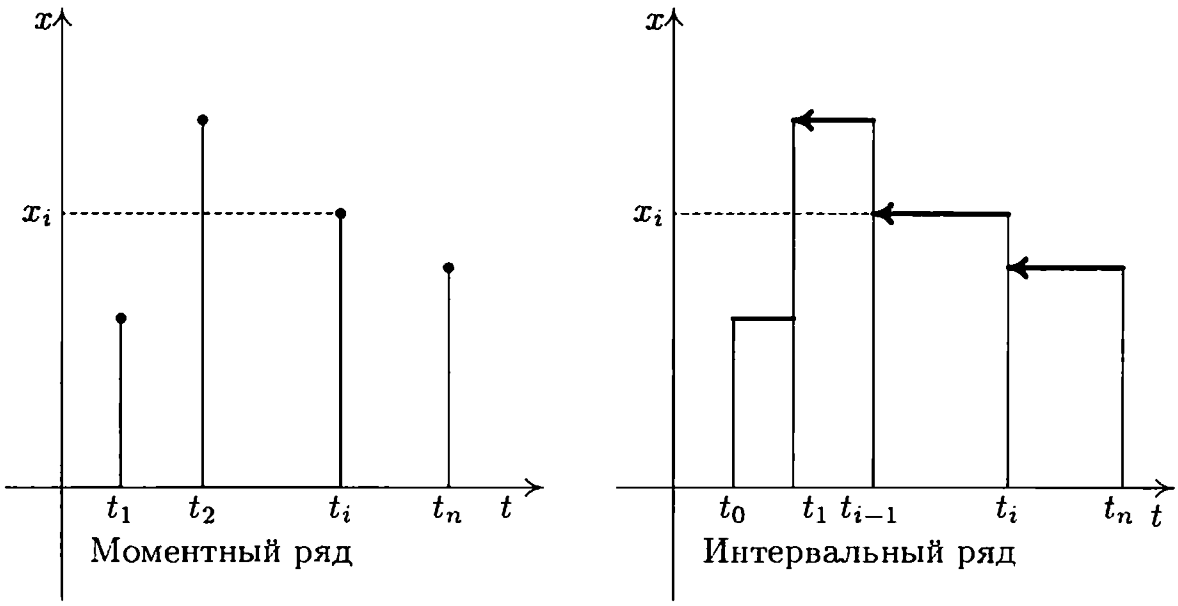


Рис. 4

В задачах прогнозирования временны́е ряды используются при наличии значительного количества реальных значений рассматриваемого показателя из прошлого и при условии, что наметившаяся в прошлом тенденция ясна и относительно стабильна. При этом неявно предполагается, что прошлое является хорошим проводником в будущее. Анализ временны́х рядов позволяет предопределить, что должно произойти при отсутствии вмешательства извне, и, значит, не может предсказать изменения тенденции. Тем самым, подобным анализом предпочтительнее пользоваться при составлении краткосрочных прогнозов.

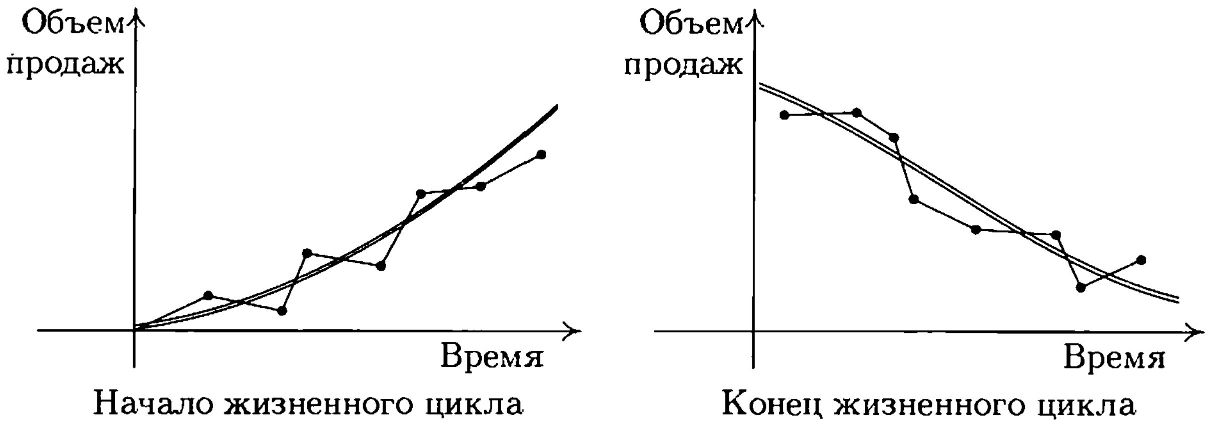


Рис. 5

Развитие процессов, реально наблюдаемых в жизни, складывается из некоторой устойчивой тенденции (тренда) и некоторой случайной составляющей, выражающейся в колебании значений показателя вокруг тренда. На рис. 5 показано, как могут зависеть объемы продаж одного и того же товара на двух стадиях его жизненного цикла (в начале и в конце продаж).

Кривые тренда сглаживают динамический ряд значений показателя, выделяя общую тенденцию. Именно выбор кривой тренда, сам по себе являющийся довольно трудной задачей, во многом определяет результаты прогнозирования.

В большинстве случаев динамический ряд, кроме тренда и случайных отклонений от него, характеризуется еще *сезонными* и *циклическими* составляющими. Циклические составляющие отличаются от сезонных большей продолжительностью и непостоянством амплитуды. Обычная продолжительность сезонной компоненты измеряется днями, неделями или месяцами, а циклической — годами или десятками лет.

В дальнейшем мы не будем учитывать циклической составляющей и, считая для простоты, что тренд является линейным, обсудим три метода анализа временны́х рядов (рис. 6).

Рис. 6

При рассмотрении того, как работают эти методы, мы будем пользоваться одним и тем же моментным временны́м рядом.

Пример. Предположим, что объемы продаж товара в течение недели описываются временны́м рядом

День недели	**Количество проданной продукции**
Понедельник	10
Вторник	6
Среда	5
Четверг	11
Пятница	9
Суббота	8
Воскресенье	7

или несколько по-иному:

t	1	2	3	4	5	6	7
x	10	6	5	11	9	8	7

10.1.1. Метод подвижного (скользящего) среднего

Метод простого скользящего среднего (simple moving average) состоит в том, что расчет показателя на прогнозируемый момент времени строится путем усреднения значений этого показателя за несколько предшествующих моментов времени.

Обратимся к заданному временно́му ряду.

Для вычисления прогнозируемого объема продаж на четверг поступим следующим образом. Возьмем фактические данные за три предыдущих дня — понедельник, вторник и среду — и найдем их среднее арифметическое:

$$f_4 = \frac{10+6+5}{3} = \frac{21}{3} = 7.$$

Прогнозируемый объем продаж на пятницу вычисляется аналогичным образом по реальным показателям за три предшествующих дня — вторник, среду и четверг:

$$f_5 = \frac{6+5+11}{3} = \frac{22}{3} \approx 7{,}33.$$

Подобным же способом рассчитываются прогнозы на субботу, воскресенье и очередной понедельник:

$$f_6 = \frac{5+11+9}{3} = \frac{25}{3} \approx 8{,}33,$$

$$f_7 = \frac{11+9+8}{3} = \frac{28}{3} \approx 9{,}33,$$

$$f_8 = \frac{9+8+7}{3} = \frac{24}{3} = 8.$$

И мы получаем следующую таблицу:

t	1	2	3	4	5	6	7	8
x	10	6	5	11	9	8	7	—
f	—	—	—	7	7,33	8,33	9,33	8

Сравнительные результаты приведены на рис. 7: темными кружками отмечены реальные значения, а светлыми — прогнозируемые.

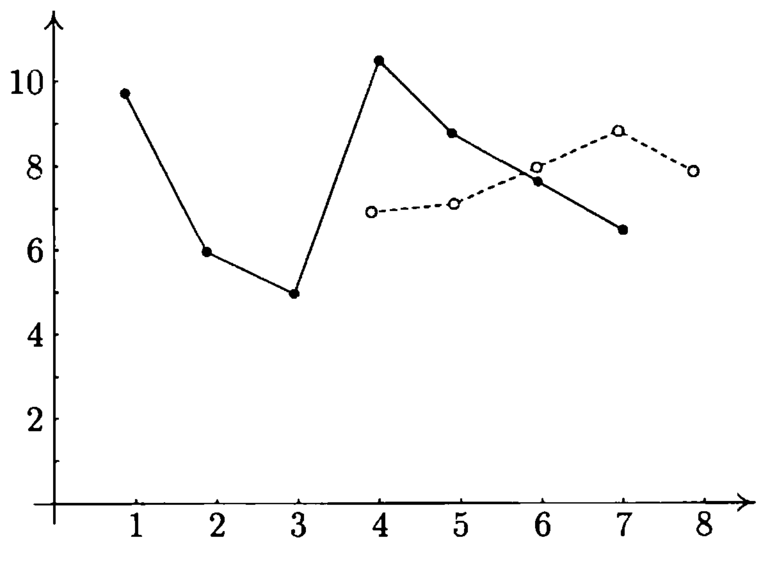


Рис. 7

Для общего случая расчетная формула выглядит так:

$$f_k = \frac{x_{k-N} + x_{k-N+1} + \cdots + x_{k-1}}{N},$$

или

$$f_k = \frac{1}{N} \sum_{i=1}^{N} x_{k-i}, \tag{1}$$

где

x_{k-i} — реальное значение показателя в момент времени t_{k-i};

N — число предшествующих моментов времени, используемых при расчете;

f_k — прогноз на момент времени t_k.

Замечание. В рассматриваемом примере $N = 3$.

Метод взвешенного подвижного (скользящего) среднего (weighted moving average). При составлении прогноза методом усреднения часто приходится наблюдать, что влияние используемых при расчете реальных показателей оказывается неодинаковым, при этом обычно более свежие данные имеют больший вес.

Математически метод взвешенного подвижного среднего можно записать так:

$$f_k = \frac{\sum_{i=1}^{N} w_{k-i} x_{k-i}}{\sum_{i=1}^{N} w_{k-i}},$$

где

x_{k-i} — реальное значение показателя в момент времени t_{k-i};

N — число предшествующих моментов времени, используемых при расчете;

f_k — прогноз на момент времени t_k;

w_{k-i} — вес, с которым используется показатель x_{k-i} при расчете.

Замечание. Вес — это всегда положительное число. В случае, когда все веса одинаковы, мы получаем формулу (1).

Для расчетов обратимся к исходному временно́му ряду, считая, что при составлении прогноза на завтрашний день объем сегодняшних продаж мы возьмем с весом 60, вчерашних — с весом 30, а позавчерашних — с весом 10.

Имеем:

$$f_4 = \frac{10 \cdot 10 + 30 \cdot 6 + 60 \cdot 5}{10 + 30 + 60} = \frac{580}{100} = 5{,}80;$$

$$f_5 = \frac{10 \cdot 6 + 30 \cdot 5 + 60 \cdot 11}{100} = \frac{870}{100} = 8{,}70;$$

$$f_6 = \frac{10 \cdot 5 + 30 \cdot 11 + 60 \cdot 9}{100} = \frac{920}{100} = 9{,}20;$$

$$f_7 = \frac{10 \cdot 11 + 30 \cdot 9 + 60 \cdot 8}{100} = \frac{860}{100} = 8{,}60;$$

$$f_8 = \frac{10 \cdot 9 + 30 \cdot 8 + 60 \cdot 7}{100} = \frac{750}{100} = 7{,}50.$$

Результаты расчетов приведены в таблице:

t	1	2	3	4	5	6	7	8
x	10	6	5	11	9	8	7	—
f	—	—	—	5,8	8,7	9,2	8,6	7,5

и на рис. 8: темными кружками отмечены реальные значения, а светлыми — прогнозируемые.

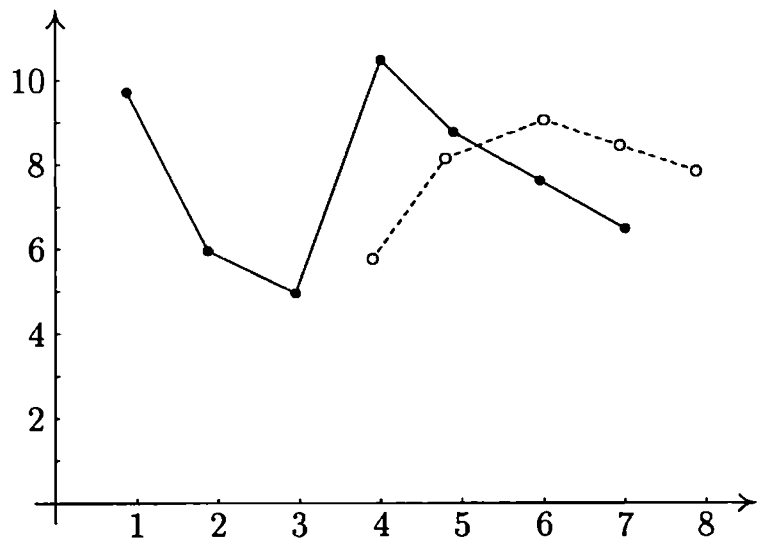


Рис. 8

10.1.2. Метод экспоненциального сглаживания

При расчете прогноза *методом экспоненциального сглаживания* (exponential smoothing) учитывается отклонение предыдущего прогноза от реального показателя, а сам расчет проводится по следующей формуле:

$$f_k = f_{k-1} + \alpha(x_{k-1} - f_{k-1}),$$

где

x_{k-1} — реальное значение показателя в момент времени t_{k-1};
f_k — прогноз на момент времени t_k;
α — постоянная сглаживания.

Замечание. Значение постоянной α, подчиненной условию $0<\alpha<1$, определяет степень сглаживания и обычно выбирается универсальным методом проб и ошибок.

Для расчетов вновь обратимся к исходному временно́му ряду, положив $\alpha = 0{,}2$ и считая, что прогноз на понедельник равен 8.

Тогда

$$\begin{aligned}
f_2 &= 8{,}00 + 0{,}2(10 - 8{,}00) = 8{,}40,\\
f_3 &= 8{,}40 + 0{,}2(\ 6\ - 8{,}40) = 7{,}92,\\
f_4 &= 7{,}92 + 0{,}2(\ 5\ - 7{,}92) \approx 7{,}34,\\
f_5 &= 7{,}34 + 0{,}2(11 - 7{,}34) \approx 8{,}07,\\
f_6 &= 8{,}07 + 0{,}2(\ 9\ - 8{,}07) \approx 8{,}26,\\
f_7 &= 8{,}26 + 0{,}2(\ 8\ - 8{,}26) \approx 8{,}21,\\
f_8 &= 8{,}21 + 0{,}2(\ 7\ - 8{,}21) \approx 7{,}93.
\end{aligned}$$

Результаты расчетов приведены в таблице:

t	1	2	3	4	5	6	7	8
x	10	6	5	11	9	8	7	–
f	–	8,4	7,92	7,34	8,07	8,26	8,21	7,93

и на рис. 9: темными кружками отмечены реальные значения, а светлыми — прогнозируемые.

Замечание. Следует иметь в виду, что при решении реальной задачи прогнозирования временно́й ряд складывается постепенно и реальное значение показателя на рассчитываемый момент времени нам заранее неизвестно. Тем не менее, прежде чем заглянуть в будущее

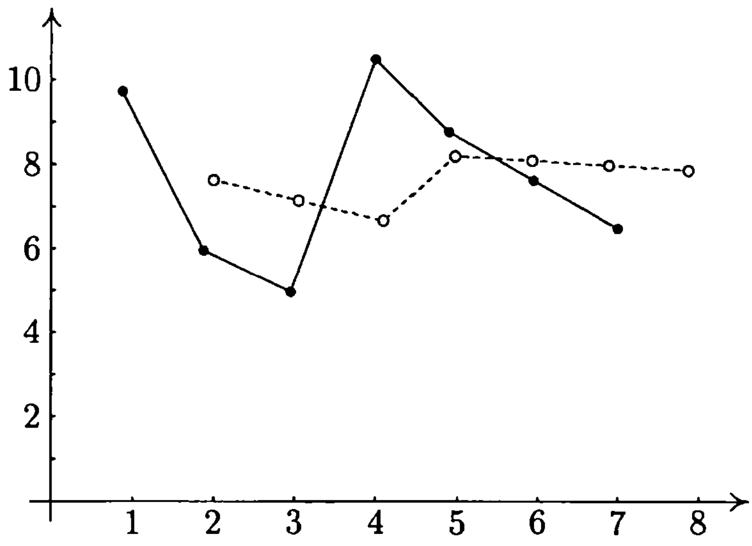


Рис. 9

посредством одного из указанных выше методов, обычно проводятся расчеты с полным временны́м рядом, описывающим некоторый промежуток времени в прошлом. Это делается для того, чтобы

подобрать подходящее значение N и сравнить результаты прогноза с реальными данными (метод простого скользящего среднего),

подобрать подходящие значения N и весов и сравнить результаты прогноза с реальными данными (метод взвешенного скользящего среднего),

подобрать подходящие значения постоянной сглаживания α и сравнить результаты прогноза с реальными данными (метод экспоненциального сглаживания).

10.1.3. Метод проецирования тренда

Основной идеей метода *проецирования (линейного) тренда* (trend projection) является построение прямой, которая "в среднем" наименее уклоняется от массива точек (t_i, x_i), $i = 1, 2, \ldots, n$, заданного временны́м рядом (рис. 10).

Эта прямая ищется в следующем виде:

$$x = at + b, \tag{2}$$

где a и b — постоянные, подлежащие определению.

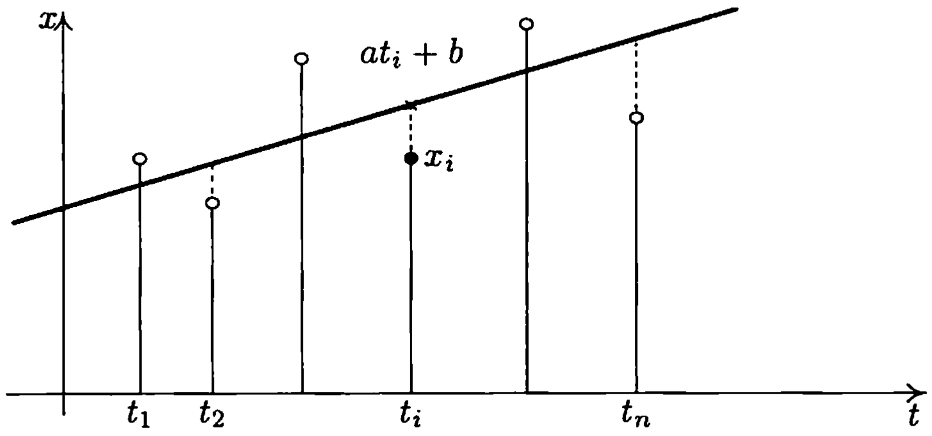


Рис. 10

Чтобы найти коэффициенты a и b, поступают так:

для каждого значения t_i переменной t, пользуясь формулой (2) вычисляют соответствующее значение переменной x:

$$at_i + b, \quad i = 1, 2, \dots, n,$$

находят разность

$$at_i + b - x_i, \quad i = 1, 2, \dots, n,$$

которую затем возводят в квадрат (чтобы не думать о знаке):

$$(at_i + b - x_i)^2, \quad i = 1, 2, \dots, n,$$

и, складывая, в итоге получают:

$$\varphi(a, b) = \sum_{i=1}^{n} (at_i + b - x_i)^2 \to \min.$$

Функция $\varphi(a, b)$ принимает минимальное значение в том случае, когда величины a и b удовлетворяют следующей линейной системе:

$$\begin{aligned} a\sum_{i=1}^{n} t_i + bn \qquad &= \sum_{i=1}^{n} x_i, \\ a\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i &= \sum_{i=1}^{n} t_i x_i. \end{aligned}$$

Эта система всегда имеет единственное решение.

Рассмотрим конкретный пример, вновь обратившись к заданному временно́му ряду.

Составим вспомогательную таблицу:

t_i	x_i	$t_i x_i$	t_i^2
1	10	10	1
2	6	12	4
3	5	15	9
4	11	44	16
5	9	45	25
6	8	48	36
7	7	49	49
$\sum t_i = 28$	$\sum x_i = 56$	$\sum t_i x_i = 223$	$\sum t_i^2 = 140$

В этом случае система уравнений для отыскания a и b записывается в следующем виде:

$$\begin{array}{rcrcr} 28a & + & 7b & = & 56, \\ 140a & + & 28b & = & 223. \end{array}$$

Решая систему, получаем:

$$a = -\frac{1}{28} \approx -0{,}04, \quad b = \frac{57}{7} \approx 8{,}14.$$

Тем самым

$$x = -0{,}04t + 8{,}14$$

— уравнение искомого тренда.

Расчет показателя на следующий день проводится так:

$$f_8 = -0{,}04 \cdot 8 + 8{,}14 = -0{,}32 + 8{,}14 = 7{,}82$$

(рис. 11).

Замечание. Точность прогноза можно оценить при помощи коэффициента корреляции.

Приведенные методы далеко не исчерпывают многообразия методов анализа временны́х рядов, большинство которых опирается не на простой подсчет при помощи калькулятора, но на основательную аналитическую и компьютерную базу. Однако наша цель состоит в том, чтобы дать определенное рабочее представление об этом типе прогнозирования.

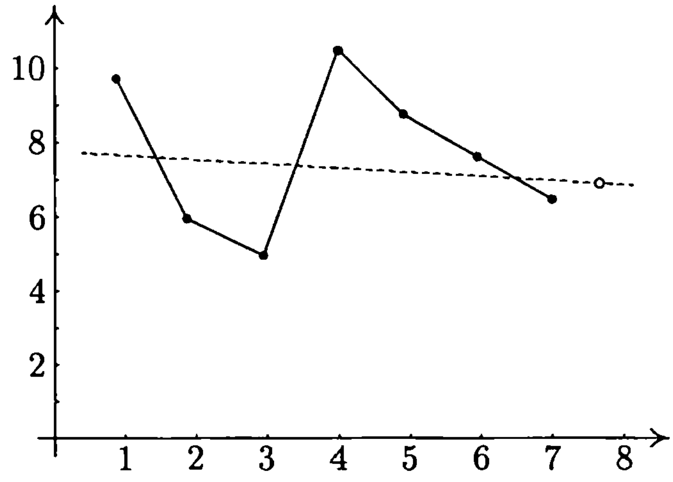


Рис. 11

Для составления среднесрочных и долгосрочных прогнозов применяются каузальные и качественные методы прогнозирования, которые значительно сложнее приведенных выше методов анализа временных рядов.

10.2. Каузальные методы прогнозирования

В случае значительных требований к точности прогноза и при наличии большого (даже огромного) массива данных используются *каузальные*, или *причинно-следственные, модели* прогнозов, в которых прогнозируемая величина является функцией большого числа переменных. Объемы продаж товара могут зависеть от цены продукта, затрат на рекламу, действий конкурентов, уровня доходов и других независимых переменных. Если связи между этими переменными удается описать математически корректно, то точность каузального прогноза может оказаться довольно высокой. Но как правило, это требует бо́льших объемов данных и существенно бо́льших интеллектуальных, временны́х и финансовых затрат, чем анализ временны́х рядов. К тому же расчет каузальных моделей связан с большими объемами вычислений, что возможно лишь при наличии мощной вычислительной техники. Мы ограничимся краткой характеристикой трех каузальных методов прогнозирования (рис. 12).

Многомерные регрессионные методы (модели) (multiple regression models), посредством которых регрессионная зависимость между величинами устанавливается по статистическим данным, являются наиболее распространенными количественными методами прогнози-

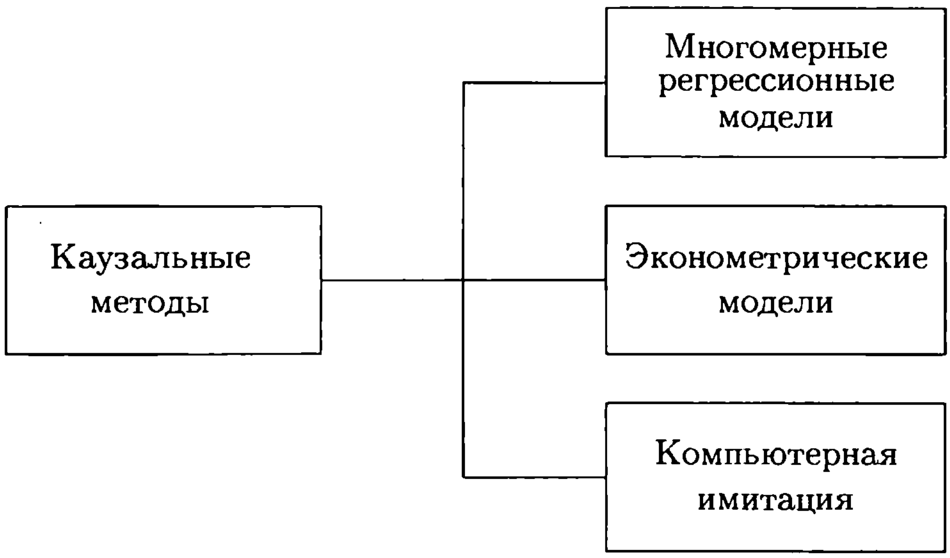


Рис. 12

рования. Простейшее представление о регрессионных моделях дает описанный выше метод проецирования тренда, в котором регрессионная зависимость устанавливается между прогнозируемым показателем и одной переменной — временем. Многомерные модели линейной регрессии можно рассматривать как естественное обобщение этого метода.

Эконометрические методы (модели) (econometric models) дают количественное описание закономерностей и взаимосвязей между экономическими объектами и процессами и разрабатываются для прогнозирования динамики экономики. Типичная эконометрическая модель представляет собой систему из тысяч уравнений, решение которой требует мощных вычислительных средств.

Компьютерная имитация (computer simulation). С появлением современных вычислительных средств уровень сложности математических моделей, при помощи которых можно делать правильные предсказания о динамике процессов, существенно вырос. Появились модели, способные создавать "иллюзию реальности". Называемые *имитационными*, эти модели являются как бы промежуточным звеном между реальностью и обычными математическими моделями. Имитационные модели находятся как бы на пределе возможностей вычислительной техники (и системного программирования).

Замечание. Всегда существуют процессы настолько сложные, что они не поддаются изучению математическими методами. Это не означает, однако, что они непознаваемы. Просто их рассматривают гуманитарными методами и средствами искусства — столь же необходимыми методами изучения реальности, как и математические методы. А подвижная граница между гуманитарными и математическими методами изучения (в том числе и прогнозирования) реальности проходит как раз по имитационными моделям в том понимании этого термина, о котором идет речь здесь.

10.3. Качественные методы прогнозирования

При отсутствии количественных данных, или когда количественная модель получается слишком дорогой, используются качественные методы прогнозирования (рис. 13), которые строятся на основе разного рода экспертных оценок.

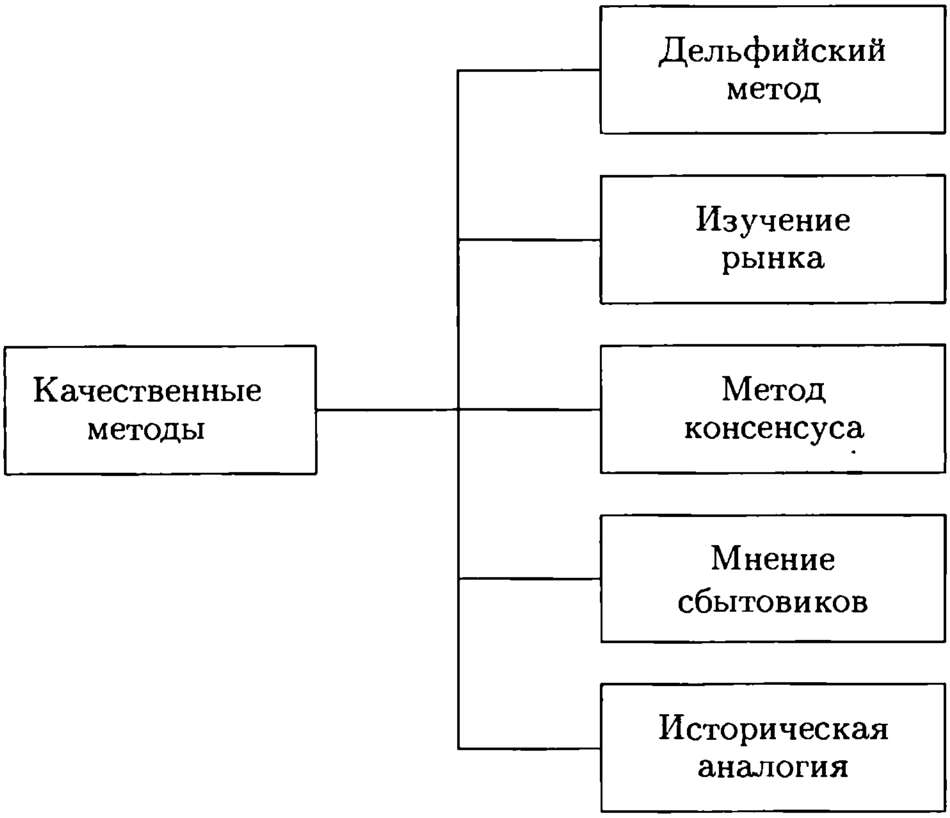


Рис. 13

Дельфийский метод (Delphi method), или *метод экспертных оценок*, представляет собой процедуру, позволяющую приходить к согласию группе экспертов из самых разных, но взаимосвязанных областей. Работа над составлением прогноза этим методом организуется так: каждому эксперту независимо рассылается вопросник по поводу рассматриваемой проблемы, ответы экспертов и их мнения кладутся в основу подготовки следующего вопросника, вновь рассылаемого экспертам, и так далее до тех пор, пока эксперты не приходят к согласию (при условии запрета на открытые дискуссии между экспертами). Обычно эта рассылка повторяется 3–4 раза.

Изучение рынка (market research), или *модель ожидания потребителя.* Прогноз строится на основании разнообразных опросов потребителей и последующей статистической обработки.

Метод консенсуса (panel consensus), или *мнение жюри*, заключается в соединении и усреднении мнений группы экспертов в процессе “мозгового штурма”.

Совокупное мнение сбытовиков (grass-roots forecasting). Метод опирается на мнение непосредственно контактирующих с потребителем торговых агентов.

Историческая аналогия (historical analogy) обычно используется в тех случаях, когда нужно дать прогноз продажи товара, по своим характеристикам близкого к выпущенному ранее (например, его модификации).

Сравнительные характеристики этих пяти методов приведены на рис. 14 и 15.

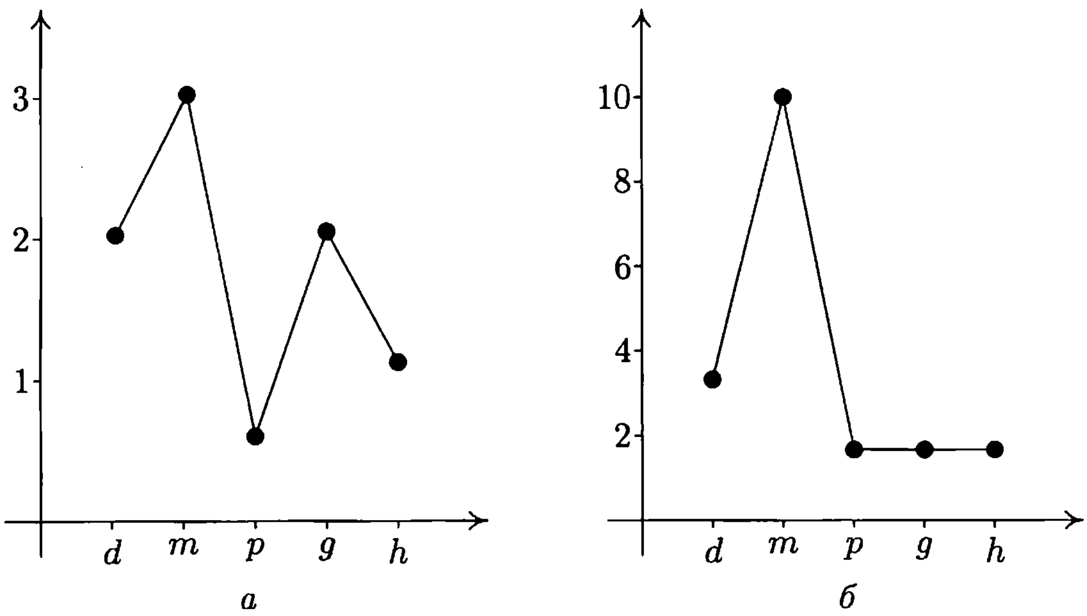


Рис. 14

Рис. 15

Пояснение к рисункам. На рис. 14*а* указано среднее время (в ме сяцах), требуемое для составления прогноза. На рис. 14*б* приведена средняя цена (в тыс. долл.) составления прогноза. На рис. 15 указана точность прогнозов в зависимости от сроков предсказания.

Обозначения:

d — Delphi method, m — market research, p — panel consensus, g — grass-roots forecasting, h — historical analogy;

пл. — плохая, *ср.* — средняя, *х.* — хорошая, *от.* — отличная, *пр.* — превосходная.

Часть II

СТОХАСТИЧЕСКИЕ МЕТОДЫ

Глава 11
СЛУЧАЙНЫЕ СОБЫТИЯ И ВЕРОЯТНОСТИ

11.1. О стохастическом моделировании

Стохастические (вероятностные) модели широко применяются в тех случаях, когда те или иные факторы носят неопределенный характер. Такие ситуации характерны для самых разных областей человеческой деятельности. Примерами могут служить погодные условия через несколько лет, спрос на какую-либо продукцию, политическая ситуация в данной стране и т. п. Для лучшего понимания рассматриваемых методов полезно иметь в виду, что неопределенность может иметь довольно разный характер. При этом логические рассуждения не создают информацию из ничего, а структурируют уже имеющуюся. В этой главе мы изучим основы методики анализа неопределенности.

11.2. Различные подходы к понятию вероятности

Понятие случайного события является основополагающим в изучении вероятностных методов и моделей. Под *случайным* будем понимать событие, которое может произойти или не произойти в результате некоторого *испытания*. При этом испытанием может быть как целенаправленное действие, так и явление, происходящее независимо от наблюдателя. В дальнейшем случайные события будем называть просто *событиями*.

Приведем несколько примеров.

Пример 1. Испытание — бросание монеты. Возможные события — выпадание герба или цифры.

Пример 2. Наступает день 12 января — испытание. "В течение дня наблюдается ясная погода" — событие.

Пример 3. Студент сдает экзамен — испытание. "Он получил оценку 5" — событие.

Каждому событию может быть поставлено в соответствие число, принадлежащее отрезку $[0, 1]$ и называемое *вероятностью* данного события. Вероятность можно понимать как меру достоверности (в том числе и субъективной) данного события. В таком смысле слово "вероятность" употребляется и в бытовой речи, где, однако, ее обычно "измеряют" в процентах — от 0 до 100%. Вероятность обычно обозначают буквой p (от англ. probability — вероятность). Чем более достоверным представляется наступление события, тем больше его вероятность. Вероятность невозможного события считается равной нулю, вероятность абсолютно достоверного события — единице.

Для определения вероятностей событий возможны различные подходы.

Начнем с рассмотрения ситуации, когда в результате испытания может произойти один из некоторого конечного множества *равновозможных исходов* (*пространства исходов*). Если общее число исходов (или, иначе говоря, элементарных событий) равно n, то каждому из них приписывается вероятность $1/n$.

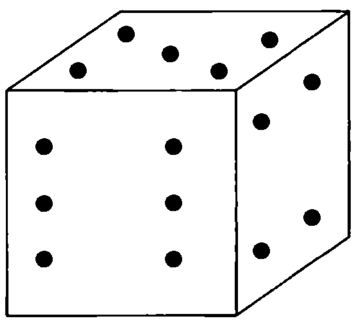


Рис. 1

Пример 4. Бросается игральный кубик, на грани которого нанесено разное число точек — от 1 до 6 включительно (рис. 1). Тогда исходов будет шесть: "выпало число 1", "выпало число 2", ..., "выпало число 6". Коротко пространство исходов можно записать следующим образом:

$$\{1, 2, 3, 4, 5, 6\}.$$

Вероятность выпадания каждого из этих чисел равна 1/6 (как говорят, "один шанс из шести").

Событием можно считать любое подмножество пространства исходов, и, наоборот, любое событие является подмножеством пространства исходов. Будем говорить, что событие A произошло, если результат (исход) испытания принадлежит множеству A. (Здесь и далее события будем обозначать, как правило, прописными латинскими буквами). Продолжая пример 4, можно заметить, что событию A_1 = "выпало четное число очков" соответствует подмножество $\{2,\ 4,\ 6\}$ пространства исходов, а событию A_2 = "выпало число очков, большее двух" соответствует подмножество $\{3,\ 4,\ 5,\ 6\}$.

Посмотрим теперь на ситуацию с более общей точки зрения.

Классическая вероятность

Пусть n — число всех равновозможных исходов, а m — число исходов, составляющих событие A. Вероятность события A (обозначение $p(A)$) определяется следующим образом:

$$p(A) = \frac{m}{n}.$$

Это так называемая *классическая вероятность*. В частности, для упомянутых выше событий A_1 и A_2 имеем:

$$p(A_1) = \frac{1}{2}, \quad p(A_2) = \frac{2}{3}.$$

Подчеркнем, что формула классической вероятности предполагает конечность числа исходов n. Обратимся теперь к случаю, когда число исходов бесконечно.

Геометрическая вероятность

Пусть на плоскости имеется фигура F, содержащая фигуру f (рис. 2). Испытание заключается в том, что в фигуру F наугад бросается точка. Тем самым пространство исходов можно отождествить с этой фигурой. Здесь число исходов бесконечно (у фигуры F бесконечно много точек), притом все исходы имеют одинаковые шансы осуществиться. Определим A как событие, заключающееся в том, что брошенная точка попала в фигуру f. Тогда вероятность события A (*геометрическая вероятность*) определяется следующим образом:

$$p(A) = \frac{S_f}{S_F},$$

где S_F и S_f — площади фигур F и f соответственно.

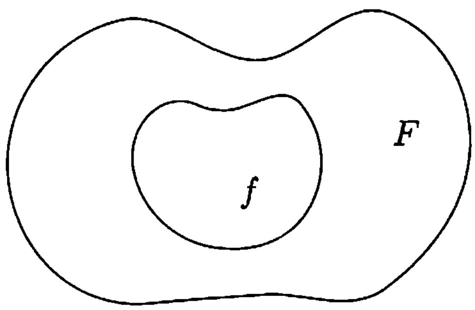


Рис. 2

Аналогично определяется геометрическая вероятность на прямой и в пространстве, только вместо площадей фигур в формуле для вероятности надо поставить соответственно длины и объемы.

Пример 5. В результате урагана был оборван телефонный кабель между 20-м и 60-м километрами линии. Какова вероятность того, что обрыв произошел между 30-м и 35-м километрами?

Здесь $l_F = 60 - 20 = 40$, а $l_f = 35 - 30 = 5$. Значит, $p = 5/40 = 1/8$.

Статистическая вероятность

Предположим, что событие A может произойти либо не произойти в результате некоторого эксперимента. Повторим эксперимент n раз и подсчитаем, сколько раз произошло событие A. Пусть это число равно m. Отношение m/n назовем *относительной частотой появления события A в n испытаниях*. Если при достаточно больших значениях n относительные частоты группируются около некоторой постоянной, то эту постоянную будем считать *статистической вероятностью* события A:

$$p(A) \approx \frac{m}{n} \qquad \text{при больших } n.$$

Пример 6. Если подбросить монету n раз и подсчитать число m выпадений герба, то при достаточно большом n отношение m/n будет близко к 0,5 (если монета симметричная — не гнутая, не смещен центр тяжести и пр.).

Субъективная вероятность

Во многих реальных ситуациях определение вероятности событий одним из приведенных выше способов невозможно. Тогда на первый план выступает отмеченное выше понимание вероятности как меры достоверности того или иного события. В этом случае следует провести экспертный опрос и на основе его результатов получить *субъективную вероятность* события.

Пример 7. Какова вероятность того, что некто станет президентом на ближайших выборах? Ясно, что здесь может идти речь о вероятности только в субъективном смысле.

Замечание. С принятием некоторого числа в качестве субъективной вероятности связаны два достаточно независимых действия. Во-первых, требуется правильно провести опрос и, во-вторых, надо правильно учесть уже высказанное мнение экспертов. При этом возникает ряд психологических и математических проблем. Их обсуждение, однако, выходит за рамки этой книги.

11.3. Формулы алгебры событий. Несовместимые и независимые события

Если *определены* вероятности элементарных событий, можно переходить к *вычислению* вероятностей более сложных событий, являющихся комбинацией определенных ранее элементарных.

Предположим, что с некоторым испытанием связаны события A и B. Назовем их *суммой* событие, заключающееся в том, что произошло хотя бы одно из событий — или A, или B (обозначение: $A+B$).

Пример 8. Пусть A = "это случилось в сентябре...", B = "это случилось в октябре...", C = "это случилось в ноябре...". Тогда $(A+B+C)$ = "это случилось осенью...".

Произведением событий A и B назовем событие, состоящее в совместном наступлении этих событий (обозначение: AB).

Пример 9. Пусть $A =$ "в аудиторию вошел студент", $B =$ "в аудиторию вошел человек в темных очках". Тогда $AB =$ "в аудиторию вошел студент в темных очках".

Событием, *противоположным* A, назовем событие, состоящее в том, что A не произошло (обозначение: $\bar{A}$, "не A").

Пример 10. Пусть испытанием является бросок баскетболиста по кольцу, $A =$ "баскетболист попал". Тогда $\bar{A} =$ "баскетболист не попал".

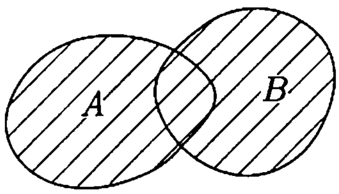


Рис. 3

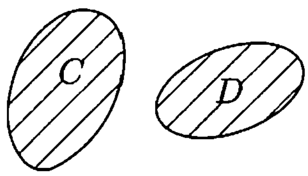


Рис. 4

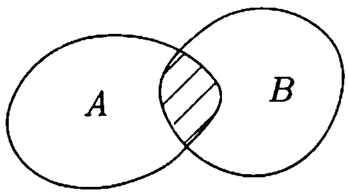


Рис. 5

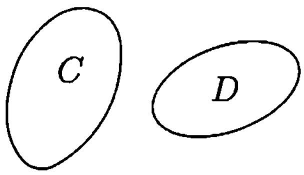


Рис. 6

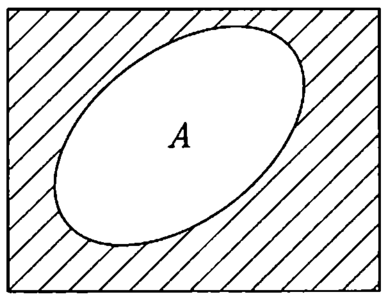


Рис. 7

Введенные понятия допускают простую геометрическую интерпретацию. Пусть испытанием является бросание точки в область на плоскости, а событиями A, B, C и D — попадание точки в области, которые мы обозначим теми же буквами — A, B, C и D соответственно. Тогда сумма событий A и B — область, заштрихованная на рис. 3, сумма $C + D$ — на рис. 4. Произведение AB иллюстрируется на рис. 5, произведение CD — на рис. 6, событие, противоположное A, — на рис. 7. Заметим, что произведение CD является невозможным событием, $CD = \varnothing$.

Перейдем теперь к вычислению вероятностей событий $A + B$, AB и $\bar{A}$, считая известными вероятности событий A и B.

Вероятность события $\bar{A}$ вычисляется легко:

$$p(\bar{A}) = 1 - p(A).$$

Для вероятности события $A + B$ справедлива следующая формула:

$$p(A + B) = p(A) + p(B) - p(AB). \tag{1}$$

В это соотношение входит пока неизвестная нам вероятность произведения AB. Впрочем, часто слагаемое $p(AB)$ оказывается равным 0. Рассмотрим эту ситуацию подробнее.

Если события A и B не могут произойти одновременно в результате одного испытания (иными словами, если AB — невозможное событие), то их называют *несовместимыми*, и тогда $p(AB) = 0$. Если же события могут произойти в результате одного испытания, то их называют *совместимыми.*

Пример 11. События A и $\bar{A}$ несовместимы.

Пример 12. События A и B на рис. 3, 5 совместимы.

Пример 13. События C и D на рис. 4, 6 несовместимы.

Для случая несовместимых событий формула (1) приобретает особенно простой вид:

$$p(A + B) = p(A) + p(B). \tag{2}$$

Для отыскания вероятности произведения событий необходимо ввести еще одно ключевое понятие.

При совместном рассмотрении двух случайных событий часто возникает вопрос: насколько наступление одного из них влияет на возможность наступления другого? Простейший вид связи — причинный. Однако бывает, что хотя причинной связи нет, но некоторая зависимость все же присутствует.

Рассмотрим испытание, состоящее в однократном бросании игрального кубика. Пусть событие A = "выпало четное число очков", событие B = "выпало число очков, большее трех". Неверно утверждать, что одно из событий с неизбежностью влечет за собой другое. Но некоторая зависимость имеется. Действительно, пространство исходов составляют шесть чисел: $\{1, 2, 3, 4, 5, 6\}$. Событие A составляют исходы $\{2, 4, 6\}$, его вероятность равна $1/2$. Событие B составляют исходы $\{4, 5, 6\}$. Из этих трех исходов ровно два (исходы 4 и 6) входят в событие A. Таким образом, если событие B произошло, то вероятность события A равна $2/3$ (рис. 8).

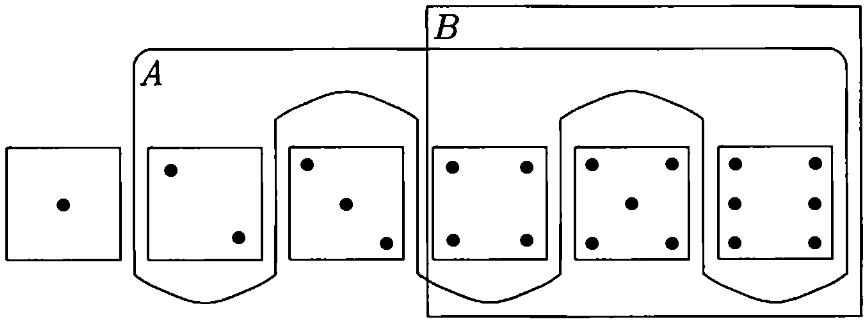


Рис. 8

Такая вероятность, называемая *условной*, обозначается $p(A|B)$, "вероятность A при условии B" (считаем, что $p(B) \neq 0$). В данном случае $p(A|B) = 2/3$.

Вообще, если наступление события B изменяет вероятность события A, то такие события называются *зависимыми*. Если же наступление одного события никак не влияет на шансы наступления другого, то такие события называются *независимыми*.

Для любых событий A и B (как независимых, так и зависимых) справедлива следующая формула:

$$p(AB) = p(A|B)p(B). \tag{3}$$

Если события A и B независимы (при этом $p(A|B) = p(A)$ и $p(B|A) = p(B)$), то формула (3) упрощается:

$$p(AB) = p(A)p(B). \tag{4}$$

Справедливо и обратное: если выполняется равенство (4), то события A и B независимы.

Пример 14. Рассмотрим еще одно событие, связанное с бросанием кубика: $C = \{1, 2, 3, 4\}$. Нетрудно проверить (например, воспользовавшись формулой (4)), что события A и C являются независимыми, а события B и C — зависимыми.

Замечание. Обычно, однако, вопрос о том, являются ли данные события независимыми (и тогда можно применить формулу (4)), решается на основании здравого смысла.

Понятия несовместимости и независимости можно обобщить на случай более чем двух событий.

События $A_1, \ldots, A_n$ называются *попарно несовместимыми*, если появление в результате испытания одного из них исключает возможность появления любого другого (рис. 9). В этом случае справедливо

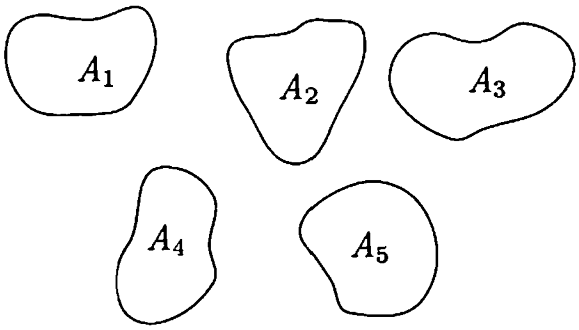


Рис. 9

следующее обобщение формулы (2):

$$p(A_1 + \cdots + A_n) = p(A_1) + \cdots + p(A_n). \tag{5}$$

События $A_1, \ldots, A_n$ называются *независимыми в совокупности*, если вероятность любого из них A_i, $1 \leqslant i \leqslant n$, не меняется при наступлении какого угодно числа событий A_j, $j \neq i, 1 \leqslant j \leqslant n$, из той же совокупности. В этом случае справедлива формула

$$p(A_1 \ldots A_n) = p(A_1) \ldots p(A_n), \tag{6}$$

являющаяся обобщением соотношения (4).

Следующая формула справедлива для независимых в совокупности событий $A_1, A_2, \ldots, A_n$:

$$p(A_1 + A_2 + \cdots + A_n) = 1 - p(\bar{A}_1)\, p(\bar{A}_2) \ldots p(\bar{A}_n). \tag{7}$$

Пример 15. Четыре стрелка одновременно стреляют по цели. Вероятности попадания в цель для каждого стрелка известны: 0,7; 0,75; 0,7 и 0,65 соответственно. Чему равна вероятность того, что цель будет поражена (хотя бы одним стрелком)?

Решение. Обозначим за A_i $(i = 1, 2, 3, 4)$ событие, состоящее в том, что i-й стрелок попал в цель. Эти события независимы в совокупности, их вероятности по условию таковы:

$$p(A_1) = 0{,}7; \qquad p(A_3) = 0{,}7;$$
$$p(A_2) = 0{,}75; \qquad p(A_4) = 0{,}65.$$

Цель будет поражена (событие A), если попадет хотя бы один из стрелков:

$$A = A_1 + A_2 + A_3 + A_4.$$

Вычисляя вероятность, получаем:

$$\begin{aligned} p(A) &= p(A_1 + A_2 + A_3 + A_4) = \\ &= 1 - p(\bar{A}_1)\, p(\bar{A}_2)\, p(\bar{A}_3)\, p(\bar{A}_4) = \\ &= 1 - 0{,}3 \cdot 0{,}25 \cdot 0{,}3 \cdot 0{,}35 = 0{,}992\,125. \end{aligned}$$

11.4. Примеры вычисления вероятностей

Перейдем к рассмотрению важного вопроса: как вычислять вероятности сложных событий, если известны вероятности простых? Подчеркнем еще раз, что вероятности простых событий определяются предварительно в классическом, геометрическом либо субъективном понимании.

Единого алгоритма решения произвольной вероятностной задачи не существует. Рассмотрим два взаимодополняющих метода — применение формул (*аналитический* метод) и применение дерева вероятностей (*графический* метод). Рассмотрение будем вести на примерах.

Пример 16. Известно, что 5% изделий некоторой фирмы бракованные. Взяли наугад на проверку два изделия. Какова вероятность того, что одно из этих двух изделий будет забраковано?

Решение 1. Обозначим за $Б_1$ ($Б_2$) событие, состоящее в том, что первое (второе) изделие оказалось бракованным. Тогда $\bar{Б}_1$ ($\bar{Б}_2$) — противоположное событие, состоящее в том, что первое (второе) изделие удовлетворяет стандартным требованиям качества. Интересующее нас событие A = {"одна деталь бракованная"} можно представить следующим образом: A = {"первая деталь бракованная" *и* "вторая деталь не бракованная" *или* "первая деталь не бракованная" *и* "вторая деталь бракованная"}. Вспомнив, что логическим *и, или, не* соответствуют в формулах алгебры событий умножение, сложение, противоположное событие, запишем:

$$A = Б_1\bar{Б}_2 + \bar{Б}_1 Б_2.$$

Теперь перейдем к вычислению вероятности события A. Заметим, что:

1) события $Б_1\bar{Б}_2$ и $\bar{Б}_1 Б_2$ несовместимы;
2) события $Б_1$ и $\bar{Б}_2$, а также $\bar{Б}_1$ и $Б_2$ независимы.

Поэтому

$$p(A) = p(\text{Б}_1\bar{\text{Б}}_2) + p(\bar{\text{Б}}_1\text{Б}_2) = p(\text{Б}_1)p(\bar{\text{Б}}_2) + p(\bar{\text{Б}}_1)p(\text{Б}_2). \qquad (8)$$

Теперь осталось подставить вероятности "простых" событий Б_1 и Б_2. По условию 5% изделий бракованные. Поэтому

$$p(\text{Б}_1) = p(\text{Б}_2) = 0{,}05;$$
$$p(\bar{\text{Б}}_1) = p(\bar{\text{Б}}_2) = 1 - 0{,}05 = 0{,}95.$$

Подставляя в формулу (8), получаем

$$p(A) = 0{,}05 \cdot 0{,}95 + 0{,}95 \cdot 0{,}05 = 0{,}095.$$

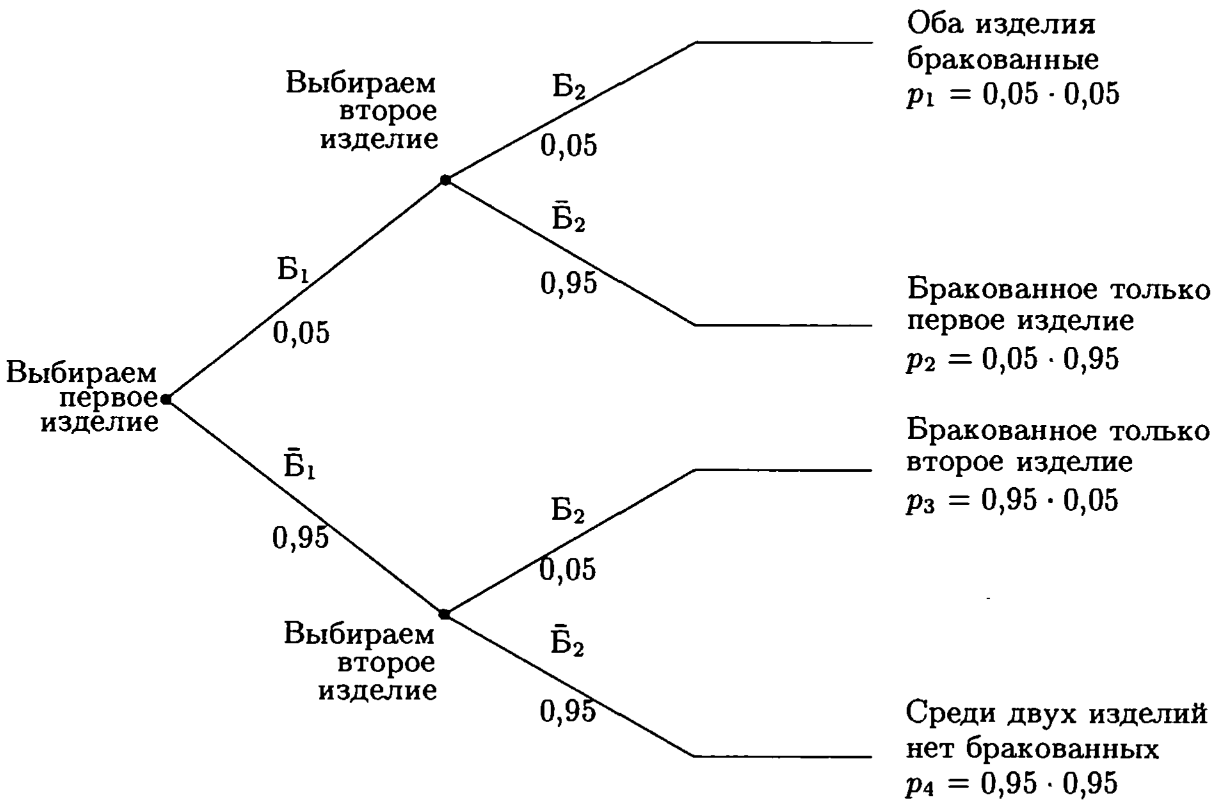


Рис. 10

Решение 2. Построим так называемое *дерево вероятностей*, учитывающее все возможные исходы (рис. 10). Здесь вершинам дерева (кроме концевых) соответствуют испытания, а ребрам — события. Сначала рассмотрим первое изделие. При этом возможны два исхода — изделие может оказаться бракованным (с вероятностью 0,05) либо качественным (с вероятностью 0,95). В каждом из этих случаев рассмотрим второе изделие, которое тоже может быть либо бракованным, либо качественным (с теми же вероятностями).

В результате получаем четыре возможности, обозначаемые концевыми вершинами дерева. К каждой из этих возможностей ведет путь из начальной точки, состоящий из двух ребер дерева. Для нахождения вероятностей p_1, p_2, p_3, p_4 перемножаются вероятности ребер соответствующего пути.

Искомая вероятность вычисляется как сумма p_2 и p_3:

$$p_2 + p_3 = 0{,}095.$$

Замечание. Поскольку все возможные исходы в сумме составляют достоверное событие, то суммарная вероятность всегда равна единице. В данном случае

$$p_1 + p_2 + p_3 + p_4 = 1.$$

Пример 17. Через остановку пролегают троллейбусный и автобусный маршруты. Троллейбус подъезжает через каждые 15 минут, автобус — через каждые 25 минут. К остановке подходит пассажир. Какова вероятность того, что в ближайшие 10 минут на остановке появится троллейбус либо автобус?

Решение 1. Пассажир подошел к остановке в некоторый случайный момент между двумя последовательными приездами троллейбуса. По условию троллейбус подъезжает через каждые 15 минут. По формуле геометрической вероятности найдем вероятность $p(T)$ того, что троллейбус появится на остановке в ближайшие 10 минут:

$$p(T) = \frac{10}{15} = \frac{2}{3}.$$

Вероятность $p(A)$ того, что в ближайшие 10 минут на остановку подъедет автобус, такова:

$$p(A) = \frac{10}{25} = \frac{2}{5}.$$

Нас интересует вероятность события $S =$ "в ближайшие 10 минут на остановку подъедет троллейбус либо автобус". Ясно, что

$$S = T + A.$$

Поскольку события T и A являются независимыми, можно воспользоваться формулой (7). Имеем:

$$p(S) = p(T + A) = 1 - p(\bar{T})p(\bar{A}) = 1 - \frac{1}{3} \cdot \frac{3}{5} = 1 - \frac{1}{5} = \frac{4}{5} = 0{,}8.$$

Таким образом, с вероятностью 0,8 пассажир уедет с остановки в ближайшие 10 минут.

Замечание 1. Вместо обращения к формуле (7) можно было рассуждать несколько иначе. Пассажир не уедет с остановки в ближайшие 10 минут, если не приедут ни троллейбус, ни автобус, т. е. если произойдет событие $\bar{T}\bar{A}$. События $\bar{T}$ и $\bar{A}$ независимы, поэтому

$$p(\bar{T}\bar{A}) = p(\bar{T})p(\bar{A}) = \frac{1}{3} \cdot \frac{3}{5} = \frac{1}{5} = 0{,}2.$$

Вероятность же того, что пассажир уедет, составляет

$$1 - 0{,}2 = 0{,}8.$$

Замечание 2. Еще один способ рассуждений состоит в применении формулы (1):

$$p(S) = p(T + A) = p(T) + P(A) - p(TA) =$$
$$= \frac{2}{3} + \frac{2}{5} - \frac{2}{3} \cdot \frac{2}{5} = \frac{16}{15} - \frac{4}{15} = \frac{12}{15} = 0{,}8.$$

(Разумеется, события T и A совместимы — могут подъехать и троллейбус, и автобус.) Заметим, однако, что в случае суммы более чем двух событий применима лишь формула (7). Хотя формула (1) и допускает обобщение на случай большего числа событий, но получающиеся выражения довольно громоздки. Например, для трех событий имеем:

$$p(A+B+C) = p(A)+p(B)+p(C)-p(AB)-p(BC)-p(AC)+p(ABC).$$

Решение 2. Построим дерево вероятностей (рис. 11). Интересующая нас вероятность вычисляется как сумма вероятностей попарно несовместимых событий:

$$p_1 + p_2 + p_3 = \frac{2}{3} \cdot \frac{2}{5} + \frac{2}{3} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{12}{15} = 0{,}8.$$

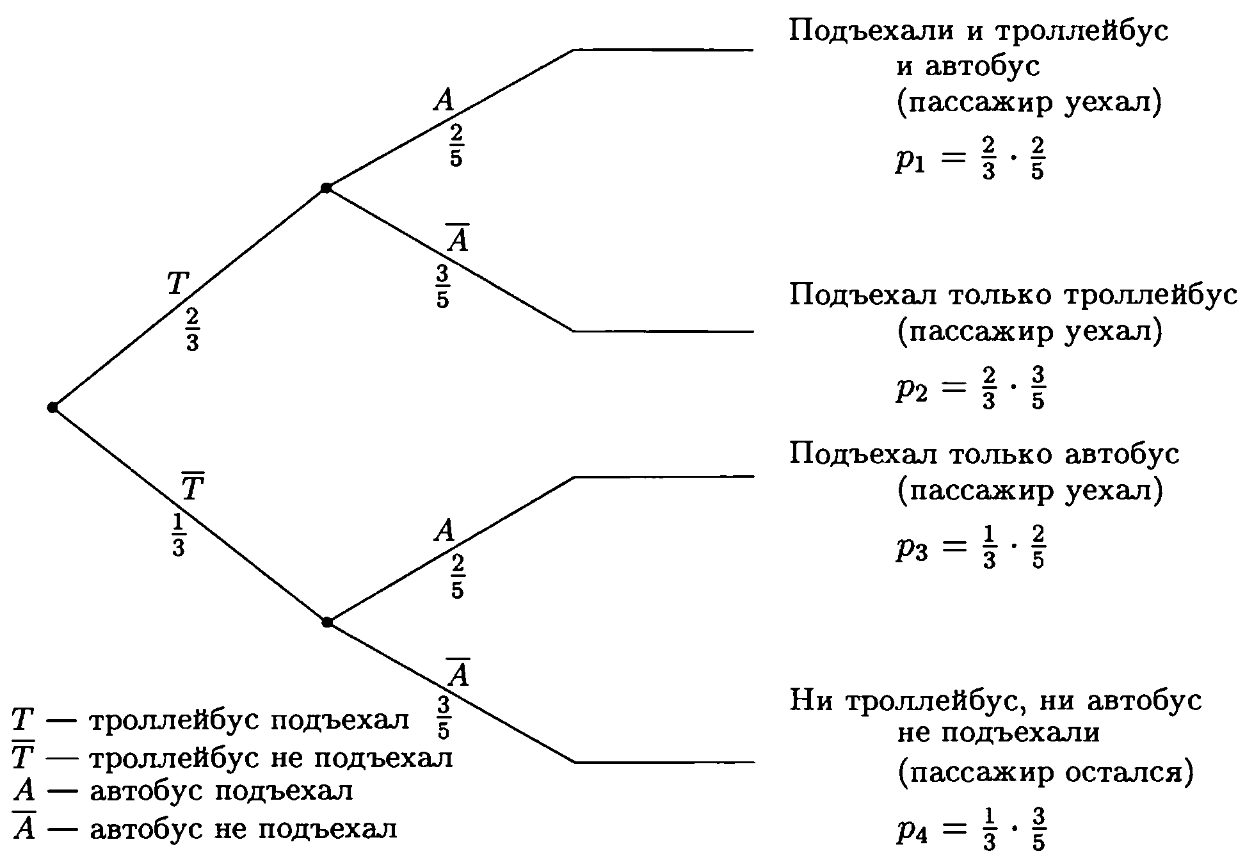


Рис. 11

Пример 18. Студент пришел на зачет, зная 15 вопросов из 20. Если студент не может ответить, ему предоставляется еще одна (но не более!) попытка. Какова вероятность сдать зачет?

Решение 1. Введем следующие обозначения:

A_1 – студент сразу вытянул знакомый билет (и сдал зачет);

$\bar{A}_1$ – студент вытянул незнакомый билет (еще одна попытка);

A_2 – студент со второго раза наконец-то вытянул знакомый билет (и сдал зачет);

$\bar{A}_2$ – студент и во второй раз вытянул незнакомый билет (и ему предстоит пересдача);

A – студент сдал зачет.

Студент сдает зачет, если он либо сразу вытянул знакомый билет, либо сначала незнакомый, а во второй раз знакомый билет. Формально это можно записать следующим образом:

$$A = A_1 + \bar{A}_1 A_2.$$

Переходя к вероятностям, получаем:

$$p(A) = p(A_1) + p(\bar{A}_1)p(A_2|\bar{A}_1) = \frac{15}{20} + \frac{5}{20} \cdot \frac{15}{19} = \frac{18}{19}.$$

(Условная вероятность $p(A_2|\bar{A}_1)$ равна 15/19, поскольку во второй попытке "участвует" 19 билетов.)

Замечание. Можно было рассуждать иначе. Студент не сдает зачет, если и в первый, и во второй раз вытянет незнакомый билет:

$$\bar{A} = \bar{A}_1\bar{A}_2.$$

Поэтому

$$p(\bar{A}) = p(\bar{A}_1)p(\bar{A}_2|\bar{A}_1) = \frac{5}{20}\cdot\frac{4}{19} = \frac{1}{19}.$$

Отсюда

$$p(A) = 1 - \frac{1}{19} = \frac{18}{19}.$$

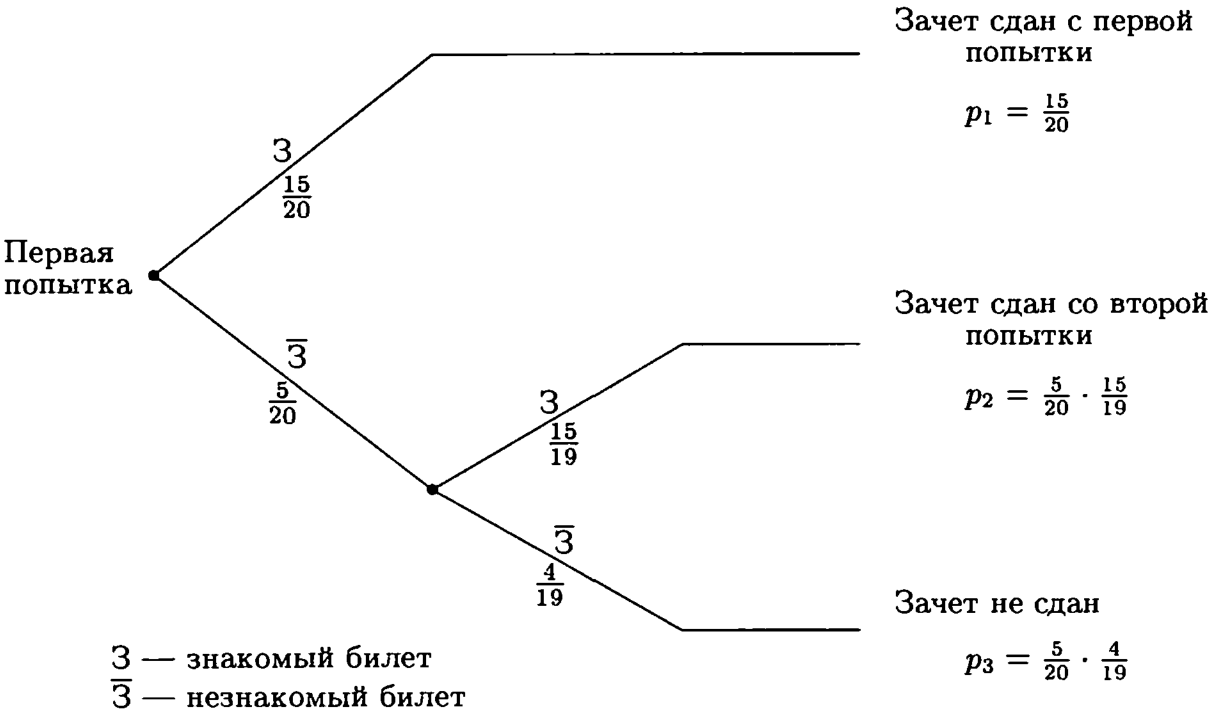


Рис. 12

Решение 2. Построим дерево вероятностей (рис. 12). Из него легко получить ответ:

$$p_1 + p_2 = \frac{15}{20} + \frac{5}{20}\cdot\frac{15}{19} = \frac{18}{19}.$$

Пример 19. Игроки A и B разыгрывают денежный приз в следующей игре. Подбрасывается монета до тех пор, пока не выпадет шесть "гербов" либо шесть "цифр". Если выпало шесть "гербов", то выигрывает игрок A, если шесть "цифр" — игрок B. Монету подбросили 8 раз. При счете 5:3 в пользу игрока A (т. е. выпало пять "гербов" и три "цифры") игра прервалась по не зависящим от игроков причинам. В каком отношении надо поделить денежный приз?

Решение 1. Если бы игра продолжалась, ситуация могла бы развиваться (начиная с девятого подбрасывания монеты) следующим образом (приведем все возможные варианты и их вероятности):

Г – игрок A выиграл, вероятность $1/2$;

Ц Г – игрок A выиграл, вероятность $1/4$;

Ц Ц Г – игрок A выиграл, вероятность $1/8$;

Ц Ц Ц – игрок B выиграл, вероятность $1/8$.

Таким образом, при счете 5:3 вероятность выигрыша игрока A составляет

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}=\frac{7}{8},$$

а игрока B — всего $1/8$. По-видимому, приз следует разделить в отношении 7:1 в пользу игрока A.

Замечание. Можно было и не перебирать все возможные варианты, а просто заметить, что игрок B выигрывает лишь в случае выпадания трех цифр подряд. Вероятность этого составляет $1/8$, а вероятность выигрыша игрока A (что является противоположным событием, ведь ничья правилами игры не предусмотрена) составляет соответственно

$$1-\frac{1}{8}=\frac{7}{8}.$$

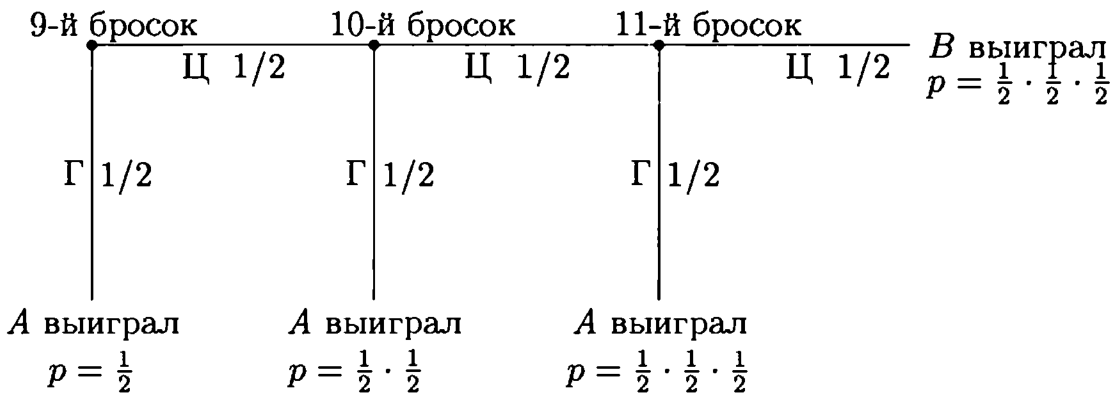


Рис. 13

Решение 2. Дерево вероятностей см. на рис. 13. Вывод тот же, что и в решении 1.

11.5. Формула полной вероятности и формула Байеса

Одним из эффективных методов подсчета вероятностей является формула полной вероятности, являющаяся следствием формул для вероятностей суммы и произведения событий.

Предположим, что событие A может наступить только вместе с одним из попарно несовместимых событий $H_1, \ldots, H_n$ (по отношению к событию A будем называть их *гипотезами*). Тогда событие A влечет за собой появление одного из событий $AH_1, \ldots, AH_n$, и его можно представить в виде

$$A = AH_1 + \cdots + AH_n$$

(см. рис. 14, где $n = 3$).

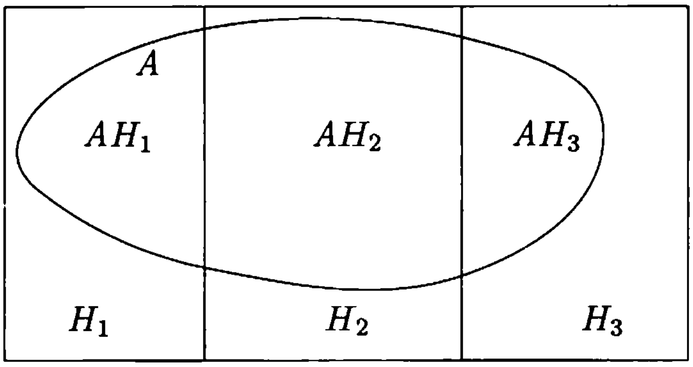


Рис. 14

Пример 20. Пусть в доме пять дверей. Событие A = "человек вошел в дом", гипотеза H_i = "человек прошел через i-ю дверь", где $i = 1, 2, 3, 4, 5$.

Поскольку события $H_1, \ldots, H_n$ попарно несовместимы, то таковыми же будут и события $AH_1, \ldots, AH_n$. Воспользовавшись формулой (5), получаем:

$$p(A) = p(AH_1 + \cdots + AH_n) = p(AH_1) + \cdots + p(AH_n).$$

Наконец, вспоминая формулу (3), можно сделать еще одно преобразование:

$$p(A) = p(A|H_1)p(H_1) + \cdots + p(A|H_n)p(H_n).$$

Это и есть *формула полной вероятности.*

С формулой полной вероятности тесно связана формула Байеса, справедливая при тех же предположениях.

Пусть опыт произведен и наступило событие A. Напомним, что событие A могло произойти только вместе с одной из гипотез $H_1, \ldots, H_n$. Поэтому можно вычислить вероятность того, что имело место именно событие H_i. Эта *апостериорная* вероятность $p(H_i|A)$ отличается, вообще говоря, от априорной вероятности $p(H_i)$ (в которой не учтен тот факт, что событие A произошло).

Записывая формулы для вероятности произведения событий, имеем:

$$p(AH_i) = p(A|H_i)p(H_i),$$

$$p(AH_i) = p(H_i|A)p(A).$$

Приравняв правые части последних формул, получаем равенство

$$p(A|H_i)p(H_i) = p(H_i|A)p(A).$$

И, далее,

$$p(H_i|A) = \frac{p(A|H_i)p(H_i)}{p(A)}.$$

Привлекая формулу полной вероятности, получаем в итоге *формулу Байеса*:

$$p(H_i|A) = \frac{p(A|H_i)p(H_i)}{p(A|H_1)p(H_1) + \cdots + p(A|H_n)p(H_n)}.$$

Пример 21. Предположим, что 5% всех мужчин и 0,25% всех женщин страдают дальтонизмом. Для простоты будем считать, что мужчин и женщин одинаковое число. Какова вероятность того, что наугад выбранное лицо страдает дальтонизмом? Если наугад выбранное лицо страдает дальтонизмом, то какова вероятность того, что это мужчина?

Решение 1. Здесь H_1 = “выбранное лицо — мужчина”, H_2 = “выбранное лицо — женщина”, A = “выбранное лицо страдает дальтонизмом”. Требуется вычислить вероятности $p(A)$ и $p(H_1|A)$. Имеем:

$$p(H_1) = p(H_2) = 0{,}5, \quad p(A|H_1) = 0{,}05, \quad p(A|H_2) = 0{,}0025.$$

Тогда по формуле полной вероятности

$$p(A) = 0{,}05 \cdot 0{,}5 + 0{,}0025 \cdot 0{,}5 = 0{,}02625$$

и по формуле Байеса

$$p(H_1|A) = \frac{0{,}05 \cdot 0{,}5}{0{,}05 \cdot 0{,}5 + 0{,}0025 \cdot 0{,}5} = \frac{20}{21} \approx 0{,}95.$$

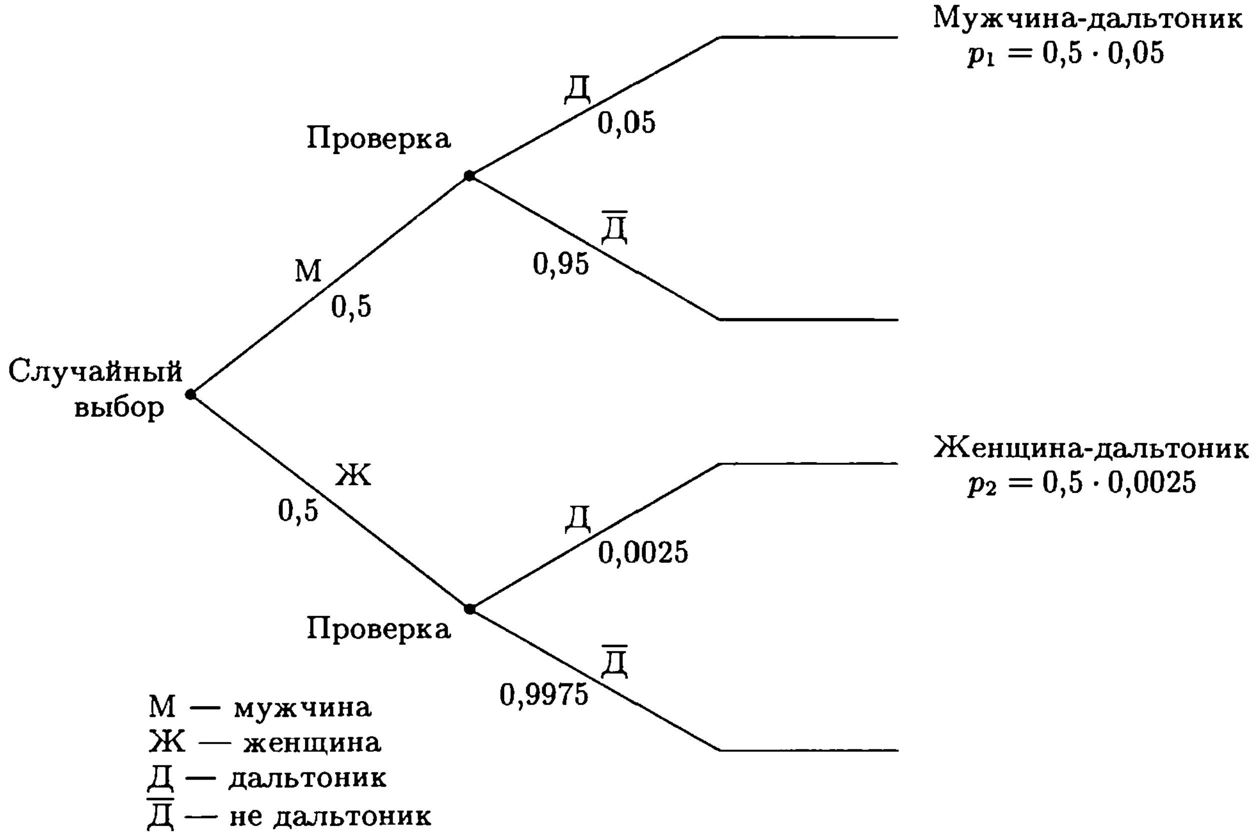


Рис. 15

Решение 2. Построим дерево вероятностей (рис. 15). Справа обозначены лишь два исхода из четырех возможных, поскольку оставшиеся два нас в данном случае не интересуют. Из рисунка видно, что вероятность того, что наугад выбранное лицо — дальтоник, составляет

$$p_1 + p_2 = 0{,}5 \cdot 0{,}05 + 0{,}5 \cdot 0{,}0025 = 0{,}02625.$$

Если известно, что наугад выбранное лицо — дальтоник, то число исходов сокращается с четырех до двух обозначенных. Вероятность того, что это мужчина (условная вероятность!), равна

$$\frac{p_1}{p_1 + p_2} = \frac{0{,}5 \cdot 0{,}05}{0{,}02625} \approx 0{,}95.$$

Пример 22. На экзамене студентам предлагается 20 билетов, 5 из которых легкие, а 15 — трудные. Два студента по очереди тянут билеты — сначала первый студент, затем второй.

а) Чему равна вероятность вытянуть легкий билет для первого студента?

б) Чему равна вероятность вытянуть легкий билет для второго студента?

в) Известно, что второй студент вытянул легкий билет. Чему равна вероятность того, что и первый вытянул легкий?

Решение. Введем обозначения:

H_1 – первый студент вытянул легкий билет;
H_2 – первый студент вытянул трудный билет;
A – второй студент вытянул легкий билет.

Тогда ответ на вопрос пункта а) дает формула классической вероятности:

$$p(H_1) = \frac{5}{20} = \frac{1}{4},$$

ответ на вопрос пункта б) — формула полной вероятности:

$$p(A) = p(A|H_1)\,p(H_1) + p(A|H_2)\,p(H_2) = \frac{4}{19}\cdot\frac{5}{20} + \frac{5}{19}\cdot\frac{15}{20} = \frac{1}{4},$$

а ответ на вопрос пункта в) — формула Байеса:

$$p(H_1|A) = \frac{p(A|H_1)\,p(H_1)}{p(A)} = \frac{\frac{4}{19}\cdot\frac{5}{20}}{\frac{4}{19}\cdot\frac{5}{20}+\frac{5}{19}\cdot\frac{15}{20}} = \frac{4}{19}.$$

Пример 23. Фирма планирует выпуск на рынок нового вида товара. Субъективные представления руководства фирмы таковы: вероятность хорошего спроса на этот товар составляет 0,7, вероятность плохого спроса — 0,3. Было проведено специальное исследование товарного рынка, которое предсказало плохой сбыт. Однако известно, что исследования такого рода дают правильный прогноз не всегда, а лишь с вероятностью 0,8. Каким образом маркетинговое исследование повлияло на вероятности хорошего и плохого сбыта?

Решение. Введем следующие обозначения:

H_1 – сбыт будет хорошим;
H_2 – сбыт будет плохим;
A – исследование рынка предсказало плохой сбыт.

Опыт и интуиция руководства фирмы дают в соответствии с условием следующие вероятности:

$$p(H_1) = 0{,}7;$$
$$p(H_2) = 0{,}3.$$

Маркетинговое исследование дает верный результат с вероятностью 0,8, поэтому

$$p(A|H_1) = 0{,}2;$$
$$p(A|H_2) = 0{,}8.$$

Подставляя все эти вероятности в формулу Байеса, получаем:

$$p(H_1|A) = \frac{p(A|H_1)\,p(H_1)}{p(A|H_1)\,p(H_1) + p(A|H_2)\,p(H_2)} = \frac{0{,}2\cdot 0{,}7}{0{,}2\cdot 0{,}7 + 0{,}3\cdot 0{,}8} = \frac{7}{19} \approx 0{,}37,$$

$$p(H_2|A) = \frac{p(A|H_2)\,p(H_2)}{p(A|H_1)\,p(H_1) + p(A|H_2)\,p(H_2)} = \frac{0{,}3\cdot 0{,}8}{0{,}2\cdot 0{,}7 + 0{,}3\cdot 0{,}8} = \frac{12}{19} \approx 0{,}63.$$

Ответ: в результате исследования вероятность хорошего сбыта уменьшилась до 0,37, а вероятность плохого увеличилась до 0,63.

11.6. Схема испытаний Бернулли

Пусть A — случайное событие, которое может произойти в результате некоторого испытания. Допустим далее, что нас интересует, наступило ли событие A: будем считать возможными лишь два события — A и $\bar{A}$. Обозначим их вероятности через p и q соответственно, $p + q = 1$.

Предположим, что испытание повторяется при одних и тех же условиях, скажем, три раза. В результате этого эксперимента (состоящего из трех испытаний) может произойти одно из следующих восьми событий:

$$AAA,\quad AA\bar{A},\quad A\bar{A}A,\quad \bar{A}AA,$$
$$A\bar{A}\bar{A},\quad \bar{A}A\bar{A},\quad \bar{A}\bar{A}A,\quad \bar{A}\bar{A}\bar{A}$$

(запись $\bar{A}\bar{A}A$ обозначает событие "в первых двух испытаниях событие A не произошло, в третьем — произошло", аналогично записаны остальные семь событий). Подсчитаем вероятности этих событий по формуле (6). Имеем:

$$
\begin{array}{ll}
p(AAA) = ppp = p^3, & p(A\bar{A}\bar{A}) = pqq = pq^2, \\
p(AA\bar{A}) = ppq = p^2q, & p(\bar{A}A\bar{A}) = qpq = pq^2, \\
p(A\bar{A}A) = pqp = p^2q, & p(\bar{A}\bar{A}A) = qqp = pq^2, \\
p(\bar{A}AA) = qpp = p^2q, & p(\bar{A}\bar{A}\bar{A}) = qqq = q^3.
\end{array}
$$

Как мы видим, вероятность каждого из исходов представима в виде p^kq^{3-k}, где k показывает число наступивших событий A, а $(3-k)$ соответственно число ненаступивших.

Вероятностные схемы такого рода называются *схемами Бернулли* или *схемами биномиальных экспериментов.* Эти схемы широко применяются при анализе реальных ситуаций в тех случаях, когда эксперимент можно считать *биномиальным*, т. е. когда

он состоит из фиксированного числа n испытаний,

в каждом из этих испытаний происходит либо не происходит некоторое событие,

вероятность этого события одинакова в каждом испытании,

испытания независимы одно от другого.

Пример 24. Тренированный стрелок совершает пять выстрелов по мишени, причем все выстрелы производятся практически в одних и тех же условиях. При этом число попаданий в "десятку" может меняться от 0 до 5.

Пример 25. В помете, состоящем из 8 мышей, происходящих от одних родителей, число мышей, имеющих прямую, а не волнистую шерстку, может равняться произвольному целому числу от 0 до 8.

Пример 26. Один за другим бросают три игральных кубика. Число выпаданий "шестерки" может принимать одно из четырех значений: от 0 до 3 включительно.

Вероятность того, что событие A произойдет ровно k раз после n испытаний, обозначим через $P(p,n,k)$ (ясно, что $0 \leqslant k \leqslant n$).

Справедлива следующая формула:

$$P(p,n,k) = C_n^k p^k q^{n-k}.$$

Здесь p — это вероятность появления события A в одном испытании, $q = 1-p$, а C_n^k (читается "C из n по k") называется *биномиальным* *коэффициентом* и вычисляется по любой из формул:

$$C_n^k = \frac{n!}{k!(n-k)!},$$

$$C_n^k = \frac{n(n-1)\dots(n-k+1)}{k!},$$

$$C_n^k = \frac{n(n-1)\dots(k+1)}{(n-k)!},$$

где $n!$ (читается "n факториал") — произведение натуральных чисел от 1 до n включительно:

$$n! = 1 \cdot 2 \cdot 3 \dots (n-2)(n-1)n.$$

Заметим также, что по определению принимается:

$$0! = 1.$$

Пример 27. Монету бросают 10 раз. Какова вероятность того, что при этом герб выпадет три раза? Выпадет меньше двух раз?

Решение. Здесь $n = 10$, $p = q = 1/2$.

$$P\{\text{герб выпал три раза}\} =$$

$$= P\left(\frac{1}{2}, 10, 3\right) = C_{10}^3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} \frac{1}{2^{10}} = \frac{15}{128}.$$

$$P\{\text{герб выпал меньше двух раз}\} =$$

$$= P\{\text{герб не выпал ни разу}\} + P\{\text{герб выпал один раз}\} =$$

$$= P\left(\frac{1}{2}, 10, 0\right) + P\left(\frac{1}{2}, 10, 1\right) = C_{10}^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} + C_{10}^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^9 =$$

$$= \frac{1}{2^{10}} + \frac{10}{2^{10}} = \frac{11}{1024}.$$

Пример 28. Студент пишет контрольную работу по теории вероятностей. У него есть предположение о том, как решить задачу, однако свою способность найти правильное решение студент оценивает невысоко — примерно числом 0,4.

Вокруг студента в аудитории сидят пять однокурсников. Можно рискнуть опросить их и принять либо отвергнуть решение на основании большинства голосов. Подготовку этих однокурсников студент оценивает так же, как и свою.

Как лучше поступить студенту — положиться на свои соображения или на большинство голосов однокурсников?

Решение. Для выбора между двумя альтернативами следует сначала выбрать какой-либо критерий. По-видимому, в данной ситуации таким критерием является вероятность правильно решить задачу. Опираясь на свои соображения, студент оценивает ее как 0,4.

Вычислим теперь вероятность того, что большинство из пяти опрошенных однокурсников дадут правильный ответ. Большинство — это либо 3, либо 4, либо 5. Поэтому искомая вероятность вычисляется следующим образом:

$$P(0{,}4;\ 5,\ 3) + P(0{,}4;\ 5,\ 4) + P(0{,}4;\ 5,\ 5) =$$
$$= C_5^3 \cdot (0{,}4)^3 \cdot (0{,}6)^2 + C_5^4 \cdot (0{,}4)^4 \cdot 0{,}6 + C_5^5 \cdot (0{,}4)^5 =$$
$$= 10 \cdot 0{,}064 \cdot 0{,}36 + 5 \cdot 0{,}0256 \cdot 0{,}6 + 0{,}01024 = 0{,}31744.$$

Вероятность снизилось с 0,4 до 0,317 44 — более чем на 20%. Вывод: опрос однокурсников в данной ситуации лучше не проводить.

11.7. Задания и ответы

1. На плоскость нанесена сетка квадратов со стороной 10 см. Найдите вероятность того, что брошенный на плоскость круг радиусом 1 см не пересечет стороны ни одного из квадратов.

Ответ: 0,64.

2. Имеются две сумки с мячами, в каждой по 5 мячей, пронумерованных от 1 до 5. Наугад вынимается по одному мячу из каждой сумки. Какова вероятность того, что это будут мячи с номерами 2 и 5 (безразлично, какой из них из какой сумки вынут)?

Ответ: 0,08.

3. Пусть испытанием является бросание точки в единичный квадрат, а событием A, B, C, D — попадание точки в соответствующую прямоугольную область (см. рис. 16, 17). Проверить совместимость и зависимость а) событий A и B (рис. 16); б) событий C и D (рис. 17).

Ответ: а) события A и B совместимы и независимы; б) события C и D несовместимы и зависимы.

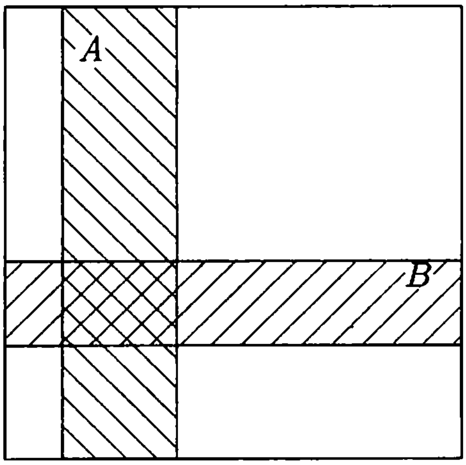


Рис. 16

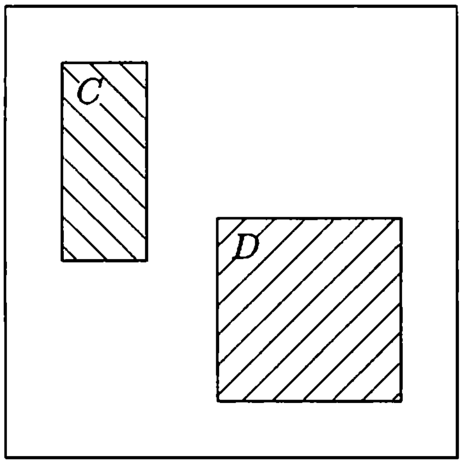


Рис. 17

4. Вероятность того, что в течение одной смены возникнет неполадка, равна 0,05. Какова вероятность того, что не произойдет ни одной неполадки за три смены?

Ответ: 0,857 375.

5. Студент пришел на зачет, зная из 30 вопросов только 24. Какова вероятность сдать зачет, если после отказа отвечать на вопрос преподаватель задает еще один вопрос?

Ответ: $\dfrac{28}{29} \approx 0{,}9655$.

6. Два охотника стреляют в волка, причем каждый делает по одному выстрелу. Для первого охотника вероятность попадания в цель 0,7, для второго — 0,8. Какова вероятность попадания в волка (хотя бы при одном выстреле)? Как изменится результат, если охотники сделают по два выстрела?

Ответ: 0,94; 0,9964.

7. Из большой связки галстуков, в которой галстуки зеленого, красного и желтого цвета находятся в пропорции 5 : 3 : 2, двое мужчин случайным образом выбирают по одному галстуку. Какова вероятность того, что они выберут галстуки одинакового цвета?

Ответ: 0,38.

8. Имеются две урны. В первой находится 1 белый шар, 3 черных и 4 красных, во второй — 3 белых, 2 черных и 3 красных. Из каждой

урны извлекают по шару. Найти вероятность того, что цвета вынутых шаров совпадают.

Ответ: $\dfrac{21}{64} \approx 0{,}328$.

9. На трех станках различной марки изготовляется определенная деталь. Производительность 1-го станка за смену составляет 40 деталей, 2-го — 35, 3-го — 25 деталей. Установлено, что 2%, 3% и 5% продукции этих станков соответственно имеют скрытые дефекты. В конце смены на контроль взята одна деталь.

а) Какова вероятность того, что эта деталь нестандартная?

б) Если деталь оказалась нестандартной, какова вероятность того, что она изготовлена на 1-м, 2-м или 3-м станке?

Ответ: а) 0,031; б) $\dfrac{8}{31} \approx 0{,}258$; $\dfrac{21}{62} \approx 0{,}339$; $\dfrac{25}{62} \approx 0{,}403$.

10. Два стрелка независимо один от другого делают по одному выстрелу по мишени. Вероятность попадания в мишень для первого стрелка 0,8, для второго — 0,4. После стрельбы в мишени обнаружена одна пробоина. Найти вероятность того, что в мишень попал первый стрелок.

Ответ: 6/7.

11. Как изменились бы вероятности хорошего и плохого сбыта в примере 23, если бы исследование рынка предсказало хороший сбыт?

Ответ: вероятность хорошего сбыта — 0,9; вероятность плохого сбыта — 0,1.

12. Игральный кубик бросают 5 раз. Найдите вероятность того, что число очков, кратное трем, появится 2 раза.

Ответ: 80/243.

13. Всхожесть семян растений данного сорта оценивается вероятностью, равной 0,8. Какова вероятность того, что из пяти посеянных семян взойдут не менее четырех?

Ответ: 0,73728.

Глава 12
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

12.1. Понятие случайной величины. Закон распределения. Биномиальная случайная величина

Мы уже знакомы с понятиями "испытание", "случайное событие", "вероятность". Теперь приступим к рассмотрению чрезвычайно важного случая, который характеризуется следующим обстоятельством: в результате испытания не только происходит событие, но есть еще и возможность наблюдать некоторое число. Причем это число нас интересует даже в большей степени, чем само событие.

Нетрудно видеть, что возможность наблюдать число часто имела место и в тех испытаниях, которые мы уже рассматривали.

Пример 1. Бросая игральный кубик, мы получаем число точек на верхней грани.

Пример 2. Бросая одновременно три монеты, фиксируем число выпадений герба (это может быть 0, 1, 2 или 3).

Если каждому событию подобным образом поставлено в соответствие некоторое число, будем говорить, что задана *случайная величина*. Иными словами, случайная величина — это величина, принимающая в результате испытания то или иное числовое значение, но заранее неизвестно, какое именно.

Будем обозначать случайные величины большими латинскими буквами X, Y и т. д.

С каждой случайной величиной связано некоторое множество чисел — значений, которые она может принимать. В результате испытания эти значения могут выпадать с различной вероятностью. Правило, устанавливающее связь между возможными значениями и их вероятностями (точнее, речь идет о вероятности события, заключающегося в том, что случайная величина приняла то или иное значение), называется *законом распределения* случайной величины.

Замечание. Случайная величина однозначно и полностью определяется своим законом распределения, подобно тому как квадрат определяется длиной стороны. Переводя эту аналогию в плоскость соотношения с реальным миром, заметим, что если квадрат со стороной 10 м может быть моделью дома, бассейна или детской площадки, то случайная величина с данным законом распределения может быть моделью числа посетителей магазина в течение дня, числа выпускаемых станком деталей и т. п.

Вследствие тесной связи между понятиями "случайная величина" и "закон распределения" (или даже просто "распределение") они часто используются как синонимы.

Перейдем теперь к рассмотрению того, каким образом может быть задан закон распределения случайной величины, когда она принимает лишь конечное число значений.

Итак, пусть случайная величина X может принимать одно из n различных значений:

$$x_1, x_2, \ldots, x_n.$$

При этом каждое из этих значений величина X принимает с определенной вероятностью — соответственно

$$p_1, p_2, \ldots, p_n.$$

Иными словами, p_1 — это вероятность события "случайная величина X приняла значение x_1" или более кратко: $X = x_1$,
p_2 — вероятность случайного события $X = x_2$,
...
p_n — вероятность случайного события $X = x_n$.

Сведем все эти значения в таблицу

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

в первой строке которой указаны значения, принимаемые случайной величиной X, во второй строке — их вероятности. Она называется *таблицей распределения* случайной величины X. Обычно числа в первой строке таблицы распределения располагают в порядке возрастания.

Замечание. Отметим следующее важное обстоятельство. Поскольку в результате испытания величина X наверняка примет одно из этих значений, сумма несовместимых событий

$$\{X = x_1\} + \{X = x_2\} + \cdots + \{X = x_n\}$$

является достоверным событием, вероятность которого равна 1. Поэтому для таблицы распределения любой случайной величины справедливо равенство

$$p_1 + p_2 + \cdots + p_n = 1.$$

Итак, для того чтобы при решении конкретной задачи заполнить таблицу распределения заданной случайной величины, надо выписать все принимаемые ею значения $x_1, x_2, \ldots, x_n$ и вычислить соответствующие вероятности $p_1, p_2, \ldots, p_n$.

Вернемся к примеру с игральным кубиком. Для упомянутой там случайной величины (обозначим ее через X) вероятности принять любое из шести значений равны между собой. Таблица распределения выглядит так:

X	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

Для случайной величины из примера с тремя монетами (обозначим ее через Y) построить таблицу распределения несколько сложнее. Вспомним, что в результате одновременного бросания трех монет возможно всего 8 равновероятных исходов: ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ. При первом исходе величина Y принимает значение 3; при втором, третьем и пятом — значение 2; при четвертом, шестом и седьмом — значение 1; при восьмом — значение 0. С учетом этого таблица распределения случайной величины Y такова:

Y	0	1	2	3
	1/8	3/8	3/8	1/8

Для более наглядного представления закона распределения часто используется координатная плоскость. По оси абсцисс отмечают значения, принимаемые случайной величиной, по оси ординат —

их вероятности. Затем на плоскости (x, p) отмечают точки (x_1, p_1), $(x_2, p_2), \ldots, (x_n, p_n)$. Для случайной величины Y из примера с тремя монетами это выглядит так, как изображено на рис. 1. Если теперь провести от отмеченных точек вертикальные отрезки до пересечения с осью абсцисс, то получится *столбиковая диаграмма* (рис. 2). Если же последовательно соединить точки отрезками, получится *полигон* (рис. 3).

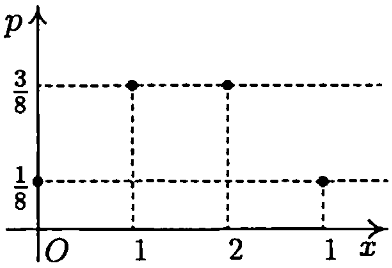


Рис. 1

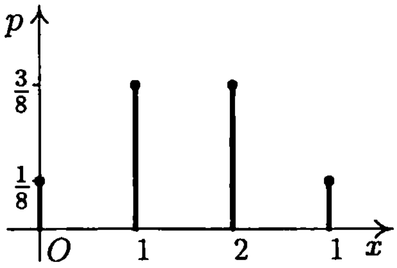


Рис. 2

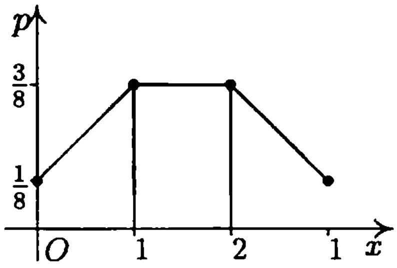


Рис. 3

Пример 3 (схема испытаний Бернулли). Испытание повторяется n раз, причем вероятность успеха в одном испытании равна p.

Общее число успехов (в n испытаниях) есть случайная величина, принимающая значения $0, 1, 2, \ldots, n$. Вероятность того, что эта случайная величина примет значение k, равна

$$P(p, n, k) = C_n^k p^k (1 - p)^{n-k}.$$

Такую случайную величину будем называть *биномиальной* и обозначать $Bi(n, p)$. Она зависит от двух параметров — n и p. Случайная величина Y из примера 2 является биномиальной случайной величиной с параметрами 3 (три монеты) и 1/2 (такова вероятность выпадения герба у одной монеты):

$$Y = Bi\left(3, \frac{1}{2}\right).$$

12.2. Операции над случайной величиной

Пусть имеется случайная величина X, принимающая в зависимости от результата испытания те или иные случайные значения. Если к каждому из этих значений прибавить одно и то же число, например 3, то в результате мы получим новые числа — значения случайной величины $X + 3$.

Таблица распределения случайной величины $X + 3$ строится по таблице распределения случайной величины X следующим образом:

X	x_1	...	x_n
	p_1	...	p_n

$\longrightarrow$

$X + 3$	$x_1 + 3$	...	$x_n + 3$
	p_1	...	p_n

Как видно, вторая строка осталась без изменений, поскольку вероятности событий $(X = x_i)$ и $(X + 3 = x_i + 3)$ равны.

Построение таблицы распределения случайной величины X^2 несколько сложнее.

Рассмотрим конкретный пример:

X	-1	0	1	2
	0,2	0,3	0,4	0,1

Таблица для случайной величины $X + 3$ строится просто:

$X + 3$	2	3	4	5
	0,2	0,3	0,4	0,1

Пытаясь действовать аналогичным образом для величины X^2, т. е. заменяя все значения x_i числами x_i^2, получаем:

	1	0	1	4
	0,2	0,3	0,4	0,1

В первой строке есть совпадающие значения. Поэтому следует объединить их в одно, сложив соответствующие вероятности:

X^2	1	0	4
	0,6	0,3	0,1

Рассмотрим еще один пример:

Z	-2	-1	0	1	2	3
	0,1	0,2	0,3	0,2	0,1	0,1

Возводя значения случайной величины Z в квадрат, получаем:

	4	1	0	1	4	9
	0,1	0,2	0,3	0,2	0,1	0,1

И наконец, в результате имеем:

Z^2	4	1	0	9
	0,2	0,4	0,3	0,1

Таблицу распределения случайной величины $Y = f(X)$ для любой функции f можно построить аналогично. Она строится в два этапа. Сначала вычисляются элементы вспомогательной таблицы

	$f(x_1)$	$f(x_2)$	$\dots$	$f(x_n)$
	p_1	p_2	$\dots$	p_n

Затем совпадающие значения $f(x_i) = f(x_j)$ для разных чисел x_i и x_j (если такие имеются) объединяются в одно, а соответствующие вероятности складываются.

12.3. Числовые характеристики случайной величины

Как было отмечено ранее, случайная величина полностью определяется своим законом распределения. При этом закон распределения может быть задан, например, при помощи таблицы распределения. Однако во многих практических задачах (в том числе в задачах принятия решения) требуется представить информацию о случайной величине в более компактном, обозримом виде. Обычно для этого применяются так называемые *числовые характеристики* случайной величины — числа, на основании которых можно судить об интересующих нас факторах.

Приведем пример. Пусть имеются две альтернативные стратегии действий, каждая из которых обещает принести определенную прибыль, причем в обоих случаях прибыль зависит от различных случайных обстоятельств и в силу этого является случайной величиной. Законы распределения обеих случайных величин будем считать известными (заданными в виде таблиц). Какую стратегию следует предпочесть?

Ниже мы рассмотрим две числовые характеристики случайной величины, позволяющие в ситуациях такого рода принимать достаточно обоснованное решение.

Математическое ожидание

Первая важная характеристика — это среднее ожидаемое значение, принимаемое случайной величиной в больших сериях испытаний.

Пусть имеется случайная величина X с заданной таблицей распределения

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

Математическое ожидание EX (ср. англ. expectation) случайной величины X определяется формулой

$$EX = x_1p_1 + x_2p_2 + \cdots + x_np_n,$$

которая в сокращенной записи выглядит так:

$$EX = \sum_{i=1}^{n} x_ip_i.$$

Поясним эту формулу. Предположим, что проведено m испытаний (m — достаточно большое число), при этом m_1 раз величина X приняла значение x_1, m_2 раза — значение x_2, ... , m_n раз — значение x_n:

$$m_1 + m_2 + \cdots + m_n = m.$$

Найдем среднее арифметическое всех этих m значений. Имеем:

$$\frac{x_1m_1 + x_2m_2 + \cdots + x_nm_n}{m} = x_1\frac{m_1}{m} + x_2\frac{m_2}{m} + \cdots + x_n\frac{m_n}{m}.$$

Дробь

$$\frac{m_k}{m}$$

представляет собой относительную частоту появления события $X = x_k$ в m испытаниях (в m испытаниях событие $X = x_k$ произошло m_k раз). При больших значениях m относительная частота примерно равна вероятности события $X = x_k$, т. е. p_k.

Во многих случаях альтернативы можно сравнивать на основании средних значений. Особенно это оправданно тогда, когда решение приходится принимать в однотипных ситуациях много раз, т. е. фактически проводить большое число испытаний (например, магазин каждый день решает, сколько батонов хлеба закупить для реализации в течение следующего дня).

Пример 4. Предприниматель размышляет над тем, куда лучше вложить деньги — в киоск для торговли мороженым или в палатку для торговли хлебобулочными изделиями.

Вложение средств в киоск с вероятностью 0,5 обеспечит годовую прибыль 5 тыс. долл., с вероятностью 0,2 — 10 тыс. долл. и с вероятностью 0,3 — 3 тыс. долл.

Для палатки прогноз таков: 5,5 тыс. долл. — с вероятностью 0,6, 5 тыс. долл. — с вероятностью 0,3 и 6,5 тыс. долл. — с вероятностью 0,1.

В каком случае (для киоска или для палатки) математическое ожидание годового дохода больше?

Решение. Для каждого из двух возможных решений годовая прибыль является случайной величиной. Обозначив эти величины через X и Y, построим таблицы распределения:

X	3000	5000	10000
	0,3	0,5	0,2

Y	5000	5500	6500
	0,3	0,6	0,1

Найдем математические ожидания:

$$EX = 3000 \cdot 0{,}3 + 5000 \cdot 0{,}5 + 10000 \cdot 0{,}2 = 5400 \text{ долл.},$$
$$EY = 5000 \cdot 0{,}3 + 5500 \cdot 0{,}6 + 6500 \cdot 0{,}1 = 5450 \text{ долл.}$$

Получается, что $EY > EX$. Таким образом, математическое ожидание для палатки больше.

Дисперсия и стандартное отклонение

Итак, математическое ожидание определяет, какое значение случайная величина принимает "в среднем". Следующий простейший пример показывает, что случайные величины с равным математическим ожиданием могут существенно различаться по степени близости к нему.

Рассмотрим случайные величины X и Y:

X	99	101
	0,5	0,5

Y	0	200
	0,5	0,5

(1)

Нетрудно видеть, что $EX = EY = 100$. Но если для величины X отклонение от значения 100 незначительно, то для величины Y оно весьма заметно.

Если выбор между величинами X и Y — это выбор между двумя альтернативными решениями, то X — это стабильная, предсказуемая прибыль, а Y — это риск. Считается, что при принятии решений в большинстве случаев пытаются уменьшить непредсказуемость, показателем которой является степень отклонения случайной величины от ее математического ожидания. Для измерения этого показателя применяется еще одна числовая характеристика случайной величины, называемая *дисперсией*. Обозначение: DX (ср. англ. dispersion). Покажем, как она вычисляется.

Вычитая из случайной величины X ее математическое ожидание, получаем новую случайную величину

$$X - EX.$$

Квадрат последней также является случайной величиной

$$(X - EX)^2,$$

математическое ожидание которой и есть дисперсия:

$$DX = E(X - EX)^2.$$

Если величина X задана таблицей

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

,

то дисперсия случайной величины X может быть вычислена по формуле

$$DX = \sum_{i=1}^{n}(x_i - EX)^2 p_i$$

или более просто:

$$DX = \sum_{i=1}^{n} x_i^2 p_i - (EX)^2.$$

Пример 5. Вычислим дисперсии случайных величин X и Y, таблицы которых приведены выше — см. (1):

$$DX = 99^2 \cdot 0{,}5 + 101^2 \cdot 0{,}5 - 100^2 = 1,$$
$$DX = 0^2 \cdot 0{,}5 + 200^2 \cdot 0{,}5 - 100^2 = 10\,000.$$

Пример 6. Для биномиальной случайной величины $X = Bi(n,p)$ справедливы соотношения

$$EX = np, \qquad DX = np(1-p).$$

Замечание. Математическое ожидание может быть любым числом, а дисперсия всегда неотрицательна.

Случайные величины, моделирующие какие-либо объекты реального мира, обычно имеют размерность. Это означает, что принимаемые ими значения могут измеряться в штуках, метрах, килограммах и т. п. При этом математическое ожидание случайной величины имеет ту же размерность, что и сама случайная величина. Размерность же дисперсии равна квадрату размерности случайной величины. Например, если случайная величина измеряется в рублях, то ее дисперсия — в рублях в квадрате.

Чтобы не иметь дело с такими причудливыми единицами измерения, вводится понятие *стандартного отклонения.* Оно обозначается греческой буквой σ (сигма) и по определению равно квадратному корню из дисперсии

$$\sigma = \sqrt{DX}.$$

Тем самым, стандартное отклонение имеет ту же размерность, что и сама случайная величина.

Выше шла речь об операциях над случайными величинами. В случае линейных преобразований случайной величины X (т. е. преобразований вида

$$Y = aX + b,$$

где a и b — некоторые числа) математическое ожидание и дисперсию получившейся случайной величины Y можно вычислить, исходя из этих же числовых характеристик величины X. Именно, справедливы следующие формулы:

$$EY = aEX + b,$$
$$DY = a^2 DY.$$

Стандартная случайная величина

Случайная величина, у которой математическое ожидание равно 0, а дисперсия равна 1, называется *стандартной.*

Пусть имеется случайная величина X с математическим ожиданием μ (мю) и стандартным отклонением σ. Нетрудно показать, что случайная величина

$$Y = \frac{X - \mu}{\sigma}$$

является стандартной.

12.4. Случайные величины с бесконечным числом значений

Выше мы рассматривали случайные величины, принимающие конечное число значений. Однако часто возникает необходимость рассмотрения случайных величин, число возможных значений которых бесконечно.

Приведем два простых примера.

Пример 7. Испытание состоит в бросании монеты до первого выпадания герба. Случайная величина Y — количество бросаний — может принимать значения 1, 2, 3 и так далее, т. е. бесконечное число значений.

Пример 8. Испытание состоит в том, что на отрезке $[0;1]$ числовой оси случайным образом отмечается точка (изначально все точки отрезка "равноправны", т. е. шансы каждой точки быть отмеченной такие же, как у любой другой). Случайная величина X — число отрезка $[0;1]$, соответствующее отмеченной точке, — может принимать любые значения из отрезка $[0;1]$.

Натуральных чисел, равно как и точек отрезка $[0;1]$, бесконечно много. Однако на числовой прямой они расположены изолированно друг от друга (дискретно). Отмеченное свойство объединяет величину Y из примера 7 со случайными величинами, принимающими конечное число значений. Все эти величины называются *дискретными* случайными величинами.

Для случайной величины Y из примера 7 можно построить "бесконечную таблицу распределения" следующего вида:

Y	1	2	3	...	k	...
	1/2	1/4	1/8	...	$(1/2)^k$	...

В отличие от дискретных случайных величин случайная величина X из примера 8 является *непрерывной* случайной величиной —

точки отрезка $[0;1]$ нельзя одну за другой выделить и записать в таблицу (хотя бы и бесконечную).

Существуют еще случайные величины *смешанного типа.*

Далее мы будем рассматривать только случайные величины, принимающие конечное число значений, и непрерывные случайные величины.

12.5. Непрерывные случайные величины

Под *непрерывной случайной величиной* мы будем понимать случайную величину, принимающую значения на прямой, луче (полупрямой) или отрезке. Описание закона распределения в непрерывном случае существенно сложнее, чем в дискретном.

Главное различие в задачах вычисления вероятностей для дискретного и непрерывного случаев состоит в следующем. В дискретном случае ищется вероятность событий типа $X = c$ (случайная величина принимает определенное значение). В непрерывном случае вероятности такого типа равны нулю, поэтому интерес представляют вероятности событий типа $a \leqslant X \leqslant b$ (случайная величина принимает значения из некоторого отрезка). При этом

$$p(a \leqslant X \leqslant b) = p(a < X \leqslant b) = p(a \leqslant X < b) = p(a < X < b),$$
$$p(X \geqslant c) = p(X > c),$$
$$p(X \leqslant c) = p(X < c).$$

Для случайной величины X, с равной вероятностью принимающей любое значение из отрезка $[0;1]$, естественно считать, что вероятность попадания в отрезок $[a;b]$ равна длине этого отрезка. Например,

$$\begin{aligned}&p(0 \leqslant X \leqslant 0{,}5) = 0{,}5,\\ &p(0{,}5 \leqslant X \leqslant 0{,}7) = 0{,}2,\\ &p(0 \leqslant X \leqslant 1) = 1,\\ &p(X \geqslant 0{,}8) = p(X \leqslant 0{,}2) = 0{,}2.\end{aligned}$$

Рассмотрим функцию

$$f(x) = \begin{cases} 0, & \text{при } x < 0;\\ 1, & \text{при } 0 \leqslant x \leqslant 1;\\ 0, & \text{при } x > 1\end{cases}$$

(рис. 4). Ее связывает со случайной величиной X следующее обстоятельство: вероятность события $a \leqslant X \leqslant b$ равна площади фигуры, ограниченной прямыми $y = 0$, $x = a$, $x = b$ и графиком функции $y = f(x)$. Иными словами, справедливо равенство

$$p(a \leqslant X \leqslant b) = \int\limits_a^b f(x)dx$$

для любых a и b, $a \leqslant b$.

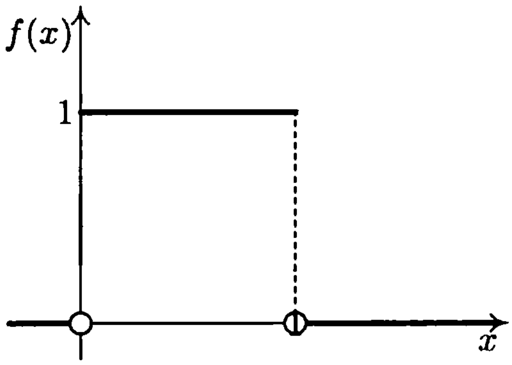


Рис. 4

Таким образом, функция $f(x)$ позволяет вычислять вероятности, связанные со случайной величиной X, т. е., по сути, задает закон распределения случайной величины X.

Посмотрим теперь на ситуацию с более общей точки зрения. Пусть имеется случайная величина X и неотрицательная функция $f(x)$ такая, что для любых чисел a и b, $a \leqslant b$, выполняется равенство

$$p(a \leqslant X \leqslant b) = \int\limits_a^b f(x)dx$$

(рис. 5). В этом случае говорят, что случайная величина X имеет *плотность распределения* $f(x)$. Записывается это следующим образом:

$$X \sim f(x).$$

Замечание 1. Интеграл от плотности распределения по всей области возможных значений случайной величины равен 1.

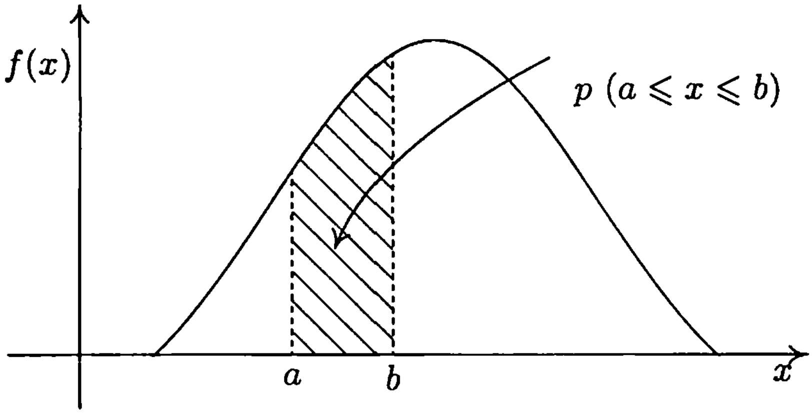


Рис. 5

Замечание 2. Описанная ситуация допускает следующую аналогию из механики. Пусть имеется сплошной стержень массой 1 кг. Масса распределена по стержню с различной, вообще говоря, плотностью. Если мы хотим найти, сколько весит некоторый отрезок стержня, надо взять по этому отрезку интеграл от плотности.

Пример 9. Пусть случайная величина X задана плотностью распределения

$$f(x) = \begin{cases} 0, & \text{при } x < 0, \\ x/2, & \text{при } 0 \leqslant x \leqslant 2, \\ 0, & \text{при } x > 2 \end{cases}$$

(рис. 6). Для нахождения вероятностей требуется вычислять соответствующие интегралы. Например:

$$P(0 \leqslant X \leqslant 1) = \int\limits_0^1 \frac{x}{2}\,dx = \frac{1}{4}x^2\Big|_0^1 = \frac{1}{4} - 0 = \frac{1}{4}.$$

Важно отметить, что

$$P(0 \leqslant X \leqslant 2) = 1,$$

поскольку все возможные значения величины X лежат в пределах от 0 до 2.

Пример 10. *Показательное (экспоненциальное) распределение* задается плотностью вида

$$f(x) = \begin{cases} 0, & \text{при } x < 0, \\ \lambda e^{-\lambda x}, & \text{при } x \geqslant 0 \end{cases}$$

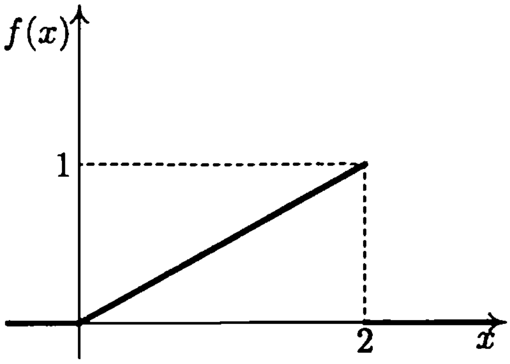


Рис. 6

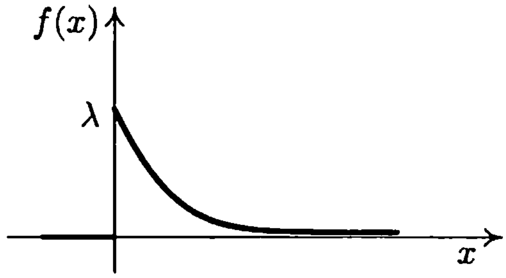


Рис. 7

(рис. 7). Здесь λ — некоторое положительное число. Показательное распределение применяется в *теории массового обслуживания.*

Для непрерывных распределений, как и для дискретных, можно рассматривать числовые характеристики: математическое ожидание, дисперсию и др. Они вычисляются при помощи плотности распределения.

12.6. Сумма случайных величин

Если в результате испытания принимают определенные значения сразу две случайные величины, X_1 и X_2, то их можно рассматривать вместе. В частности, определим их сумму $X_1 + X_2$ следующим образом: если в результате испытания величина X_1 принимает значение x_1, а величина X_2 — значение x_2, то случайная величина $X_1 + X_2$ принимает значение $x_1 + x_2$.

Аналогично определяется сумма n случайных величин.

Пример 11. Студент сдает в сессию три экзамена. Оценка по i-му экзамену ($i = 1, 2, 3$) — случайная величина X_i. Тогда общая сумма оценок за все три экзамена — случайная величина $X_1 + X_2 + X_3$.

Приведем пример построения закона распределения суммы двух случайных величин.

Пример 12. Пусть случайные величины X_1 и X_2 задаются таблицами распределения следующего вида:

X_1	-1	0	1
	0,2	0,5	0,3

X_2	-1	1
	0,4	0,6

Так как X_1 принимает три различных значения, а X_2 — два, то для суммы $X_1 + X_2$ получаем шесть возможностей. Выпишем их, вычислив попутно вероятности.

Итак, определим:

$$p(X_1 = -1)\, p(X_2 = -1) = 0{,}2 \cdot 0{,}4 = 0{,}08,$$

при этом $\quad X_1 + X_2 = -2;$

$$p(X_1 = -1)\, p(X_2 = 1) = 0{,}2 \cdot 0{,}6 = 0{,}12,$$

при этом $\quad X_1 + X_2 = 0;$

$$p(X_1 = 0)\, p(X_2 = -1) = 0{,}5 \cdot 0{,}4 = 0{,}2,$$

при этом $\quad X_1 + X_2 = -1;$

$$p(X_1 = 0)\, p(X_2 = 1) = 0{,}5 \cdot 0{,}6 = 0{,}3,$$

при этом $\quad X_1 + X_2 = 1;$

$$p(X_1 = 1)\, p(X_2 = -1) = 0{,}3 \cdot 0{,}4 = 0{,}12,$$

при этом $\quad X_1 + X_2 = 0;$

$$p(X_1 = 1)\, p(X_2 = 1) = 0{,}3 \cdot 0{,}6 = 0{,}18,$$

при этом $\quad X_1 + X_2 = 2.$

Таким образом, для величины $X_1 + X_2$ получаем следующую таблицу распределения:

$X_1 + X_2$	-2	-1	0	1	2
	0,08	0,2	0,24	0,3	0,18

Замечание 1. Если X_1 и X_2 — непрерывные случайные величины, то закон распределения суммы $X_1 + X_2$ строится гораздо сложнее.

В случае суммирования n дискретных случайных величин, где $n > 2$, таблица распределения строится аналогичным образом. Как в дискретном, так и в непрерывном случаях для математического ожидания суммы $X_1 + X_2 + \cdots + X_n$ справедлива следующая простая формула:

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n$$

или, применяя знак сокращенного суммирования:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i.$$

Аналогичная формула для дисперсии справедлива не всегда, а только в случае *независимых* случайных величин.

Случайные величины $X_1, X_2, \ldots, X_n$ *называются независимыми, если закон распределения каждой из них не зависит от того, какие значения приняли другие величины.*

Для дисперсии суммы независимых случайных величин справедливо соотношение

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} DX_i.$$

Замечание 2. Отметим, что величины $X_1, X_2, \ldots, X_n$ могут иметь один и тот же закон распределения. В этом случае вместо термина "независимые случайные величины" употребляют термин "независимые наблюдения (испытания)".

12.7. Нормальное распределение

Наиболее часто применяется для анализа реальных ситуаций так называемое *нормальное* (или *гауссово*) распределение. Оно зависит от двух параметров, μ и σ, и задается плотностью вида

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(рис. 8). Случайную величину, распределенную нормально с параметрами μ и σ, будем обозначать $N(\mu, \sigma)$. Параметры μ и σ имеют вполне ясный смысл: это соответственно математическое ожидание и стандартное отклонение. Зависимость нормального распределения от параметров мы обсудим несколько позже, сначала же рассмотрим важный частный случай.

При $\mu = 0$ и $\sigma = 1$ получается *стандартное нормальное распределение* $N(0, 1)$ (напомним, что стандартной называется случайная величина с нулевым математическим ожиданием и единичной дисперсией). Его плотность будем обозначать особым образом, $\varphi(x)$, а

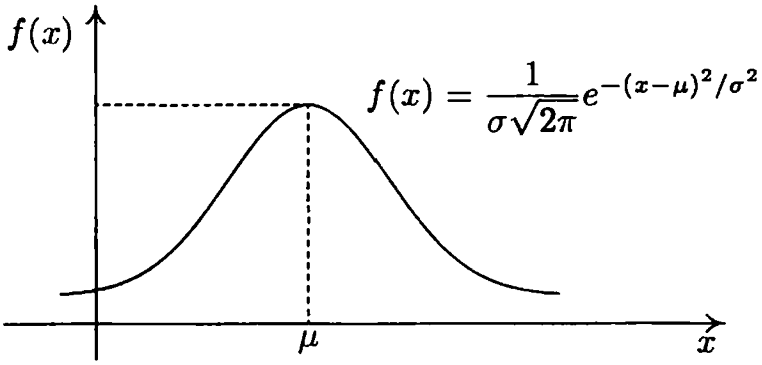


Рис. 8

саму случайную величину — U. Ясно, что

$$U \sim \varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Функция $\varphi(x)$ четная, т. е. $\varphi(-x) = \varphi(x)$, и, следовательно, график симметричен относительно оси ординат (рис. 9). Поэтому для

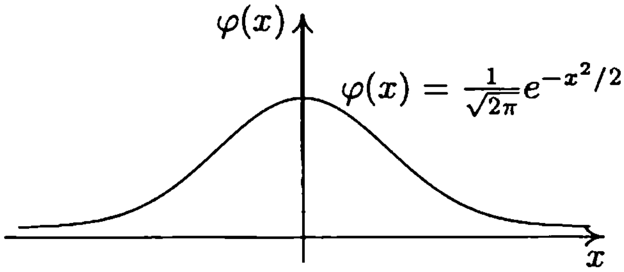


Рис. 9

исследования функции $\varphi(x)$ достаточно рассмотреть ее для значений $x \geqslant 0$. В точке $x = 0$ функция $\varphi(x)$ имеет максимум

$$\varphi(0) = \frac{1}{\sqrt{2\pi}},$$

а с увеличением аргумента x убывает, причем это убывание происходит довольно быстро:

x	0,0	0,5	1,0	1,5	2,0	2,5	3,0
$\varphi(x)$	0,399	0,352	0,242	0,130	0,054	0,018	0,001

Перейдем к рассмотрению основной задачи — вычислению вероятностей типа

$$p(a \leqslant U \leqslant b).$$

С этой целью определим для неотрицательных чисел x функцию $\Phi(x)$. Имеем:

$$\Phi(x) = \int_0^x \varphi(\tau)\, d\tau.$$

Площадь фигуры, заштрихованной на рис. 10, равна $\Phi(x)$. Иначе говоря, функция $\Phi(x)$ — первообразная функции $\varphi(x)$.

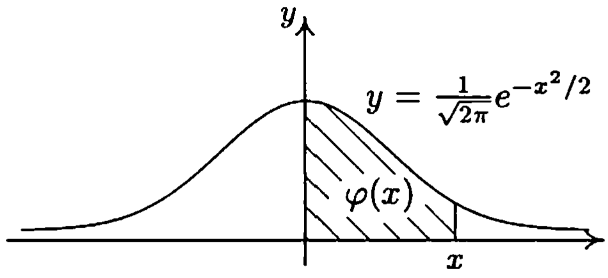


Рис. 10

Для вычислений, связанных с функцией $\Phi(x)$, обычно пользуются таблицами различной степени подробности. Мы будем пользоваться следующей таблицей (для удобства она повторена в приложении в конце книги):

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,1	0,040	1,1	0,364	2,1	0,482
0,2	0,079	1,2	0,385	2,2	0,486
0,3	0,118	1,3	0,403	2,3	0,489
0,4	0,155	1,4	0,419	2,4	0,492
0,5	0,192	1,5	0,433	2,5	0,494
0,6	0,226	1,6	0,445	2,6	0,495
0,7	0,258	1,7	0,455	2,7	0,497
0,8	0,288	1,8	0,464	2,8	0,497
0,9	0,316	1,9	0,471	2,9	0,498
1,0	0,341	2,0	0,477	3,0	0,499

Можно считать, что при $x > 3$ имеем $\Phi(x) \approx 0{,}5$.

При помощи этой таблицы можно сразу найти, например, следующие вероятности:

$$p(0 \leqslant U \leqslant 0{,}5) = \Phi(0{,}5) = 0{,}192,$$
$$p(0 \leqslant U \leqslant 0{,}3) = \Phi(0{,}3) = 0{,}118.$$

Из симметричности графика плотности распределения (рис. 9), легко видеть, что

$$p(U \geqslant 0) = p(U \leqslant 0) = 0,5.$$

Поэтому

$$p(U \geqslant 0{,}7) = 0{,}5 - p(0 \leqslant U \leqslant 0{,}7) = 0{,}5 - \Phi(0{,}7) = 0{,}5 - 0{,}258 = 0{,}242,$$
$$p(U \geqslant 0{,}4) = 0{,}5 - p(0 \leqslant U \leqslant 0{,}4) = 0{,}5 - \Phi(0{,}4) = 0{,}5 - 0{,}155 = 0{,}345$$

(рис. 11).

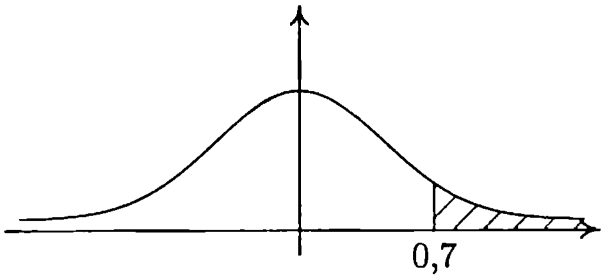


Рис. 11

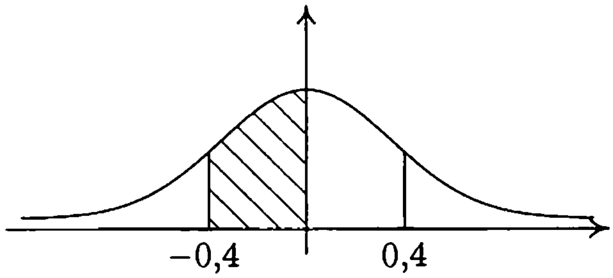


Рис. 12

Поскольку график функции $\varphi(x)$ симметричен, справедливы следующие соотношения:

$$p(-0{,}4 \leqslant U \leqslant 0) = p(0 \leqslant U \leqslant 0{,}4) = \Phi(0{,}4) = 0{,}155,$$
$$p(-0{,}8 \leqslant U \leqslant 0) = p(0 \leqslant U \leqslant 0{,}8) = \Phi(0{,}8) = 0{,}288$$

(рис. 12).

Наконец, вот еще три случая:

$$p(0{,}1 \leqslant U \leqslant 0{,}3) = p(-0{,}3 \leqslant U \leqslant -0{,}1) =$$
$$= \Phi(0{,}3) - \Phi(0{,}1) = 0{,}118 - 0{,}040 = 0{,}078,$$
$$p(-0{,}1 \leqslant U \leqslant 0{,}3) = \Phi(0{,}3) + \Phi(0{,}1) = 0{,}118 + 0{,}040 = 0{,}158.$$

(рис. 13, 14, 15).

Замечание. Во всех этих формулах знак $\leqslant$ можно заменить на знак $<$, а знак $\geqslant$ — на знак $>$.

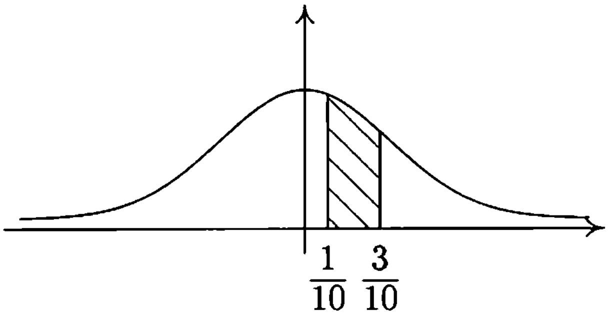


Рис. 13

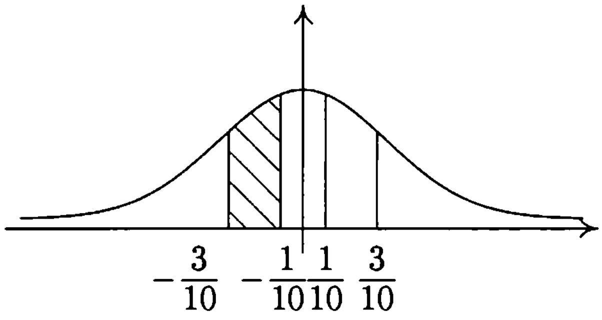


Рис. 14

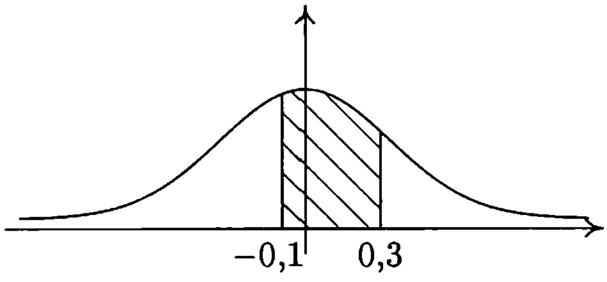


Рис. 15

Перейдем теперь к рассмотрению нормального распределения общего вида. Напомним, что нормальное распределение с математическим ожиданием μ и стандартным отклонением σ обозначается через $N(\mu, \sigma)$.

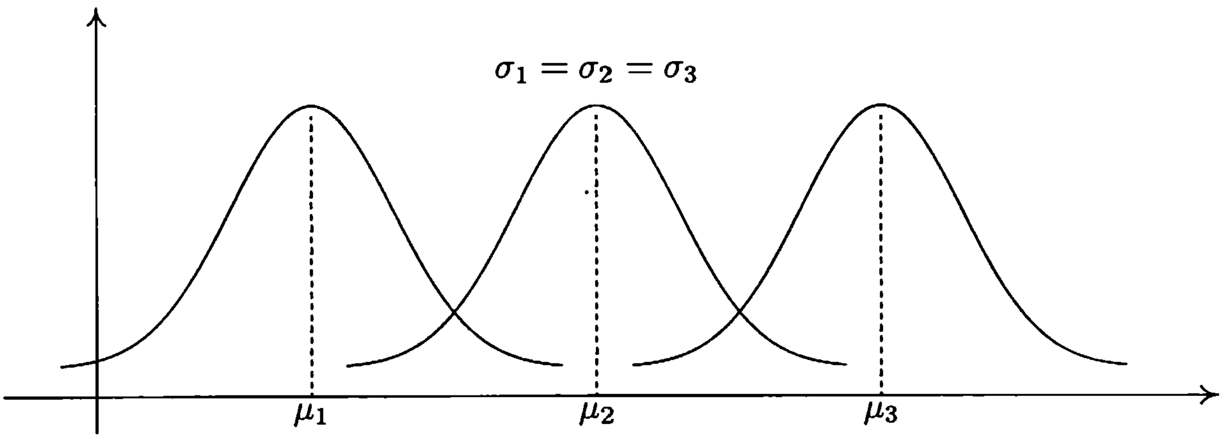


Рис. 16

На рис. 16, 17 показано, как график плотности нормального распределения зависит от параметров μ и σ. Чем больше μ, тем правее расположена наивысшая точка графика (при одинаковых σ). Чем больше σ, тем более пологим является график (при одинаковых μ).

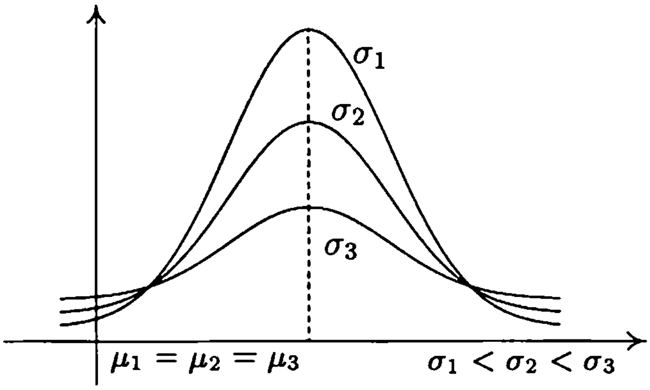


Рис. 17

Пусть имеется случайная величина $X = N(\mu, \sigma)$. Вычисление вероятностей типа $p(a \leqslant X \leqslant b)$ сводится к вычислению аналогичных вероятностей для стандартной величины $U = N(0, 1)$. Именно: оказывается, что линейные операции над нормальной случайной величиной приводят опять к нормальной случайной величине (с другими числовыми характеристиками). Для нас сейчас важно то, что, *отнимая от нормальной случайной величины ее математическое ожидание и деля на стандартное отклонение, получаем в результате стандартную нормальную величину* U:

$$\frac{X-\mu}{\sigma} = U.$$

Поэтому

$$p(a \leqslant X \leqslant b) = p\left(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\right) = \\ = p\left(\frac{a-\mu}{\sigma} \leqslant U \leqslant \frac{b-\mu}{\sigma}\right).$$

Аналогично

$$p(X \leqslant c) = p\left(\frac{X-\mu}{\sigma} \leqslant \frac{c-\mu}{\sigma}\right) = p\left(U \leqslant \frac{c-\mu}{\sigma}\right),$$

$$p(X \geqslant c) = p\left(\frac{X-\mu}{\sigma} \geqslant \frac{c-\mu}{\sigma}\right) = p\left(U \geqslant \frac{c-\mu}{\sigma}\right).$$

В частности, для некоторых конкретных значений получаем:

$$p(6 \leqslant N(7{,}2) \leqslant 8) = p\left(\frac{6-7}{2} \leqslant U \leqslant \frac{8-7}{2}\right) =$$

$$= p\left(-\frac{1}{2} \leqslant U \leqslant \frac{1}{2}\right) = 2\Phi(0{,}5) = 2 \cdot 0{,}192 = 0{,}384,$$

$$p(N(9{,}6) \geqslant 10) = p\left(U \geqslant \frac{10-9}{6}\right) = p\left(U \geqslant \frac{1}{6}\right) \approx$$

$$\approx p(U \geqslant 0{,}17) = 0{,}5 - \Phi(0{,}2) = 0{,}5 - 0{,}079 = 0{,}421,$$

$$p(N(8{,}5) \leqslant -2) = p\left(U \leqslant \frac{-2-8}{5}\right) = p(U \leqslant -2) =$$

$$= p(U \geqslant 2) = 0{,}5 - \Phi(2) = 0{,}421.$$

Приведем еще один пример.

Пример 13. Вес пачек с рисом, расфасованных фирмой “Новый рис”, является нормально распределенной случайной величиной с математическим ожиданием 1 кг и стандартным отклонением 10 г. Какой процент пачек имеет вес:

а) от 990 г до 1020 г;

б) менее 990 г;

в) более 1020 г?

Решение. Вес пачки риса является нормально распределенной величиной $X = N(1000, 10)$. Поэтому

$$\text{а) } p\,(990 < X < 1020) = p\left(\frac{990-1000}{10} < U < \frac{990+1000}{10}\right) =$$

$$= p\,(-1 < U < 2) = \Phi(1) + \Phi(2) = 0{,}341 + 0{,}477 = 0{,}818;$$

$$\text{б) } p\,(X < 990) = p\left(U < \frac{990-1000}{10}\right) =$$

$$= p\,(U < -1) = 0{,}5 - \Phi(1) = 0{,}5 - 0{,}341 = 0{,}159;$$

$$\text{в) } p\,(X > 1020) = p\left(U > \frac{1020-1000}{10}\right) =$$

$$= p\,(U > 2) = 0{,}5 - \Phi(2) = 0{,}5 - 0{,}477 = 0{,}023.$$

Ответ: а) 81,8%; б) 15,9%; в) 2,3%.

Наряду с вычисленем вероятностей, связанных с нормальным распределением, можно решать задачи и другого типа.

Пример 14. Найти такое число C, что

$$p(0 < U < C) = 0{,}2.$$

Решение. Имеем:

$$p(0 < U < C) = \Phi(C).$$

По условию

$$\Phi(C) = 0{,}2.$$

Обратимся к таблице значений функции Φ. Среди ее значений нет числа 0,2. Поэтому ищем ближайшее к нему. Это число 0,192, соответствующее значению $C = 0{,}5$.

Ответ: $C = 0{,}5$.

Пример 15. Найти такое число C, что

$$p(U < C) = 0{,}1.$$

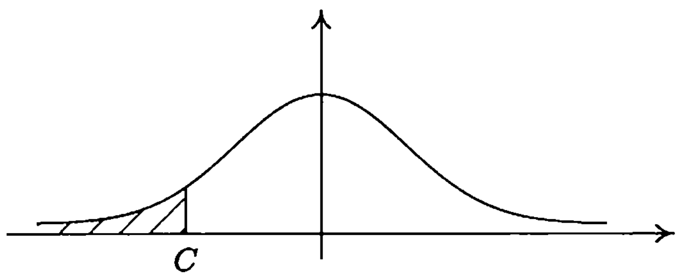


Рис. 18

Решение. Для любого положительного числа C выполняется неравенство

$$p(U < C) > 0{,}5.$$

Поэтому искомое значение C — число отрицательное (рис. 18). С учетом этого получаем, что

$$p(U < C) = 0{,}5 - \Phi(-C)$$

(в качестве аргумента функции Φ указано положительное число $-C$). Далее,

$$0{,}5 - \Phi(-C) = 0{,}1,$$
$$\Phi(-C) = 0{,}4.$$

Среди значений функции Ф ищем ближайшее к 0,4. Это число 0,445, соответствующее значению аргумента 1,6. Поэтому

$$-C = 1{,}6.$$

Ответ: $C = -1{,}6$.

Пример 16. Вес пачек с рисом, расфасованных фирмой "Новый рис", является нормально распределенной случайной величиной с математическим ожиданием 1 кг и стандартным отклонением 10 г. Проверка показала, что 2,5% выпускаемых пачек имеют вес меньше минимально допустимого стандартом. Каков этот минимальный вес?

Решение. Вес пачки риса является нормально распределенной величиной $X = N(1000, 10)$. Для ответа на вопрос достаточно решить уравнение

$$p(X < C) = 0{,}025.$$

Преобразуем его:

$$p\left(U < \frac{C - 1000}{10}\right) = 0{,}025.$$

Ясно, что

$$\frac{C - 1000}{10} < 0.$$

С учетом этого получаем

$$0,5 - \Phi\left(\frac{1000 - C}{10}\right) = 0{,}025,$$

$$\Phi\left(\frac{1000 - C}{10}\right) = 0{,}475,$$

$$\frac{1000 - C}{10} = 2,$$

$$C = 980.$$

Ответ: Минимально допустимый стандартный вес составляет 980 г.

12.8. Формула Муавра–Лапласа

Напомним, что биномиальное распределение $Bi(n,p)$ имеет два параметра — n и p. Если случайная величина X распределена по закону $Bi(n,p)$, то это означает следующее:

$$p(X=k) = C_n^k p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

При этом

$$EX = np, \qquad DX = np(1-p).$$

Между биномиальным и нормальным распределениями существует простая связь. При больших значениях n справедлива следующая приближенная формула:

$$Bi(n,p) \approx N\left(np, \sqrt{np(1-p)}\right). \tag{2}$$

Эта формула называется *формулой Муавра–Лапласа.* Мы будем применять ее для вычисления вероятностей, связанных с биномиальным распределением, при выполнении условий

$$n > 90, \qquad .np(1-p) > 9.$$

Пример 17. Игральный кубик бросают 120 раз. Какова вероятность того, что число выпаданий "единицы" будет лежать в пределах от 20 до 30?

Решение. Здесь мы имеем дело с биномиальной случайной величиной $Bi(120, 1/6)$. Проверим, можно ли применить формулу Муавра–Лапласа:

$$n = 120 > 90,$$
$$np(1-p) = 120 \cdot \frac{1}{6} \cdot \frac{5}{6} \approx 16{,}7 > 9.$$

Условия выполнены. Поэтому случайную величину $Bi(120, 1/6)$ можно заменить нормальным распределением $N(\mu, \sigma)$ с параметрами

$$\mu = np = 120 \cdot \frac{1}{6} = 20,$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{120 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 4{,}1.$$

Имеем:

$$p\left(20 \leqslant Bi\left(120, \frac{1}{6}\right) \leqslant 30\right) \approx p\,(20 \leqslant N(20; 4{,}1) \leqslant 30) =$$

$$= p\left(\frac{20-20}{4{,}1} \leqslant U \leqslant \frac{30-20}{4{,}1}\right) = p(0 \leqslant U \leqslant 2{,}4) = \Phi(2,4) = 0,492.$$

Пример 18. На некотором производстве регулярно проводится проверка качества продукции, для чего выбираются случайным образом 500 изделий и проверяются. Опыт показывает, что в среднем 11 изделий оказываются бракованными. Найти вероятность того, что во время очередной проверки число бракованных изделий окажется не менее 20.

Решение. Число бракованных изделий является биномиальной случайной величиной с параметрами $n = 500$ и $p = \dfrac{11}{500} = 0{,}022$. Проверим условия применимости формулы Муавра–Лапласа. Имеем:

$$n = 500 > 90,$$
$$np(1-p) = 11 \cdot (1 - 0{,}022) = 10{,}758 > 9.$$

Условия выполнены, поэтому величину $Bi(500; 0{,}022)$ можно заменить нормальной случайной величиной с параметрами:

$$\mu = np = 11,$$
$$\sigma = \sqrt{np(1-p)} = 3{,}3.$$

Имеем:

$$p\,(Bi(500; 0{,}022) \geqslant 20) \approx p\,(N(11; 3{,}3) \geqslant 20) = p\left(\cdot U \geqslant \frac{20-11}{3{,}3}\right) =$$

$$= p(U \geqslant 2{,}7) = 0{,}5 - \Phi(2{,}7) = 0{,}5 - 0{,}497 = 0{,}003.$$

Замечание (*центральная предельная теорема*). Формула Муавра–Лапласа является частным случаем следующего удивительного утверждения:

сумма большого числа почти произвольных случайных величин распределена приближенно по нормальному закону.

Этот факт является теоретическим обоснованием того, что многие реально наблюдаемые величины, испытывающие влияние множества случайных факторов, распределены по нормальному закону.

12.9. Задания и ответы

1. Монету бросают 5 раз. Случайная величина X — число выпадений герба. Составьте таблицу распределения, найдите математическое ожидание и дисперсию случайной величины X.

Ответ: 2,5; 1,25.

2. В лотерее на каждые 100 билетов приходится 15 выигрышей. Количество и размер выигрышей заданы в таблице:

Размер выигрыша	20	5	1
Количество выигрышей	1	4	10

Требуется составить закон распределения случайной величины — размера выигрыша в лотерее, приходящегося на один билет, а также найти математическое ожидание и дисперсию этой случайной величины.

Ответ: 0,5; 4,85.

3. Найдите вероятности (здесь и в следующей задаче $U = N(0,1)$ — стандартная нормальная случайная величина):

а) $p(0 \leqslant U \leqslant 0{,}7)$;

б) $p(U \geqslant 0{,}7)$;

в) $p(0{,}3 \leqslant U \leqslant 0{,}4)$;

г) $p(-0{,}1 \leqslant U \leqslant 0{,}2)$;

д) $p(-1{,}9 \leqslant U \leqslant -1{,}3)$;

е) $p(U \geqslant -2)$.

Ответ: а) 0,258; б) 0,242; в) 0,037; г) 0,119; д) 0,068; е) 0,977.

4. Найдите число x такое, что

а) $p(0 \leqslant U \leqslant x) = 0{,}4$;

б) $p(|U| < x) = 0{,}95$.

Ответ: а) 1,3; б) 2.

5. Найдите следующие вероятности:

а) $p(0 \leqslant N(1,3) \leqslant 2)$;

б) $p(N(-20,10) < -10)$;

в) $p(11,3 \leqslant N(15,7) \leqslant 16)$;

г) $p(3 \leqslant N(0,30) \leqslant 9)$;

д) $p(N(37,37) > 100)$.

Ответ: а) 0,236; б) 0,841; в) 0,232; г) 0,078; д) 0,045.

6. Производительность труда рабочих некоторого цеха является нормально распределенной случайной величиной с математическим ожиданием 90 кг за смену и стандартным отклонением 15 кг за смену. Вычислите долю рабочих, производительность которых:

а) лежит в промежутке от 80 до 110 кг за смену;

б) превышает 110 кг за смену;

в) менее 80 кг за смену.

г) Какой следует установить норму дневной выработки, чтобы 90% рабочих ее выполняли?

Ответ: а) 0,66; б) 0,1; в) 0,24; г) 70,5 кг.

7. Производство дает 1% брака. Какова вероятность того, что из взятых на исследование 1100 изделий забраковано будет не более 17?

Ответ: 0,964.

8. Какова вероятность того, что при 100-кратном бросании монеты число выпаданий герба будет от 45 до 55?

Ответ: 0,682.

Глава 13
О МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

13.1. Вводные замечания о математической статистике

Математическая статистика изучает различные методы обработки и осмысления результатов многократно повторяемых случайных событий. Понятие случайного события определяется в теории вероятностей, обработка результатов также производится при помощи теоретически разработанных вероятностных методов. Задачей математической статистики является построение и оценка адекватности идеальных вероятностных моделей реальных процессов.

В качестве иллюстрации рассмотрим следующий пример. Пусть имеется "реальный" игральный кубик. Обычно рассматривается модель, в соответствии с которой при бросании кубика вероятность выпадания любого числа очков от 1 до 6 одинакова и равна 1/6. Однако для "реального" кубика может случиться так, что при бросании его, например, 1000 раз шестерка выпала в 300 случаях. В принципе такое может произойти и в рамках модели с вероятностью 1/6, однако здравый смысл подсказывает, что скорее всего у кубика смещен центр тяжести. Это означает, что в дальнейшем имеет смысл использовать другую гипотезу относительно вероятности выпадания шестерки при бросании данного кубика. Например, довольно логично предположить, что эта вероятность близка к 0,3.

Для процесса построения и применения моделей характерно следующее обстоятельство: чем больше данных, тем точнее, адекватнее модель. В полной мере это относится к статистическим моделям. И в приведенном выше примере с кубиком вывод делается на основе многократного повторения испытания (бросания кубика).

Поскольку мы имеем дело со случайными событиями, то рекомендации, полученные на основе статистических соображений, всегда носят вероятностный характер. Однако это ни в коей мере не снижает их ценности. Напротив, вероятностный характер модели является показателем близости к описываемой реальной ситуации, которая зачастую слишком сложна для детерминированного описания.

Приведем несколько примеров практических задач, требующих применения статистических методов.

1. Пусть по некоторому вопросу в городе опрошено N человек, из них M дали положительный ответ и $(N-M)$ — отрицательный. Как по этим данным оценить долю горожан, дающих положительный ответ? Понятно, что эта доля близка к M/N, но в какой степени?

Замечание. Имеется еще немаловажный вопрос о том, каким образом правильно организовать опрос (чтобы учесть мнение различных групп населения), но здесь мы этой проблематики касаться не будем.

2. Предположим, что на некоторой фирме применили нововведение: внедрили новую технологию, перешли на выпуск новой продукции и т. п. Для простоты будем считать, что регистрируемыми данными являются значения производительности труда в течение дня. Эта производительность зависит от ряда случайных факторов и, следовательно, является случайной величиной.

Пусть имеется последовательность чисел — производительность за некоторый срок. Например: 7, 11, 10, 6, 6, 9, 9, 10, 5, 6, 10, 7 (в этот момент процесс производства был изменен, последующие данные относятся к новой ситуации), 11, 7, 9, 8, 10, 9.

Как по имеющимся числовым данным определить, достигнуто ли повышение производительности труда?

3. Требуется дать прогноз на изменение значения некоторой величины — курса доллара, спроса на продукцию, числа зрителей на стадионе и т. п. Основными исходными данными здесь являются значения, которые принимала изучаемая величина в прошлом.

13.2. Первичная обработка данных

Обработка данных и получение на ее основе каких-либо рекомендаций относительно принятия того или иного управленческого решения — процесс, вообще говоря, многоэтапный. Обычно полученные в результате наблюдений данные представляют собой набор чисел.

Просматривая этот набор, как правило, трудно выявить какую-либо закономерность. Поэтому данные подвергают некоторой первичной обработке, целью которой является упрощение дальнейшего анализа. Мы рассмотрим подробно один из возможных способов.

Рассмотрим данные, полученные в результате регистрации значений некоторой случайной величины, — набор чисел:

$$x_1, x_2, \dots, x_n$$

(отметим, что некоторые значения могут совпадать). Этот набор чисел называется *выборкой.*

Дальнейшие действия зависят от того, как много в выборке *различных* чисел. Если мы имеем дело с дискретной случайной величиной, то различных чисел немного; если с непрерывной случайной величиной, то могут и все числа оказаться различными. Поэтому далее рассмотрим два этих случая по отдельности.

Дискретный случай

Первый этап обработки выборки — это составление *вариационного ряда.* Его получают так: среди всех чисел x_i отбирают различные и располагают в порядке возрастания —

$$\alpha_1, \alpha_2, \dots, \alpha_m,$$

где $\alpha_1 < \alpha_2 < \dots < \alpha_m$.

Следующий этап обработки выборки — составление *дискретной таблицы частот*:

α_1	α_2	...	α_m
k_1	k_2	...	k_m
$n_1 = k_1/n$	$n_2 = k_2/n$	...	$n_m = k_m/n$

Здесь n — число всех измерений, k_i — число измерений, в которых наблюдалось значение α_i. Величины k_i называются *частотами*, а величины $n_i = k_i/n$ — *относительными частотами.*

Графической иллюстрацией дискретной таблицы частот является *столбиковая диаграмма* (рис. 1).

Замечание 1. Частоты и относительные частоты пропорциональны, поэтому при построении столбиковой диаграммы на вертикальной

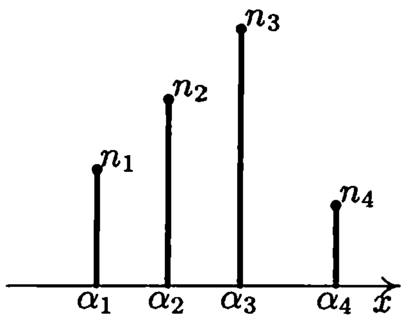


Рис. 1

оси можно указывать значения либо относительных частот, либо частот — визуальное восприятие от этого не зависит.

Пример 1. Пусть нашей задачей является выявление картины успеваемости студентов, сдавших экзамен по курсу "Математические модели в управлении". На курсе 56 человек. Полученные студентами оценки представляют собой (в порядке алфавитного списка) следующий набор чисел:

$$3,\ 4,\ 5,\ 4,\ 3,\ 3,\ 5,\ 4,\ 3,\ 5,\ 5,\ 2,\ 3,\ 5,\ 3,\ 5,\ 3,\ 5,\ 4,\ 4,\ 3,\ 3,\ 4,\ 3,\ 4,\ 3,\ 3,\ 5,$$
$$3,\ 3,\ 4,\ 3,\ 4,\ 3,\ 5,\ 3,\ 4,\ 4,\ 3,\ 5,\ 3,\ 3,\ 5,\ 4,\ 2,\ 5,\ 3,\ 4,\ 2,\ 3,\ 5,\ 4,\ 3,\ 5,\ 3,\ 5.$$

Это и есть исходные данные — выборка. Числа, составляющие выборку, представляют собой реализации случайной величины — оценки на экзамене.

Составление вариационного ряда не представляет сложностей. Вот он:

$$2,\ 3,\ 4,\ 5.$$

Теперь надо подсчитать, сколько раз встречается каждая из оценок. Это можно сделать непосредственно, однако существует и другой способ. Выписываются значения 2, 3, 4, 5 по одному на каждой строке. После этого выборка просматривается: одно число за другим, и для каждого значения ставится вертикальный отрезок в соответствующей строке. После этого подсчитывается число отрезков в каждой строке.

В данном случае имеем:

2: | | | — 3 значения,
3: | | | | | | | | | | | | | | | | | | | | | | | | — 24 значения,
4: | | | | | | | | | | | | | | — 14 значений,
5: | | | | | | | | | | | | | | | — 15 значений.

Таблица частот выглядит следующим образом:

2	3	4	5
3	24	14	15
$3/56 \approx 0{,}05$	$24/56 \approx 0{,}43$	$14/56 \approx 0{,}25$	$15/56 \approx 0{,}27$

Здесь в последней строке — относительные частоты, полученные при делении частот на число измерений $n = 56$.

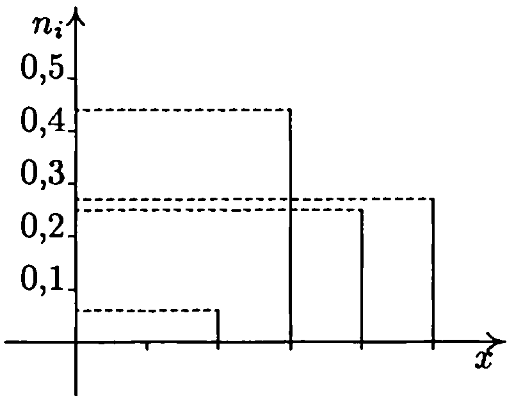


Рис. 2

Столбиковая диаграмма, иллюстрирующая полученную таблицу, изображена на рис. 2.

Непрерывный случай

Если число различных значений в выборке велико, вычислять частоту каждого из них не имеет большого смысла. Например, если *все* значения в выборке различны, то дискретная таблица частот имеет вид

α_1	α_2	$\dots$	α_m
1	1	$\dots$	1
$1/n$	$1/n$	$\dots$	$1/n$

Понятно, что такая таблица не добавляет наглядности.

Поэтому поступают следующим образом. Весь промежуток изменения значений выборки, от минимального до максимального, разбивают на интервалы. После этого подсчитывают число значений из выборки, попадающих в каждый интервал (частоты), а затем — относительные частоты. В результате получаем *интервальную та-*

блицу частот:

$[\mu_1; \mu_2]$	$(\mu_2; \mu_3]$	...	$(\mu_m; \mu_{m+1}]$
k_1	k_2	...	k_m
$n_1 = k_1/n$	$n_2 = k_2/n$	...	$n_m = k_m/n$

Здесь n — число всех измерений, m — число интервалов, k_i — количество чисел, приходящихся на i-й интервал, $n_i = k_i/n$ — относительная частота попадания в i-й интервал. Интервалы обычно берут одинаковой длины, хотя это и не обязательно.

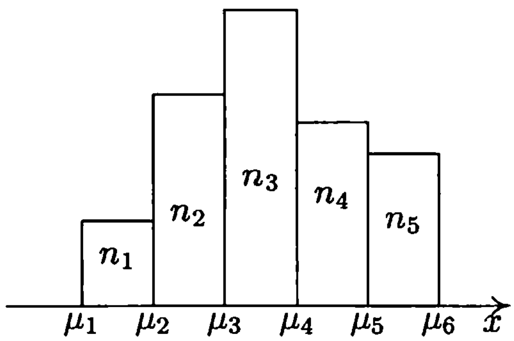


Рис. 3

Графической иллюстрацией интервальной таблицы частот является *гистограмма* (рис. 3). Гистограмма представляет собой ступенчатую линию; основанием i-й ступеньки является интервал $(\mu_i; \mu_{i+1}]$, а площадь этой ступеньки равна n_i.

Замечание 2. Число интервалов m выбирают из соображений наглядности получающейся гистограммы. Обычно m лежит в пределах от 5 до 15.

Замечание 3. Если интервалы $(\mu_i; \mu_{i+1}]$ выбраны одинаковой длины, то площади ступенек гистограммы пропорциональны их высотам, и в этом случае можно отмечать на оси ординат просто частоты k_i.

Пример 2. Предположим, что студенты некоторой группы, состоящей из 25 человек, написали контрольную работу. Каждый студент получил определенное количество баллов. Приведем эти баллы (в порядке алфавитного списка группы):

75, 145, 150, 180, 125, 150, 150, 165, 95, 135, 130, 70, 130, 105, 135, 135, 100, 160, 60, 85, 120, 60, 145, 150, 135.

Требуется построить интервальную таблицу частот и гистограмму.

Нетрудно найти среди приведенных чисел минимальное и максимальное — это числа 60 и 180. Таким образом, все значения лежат на отрезке [60;180]. Разобьем этот отрезок, например, на $m = 6$ равных частей. После этого подсчитаем число значений, попавших в каждый интервал (воспользуемся методом, описанным в примере 1):

[60;80]: | | | | — 4 значения,
(80;100]: | | | — 3 значения,
(100;120]: | | — 2 значения,
(120;140]: | | | | | | | — 7 значений,
(140;160]: | | | | | | | — 7 значений,
(160;180]: | | — 2 значения.

Построим теперь интервальную таблицу частот:

[60; 80]	(80; 100]	(100; 120]	(120; 140]	(140; 160]	(160; 180]
4	3	2	7	7	2
$\frac{4}{25} = 0{,}16$	$\frac{3}{25} = 0{,}12$	$\frac{2}{25} = 0{,}08$	$\frac{7}{25} = 0{,}28$	$\frac{7}{25} = 0{,}28$	$\frac{2}{25} = 0{,}08$

Соответствующая гистограмма изображена на рис. 4. На вертикальной оси проставлены частоты — см. замечание 3.

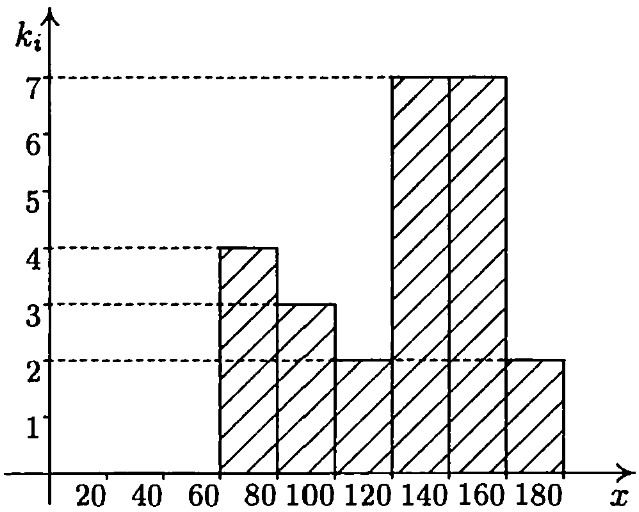


Рис. 4

Глава 14
ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ

Выше были рассмотрены некоторые методы первичной обработки данных, в ходе которой имеющиеся "сырые" результаты наблюдений преобразуются для достижения большей наглядности. Теперь мы рассмотрим методы, позволяющие, исходя из тех же данных, делать предположения относительно числовых характеристик наблюдаемой случайной величины — математического ожидания и дисперсии.

14.1. Точечные оценки

При помощи таблицы частот либо гистограммы можно судить (хотя бы примерно) о характере распределения наблюдаемой случайной величины. Однако часто интерес представляют численные характеристики распределения — математическое ожидание, дисперсия и др. Это обусловлено тем, что в процессе принятия решения удобнее опираться на небольшое число значимых параметров.

В связи с этим возникает задача: исходя из набора значений (выборки)

$$x_1, x_2, \ldots, x_n$$

величины X, полученного в результате n независимых наблюдений, оценить значение математического ожидания EX, дисперсии DX либо еще какого-нибудь параметра. Пока наши рассуждения носят общий характер, мы будем обозначать оцениваемый параметр буквой θ (тэта).

Повторим сказанное иными словами. Имеется случайная величина X, значения (реализации) которой $x_1, x_2, \ldots, x_n$ каким-либо образом становятся нам известными. (Этой случайной величиной может быть число посетителей данного магазина в течение дня, рост в сантиметрах студента данного вуза, годовой доход гражданина данной страны и пр.) У величины X имеется, скажем, математическое ожидание, которое нам неизвестно. Требуется найти способ, при помощи которого по известным реализациям величины X можно разумным образом оценить неизвестное математическое ожидание.

Сейчас мы рассмотрим, каким образом неизвестный параметр оценивается одним числом. Такая оценка называется *точечной.*

Любая оценка для θ — обозначим ее $\tilde{\theta}$ — будет представлять собой некоторое выражение, составленное из символов $x_1, x_2, \ldots, x_n$:

$$\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n).$$

Тем самым $\tilde{\theta}$ будет случайной величиной (принимающей свои значения в результате n опытов над X). Ее закон распределения будет зависеть от закона распределения случайной величины X (последнему подчинена каждая из величин $x_1, x_2, \ldots, x_n$) и от числа наблюдений n.

Замечание. Подчеркнем, что на момент принятия решения о значении параметра θ (т. е. после проведения испытаний (наблюдений)) величины $x_1, x_2, \ldots, x_n$ являются конкретными числами. Однако на момент, когда выбирается сам метод (алгоритм, формула) вычисления $\tilde{\theta}$, величины $x_1, x_2, \ldots, x_n$ являются случайными, ведь еще неизвестно, какие именно значения они примут.

Пример 1. Студент размышляет, ехать ли ему домой или идти в библиотеку. Не находя решающих аргументов в пользу того или другого, он решает провести испытание — подбросить монету, и если выпадет герб, то ехать домой (в противном случае — идти в библиотеку). В сущности, решение тем самым принято (определен алгоритм действий). Однако до подбрасывания монеты неизвестно, что же выпадет, герб или цифра (соответственно неизвестно и направление дальнейших передвижений студента).

Естественно предъявить к оценке $\tilde{\theta}$ некоторые требования. Упомянем два важнейших.

1) Желательно, чтобы при использовании величины $\tilde{\theta}$ вместо неизвестного параметра θ мы не делали систематических ошибок ни в

сторону завышения, ни в сторону занижения, т. е. чтобы выполнялось равенство

$$E\tilde{\theta} = \theta.$$

Оценка, удовлетворяющая такому условию, называется *несмещенной.*

2) Желательно, чтобы с увеличением числа n опытов значения случайной величины $\tilde{\theta}$ концентрировались около θ все более тесно, т. е. чтобы с ростом n точность оценки возрастала. Поскольку мерой рассеяния случайной величины вокруг ее математического ожидания является дисперсия, это требование можно сформулировать так:

$$D\tilde{\theta} \longrightarrow 0 \qquad \text{при} \qquad n \longrightarrow \infty$$

(дисперсия оценки стремится к нулю при неограниченном возрастании числа наблюдений). Оценка, удовлетворяющая такому условию, называется *состоятельной.*

Рассмотрим важный пример — точечную оценку для математического ожидания EX. Таким образом, в данном случае $\theta = EX$. В качестве такой оценки примем так называемое *эмпирическое среднее*, т. е. среднее арифметическое величин $x_1, x_2, \ldots, x_n$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Случайные величины $x_1, x_2, \ldots, x_n$ имеют один и тот же закон распределения, совпадающий с законом распределения величины X. Поэтому

$$E\bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}nEX = EX.$$

Таким образом, оценка $\bar{X}$ является несмещенной. Дисперсия этой оценки вычисляется следующим образом:

$$D\bar{X} = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{1}{n^2}nDX = \frac{DX}{n}.$$

Выражение DX/n, очевидно, стремится к нулю при неограниченном возрастании n. Отсюда вытекает состоятельность оценки $\bar{X}$.

Соотношения

$$E\bar{X} = EX, \qquad D\bar{X} = \frac{1}{n}DX \tag{1}$$

понадобятся нам в дальнейших рассмотрениях.

14.2. Интервальные оценки

Рассмотренная в предыдущем пункте точечная оценка часто бывает достаточной для практических выводов. Однако если есть необходимость в более детальном анализе, то надо оценить, насколько истинное значение параметра расходится с точечной оценкой этого значения.

В этом случае можно поступить следующим образом. Выберем интервал (θ_1, θ_2) таким образом, чтобы вероятность включения в этот интервал параметра θ была достаточно велика (близка к единице). Говоря более строго, это означает, что вероятность выполнения двойного неравенства

$$\theta_1(x_1, x_2, \ldots, x_n) < \theta < \theta_2(x_1, x_2, \ldots, x_n)$$

не меньше заданного числа γ.

Вероятность γ называется *доверительной вероятностью* или *надежностью*, а интервал (θ_1, θ_2) — *доверительным интервалом* (соответствующим доверительной вероятности γ). Обычно значение γ выбирают равным 0,95 либо 0,99.

Рассмотренная оценка называется *интервальной.* Еще раз подчеркнем, что интервальная оценка зависит не только от имеющихся данных, но и от требуемой надежности γ.

Методы построения интервальных оценок будут обсуждаться в последующих разделах. Сейчас же приведем неформальный пример, поясняющий различие точечной и интервальной оценок.

Когда о каком-либо человеке говорят: “Ему примерно 38 лет”, это ни что иное, как точечная оценка возраста. Когда же говорят: “Ему лет 35–40”, это интервальная оценка, доверительный интервал — (35;40). Надежность оценки при этом в явном виде не указывается, но предполагается довольно близкой к единице.

Иногда можно слышать и высказывания такого рода: “Ему лет 35–40, по крайней мере не больше 45”. Очевидно, что доверительный интервал (35;45) имеет бо́льшую доверительную вероятность, чем интервал (35;40). Однако интервальная оценка (35;40) более информативна, чем оценка (35;45).

14.3. Оценки математического ожидания нормального распределения

В этом разделе мы рассмотрим конкретный пример построения точечной и интервальной оценок. И здесь, как и ранее, исходными данными являются результаты наблюдений

$$x_1, x_2, \ldots, x_n.$$

Кроме того, известно, что эти числа являются реализациями нормальной случайной величины $N(\mu, \sigma)$.

Далее вопрос может быть поставлен различным образом:

требуется оценить неизвестный параметр μ;

требуется оценить неизвестный параметр σ;

известен параметр σ, требуется оценить параметр μ.

Из перечисленных трех задач мы рассмотрим лишь одну — последнюю.

Итак, требуется найти точечную и интервальную оценки математического ожидания μ случайной величины X, распределенной по нормальному закону с известным стандартным отклонением σ. Исходными данными являются $\sigma; x_1, x_2, \ldots, x_n$ (для интервальной оценки требуется еще задать доверительную вероятность γ).

Точечная оценка

Точечную оценку $\tilde{\mu}$ параметра μ определим как эмпирическое среднее:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Как уже отмечалось, такая оценка является несмещенной и состоятельной для любой случайной величины, в том числе, разумеется, и в данном случае.

Интервальная оценка

Для построения интервальной оценки необходимо сначала выбрать доверительную вероятность γ, которая после этого считается заданной.

Справедливо следующее утверждение: *сумма независимых случайных величин, каждая из которых распределена по нормальному* *закону, также распределена по нормальному закону.* Отсюда вытекает, что величина $\bar{X}$ распределена по нормальному закону.

Выше было показано (см. (1)), что

$$E\bar{X} = EX, \quad D\bar{X} = \frac{DX}{n}.$$

Применительно к данному случаю имеем:

$$E\bar{X} = \mu, \quad D\bar{X} = \frac{\sigma^2}{n},$$

так что $\bar{X} = N(\mu, \sigma/\sqrt{n})$.

Рассмотрим случайную величину

$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Ее распределение является нормальным с математическим ожиданием 0 и дисперсией 1. Пользуясь этим обстоятельством, можно по данному γ найти такое число u_γ, чтобы выполнялось соотношение

$$p(-u_\gamma < U < u_\gamma) = \gamma$$

(площадь области, заштрихованной на рис. 1, равна γ).

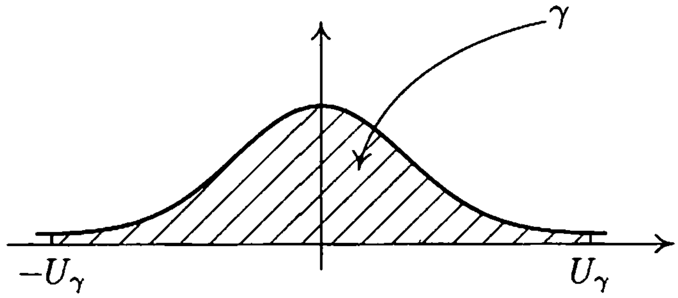


Рис. 1

Замечание. Индекс γ в выражении u_γ подчеркивает зависимость u от γ. Если по контексту понятно, какое γ имеется в виду, то, опуская индекс, мы будем писать просто u.

Справедливо равенство

$$p(-u_\gamma < U < u_\gamma) = 2p(0 < U < u_\gamma) = 2\Phi(u_\gamma).$$

Напомним, что функция $\Phi(x)$ служит для вычисления вероятностей, связанных с нормальным распределением. Таблица ее значений имеется в приложении (в конце книги).

Для нахождения числа u_γ достаточно решить уравнение

$$2\Phi(u_\gamma) = \gamma.$$

Число u_γ ищется по заданному числу γ при помощи таблицы значений функции $\Phi(x)$.

Получив u_γ, мы можем утверждать, что вероятность события

$$-u_\gamma < U < u_\gamma,$$

или, более подробно,

$$-u_\gamma < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < u_\gamma,$$

равна γ. Последнее неравенство равносильно следующему:

$$\bar{X} - u_\gamma \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_\gamma \frac{\sigma}{\sqrt{n}}.$$

Таким образом, с доверительной вероятностью γ математическое ожидание μ лежит в интервале (μ_1, μ_2), где

$$\mu_1 = \bar{X} - u_\gamma \frac{\sigma}{\sqrt{n}}, \tag{2}$$

$$\mu_2 = \bar{X} + u_\gamma \frac{\sigma}{\sqrt{n}}. \tag{3}$$

Число

$$\varepsilon = u_\gamma \frac{\sigma}{\sqrt{n}} \tag{4}$$

называется *погрешностью* оценки математического ожидания. С доверительной вероятностью γ можно считать, что оценка $\bar{X}$ отклоняется от истинного значения μ не более чем на ε (рис. 2).

Из формулы (4) видно, что с ростом числа данных n погрешность становится все меньше, т. е. доверительный интервал сужается. В

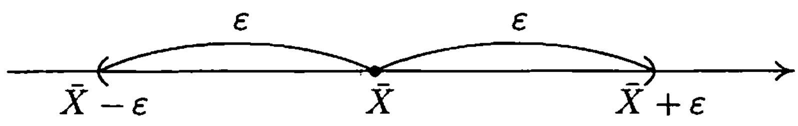


Рис. 2

свою очередь, чем меньше погрешность, тем больше данных необходимо собрать для ее достижения. Если требуемая погрешность ε задана, то можно найти соответствующий объем данных:

$$n \geqslant \left(\frac{u_\gamma\,\sigma}{\varepsilon}\right)^2. \tag{5}$$

При этом, разумеется, n должно быть целым числом. Так как сбор данных часто связан с некоторыми затратами, обычно берут минимальное целое n, удовлетворяющее неравенству (5).

Пример 2. При измерении нормальной случайной величины со стандартным отклонением, равным 5, и неизвестным математическим ожиданием μ получена следующая выборка:

$$3,\ 12,\ 8,\ 14,\ 15,\ 6,\ 19,\ 10,\ 15,\ 6.$$

Требуется найти: 1) интервал, содержащий параметр μ с доверительной вероятностью $\gamma = 0{,}95$; 2) погрешность оценки $\bar{X}$, соответствующую этой доверительной вероятности.

В данном случае

$$\sigma = 5, \quad n = 10, \quad \gamma = 0{,}95.$$

Для точечной оценки получаем

$$\bar{X} = \frac{3+12+8+14+15+6+19+10+15+6}{10} = 10{,}8.$$

Решим уравнение

$$2\Phi(u) = 0{,}95,$$

$$\Phi(u) = \frac{0{,}95}{2} = 0{,}475.$$

Для этого следует найти число 0,475 или ближайшее к нему в таблице среди значений функции $\Phi(x)$ (таблица дана в приложении). Ближайшее число — 0,477. Соответствующее ему значение аргумента

$$u = 2{,}0.$$

Теперь осталось воспользоваться формулами (2) и (3), подставив туда данные в условии σ, n и найденные $\bar{X}$, u_γ. Имеем:

$$\mu_1 = 10{,}8 - 2 \cdot \frac{5}{\sqrt{10}} \approx 10{,}8 - 3{,}16 \approx 7{,}64,$$

$$\mu_2 = 10{,}8 + 2 \cdot \frac{5}{\sqrt{10}} \approx 10{,}8 + 3{,}16 \approx 13{,}96.$$

Таким образом, искомым доверительным интервалом является интервал $(7{,}64; 13{,}96)$.

Для погрешности ε получаем следующее значение:

$$\varepsilon = 2 \cdot \frac{5}{\sqrt{10}} \approx 3{,}16.$$

14.4. Оценки вероятности события

Пусть некоторое событие происходит в результате единичного испытания с вероятностью P, которая нам неизвестна. Однако ситуация позволяет многократно повторить испытание и подсчитать, сколько раз произошло указанное событие. Более точно: пусть произведено n испытаний, в которых событие произошло m раз. Здесь в отличие от предыдущих рассмотрений исходными данными для анализа будут всего два числа — n и m.

Задача, которую мы будем рассматривать, заключается в отыскании оценки неизвестной вероятности P по имеющимся данным n, m и доверительной вероятности γ.

Точечная оценка

Не вызывает сомнений, что точечная оценка $\tilde{P}$ вероятности P определяется следующим соотношением:

$$\tilde{P} = \frac{m}{n}. \tag{6}$$

Докажем несмещенность и состоятельность этой оценки.

При любом фиксированном n величина m является случайной величиной с биномиальным законом распределения:

$$m = Bi(n, P).$$

Напомним, что в этом случае

$$Em = nP, \qquad Dm = nP(1-P).$$

Докажем несмещенность оценки $\tilde{P}$. Имеем:

$$E\tilde{P} = E\left(\frac{m}{n}\right) = \frac{1}{n}Em = \frac{1}{n}nP = P.$$

Теперь вычислим величину $D\tilde{P}$:

$$D\tilde{P} = D\left(\frac{m}{n}\right) = \frac{1}{n^2}Dm = \frac{1}{n^2}nP(1-P) = \frac{P(1-P)}{n}.$$

Таким образом,

$$D\tilde{P} = \frac{P(1-P)}{n} \longrightarrow 0 \qquad \text{при} \qquad n \longrightarrow \infty,$$

что доказывает состоятельность точечной оценки $\tilde{P}$.

Интервальная оценка

Для построения интервальной оценки вспомним (см. формулу (2) на с. 262), что биномиальная случайная величина $Bi(n,p)$ может быть приближена нормальной случайной величиной $N\left(np, \sqrt{np(1-p)}\right)$ при $n > 90$ и $np(1-p) > 9$ (эти условия будем далее считать выполненными). Поэтому можно считать, что

$$m \approx N\left(nP, \sqrt{nP(1-P)}\right).$$

Пользуясь этим обстоятельством, можно найти границы доверительного интервала (P_1, P_2) методом, сходным с методом вычисления доверительного интервала для математического ожидания нормального распределения. Выпишем сразу результат:

$$P_1 = \tilde{P} - u_\gamma\sqrt{\frac{\tilde{P}(1-\tilde{P})}{n}}, \tag{7}$$

$$P_2 = \tilde{P} + u_\gamma\sqrt{\frac{\tilde{P}(1-\tilde{P})}{n}}, \tag{8}$$

где u_γ — решение уравнения

$$\Phi(u_\gamma) = \frac{\gamma}{2}. \tag{9}$$

Пример 3. Из 200 случайным образом отобранных изделий некоторой фирмы 10 оказались бракованными. Найти доверительный интервал, содержащий с надежностью $\gamma = 0{,}99$ долю бракованных изделий среди всей продукции фирмы.

Решение. Здесь

$$n = 200, \quad m = 10, \quad \gamma = 0{,}99.$$

Применяя формулы (9), (6), (7), (8), получаем, что

$$\Phi(u) = \frac{0{,}99}{2} = 0{,}495,$$
$$u = 2{,}6$$

(напомним, что это значение находится по таблице),

$$\tilde{P} = \frac{10}{200} = 0{,}05$$

(это точечная оценка),

$$P_1 = 0{,}05 - 2{,}6 \cdot \sqrt{\frac{0{,}05 \cdot 0{,}95}{200}} \approx 0{,}05 - 0{,}04 = 0{,}01,$$

$$P_2 = 0{,}05 + 2{,}6 \cdot \sqrt{\frac{0{,}05 \cdot 0{,}95}{200}} \approx 0{,}05 + 0{,}04 = 0{,}09.$$

Таким образом, доверительный интервал оказался следующим: (0,01; 0,09). Иными словами, с доверительной вероятностью 0,99 доля бракованных изделий лежит в промежутке между 1% и 9%.

14.5. Задания и ответы

1. Выборка из большой партии электроламп содержит 100 ламп. Средняя продолжительность работы лампы из выборки оказалась равной 1000 ч. Найти 95%-й доверительный интервал для средней

продолжительности работы лампы, случайно выбранной из всей партии, если время работы является нормально распределенной случайной величиной со стандартным отклонением 40 ч.

Ответ: (992, 1008).

2. У нормально распределенной случайной величины стандартное отклонение равно 2. Найти минимальный объем выборки, при котором с надежностью, не меньшей 0,8, погрешность оценки математического ожидания будет меньше 0,3.

Ответ: 76.

3. Из 400 случайным образом отобранных изделий некоторой фирмы 20 оказались бракованными. Найти доверительный интервал, содержащий с надежностью $\gamma = 0{,}99$ долю бракованных изделий среди всей продукции фирмы.

Сравните полученный результат с тем, который был получен при разборе примера 3.

Ответ: (0,02; 0,08).

Глава 15
КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Одним из важных приложений методов математической статистики является установление зависимости между двумя или более наблюдаемыми величинами. При этом наряду с раздельным анализом выборок, составленных из значений этих величин, возможен и совместный анализ. Ниже рассматриваются некоторые методы такого анализа.

15.1. Корреляция

Рассмотрим ситуацию, когда в результате эксперимента измеряется не одна, а сразу две случайные величины, скажем X и Y. Примерами здесь могут служить врачебный осмотр, где у каждого пациента измеряют рост и вес; измерение средней температуры воздуха в двух городах в течение определенного дня; проверка квалификации рабочих, когда фиксируются производительность и стаж работы.

Итак, исходными данными являются пары чисел (точки)

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n), \tag{1}$$

где n — число испытаний. Наряду с анализом величин X и Y по отдельности представляет интерес исследование возможной зависимости между ними. Являются ли величины X и Y независимыми? Если же между ними имеется некоторая зависимость, то какова она?

Обратимся к рис. 1–4. На них изображены различные виды *графиков* (или *диаграмм*) *рассеяния*, т. е. нанесены точки (1). Величины X и Y на рис. 1, по-видимому, независимы: зная, какое значение приняла величина X, ничего нельзя сказать о значении Y. На рис. 2–4 зависимость налицо: зная значение, которое приняла величина X в результате испытания, можно довольно точно сказать, каково значение Y.

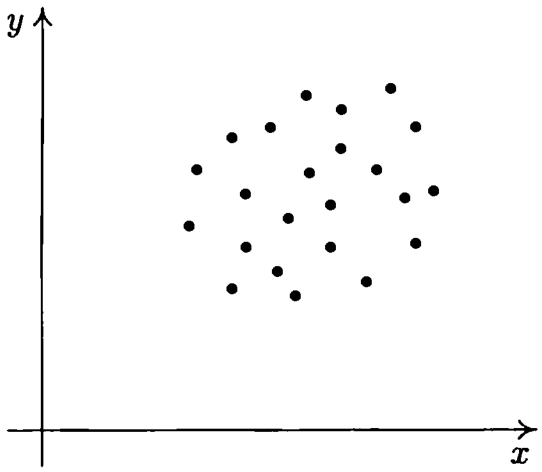


Рис. 1

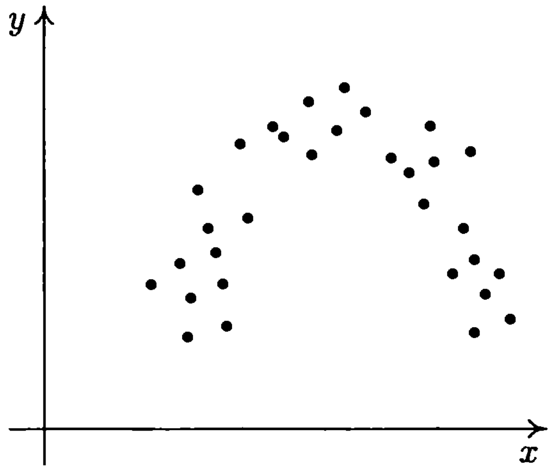


Рис. 2

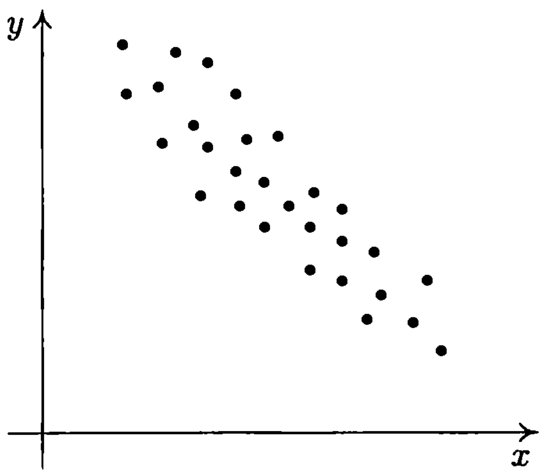


Рис. 3

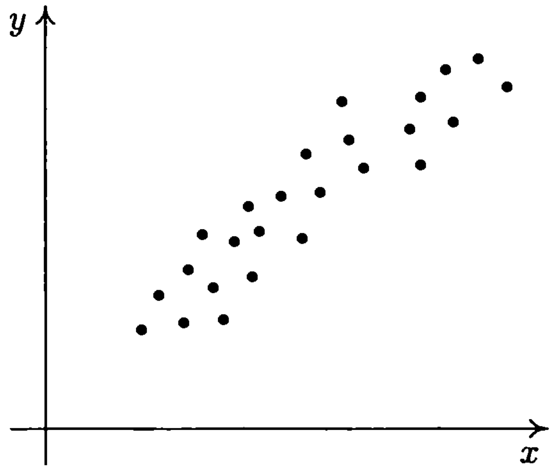


Рис. 4

Зависимость на рис. 3 и 4 близка к линейной, т. е. точки заметным образом группируются вокруг некоторой прямой. В таких случаях говорят, что величины X и Y *коррелированы*. Существует простой способ определения степени коррелированности случайных величин. Он основан на вычислении *коэффициента корреляции* r_{xy}. Если понятно, о каких случайных величинах идет речь, будем вместо r_{xy} писать просто r.

Коэффициент корреляции обладает следующим свойством:

$$-1 \leqslant r \leqslant 1.$$

При этом чем ближе r к нулю, тем слабее корреляция. И наоборот, чем ближе r к 1 или -1, тем сильнее корреляция, т. е. зависимость между X и Y близка к линейной. Если r в точности равно 1 или -1, то точки (1) лежат на одной прямой.

Подчеркнем, что коэффициент корреляции отражает степень *линейной* зависимости между величинами. При наличии ярко выраженной зависимости другого вида (например, квадратичной) он может быть близок к нулю.

Приведем формулы для вычисления r_{xy}:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{2}$$

$$s_x^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2, \qquad s_y^2 = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - \bar{y}^2, \tag{3}$$

$$s_{xy} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\bar{y}, \tag{4}$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}. \tag{5}$$

Пример. Рассмотрим проблему, которая стоит перед администрацией некоторого крытого стадиона, где проходят матчи, концерты и другие развлекательные мероприятия. Перед каждым таким мероприятием требуется оценить, какое количество зрителей придет, это необходимо для оптимальной организации работы различных вспомогательных служб. Один из подходов к решению этой проблемы — учет предыдущего опыта. В частности, можно предположить, что окончательное число зрителей сильно зависит от того, сколько билетов продано за день до мероприятия (как раз за сутки определяется план работы вспомогательных служб).

Пусть опыт первых пяти мероприятий этого года таков:

Число билетов, проданных накануне (в тыс.)	3,5	4,6	5,8	4,2	5,2
Число зрителей (в тыс.)	8,1	9,4	11,3	6,9	9,7

Каков коэффициент корреляции между числом проданных накануне билетов и числом зрителей?

Решение. Примем число билетов за X, а число зрителей за Y. В таблице даны пять реализаций пары случайных величин — пары чисел (x_i, y_i), $i = 1, \ldots, 5$. Для расчета коэффициента корреляции удобно найти сначала суммы

$$\sum_{i=1}^{5} x_i = 3{,}5 + 4{,}6 + 5{,}8 + 4{,}2 + 5{,}2 = 23{,}3,$$

$$\sum_{i=1}^{5} y_i = 8{,}1 + 9{,}4 + 11{,}3 + 6{,}9 + 9{,}7 = 45{,}4,$$

$$\sum_{i=1}^{5} x_i^2 = (3{,}5)^2 + (4{,}6)^2 + (5{,}8)^2 + (4{,}2)^2 + (5{,}2)^2 = 111{,}73,$$

$$\sum_{i=1}^{5} y_i^2 = (8{,}1)^2 + (9{,}4)^2 + (11{,}3)^2 + (6{,}9)^2 + (9{,}7)^2 = 423{,}36,$$

$$\sum_{i=1}^{5} x_i y_i = 3{,}5 \cdot 8{,}1 + 4{,}6 \cdot 9{,}4 + 5{,}8 \cdot 11{,}3 + 4{,}2 \cdot 6{,}9 + 5{,}2 \cdot 9{,}7 = 216{,}55.$$

Эти суммы подставим затем в формулы (2)–(5). Имеем:

$$\bar{x} = \frac{23{,}3}{5} = 4{,}66, \quad \bar{y} = \frac{45{,}4}{5} = 9{,}08,$$

$$s_x^2 = \frac{111{,}73}{5} - (4{,}66)^2 = 0{,}6304, \quad s_y^2 = \frac{423{,}36}{5} - (9{,}08)^2 = 2{,}2256,$$

$$s_{xy} = \frac{216{,}55}{5} - 4{,}66 \cdot 9{,}08 = 0{,}9972,$$

$$r_{xy} = \frac{0{,}9972}{\sqrt{0{,}6304}\,\sqrt{2{,}2256}} \approx 0{,}842.$$

Таким образом, коэффициент корреляции r оказался довольно близким к единице. Этим обстоятельством можно воспользоваться для прогнозирования числа зрителей по имеющейся накануне информации. О том, каким образом это делается, см. в продолжении этого примера на с. 290.

15.2. Регрессия

Предположим, что зависимость между случайными величинами X и Y близка к линейной (в этом случае коффициент корреляции r близок к 1 или -1). Тогда естественно ставить вопрос об отыскании функции

$$y = ax + b, \tag{6}$$

которая наилучшим образом выражает зависимость Y от X. Для нахождения такой функции пользуются *методом наименьших квадратов.*

Итак, пусть даны n пар чисел (иначе говоря, n точек):

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n).$$

Требуется найти такую прямую, чтобы сумма квадратов "отклонений" этих точек от прямой (6) была как можно меньше. Это означает, что выражение

$$\sum_{i=1}^{n} [y_i - (ax_i + b)]^2 \tag{7}$$

должно быть минимальным (на рис. 5 отклонения изображены в виде вертикальных отрезков).

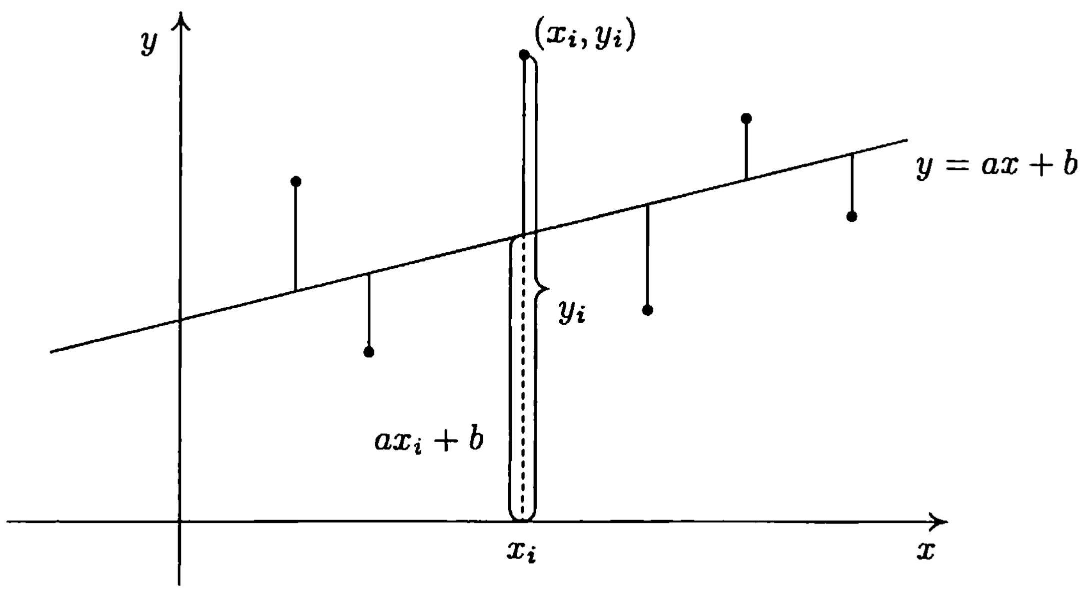


Рис. 5

Выражение (7) является функцией двух переменных a и b (поскольку результаты наблюдений заданы). Можно показать, что выражение (7) принимает минимальное значение, если величины a и b связаны соотношениями

$$\begin{cases} a \sum\limits_{i=1}^{n} x_i^2 + b \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ a \sum\limits_{i=1}^{n} x_i + n\, b = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Эта система имеет единственное решение:

$$a = \frac{s_{xy}}{s_x^2}, \qquad b = \bar{y} - a\bar{x}. \tag{8}$$

Найдя значения неизвестных параметров a и b, мы найдем тем самым прямую (6), наилучшим образом выражающую статистическую связь между величинами X и Y. Полученная прямая называется *прямой регрессии Y на X*.

Продолжение примера (начало см. на с. 287). Для прогнозирования числа зрителей надо найти прямую регрессии Y на X. Подставим найденные значения $s_{xy}, s_x^2, \bar{x}, \bar{y}$ в формулы (8). Получаем

$$a = \frac{0{,}9972}{0{,}6304} \approx 1{,}58, \qquad b \approx 9{,}08 - 1{,}58 \cdot 4{,}66 \approx 1{,}72.$$

Таким образом, прямая регрессии имеет уравнение

$$y = 1{,}58 \cdot x + 1{,}72.$$

Если, например, за день до мероприятия продано 4300 билетов, то предполагаемое число зрителей составляет

$$y = 1{,}58 \cdot 4{,}3 + 1{,}72 = 8{,}514 \approx 8{,}5 \text{ тыс.}$$

Если прямая регрессии найдена, можно оценить, насколько хорошо она приближает результаты наблюдений. Подставляя в выражение (7) найденные значения a и b, вычислим так называемую *среднюю квадратическую погрешность* или *ошибку* уравнения регрессии, которую будем обозначать буквой δ (дельта):

$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [y_i - (ax_i + b)]^2}.$$

Величина δ дает представление о том, какую в среднем длину имеют вертикальные отрезки на рис. 5. И чем меньше δ, тем ближе результаты наблюдений к найденной регрессионной прямой.

Замечание 1 (нелинейная регрессия). Вычисление средней квадратической погрешности имеет большое значение тогда, когда наряду с моделью линейной зависимости Y от X рассматриваются и другие модели, другие уравнения зависимости. (Например, в случае, приведенном на рис. 2, в качестве уравнения зависимости Y от X уместно подбирать не прямую, а параболу.) Для каждого из уравнений следует найти свою среднюю квадратическую погрешность, после чего выбрать из них минимальную. Соответствующая модель и является наилучшей (конечно, при отсутствии каких-либо дополнительных соображений).

Если, например, регрессионная кривая является параболой

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

то средняя квадратическая погрешность δ вычисляется по формуле

$$\delta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - (ax_i^2 + bx_i + c)]^2}.$$

Замечание 2 (множественная регрессия). Возможно рассмотрение зависимости параметра Y от нескольких параметров $X_1, X_2, \ldots, X_k$, например, в таком виде:

$$y = a_1X_1 + a_2X_2 + \cdots + a_kX_k + b,$$

где числа $a_1, a_2, \ldots, a_k, b$ подлежат определению.

15.3. Задания и ответы

1. При измерении в баллах результатов тестирования по истории (X) и географии (Y) получены следующие пары чисел для четырех школьников: (2, 2), (4, 5), (6, 7), (8, 10). Найдите коэффициент корреляции и прямую регрессии Y на X.

Ответ: $r_{xy} = 0{,}997$, $y = 1{,}3\,x - 0{,}5$.

2. Проводится исследование спроса на некоторый вид товара. Пробные продажи показали следующие данные о зависимости дневного спроса от цены:

Цена, руб.	10	12	14	16	18
Спрос, ед. товара	91	76	68	59	53

Требуется:

а) определить коэффициент корреляции между ценой X и спросом Y, построить прямую регрессии Y на X;

б) исходя из данных пункта а) определить спрос при цене 15 руб. за ед. товара.

Ответ: а) $r_{xy} = -0{,}986$; $y = -4{,}65x + 134{,}5$; б) 65.

3. В ситуации, описанной в предыдущей задаче, была предложена другая модель зависимости спроса от цены:

$$y = \frac{850}{x} + 6.$$

Какая модель является более адекватной экспериментальным данным?

Глава 16
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

16.1. Основные понятия. Примеры

Во многих случаях возникает необходимость на основе имеющихся данных решить, справедливо ли то или иное суждение. Например, верно ли, что A более меткий стрелок, чем B, что от дома до работы быстрее доехать на метро, а не на автобусе, что новое лекарство лучше, чем старое.

Если исходные данные таких суждений носят случайный характер (а во всех приведенных примерах это именно так!), то и ответы можно давать лишь с определенной степенью уверенности. Однако если вероятность ошибки мала, то суждение можно считать практически достоверным, как обычно и решают на практике.

Итак, *статистическая гипотеза — это предположение о случайной величине, проверяемое по выборке (по результатам наблюдений). Процедура сопоставления высказанной гипотезы с выборочными данными называется проверкой гипотезы.*

Пример 1. Подбросим монету 10 раз. На основании полученных при этом результатов можно судить о том, является ли монета правильной, т. е. одинаковы ли вероятности выпадения герба и цифры. Возможные гипотезы: 1) монета правильная, 2) чаще выпадает герб, 3) чаще выпадает цифра и т. п.

Важно, что исходное предположение должно быть формализовано в терминах параметров, участвующих в задаче. Выбрав в качестве параметра вероятность p выпадения герба, приведенные выше гипотезы можно записать так: 1) $p = 1/2$, 2) $p > 1/2$, 3) $p < 1/2$.

Пример 2. Стрелки A и B сделали по 3 выстрела в мишень. Гипотезы: 1) A стреляет лучше, 2) B стреляет не хуже, 3) A и B стреляют с одинаковой меткостью и т. п. Читателю предлагается подумать над вопросом: как в данном случае можно формализовать гипотезы?

Обозначим через Ω множество всевозможных результатов наблюдения (выборок) ω. Зафиксируем, далее, гипотезу H_0 (по каким-либо причинам наиболее важную) и будем называть ее *основной* или *нулевой*. Пусть мы хотим проверить ее, сопоставив с другой, *альтернативной гипотезой* H_1 (альтернативных гипотез может быть и несколько, и даже бесконечно много).

Выделим в Ω область S исходя из следующих соображений: если гипотеза H_0 верна, то наступление события $\omega \in S$ маловероятно. Записывается это так:

$$p\{\omega \in S | H_0\} = \alpha,$$

где число α близко к нулю. Иными словами, вероятность события $\omega \in S$ при условии, что верна гипотеза H_0, равна близкому к нулю числу α. Это число α называется *уровнем значимости* гипотезы H_0, а область S — *критической областью* гипотезы H_0.

Замечание. Слова "число α близко к нулю" могут быть уточнены лишь применительно к конкретной ситуации. Реальная ситуация принятия решения существенно влияет на критерий, какую вероятность считать малой. Например, при приеме на работу того или иного сотрудника вероятность ошибки $\alpha = 0{,}01$ следует в большинстве случаев считать малой. Если же речь идет о вероятности отказа тормозной системы автомобиля, то та же вероятность $\alpha = 0{,}01$ недопустимо высока (в среднем один отказ на каждые сто торможений).

Мы будем обычно полагать α равным 0,01 либо 0,05.

Будем считать, что если событие $\omega \in S$ все-таки произошло, то гипотеза H_0 отвергается. При этом, разумеется, можно допустить ошибку, заключающуюся в том, что гипотеза H_0 отвергается, хотя она верна. Это так называемая *ошибка первого рода.* Ее вероятность равна α.

Возможна и *ошибка второго рода*, которая состоит в том, что гипотеза H_0 принимается, хотя она неверна, а верна одна из альтернативных гипотез. В случае когда альтернативная гипотеза H_1 единственна и при этом однозначно определено вероятностное распределение на Ω, можно вычислить вероятность β ошибки второго рода

рода (в предположении, что верна гипотеза H_1!):

$$\beta = p\{\omega \notin S | H_1\}.$$

Рассмотрим процесс проверки статистической гипотезы на примере диагностики туберкулеза (пример принадлежит одному из основоположников математической статистики — Ю. Нейману).

Пример 3. Представим себе, что в некоторой клинике при обследовании делается несколько рентгеновских снимков пациента с целью обнаружения возможных признаков начинающегося туберкулеза. Допустим, что толкование нескольких снимков, принадлежащих одному и тому же пациенту, производится так, чтобы обеспечить независимость диагноза по каждому снимку от выводов, сделанных по предыдущим снимкам. Этого можно достигнуть, предлагая рентгенологу для просмотра в случайном порядке снимки, принадлежащие разным людям.

В дальнейшем мы будем употреблять термин “рентгеновский анализ” для обозначения комбинации из двух операций: производства снимка и толкования его рентгенологом. Таким образом, когда мы говорим, что проделано n рентгеновских анализов индивидуума, мы имеем в виду, что сделано n снимков и по каждому снимку поставлен диагноз.

Пусть прошлый опыт работы клиники выражается в следующем:

1) если пациент не несет никаких следов туберкулеза, то вероятность того, что отдельный рентгеновский снимок будет ошибочным и у пациента будет признан туберкулез, равна $p_1 = 0{,}01$;

2) если данный пациент страдает туберкулезом в умеренной форме, то вероятность того, что отдельный рентгеновский анализ определит болезнь, равна $p_2 = 0{,}6$. (Для простоты мы игнорируем категорию пациентов, страдающих туберкулезом в тяжелой форме.)

Предположим, что для каждого пациента делается $n = 5$ независимых анализов. Обозначим через ω число анализов, по которым вынесен положительный ответ. Предположим далее, что если $\omega = 0$, то пациент считается совершенно здоровым, в противном же случае он подвергается более тщательному обследованию по поводу туберкулеза.

Только что описанная процедура — это пример испытания статистической гипотезы.

Гипотез в данном случае две: первая — пациент страдает туберкулезом, вторая — пациент здоров. По-видимому, в данном случае важнее не проглядеть начинающуюся болезнь, чем не беспокоить пациента лишним анализом. Поэтому в качестве основной гипотезы H_0 примем первую — "пациент страдает туберкулезом". Вторая гипотеза является отрицанием основной, обозначим ее H_1 — "пациент здоров".

Множеством Ω всевозможных результатов наблюдений будет

$$\{0, 1, 2, 3, 4, 5\}$$

(рис. 1), поскольку положительный результат могут иметь $0, 1, \ldots, 5$ анализов.

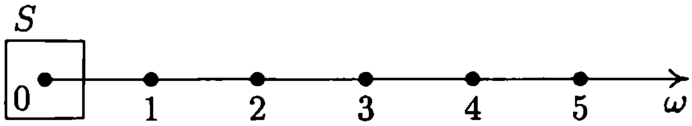


Рис. 1

Установившаяся практика состоит в том, что туберкулез подозревается во всех случаях, когда хотя бы один рентгеновский анализ дал положительный ответ. Это означает, что критическая область S состоит из одной выборочной точки $\omega = 0$. То есть лишь если среди $n = 5$ рентгеновских анализов нет ни одного положительного, гипотеза о том, что пациент болен, отвергается.

Найдем вероятность α ошибки первого рода. Она будет иметь место, если гипотеза H_0 истинна, но выборочная точка ω займет положение $\omega = 0$:

$$\alpha = p\{\omega = 0 | H_0\} = (1 - p_2)^n = (0{,}4)^5 = 0{,}01024.$$

Ошибка второго рода будет иметь место, если истинна гипотеза H_1, а выборочная точка ω не займет положения $\omega = 0$. Вероятность β ошибки второго рода равна

$$\beta = p\{\omega > 0 | H_1\} = 1 - p\{\omega = 0 | H_1\} = 1 - (1 - p_1)^n = \\ = 1 - (0{,}99)^5 \approx 0{,}04901.$$

Полученные формулы можно интерпретировать следующим образом. Если все сделанные предположения приближенно выполняются и в клинике постоянно используется описанная процедура многократных рентгеновских анализов, то окончательный диагноз, основанный на пяти анализах, будет отрицательным приблизительно для одного процента всех больных туберкулезом пациентов

($\alpha \approx 0{,}01 = 1\%$) и приблизительно для 95% всех пациентов, которых не коснулась болезнь ($1 - \beta \approx 0{,}05 = 5\%$).

Рассмотрим еще один пример процедуры проверки статистической гипотезы.

Пример 4. Крупная партия товара может содержать дефектные изделия. Поставщик полагает, что их доля составляет 3%, покупатель — 20%. Достигнута следующая договоренность: поставка состоится, если при проверке 10 случайно отобранных изделий будет обнаружено не более одного дефектного.

Сформулируйте нулевую и альтернативную гипотезы. Определите критическую область и область принятия нулевой гипотезы. Сформулируйте, в чем состоят ошибки первого и второго рода, найдите их вероятности.

Решение. Если смотреть на ситуацию с точки зрения покупателя, то нулевой гипотезой H_0 следует, по-видимому, считать гипотезу о 20% дефектных изделий. Альтернативная гипотеза H_1 соответствует версии поставщика — дефектных изделий 3%.

Поскольку отбирается 10 изделий и затем фиксируется число дефектных, то множеством всевозможных результатов испытаний будет:

$$\Omega = \{0, 1, 2, \ldots, 10\},$$

так как может оказаться 0, 1, 2, ..., 10 дефектных изделий. По условиям поставки гипотеза покупателя отвергается в случае $\omega \leqslant 1$, так что критическая область такова:

$$S = \{0, 1\}.$$

Соответственно область принятия нулевой гипотезы:

$$\{2, 3, \ldots, 10\}.$$

Ошибка первого рода состоит, напомним, в следующем: нулевая гипотеза отвергается, хотя она верна. В данном случае это означает: партия закупается, хотя в ней 20% дефектных изделий.

Ошибка второго рода состоит в принятии нулевой гипотезы, в то время как верна альтернативная. Это означает, что поставка не состоится, хотя в партии лишь 3% дефектных изделий.

Найдем теперь вероятности этих ошибок.

Если нулевая гипотеза H_0 верна, то вероятность того, что одно случайно выбранное изделие окажется дефектным, составляет 0,2.

Ошибка первого рода произойдет, если из 10 изделий дефектным будет не более чем одно (т. е. 0 либо 1).

Заметим, что в случае истинности гипотезы H_0 число дефектных изделий ω является биномиальной случайной величиной $Bi(10; 0{,}2)$. Поэтому

$$p(\omega = 0) = (0{,}8)^{10} \approx 0{,}107,$$
$$p(\omega = 1) = 10 \cdot (0{,}8)^9 \cdot 0{,}2 \approx 0{,}268.$$

Теперь можно вычислить вероятность α ошибки первого рода:

$$\alpha = p\{\omega \leqslant 1|H_0\} = p\{\omega = 0|H_0\} + p\{\omega = 1|H_0\} = \\ = 0{,}107 + 0{,}268 = 0{,}375.$$

Если верна альтернативная гипотеза H_1, то вероятность выбрать дефектное изделие составляет 0,03. Ошибка второго рода произошла, если из 10 изделий дефектными окажутся два или более.

В случае истинности гипотезы H_1 число дефектных изделий ω также является биномиальной случайной величиной, но с другим параметром — $Bi(10; 0{,}03)$. Поэтому

$$p(\omega = 0) = (0{,}97)^{10} \approx 0{,}737,$$
$$p(\omega = 1) = 10 \cdot (0{,}97)^9 \cdot 0{,}03 \approx 0{,}228.$$

Таким образом, для вероятности β ошибки второго рода получаем:

$$\beta = p\{\omega > 1|H_1\} = 1 - p\{\omega \leqslant 1|H_1\} = \\ 1 - p\{\omega = 0|H_1\} - p\{\omega = 1|H_1\} = 1 - 0{,}737 - 0{,}268 = 0{,}035.$$

Замечание. Из сравнения вероятностей α и β можно заключить, что оговоренная процедура проверки выгодна скорее поставщику.

16.2. Проверка биномиальных гипотез

В этом разделе мы рассмотрим один частный вид гипотез — гипотезы, связанные с вероятностью события. Проверка таких гипотез связана с биномиальным распределением, поэтому их называют биномиальными гипотезами. Основное отличие от предыдущих рассмотрений в том, что там вероятность ошибки определялась исходя из критической области и области принятия нулевой гипотезы. Теперь мы, напротив, будем искать критическую область исходя из заданного уровня значимости α.

Пример 5. О некотором игральном кубике было высказано мнение, что при его бросании "шестерка" выпадает слишком часто. Требуется проверить, так ли это?

Понятно, что в такой формулировке задача поставлена не вполне четко. Для уточнения определим сначала нулевую и альтернативную гипотезы. Поскольку вероятность выпадания "шестерки" у правильного кубика составляет 1/6, то имеем нулевую гипотезу

$$H_0: \quad p = \frac{1}{6}.$$

Альтернативную гипотезу естественно выбрать следующим образом:

$$H_1: \quad p > \frac{1}{6},$$

т. е. "шестерка" выпадает чаще, чем это должно быть. Такая альтернативная гипотеза называется *правосторонней*. Другие два вида альтернативных гипотез рассмотрены в последующих примерах.

Другое обстоятельство, требующее уточнения: в чем будет состоять испытание, на основании которого нулевая гипотеза будет принята либо отвергнута? Предположим, что испытание состоит в 120-кратном бросании кубика. При этом фиксируется число выпаданий "шестерки", которое обозначим ω.

Далее, необходимо выбрать уровень значимости α. Напомним, что это вероятность отвергнуть нулевую гипотезу при условии, что она верна. Положим, например, $\alpha = 0{,}01$.

В данном случае множество всевозможных результатов испытаний Ω будет таким:

$$\{0, 1, \ldots, 120\},$$

поскольку при 120-кратном бросании кубика "шестерка" может выпасть от 0 до 120 раз включительно.

Теперь осталось определить критическую область S. Эта задача уже формализована. Действительно, если "шестерка" выпала, скажем, 50 раз, то нулевую гипотезу следует отвергнуть. А если 40 раз? Или 30? Ясно, что критическая область S здесь имеет вид

$$\{\omega_0,\ \omega_0 + 1,\ \omega_0 + 2,\ \ldots,\ 120\},$$

где ω_0 — число, подлежащее определению. Такая критическая область называется *правосторонней*, она как бы "прижата" к правой границе множества Ω.

Для нахождения критической области воспользуемся тем, что число ω выпаданий "шестерки" в 120 испытаниях — биномиальная случайная величина, которая может быть приближена нормальной случайной величиной. Если $p = 1/6$ (верна гипотеза H_0), то

$$\omega = Bi\left(120, \frac{1}{6}\right) \approx N\left(120 \cdot \frac{1}{6}, \sqrt{120 \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right)}\right) \approx N(20, 4).$$

Найдем теперь число $x_{\text{пр}}$, для которого

$$p\left(N(20, 4) > x_{\text{пр}}\right) = 0{,}01,$$

где $x_{\text{пр}}$ — критическое значение, которое разделяет критическую область и область принятия гипотезы. Как уже отмечалось, критическая область здесь является правосторонней (рис. 2).

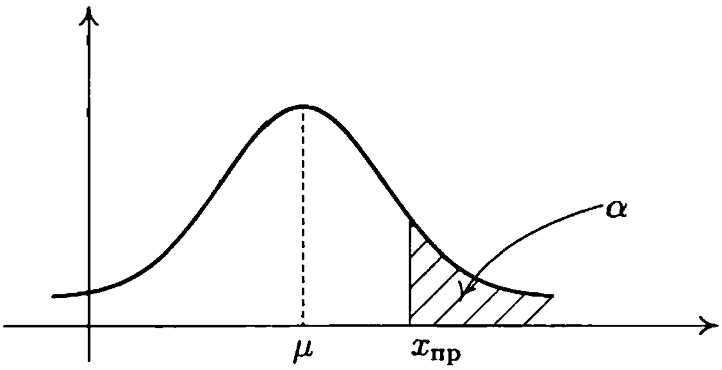


Рис. 2

Можно убедиться, что правостороннее критическое значение $x_{\text{пр}}$ вычисляется по формуле

$$x_{\text{пр}} = \mu + \sigma u_{1-2\alpha}, \tag{1}$$

где $u_{1-2\alpha}$ — решение уравнения

$$2\Phi(u_{1-2\alpha}) = 1 - 2\alpha.$$

Имеем:

$$u_{1-2\alpha} = u_{0,98} = 2{,}3,$$
$$x_{\text{пр}} = 20 + 4 \cdot 2{,}3 = 29{,}2.$$

Таким образом, если нулевая гипотеза верна, то вероятность выпадания "шестерки" более чем 29 раз очень мала. В итоге получаем

$$S = \{30, 31, \ldots, 100\}$$

(т. е. $\omega_0 = 30$). Соответственно область принятия нулевой гипотезы такова:

$$\{0, 1, \ldots, 29\}.$$

Замечание. Напомним, что α называется также вероятностью ошибки первого рода. Вероятность ошибки второго рода здесь является не числом, а функцией $\beta(p)$, поскольку гипотеза H_1 не задает однозначно закон распределения. Предоставляем заинтересованному читателю самостоятельно определить вид функции $\beta(p)$.

В рассмотренном примере критическая область была *односторонней*, в частности правосторонней. *Левосторонняя* критическая область соответствует альтернативным гипотезам вида $p < p_0$. Левостороннее критическое значение $x_{\text{лев}}$ находится по формуле

$$x_{\text{лев}} = \mu - \sigma u_{1-2\alpha} \tag{2}$$

(рис. 3).

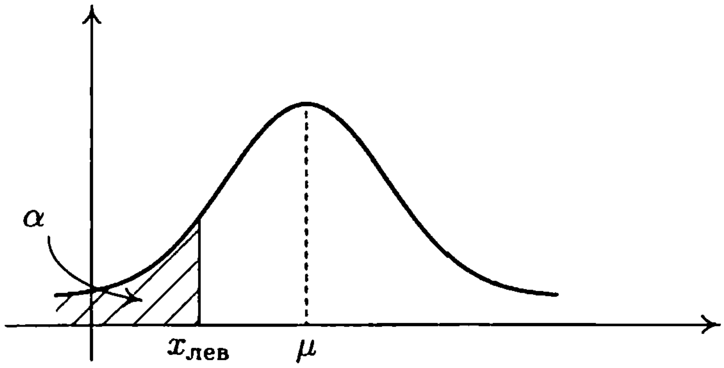


Рис. 3

В следующем примере критическая область *двусторонняя*.

Пример 6. О некоторой монете утверждается, что она правильная, т. е. вероятности выпадания герба и цифры равны между собой. Монету подбросили 100 раз, при этом 70 раз выпал герб, а 30 раз — цифра. Проверить утверждение о правильности монеты на 5%-м уровне значимости (т. е. на уровне значимости $\alpha = 0{,}05$).

Здесь нулевая и альтернативная гипотезы формулируются следующим образом:

$$H_0 : \ p = \frac{1}{2},$$

$$H_1 : \ p \neq \frac{1}{2},$$

где p — вероятность выпадания герба. Здесь альтернативная гипотеза, как и критическая область, является двусторонней. Это обусловлено тем, что нет предположений относительно более частого либо более редкого выпадания герба.

Множество всевозможных результатов испытаний Ω (в данном случае это количество выпаданий герба) составляют целые числа от 0 до 100:

$$\Omega = \{0, 1, \ldots, 100\}.$$

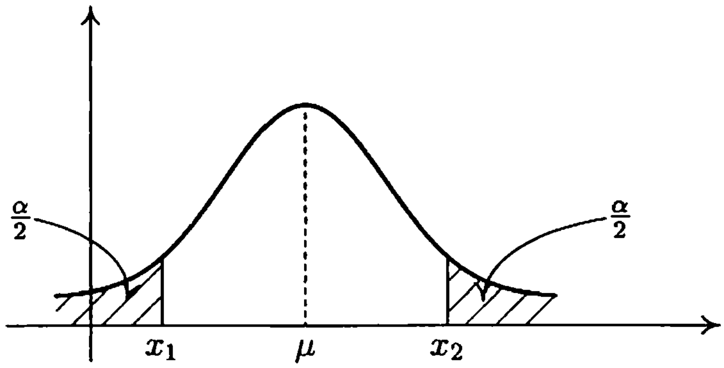


Рис. 4

Число выпаданий герба ω — это биномиальная случайная величина $Bi\left(100, \frac{1}{2}\right)$. Следовательно,

$$\omega \approx N\left(100 \cdot \frac{1}{2}, \sqrt{100 \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right)}\right) = N(50, 5).$$

На рис. 4 изображена двусторонняя критическая область. Критические значения x_1 и x_2 вычисляются по формулам

$$x_1 = \mu - \sigma u_{1-\alpha}, \qquad x_2 = \mu + \sigma u_{1-\alpha}, \tag{3}$$

где $u_{1-\alpha}$ — решение уравнения

$$2\Phi(u_{1-\alpha}) = 1 - \alpha.$$

В данном случае получаем

$$u_{1-\alpha} = u_{0,95} = 2,$$
$$x_1 = 50 - 5 \cdot 2 = 40, \qquad x_2 = 50 + 5 \cdot 2 = 60.$$

Таким образом, критическую область составляют значения ω до 40 и больше 60:

$$S = \{0, 1, \ldots, 40\} \cup \{60, 61, \ldots, 100\}.$$

Отметим, что при $\omega = 40$ либо $\omega = 60$ имеет смысл, по-видимому, не принимать окончательного решения о принятии либо отвержении нулевой гипотезы. Но в данном случае испытание дало результат $\omega = 70$, поэтому гипотеза H_0 отвергается в пользу гипотезы H_1.

Окончательный вывод таков: монета правильной не является.

В общем случае процедура проверки биномиальной гипотезы состоит из следующих шагов:

1) формулировка нулевой и альтернативной гипотез;

2) выбор уровня значимости α;

3) определение объема выборки n;

4) описание множества всевозможных результатов наблюдений Ω;

5) вычисление критической области S и области принятия нулевой гипотезы (по одной из формул (1), (2) либо (3));

6) выполнение эксперимента (сбор данных) и принятие решения.

Пример 7. Производственная линия выпускала 5% бракованных товаров. Было предложено усовершенствование, призванное снизить процент брака. После переналадки линии на осмотр поступило 300 единиц товара, из которых бракованными оказались 9 единиц. Можно ли на 1%-м уровне значимости считать, что качество продукции производственной линии улучшилось?

Решение. Здесь нулевая гипотеза состоит в том, что линия по-прежнему выпускает 5% брака:

$$H_0: \quad p = 0{,}05.$$

Альтернативная гипотеза H_1 является левосторонней — процент брака уменьшился:

$$H_1: \quad p < 0{,}05.$$

Уровень значимости и объем выборки даны в условии:

$$\alpha = 0{,}01, \quad n = 300.$$

Если при проверке фиксируется число бракованных изделий ω, то эта величина может принимать значения из множества

$$\Omega = \{0, 1, \ldots, 300\}.$$

В случае истинности гипотезы H_0 величина ω является биномиальной случайной величиной $Bi\,(300; 0{,}05)$. Следовательно,

$$\omega \approx N\left(300 \cdot 0{,}05;\ \sqrt{300 \cdot 0{,}05 \cdot 0{,}95}\right) \approx N(15; 3{,}8).$$

Найдем границу критической области по формуле (2):

$$u_{1-2\alpha} = u_{0{,}98} = 2{,}3,$$
$$x_{\text{лев}} = 15 - 3{,}8 \cdot 2{,}3 = 6{,}26.$$

Таким образом, критическая область может быть записана в виде

$$S = \{0, 1, \ldots, 6\}.$$

Область принятия нулевой гипотезы соответственно такова:

$$\{7, 8, \ldots, 300\}.$$

Если в выборке бракованными оказались 9 единиц товара, то следует принять нулевую гипотезу. Окончательный вывод: имеющиеся данные не дают оснований считать, что качество продукции улучшилось.

16.3. Критерий согласия χ^2 (хи-квадрат)

При проверке биномиальных гипотез требовалось проверить гипотезу о равенстве неизвестной вероятности некоторому числу. Подчеркнем, что речь шла об уточнении значения *одного* параметра — вероятности. Иной характер имеет ситуация, когда требуется проверить гипотезу о равенстве определенным значениям *нескольких* вероятностей (иначе говоря, о законе распределения в целом). В таких случаях применяются так называемые *критерии согласия*, один из которых мы и рассмотрим.

Пусть в результате некоторого испытания может произойти одно из r событий $A_1, A_2, \cdots, A_r$. Нулевая гипотеза H_0 имеет следующий вид:

$$p(A_1) = p_1, \quad p(A_2) = p_2, \quad \ldots, \quad p(A_r) = p_r,$$

где $p_1, p_2, \cdots, p_r$ — некоторые положительные числа, сумма которых равна 1. Альтернативной гипотезой является невыполнение хотя бы одного из этих равенств.

Исходными данными для проверки гипотезы H_0 являются результаты n независимых испытаний. Пусть эти результаты следующие:

событие A_1 произошло m_1 раз,
событие A_2 произошло m_2 раз,
....
событие A_r произошло m_r раз.

Очевидно, что
$$m_1 + m_2 + \cdots + m_r = n.$$

Подсчитаем величину χ^2 (хи-квадрат) по формуле
$$\chi^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i}.$$

Это значение будет решающим при проверке гипотезы. Величина χ^2 показывает, насколько экспериментальные значения m_i расходятся с теоретически наиболее вероятными значениями np_i. Отметим, что величина χ^2 может принимать лишь неотрицательные значения (отсюда и обозначение "хи-*квадрат*").

Сама проверка гипотезы осуществляется следующим образом. Выбирается уровень значимости $\alpha = 0{,}05$ либо $\alpha = 0{,}01$ (возможны, разумеется, и другие значения). Затем вычисляется
$$k = r - 1,$$
после чего находится критическое значение $\chi^2_{\text{кр}}$ (зависящее от α и k) по таблице:

$\alpha \backslash k$	1	2	3	4	5	6	7	8	9	10
0,01	6,8	9,2	11,3	13,3	15,1	16,8	18,5	20,1	21,7	23,2
0,05	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9	18,3

Если
$$\chi^2 > \chi^2_{\text{кр}},$$
то гипотеза H_0 отвергается, в противном случае — принимается.

Замечание 1. Таблица критических значений $\chi^2_{\text{кр}}$ (для удобства она повторена в приложении — в конце книги) получена из некоторых теоретических соображений. Дело в том, что при истинности нулевой гипотезы величина χ^2 оказывается распределенной (приближенно) в соответствии с законом распределения случайной величины $\chi^2(k)$ — "*хи-квадрат с k степенями свободы*" (распределение $\chi^2(k)$ называется также распределением К. Пирсона). При этом критическое значение $\chi^2_{\text{кр}}$ определяется из условия

$$p\left(\chi^2(k) > \chi^2_{\text{кр}}\right) = \alpha.$$

Что касается самой величины $\chi^2(k)$, то она определяется как сумма квадратов k независимых стандартных нормальных случайных величин:

$$\chi^2(k) = U_1^2 + U_2^2 + \cdots + U_k^2,$$

где

$$U_i = N(0,1), \quad i = 1, 2, \ldots, k.$$

Отметим также, что

$$E\left[\chi^2(k)\right] = k, \qquad D\left[\chi^2(k)\right] = 2k.$$

Замечание 2. Критерий χ^2 применяют тогда, когда все величины np_i достаточно велики, например при

$$np_i > 5, \quad i = 1, 2, \ldots, r.$$

Пример 8. При 4040 бросаниях монеты французский естествоиспытатель Бюффон получил 2048 выпаданий герба и 1992 выпадания цифры. На уровне значимости $\alpha = 0{,}05$ проверим гипотезу о том, что монета была правильной.

Решение. Здесь в результате испытания может произойти одно из двух событий — выпадание герба либо выпадание цифры. Поэтому имеем:

$A_1 = \{$выпадание герба$\}$, $A_2 = \{$выпадание цифры$\}$,

$n = 4040$, $m_1 = 2048$, $m_2 = 1992$.

Нулевая гипотеза —

$$H_0: \ p(A_1) = p(A_2) = \frac{1}{2},$$

т. е.

$$p_1 = p_2 = \frac{1}{2}.$$

Вычислим величину χ^2. Имеем:

$$\chi^2 = \frac{(m_1 - np_1)^2}{np_1} + \frac{(m_2 - np_2)^2}{np_2} = \frac{(2048-2020)^2}{2020} + \frac{(1992-2020)^2}{2020} \approx 0{,}776.$$

Число степеней свободы k в данном случае равно $r - 1 = 2 - 1 = 1$. По известным значениям $\alpha = 0{,}05$, $k = 1$ находим в таблице

$$\chi^2_{\text{кр}} = 3{,}8.$$

Так как

$$\chi^2 < \chi^2_{\text{кр}},$$

то нулевая гипотеза принимается — монета была правильной.

Пример 9. Некоторая фирма владеет тремя магазинами, расположенными недалеко друг от друга. Руководство фирмы решило выяснить, посещают ли покупатели все три магазина одинаково охотно либо имеется некоторое различие. (Это различие может быть, по мнению руководства, связано с более или менее выгодным расположением магазинов, разной квалификацией персонала и т. п.)

Для проверки была собрана информация о количестве покупателей, сделавших покупки в течение недели. Оказалось, что в первом магазине это число составляет 160 человек, во втором — 225, в третьем —215.

Решение. Здесь нулевой гипотезой будет равенство вероятностей посещения покупателем первого (p_1), второго (p_2) и третьего (p_3) магазинов:

$$H_0: \quad p_1 = p_2 = p_3 = \frac{1}{3}.$$

В результате испытания получаем

$$m_1 = 160,\ m_2 = 225,\ m_3 = 215,\ n = 160 + 225 + 215 = 600.$$

Вычислим величину

$$\chi^2 = \frac{(160-200)^2}{200} + \frac{(225-200)^2}{200} + \frac{(215-200)^2}{200} = 12{,}25.$$

Обратимся теперь к таблице критических значений (при $k = 2$). Даже на уровне значимости $\alpha = 0{,}01$ имеем $\chi^2_{\text{кр}} = 9{,}2$. Таким образом,

$$\chi^2 > \chi^2_{\text{кр}}.$$

Поэтому, видимо, разницу в посещаемости магазинов в течение недели нельзя объяснить случайными колебаниями.

16.4. Задания и ответы

1. Ваш друг утверждает, что он умеет различать на вкус два близких сорта вина если и не всегда, то хотя бы в четырех случаях из пяти. Вы же склонны считать, что он просто угадывает.

Сформулируйте оба этих мнения в виде статистических гипотез и предложите какую-либо процедуру проверки. В чем состоят ошибки первого и второго рода?

2. Урна содержит большое количество белых и черных шаров. 100 раз производится следующее действие: из урны наугад достается шар, фиксируется его цвет, затем шар опускается обратно в урну, после чего шары перемешиваются. Оказалось, что 67 раз достали белый шар, 33 раза — черный. Можно ли на 5%-м уровне значимости принять гипотезу о том, что доля белых шаров в урне составляет 0,6?

Ответ: да.

3. Обычно применяемое лекарство снимает послеоперационные боли у 80% пациентов. Новое лекарство, применяемое для тех же целей, помогло 90 пациентам из первых 100 оперированных. Можно ли на уровне значимости $\alpha = 0{,}05$ считать, что новое лекарство лучше? А на уровне $\alpha = 0{,}01$?

Ответ: да.

4. Игральный кубик бросили 60 раз, при этом числа 1, 2, 3, 4, 5, 6 выпали соответственно 12, 9, 13, 11, 8, 7 раз. Можно ли на 5%-м уровне значимости отвергнуть гипотезу о симметричности кубика?

Ответ: нет.

5. Трое рабочих работают на трех одинаковых станках. В конце смены первый рабочий изготовил 60 деталей, второй — 80, третий — 100 деталей. Можно ли на уровне значимости $\alpha = 0{,}01$ принять гипотезу о том, что производительности труда первых двух рабочих равны между собой и в 2 раза меньше производительности третьего рабочего?

Ответ: да.

Часть III

ИГРОВЫЕ МЕТОДЫ

В практической деятельности весьма часто приходится рассматривать явления и ситуации, в которых участвуют две или более стороны, имеющие различные интересы и обладающие возможностями применять для достижения своих целей разнообразные действия. Подобные явления и ситуации принято называть *конфликтными* или просто *конфликтами.*

Типичный конфликт характеризуется тремя основными составляющими:

1) заинтересованными сторонами,
2) возможными действиями этих сторон,
3) интересами сторон.

Конфликтная ситуация, взятая из реальной жизни, как правило, довольно сложна. К тому же ее изучение затруднено наличием разных обстоятельств, часть из которых не оказывает сколько-нибудь существенного влияния ни на развитие конфликта, ни на его исход. Поэтому для того чтобы анализ конфликтной ситуации оказался возможным, необходимо отвлечение от этих второстепенных факторов, что при удачном стечении обстоятельств позволяет построить упрощенную формализованную модель конфликта, которую и принято называть *игрой.* От реальной конфликтной ситуации игра отличается еще и тем, что ведется по вполне определенным правилам.

Необходимость изучения и анализа конфликтов, представляемых в виде упрощенных математических моделей (игр), вызвала к жизни специальный математический аппарат — *теорию игр.*

Опишем некоторые основные понятия, используемые в этой теории.

Заинтересованные стороны называются *игроками.* Любое возможное для игрока действие (в рамках заданных *правил игры*) называется его *стратегией.* В условиях конфликта каждый игрок выбирает свою стратегию, в результате чего складывается набор стратегий, называемый *ситуацией*. Заинтересованность игроков в ситуации проявляется в том, что каждому игроку в каждой ситуации приписывается число, выражающее степень удовлетворения его интересов в этой ситуации и называемое его *выигрышем* в ней.

В этих условиях протекание конфликта состоит в выборе каждым игроком своей стратегии и получении им в сложившейся ситуации выигрыша из некоторого источника.

На этом пути создается *теория игр с выигрышами*.

Однако оценка игроком ситуации путем предположения о своем выигрыше, вообще говоря, не всегда возможна практически и даже не всегда имеет смысл. В подобных случаях иногда удается вместо прямых численных оценок ситуаций указывать на их сравнительную предпочтительность для отдельных игроков.

На этом пути создается *теория игр с предпочтениями*, включающая в себя и теорию игр с выигрышами.

В дальнейшем мы ограничимся рассмотрением только игр с выигрышами.

Изучение игр можно проводить с различных точек зрения.

Мы будем стремиться

к выработке принципов оптимальности, т. е. того критерия, по которому поведение игроков следует считать оптимальным (разумным, целесообразным);

к выяснению реализуемости этих принципов, т. е. установлению существования оптимальных в выработанном смысле ситуаций, и отысканию этих реализаций.

Одной из плодотворных форм реализации представлений об оптимальности можно считать понятие *равновесия*, при котором складывается такая (равновесная) ситуация, в нарушении которой не заинтересован ни один из игроков.

Именно ситуации равновесия могут быть предметом устойчивых договоров между игроками (ни у одного из игроков не будет мотивов к нарушению договора).

Кроме того, ситуации равновесия являются выгодными для каждого игрока: в равновесной ситуации каждый игрок получает наибольший выигрыш (разумеется, в той мере, в какой это от него зависит).

Если в игре ситуации равновесия (в пределах отпущенных возможностей) нет, то, оставаясь в условиях стратегий, имеющихся у игроков, мы сталкиваемся с неразрешимой задачей.

При возникновении подобных случаев естественно ставить вопрос о таком расширении первоначального понятия стратегии, чтобы среди ситуаций, составленных из новых, обобщенных стратегий, заведомо нашлись равновесные.

Если такие обобщенные стратегии существуют, то обычно они представляются некоторыми комбинациями исходных стратегий.

Для того чтобы отличать прежние стратегии от новых, первые называют *чистыми*, а вторые — *смешанными* стратегиями.

Весьма плодотворным является представление смешанной стратегии как случайного выбора игроками их чистых стратегий, при котором случайные выборы различных игроков независимы в совокупности, а выигрыш каждого из них определяется как математическое ожидание случайного выигрыша.

Таким образом, преобразованная игра обычно называется *смешанным расширением* исходной игры.

Проиллюстрируем сказанное на примере одного из самых простых, но одновременно и наиболее изученных и продвинутых классов игр, на так называемых *матричных играх*. Исследование матричных игр интересно еще и потому, что к ним могут быть приближенно сведены многие игры более общего вида.

Затем мы рассмотрим позиционные и биматричные игры, а также пример игры со многими участниками.

Глава 17
МАТРИЧНЫЕ ИГРЫ

Рассмотрим игру, в которой участвуют два игрока, причем каждый из них имеет конечное число стратегий.

Обозначим для удобства одного из игроков через A, в другого — через B.

Предположим, что игрок A имеет m стратегий: A_1, A_2, ..., A_m, а игрок B — n стратегий: B_1, B_2, ..., B_n.

Пусть игрок A выбрал стратегию A_i, а игрок B — стратегию B_k.

Будем считать, что выбор игроками стратегий A_i и B_k однозначно определяет исход игры — выигрыш a_{ik} игрока A и выигрыш b_{ik} игрока B, причем эти выигрыши связаны равенством

$$b_{ik} = -a_{ik}.$$

Последнее условие показывает, что в рассматриваемых обстоятельствах выигрыш одного из игроков равен выигрышу другого, взятому с противоположным знаком. Поэтому при анализе такой игры можно рассматривать выигрыши только одного из игроков. Пусть это будут, например, выигрыши игрока A.

Если нам известны значения a_{ik} выигрыша при каждой паре стратегий (в каждой ситуации) $\{A_i, B_k\}$, $i = 1, 2, \ldots, m$, $k = 1, 2, \ldots, n$, то их удобно записывать

или в виде прямоугольной таблицы, строки которой соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B,

	B_1	B_2	...	B_n
A_1	a_{11}	a_{12}	...	a_{1n}
A_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}

или в виде матрицы

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Полученная матрица имеет размер $m \times n$ и называется *матрицей игры* или *платежной матрицей* (отсюда и название игры — *матричная*).

Рассматриваемую игру часто называют *игрой* $m \times n$ или $m \times n$-*игрой*.

Замечание. Матричные игры относятся к разряду так называемых антагонистических игр, т. е. игр, в которых интересы игроков прямо противоположны.

Пример 1. Каждый из двух игроков A и B одновременно и независимо друг от друга записывает на листе бумаги любое целое число. Если выписанные числа имеют одинаковую четность, то игрок A получает от игрока B 1 рубль, а если разную, то, наоборот, игрок A платит 1 рубль игроку B.

У игрока A две стратегии: A_1 — записать четное число и A_2 — записать нечетное число.

У игрока B такие же две стратегии: B_1 — записать четное число и B_2 — записать нечетное число.

Выбор игроками соответственно стратегий A_i и B_k однозначно определяет исход игры: a_{ik} — выигрыш игрока A.

Матрица этой 2×2-игры имеет следующий вид:

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

(здесь строки соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B).

17.1. Равновесная ситуация

Рассмотрим следующий пример.

Пример 2. Два игрока A и B, не глядя друг на друга, кладут на стол по картонному кружку красного (r), зеленого (g) или синего (b)

цвета, сравнивают цвета кружков и расплачиваются друг с другом так, как показано в матрице игры:

$$\begin{pmatrix} -2 & 2 & -1 \\ 2 & 1 & 1 \\ 3 & -3 & 1 \end{pmatrix}$$

(напомним, что у этой 3×3-матрицы строки соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B).

Считая, что эта 3×3-игра повторяется многократно, попробуем определить оптимальные стратегии каждого из игроков.

Начнем с последовательного анализа стратегий игрока A, не забывая о том, что, выбирая стратегию игрока A, должно принимать в расчет, что его противник B может ответить на нее той из своих стратегий, при которой выигрыш игрока A будет минимальным.

Так, на стратегию A_r он ответит стратегией B_r (минимальный выигрыш равен -2, что на самом деле означает проигрыш игрока A, равный 2), на стратегию A_g — стратегией B_g или B_b (минимальный выигрыш игрока A равен 1), а на стратегию A_b — стратегией B_g (минимальный выигрыш игрока A равен -3).

Запишем эти минимальные выигрыши в правом столбце таблицы:

A_r	-2	2	-1	-2
A_g	2	1	1	1
A_b	3	-3	1	-3

Максимин (*maxmin*). Неудивительно, что игрок A останавливает свой выбор на стратегии A_g, при которой его минимальный выигрыш максимален (из трех чисел -2, 1 и -3 максимальным является 1), $maxmin = 1$.

Если игрок A будет придерживаться этой стратегии, то ему гарантирован выигрыш, не меньший 1, при любом поведении противника.

Аналогичные рассуждения можно провести и за игрока B. Так как игрок B заинтересован в том, чтобы обратить выигрыш игрока A в минимум, то ему нужно проанализировать каждую свою стратегию с точки зрения максимального выигрыша игрока A.

Выбирая свою стратегию, игрок B должен учитывать, что при этом стратегией его противника A может оказаться та, при которой выигрыш игрока A будет максимальным.

Так, на стратегию B_r он ответит стратегией A_b (максимальный выигрыш игрока A равен 3), на стратегию B_g — стратегией A_r (максимальный выигрыш игрока A равен 2), а на стратегию B_b — стратегией A_g или A_b (максимальный выигрыш игрока A равен 1).

Эти максимальные выигрыши записаны в нижней строке таблицы

	B_r	B_g	B_b	
A_r	−2	2	−1	−2
A_g	2	1	1	1
A_b	3	−3	1	−3
	3	2	1	

Минимакс ($minmax$). Неудивительно, если игрок B остановит свой выбор на стратегии B_b, при которой максимальный выигрыш игрока A минимален (из трех чисел 3, 2 и 1 минимальным является 1), $minmax = 1$.

Если игрок B будет придерживаться этой стратегии, то при любом поведении противника он проиграет не больше 1.

Итак, в рассматриваемой игре числа $maxmin$ и $minmax$ совпали:

$$maxmin = minmax = 1$$

(соответствующие элементы в таблице выделены жирным шрифтом):

	B_r	B_g	$\boldsymbol{B_b}$
A_r	−2	2	−1
$\boldsymbol{A_g}$	2	1	**1**
A_b	3	−3	1

Выделенные стратегии A_g и B_b являются оптимальными для игроков A и B,

$$A_g = A_{opt}, \quad B_b = B_{opt},$$

в следующем смысле:

при многократном повторении игры отказ от выбранной стратегии любого из игроков уменьшает его шансы на выигрыш (увеличивает шансы на проигрыш).

В самом деле, если игрок A не будет придерживаться стратегии A_{opt}, а выберет иную стратегию, например A_r, то вряд ли стоит рассчитывать на то, что игрок B этого не заметит. Конечно, заметит и не преминет воспользоваться своим наблюдением. Ясно, что в этом случае он отдаст предпочтение стратегии B_r. А на выбор A_b игрок B ответит, например, так: B_g. В результате отказа от стратегии A_g выигрыш игрока A уменьшится.

Если же от оптимальной стратегии отказывается игрок B, выбирая, например, стратегию B_r, то игрок A может ответить на это стратегией A_b и тем самым увеличить свой выигрыш. В случае стратегии B_g ответ игрока A — A_r.

Тем самым ситуация $\{A_g, B_b\}$ оказывается равновесной.

Еще раз подчеркнем, что элементами матрицы игры являются числа, описывающие выигрыш игрока A. Более точно, выигрыш соответствует положительному элементу платежной матрицы, а отрицательный указывает на проигрыш игрока A.

Матрица выплат игроку B получается из матрицы игры заменой каждого ее элемента на противоположный.

Рассмотрим теперь произвольную матричную игру:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

(строки заданной $m \times n$-матрицы соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B) и опишем общий алгоритм, посредством которого можно определить, есть ли в этой игре ситуация равновесия или ее нет.

В теории игр предполагается, что оба игрока действуют разумно, т. е. стремятся к получению максимального выигрыша, считая, что соперник действует наилучшим (для себя) образом.

А. *Действия игрока А.*

1-й шаг. В каждой строке матрицы **A** находится минимальный элемент

$$\alpha_i = \min_k a_{ik}, \quad i = 1, 2, \dots, m.$$

Полученные числа

$$\alpha_1, \alpha_2, \dots, \alpha_m$$

приписываются к заданной таблице в виде правого добавочного столбца:

$$\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & \alpha_1 \\ a_{21} & a_{22} & \dots & a_{2n} & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & \alpha_m \end{array}$$

Пояснение. Выбирая стратегию A_i, игрок A должен рассчитывать на то, что в результате разумных действий противника (игрока B) он выиграет не меньше чем α_i.

2-й шаг. Среди чисел

$$\alpha_1, \alpha_2, \ldots, \alpha_m$$

выбирается максимальное число

$$\alpha = \max_i \alpha_i,$$

или подробнее:

$$\alpha = \max_i \min_k a_{ik}.$$

Специально отметим, что выбранное число α является одним из элементов заданной матрицы **A**.

Пояснение. Действуя наиболее осторожно и рассчитывая на наиболее разумное поведение противника, игрок A должен остановиться на той стратегии A_i, для которой число α_i является максимальным.

Если игрок A будет придерживаться стратегии, выбранной описанным выше способом, то при любом поведении игрока B игроку A гарантирован выигрыш, не меньший α.

Число α называется *нижней ценой игры.*

Принцип построения стратегии игрока A, основанный на максимизации минимальных выигрышей, называется *принципом максимина*, а выбираемая в соответствии с этим принципом стратегия A_{i^0} – *максиминной* стратегией игрока A.

B. *Действия игрока B.*

1-й шаг. В каждом столбце матрицы **A** ищется максимальный элемент

$$\beta_k = \max_i a_{ik}, \quad k = 1, 2, \ldots, n.$$

Полученные числа

$$\beta_1, \beta_2, \ldots, \beta_n$$

приписываются к заданной таблице в виде нижней добавочной строки:

$$\begin{array}{cccc|c}
a_{11} & a_{12} & \ldots & a_{1n} & \alpha_1 \\
a_{21} & a_{22} & \ldots & a_{2n} & \alpha_2 \\
\multicolumn{4}{c|}{\ldots\ldots\ldots\ldots\ldots\ldots} & \ldots \\
a_{m1} & a_{m2} & \ldots & a_{mn} & \alpha_m \\
\hline
\beta_1 & \beta_2 & \ldots & \beta_n &
\end{array}$$

Пояснение. Выбирая стратегию B_k, игрок B должен рассчитывать на то, что в результате разумных действий противника (игрока A) он проиграет не больше, чем β_k.

2-й шаг. Среди чисел

$$\beta_1, \beta_2, \ldots, \beta_n$$

выбирается минимальное число

$$\beta = \min_k \beta_k,$$

или подробнее:

$$\beta = \min_k \max_i a_{ik}.$$

Выбранное число β также является одним из элементов заданной матрицы $\mathbf{A}$.

Пояснение. Действуя наиболее осторожно и рассчитывая на наиболее разумное поведение противника, игрок B должен остановиться на той стратегии B_k, для которой число β_k является минимальным.

Если игрок B будет придерживаться стратегии, выбранной описанным выше способом, то при любом поведении игрока A игроку B гарантирован проигрыш, не больший β.

Число β называется *верхней ценой игры.*

Принцип построения стратегии игрока B, основанный на минимизации максимальных потерь, называется *принципом минимакса*, а выбираемая в соответствии с этим принципом стратегия B_{k^0} — *минимаксной* стратегией игрока B.

Нижняя цена игры α и верхняя цена игры β всегда связаны неравенством

$$\alpha \leqslant \beta.$$

Замечание. Реализация описанного алгоритма требует $2mn-1$ сравнений элементов матрицы $\mathbf{A}$:

$$(n-1)m + m - 1 = mn - 1$$

сравнений для определения α,

$$(m-1)n + n - 1 = mn - 1$$

сравнений для определения β и одно сравнение полученных чисел α и β.

Если

$$\alpha = \beta,$$

или подробнее:

$$\max_i \min_k a_{ik} = a_{i^0k^0} = \min_k \max_i a_{ik},$$

то ситуация $\{A_{i^0}, B_{k^0}\}$ оказывается равновесной и ни один из игроков не заинтересован в том, чтобы ее нарушить (в этом нетрудно убедиться путем рассуждений, подобных проведенным при анализе игры в примере 2).

В том случае, когда нижняя цена игры равна верхней цене игры, их общее значение называется просто *ценой игры* и обозначается через ν.

Цена игры совпадает с элементом $a_{i^0k^0}$ матрицы игры $\mathbf{A}$, расположенным на пересечении i^0-й строки (стратегия A_{i^0} игрока A) и k^0-го столбца (стратегия B_{k^0} игрока B), — минимальным в своей строке и максимальным в своем столбце.

Этот элемент называют *седловой точкой матрицы* $\mathbf{A}$ или *точкой равновесия*, а про игру говорят, что она *имеет седловую точку*.

Стратегии A_{i^0} и B_{k^0}, соответствующие седловой точке, называются *оптимальными*, а совокупность оптимальных ситуаций и цена игры — *решением матричной игры с седловой точкой.*

Замечание. Седловых точек в матричной игре может быть несколько, но все они имеют одно и то же значение.

Матричные игры с седловой точкой важны и интересны, однако более типичным является случай, когда применение описанного алгоритма приводит к неравенству

$$\alpha < \beta.$$

Как показывает следующий пример, в этом случае предложенный выбор стратегий уже, вообще говоря, к равновесной ситуации не приводит и при многократном ее повторении у игроков вполне могут возникнуть мотивы к нарушению рекомендаций, основанных на описанной последовательности действий игроков A и B.

Пример 3. Рассмотрим 3×3-игру, заданную матрицей

$$\begin{pmatrix} 4 & -1 & -3 \\ -2 & 1 & 3 \\ 0 & 2 & -3 \end{pmatrix}.$$

Применив предложенный алгоритм

$$\begin{array}{rrr|r} 4 & 1 & -3 & -3 \\ -2 & 1 & 3 & -2 \\ 0 & 2 & -3 & -3 \\ \hline 4 & 2 & 3 & \end{array}$$

находим нижнюю $\alpha = -2$ и верхнюю $\beta = 2$ цену игры и соответствующие им стратегии A_2 и B_2.

Нетрудно убедиться в том, что, пока игроки придерживаются этих стратегий, средний выигрыш при многократном повторении игры будет равен 1. Он больше нижней цены игры, но меньше верхней цены.

Однако если игроку B станет известно, что игрок A придерживается стратегии A_2, он немедленно ответит стратегией B_1 и сведет его выигрыш к проигрышу -2. В свою очередь, на стратегию B_1 у игрока A имеется ответная стратегия A_1, дающая ему выигрыш 4.

Тем самым ситуация $\{A_2, B_2\}$ равновесной не является.

17.2. Смешанные стратегии

В случае когда нижняя цена игры α и верхняя цена игры β не совпадают,

$$\alpha < \beta,$$

игрок A может обеспечить себе выигрыш, не меньший α, а игрок B имеет возможность не дать ему больше, чем β.

Возникает вопрос: а как разделить между игроками разность $\beta - \alpha$?

Предыдущие построения на этот вопрос ответа не дают — тесны рамки возможных действий игроков.

Поэтому довольно ясно, что механизм, обеспечивающий получение каждым из игроков как можно большей доли этой разности, следует искать в определенном расширении стратегических возможностей, имеющихся у игроков изначально.

Оказывается, что компромиссного распределения разности $\beta - \alpha$ между игроками и уверенного получения каждым игроком своей доли при многократном повторении игры можно достичь путем случайного применения ими своих первоначальных, чистых стратегий.

Такие действия,

во-первых, обеспечивают наибольшую скрытность выбора стратегии (результат выбора не может стать известным противнику, поскольку он неизвестен самому игроку), во-вторых, при разумном построении механизма случайного выбора стратегий последние оказываются оптимальными.

Случайная величина, значениями которой являются стратегии игрока, называется его *смешанной стратегией.*

Тем самым задание смешанной стратегии игрока состоит в указании тех вероятностей, с которыми выбираются его первоначальные стратегии.

Рассмотрим произвольную $m \times n$-игру, заданную $m \times n$-матрицей

$$\mathbf{A} = (a_{ik}).$$

Так как игрок A имеет m чистых стратегий, то его смешанная стратегия может быть описана набором m неотрицательных чисел:

$$p_1 \geqslant 0,\ p_2 \geqslant 0,\ \ldots,\ p_m \geqslant 0,$$

сумма которых равна 1,

$$\sum_{i=1}^{m} p_i = 1.$$

Смешанная стратегия второго игрока B, имеющего n чистых стратегий, описывается набором n неотрицательных чисел:

$$q_1 \geqslant 0,\ q_2 \geqslant 0,\ \ldots,\ q_n \geqslant 0,$$

сумма которых равна 1,

$$\sum_{k=1}^{n} q_k = 1.$$

Замечание. Каждая чистая стратегия является частным случаем смешанной стратегии: например, чистая стратегия A_i является смешанной стратегией, описываемой набором чисел $p_1,\ p_2,\ \ldots,\ p_m$, в котором

$$p_i = 1, \quad p_j = 0 \quad (j \neq i).$$

Подчеркнем, что для соблюдения секретности каждый из игроков применяет свои стратегии независимо от другого игрока.

Таким образом, задав два набора:

$$\mathrm{P} = \{p_1, p_2, \ldots, p_m\}, \quad \mathrm{Q} = \{q_1, q_2, \ldots, q_n\},$$

мы оказываемся *в ситуации в смешанных стратегиях.*

В этих условиях каждая обычная ситуация (в чистых стратегиях) $\{A_i, B_k\}$ по определению является случайным событием и ввиду независимости P и Q реализуется с вероятностью $p_i q_k$. Поскольку в этой ситуации игрок A получает выигрыш a_{ik}, то математическое ожидание выигрыша в условиях ситуации в смешанных стратегиях $\{\mathrm{P}, \mathrm{Q}\}$ равно

$$\mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}) = \sum_{i=1}^{m} \sum_{k=1}^{n} a_{ik} p_i q_k.$$

Это число и принимается за средний выигрыш игрока A в ситуации в смешанных стратегиях $\{\mathrm{P}, \mathrm{Q}\}$.

Стратегии

$$\mathrm{P}^0 = \{p_1^0, p_2^0, \ldots, p_m^0\} \quad \text{и} \quad \mathrm{Q}^0 = \{q_1^0, q_2^0, \ldots, q_n^0\}$$

называются *оптимальными смешанными стратегиями* игроков A и B соответственно, если выполнено следующее соотношение:

$$\mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}^0) \leqslant \mathrm{E}(\mathbf{A}, \mathrm{P}^0, \mathrm{Q}^0) \leqslant \mathrm{E}(\mathbf{A}, \mathrm{P}^0, \mathrm{Q}).$$

Последнее равносильно тому, что

$$\max_{\mathrm{P}} \min_{\mathrm{Q}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}) = \mathrm{E}(\mathbf{A}, \mathrm{P}^0, \mathrm{Q}^0) = \min_{\mathrm{Q}} \max_{\mathrm{P}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}).$$

Величина

$$\nu = \mathrm{E}(\mathbf{A}, \mathrm{P}^0, \mathrm{Q}^0),$$

определяемая последней формулой, называется *ценой игры*.

Набор $(\mathrm{P}^0, \mathrm{Q}^0, \nu)$, состоящий из оптимальных смешанных стратегий игроков A и B и цены игры, называется *решением матричной игры*.

Естественно, возникают два *ключевых вопроса*:

1) какие матричные игры имеют решение в смешанных стратегиях?

2) как находить решение матричной игры, если оно существует?

Ответы на эти вопросы дают следующие две теоремы.

Основная теорема теории матричных игр

ТЕОРЕМА 1 (Дж. фон Нейман). *Для матричной игры с любой матрицей* $\mathbf{A}$ *величины*

$$\max_{\mathrm{P}} \min_{\mathrm{Q}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}), \quad \min_{\mathrm{Q}} \max_{\mathrm{P}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q})$$

существуют и равны между собой:

$$\max_{\mathrm{P}} \min_{\mathrm{Q}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}) = \min_{\mathrm{Q}} \max_{\mathrm{P}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}).$$

Более того, существует хотя бы одна ситуация в смешанных стратегиях $\{\mathrm{P}^0, \mathrm{Q}^0\}$, *для которой выполняется соотношение*

$$\mathrm{E}(\mathbf{A}, \mathrm{P}^0, \mathrm{Q}^0) = \max_{\mathrm{P}} \min_{\mathrm{Q}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}) = \min_{\mathrm{Q}} \max_{\mathrm{P}} \mathrm{E}(\mathbf{A}, \mathrm{P}, \mathrm{Q}).$$

Основные свойства оптимальных смешанных стратегий

ТЕОРЕМА 2. *Пусть*

$$\mathrm{P}^0 = \{p_1^0, p_2^0, \ldots, p_m^0\} \quad и \quad \mathrm{Q}^0 = \{q_1^0, q_2^0, \ldots, q_n^0\}$$

— оптимальные смешанные стратегии и ν *— цена игры.*

Оптимальная смешанная стратегия P^0 *игрока* A *смешивается только из тех чистых стратегий* A_i, $i = 1, 2, \ldots, m$ *(т. е. только те вероятности* p_i, $i = 1, 2, \ldots, m$, *могут быть отличны от нуля), для которых*

$$\sum_{k=1}^{n} a_{ik} q_k^0 = \nu.$$

Аналогично, только те вероятности q_k, $k = 1, 2, \ldots, n$, *могут быть отличны от нуля, для которых*

$$\sum_{i=1}^{m} a_{ik} p_i^0 = \nu.$$

Имеют место соотношения

$$\nu = \min_{1 \leqslant k \leqslant n} \sum_{i=1}^{m} a_{ik} p_i^0 = \max_{\mathrm{P}} \min_{1 \leqslant k \leqslant n} \sum_{i=1}^{m} a_{ik} p_i =$$

$$= \min_{\mathrm{Q}} \max_{1 \leqslant i \leqslant m} \sum_{k=1}^{n} a_{ik} q_k = \max_{1 \leqslant i \leqslant m} \sum_{k=1}^{n} a_{ik} q_k^0 = \nu.$$

В этом последнем скоплении равенств, по существу, и лежат истоки, питающие методы построения решений матричных игр.

Опишем некоторые из них.

17.3. Методы решения матричных игр

Наши рассмотрения мы начнем с матричных игр, в которых число стратегий хотя бы одного из игроков равно двум.

Для построения решений $2\times n$- и $m\times 2$-игр существует эффективный метод, основанный на простых геометрических соображениях. Этот метод называют *графическим*.

17.3.1. $2 \times n$-игры

Пусть

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \end{pmatrix}$$

— платежная матрица $2 \times n$-игры.

Согласно теореме о двойном описании игры (теорема 2), нахождение цены игры и оптимального значения p^0 для игрока A равносильно решению уравнения

$$\nu = \min_{1\leqslant k\leqslant n}(a_{1k}p^0 + a_{2k}(1-p^0)) = \max_p \min_{1\leqslant k\leqslant n}(a_{1k}p + a_{2k}(1-p)).$$

Опишем общую схему, приводящую к искомому результату.

Максимум функции

$$\min_{1\leqslant k\leqslant n}(a_{1k}p + a_{2k}(1-p)) \tag{1}$$

проще всего найти, построив ее график.

Для этого поступают следующим образом.

Предположим, что игрок A выбрал смешанную стратегию $\mathrm{P} = \{p, 1-p\}$, а игрок B — k-ю чистую стратегию, $k = 1, 2, \dots, n$. Тогда средний выигрыш игрока A в ситуации $\{\mathrm{P},\ k\}$ оказывается равным

$$(k):\ \ w = a_{1k}p + a_{2k}(1-p).$$

На плоскости (p, w) уравнение (k) описывает прямую. Тем самым каждой чистой стратегии игрока B на этой плоскости соответствует своя прямая.

Поэтому сначала на плоскости (p, w) последовательно и аккуратно рисуются все прямые

$$(k): \ w = a_{1k}p + a_{2k}(1-p), \quad k = 1, 2, \ldots, n$$

(рис. 1). Затем для каждого значения p, $0 \leqslant p \leqslant 1$, путем визуального сравнения соответствующих ему значений w на каждой из построенных прямых определяется и отмечается наименьшее из них (рис. 2). В результате описанной поцедуры получается ломаная, которая и является графиком функции (1) (жирная линия на рис. 3).

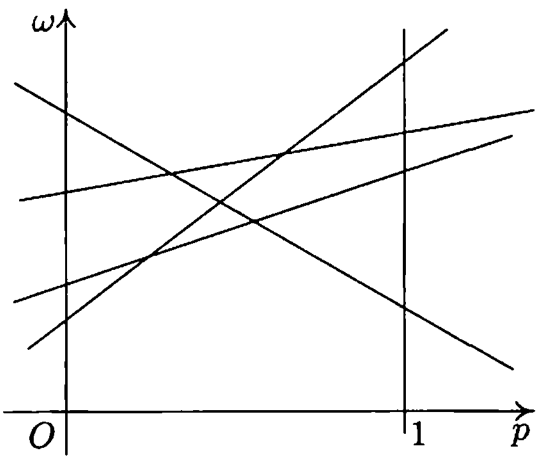


Рис. 1

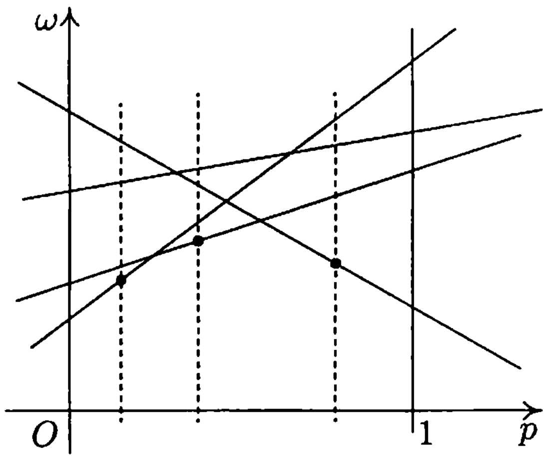


Рис. 2

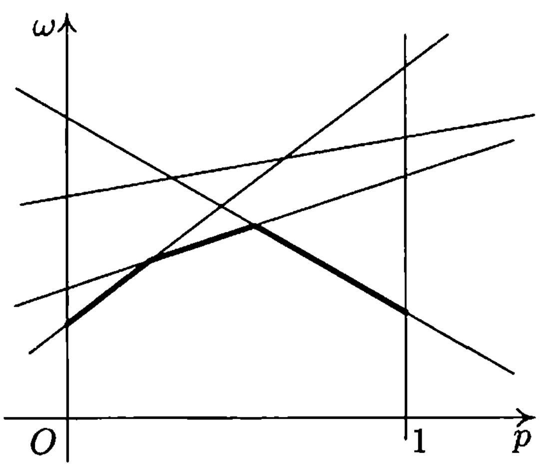


Рис. 3

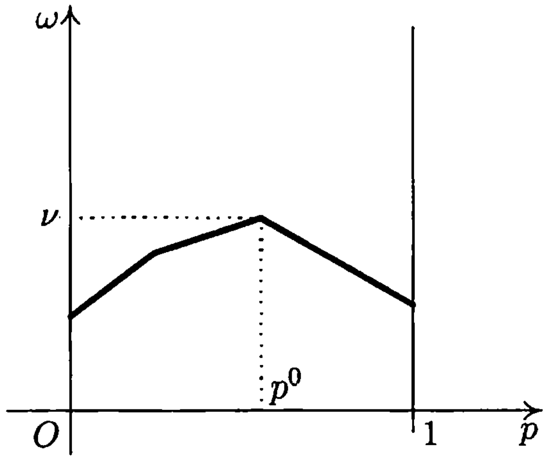


Рис. 4

Эта ломаная как бы огибает снизу все семейство построенных прямых, и по этой причине ее принято называть *нижней огибающей* этого семейства.

Верхняя точка построенной нижней огибающей определяет и цену игры — ν, и оптимальную стратегию игрока A — $\mathrm{P}^0 = \{p^0, 1 - p^0\}$ (рис. 4).

Опробуем описанную схему решения $2 \times n$-игры на конкретном примере.

Пример 4. Рассмотрим игру, заданную 2×6-матрицей

$$\begin{pmatrix} 6 & 4 & 3 & 1 & -1 & 0 \\ -2 & -1 & 1 & 0 & 5 & 4 \end{pmatrix}.$$

Решение.

1-й шаг. Анализ игры на наличие седловой точки.

Нижняя цена игры равна -1, верхняя — равна 1. Седловой точки нет. Решение игры нужно искать в смешанных стратегиях.

2-й шаг. Вычисление средних выигрышей игрока A (проводится при условии, что игрок B выбирает только чистые стратегии).

Из таблицы

$$\begin{array}{r|rrrrrr} p & 6 & 4 & 3 & 1 & -1 & 0 \\ 1-p & -2 & -1 & 1 & 0 & 5 & 4 \end{array}$$

легко получаем

$$\begin{array}{lcccccc} (1): & w & = & 6p & - & 2(1-p), \\ (2): & w & = & 4p & - & (1-p), \\ (3): & w & = & 3p & + & (1-p), \\ (4): & w & = & p, & & \\ (5): & w & = & -p & + & 5(1-p), \\ (6): & w & = & & & 4(1-p). \end{array}$$

3-й шаг. Построение нижней огибающей.

Строим на координатной плоскости (p, w) все шесть прямых, уравнения которых получены на 2-м шаге (рис. 5, масштаб по осям разный), и находим их нижнюю огибающую.

4-й шаг. Отыскание цены игры и оптимальной смешанной стратегии игрока A.

При построении нижней огибающей нетрудно определить, на пересечении каких двух из шести прямых лежит ее наивысшая точка. В данном случае это прямые (4) и (5), заданные уравнениями $w = p$ и $w = -p + 5(1-p)$ соответственно. Решая систему уравнений

$$w = p,$$
$$w = -p + 5(1-p),$$

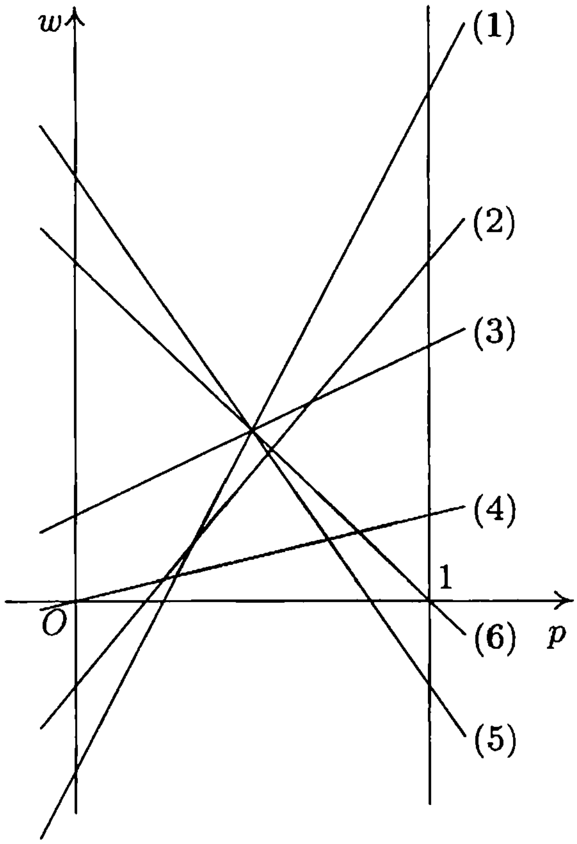


Рис. 5

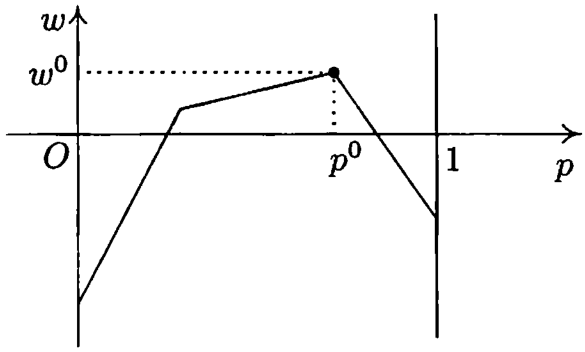


Рис. 6

получаем

$$p^0 = \frac{5}{7}, \quad w^0 = \frac{5}{7}$$

(рис. 6).

Тем самым, цена игры ν и оптимальная стратегия P^0 игрока A соответственно равны:

$$\nu = \frac{5}{7}, \quad \mathrm{P}^0 = \left\{\frac{5}{7}, \frac{2}{7}\right\}.$$

Собственно, этим и заканчивается решение игры для игрока A, поскольку его в первую очередь интересует отыскание собственной оптимальной стратегии и ожидаемого наилучшего гарантированного результата.

Замечание. Решающий матричную игру обычно отождествляет себя с одним из игроков (как правило, это игрок A), считая другого своим противником. Это связано еще и с тем, что в некоторых случаях основное внимание уделяется поиску оптимальных стратегий только игрока A, а стратегии противника могут вообще не интересовать исследователя.

Однако в целом ряде случаев оказывается важным знать оптимальные смешанные стратегии обоих игроков.

Как находится оптимальная смешанная стратегия игрока A, мы уже описали. Покажем теперь, как отыскать оптимальную смешанную стратегию игрока B.

Здесь в зависимости от формы нижней огибающей может представиться несколько случаев.

А. Нижняя огибающая имеет ровно одну наивысшую точку (p^0, w^0).

1) Если $p^0 = 0$ (оптимальная стратегия игрока A — чистая стратегия A_2), то игроку B выгодно применить чистую стратегию, соответствующую прямой, проходящей через точку $(0, w^0)$ и имеющей наибольший отрицательный наклон (рис. 7).

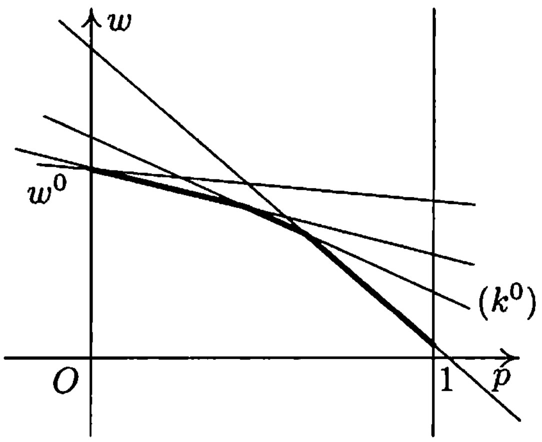


Рис. 7

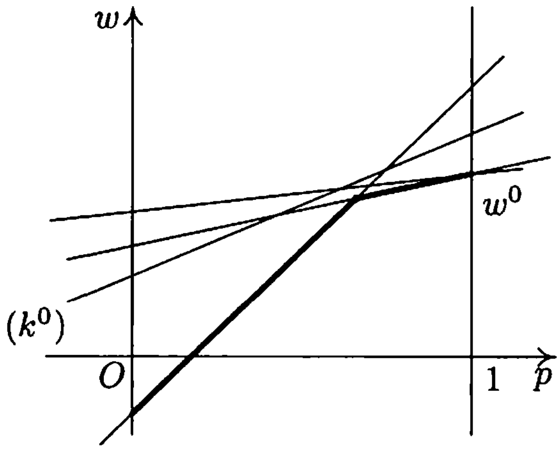


Рис. 8

2) Если $p^0 = 1$ (оптимальная стратегия игрока A — чистая стратегия A_1), то оптимальной для игрока B является чистая стратегия, соответствующая прямой, проходящей через точку $(1, w^0)$ и имеющей наименьший положительный наклон (рис. 8).

3) Если $0 < p^0 < 1$, то в наивысшей точке нижней огибающей пересекаются по меньшей мере две прямые, одна из которых (k-я) имеет положительный наклон, а другая (l-я) — отрицательный (рис. 9), и оптимальная смешанная стратегия игрока B получается, если положить

$$q_k = q, \quad q_l = 1 - q, \quad q_j = 0, \quad j \neq k, l,$$

где q — решение уравнения

$$a_{1k}q + a_{1l}(1 - q) = a_{2k}q + a_{2l}(1 - q).$$

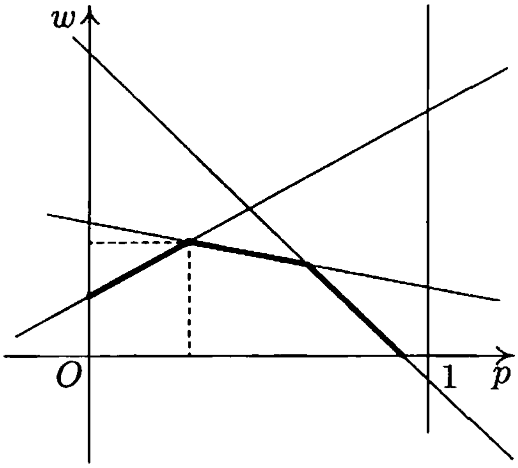


Рис. 9

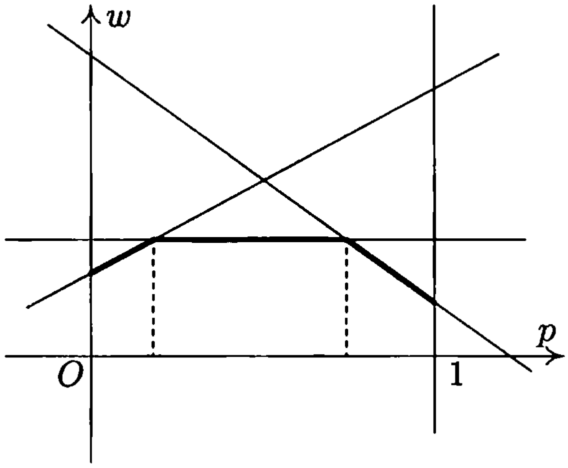


Рис. 10

Б. Нижняя огибающая содержит горизонтальный участок, соответствующий чистой стратегии k^0 игрока B, которая и является оптимальной для него (рис. 10).

Покажем теперь, как найти полное решение игры из примера 4, т. е. еще и оптимальную смешанную стратегию

$$Q^0 = \{q_1^0, q_2^0, q_3^0, q_4^0, q_5^0, q_6^0\}$$

игрока B.

Для этого поступают так:

1) полагают

$$q_1^0 = 0, \quad q_2^0 = 0, \quad q_3^0 = 0, \quad q_4^0 = q, \quad q_5^0 = 1 - q, \quad q_6^0 = 0$$

(выделяя тем самым из шести чистых стратегий игрока B стратегии B_4 и B_5, которые соответствуют прямым (4) и (5), определяющим наивысшую точку нижней огибающей);

2) приравнивают любой из двух средних выигрышей игрока B (игрок A выбирает только чистые стратегии), отвечающий предложенной смешанной стратегии

0	0	0	q	$1-q$	0
6	4	3	1	-1	0
-2	-1	1	0	5	4

к цене игры:

$$q - (1 - q) = \frac{5}{7}, \qquad 5(1 - q) = \frac{5}{7}$$

и

3) получают (в обоих случаях), что

$$q^0 = \frac{6}{7}.$$

Полное решение игры имеет следующий вид:

$$\mathrm{P}^0 = \left\{\frac{5}{7}, \frac{2}{7}\right\}, \quad \mathrm{Q}^0 = \left\{0, 0, 0, \frac{6}{7}, \frac{1}{7}, 0\right\}, \quad \nu = \frac{5}{7}.$$

Замечание. Ситуацию с наличием лишь двух конкурирующих стратегий игрока A нельзя считать надуманной. Она возникает сравнительно часто. Например, в случае, если нужно сравнить два образца некоторого изделия (скажем, старого и модернизированного) с целью выяснения возможности замены, это весьма удобно сделать при помощи платежной матрицы $2 \times n$.

17.3.2. $m \times 2$-игры

Пусть теперь в матричной игре две чистые стратегии имеет игрок B, а число чистых стратегий у игрока A произвольно (равно m).

Это означает, что платежная матрица такой игры имеет вид

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \ldots & \ldots \\ a_{m1} & a_{m2} \end{pmatrix}.$$

Анализ такой игры во многом напоминает рассуждения, описанные для игры $2 \times n$.

Пусть $\mathrm{Q} = \{q, 1-q\}$ — произвольная смешанная стратегия игрока B. Если игрок A выбирает i-ю чистую стратегию, $i = 1, 2, \ldots, m$, то средний выигрыш игрока B в ситуации $\{i,\mathrm{Q}\}$ будет равным

$$(i): \ w = a_{i1}q + a_{i2}(1-q), \qquad i = 1, 2, \ldots, m. \tag{2}$$

Зависимость этого выигрыша от переменной q описывается прямой.

Графиком функции

$$\max_{1 \leqslant i \leqslant m} \left(a_{i1}q + a_{i2}(1-q)\right)$$

является верхняя огибающая семейства прямых (2), соответствующих чистым стратегиям игрока A (рис. 11).

Абсциссой нижней точки полученной ломаной будет значение q^0, определяющее оптимальную смешанную стратегию игрока B, а ординатой w^0 — цена игры.

Замечание. Отыскание оптимальной смешанной стратегии игрока A проводится по той же схеме, которая позволяет находить оптимальную смешанную стратегию игрока B в игре $2 \times n$.

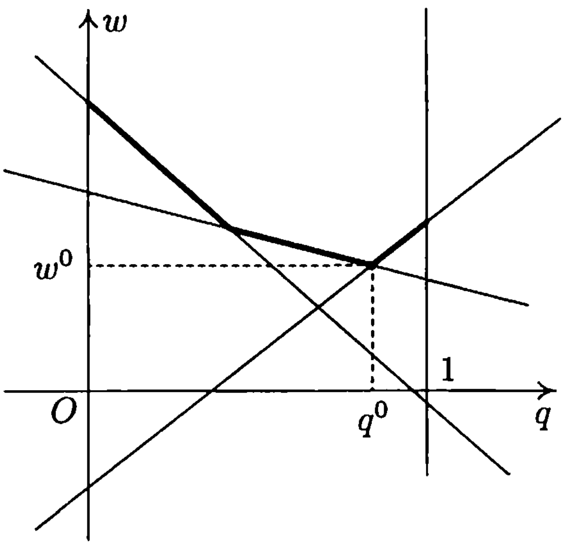


Рис. 11

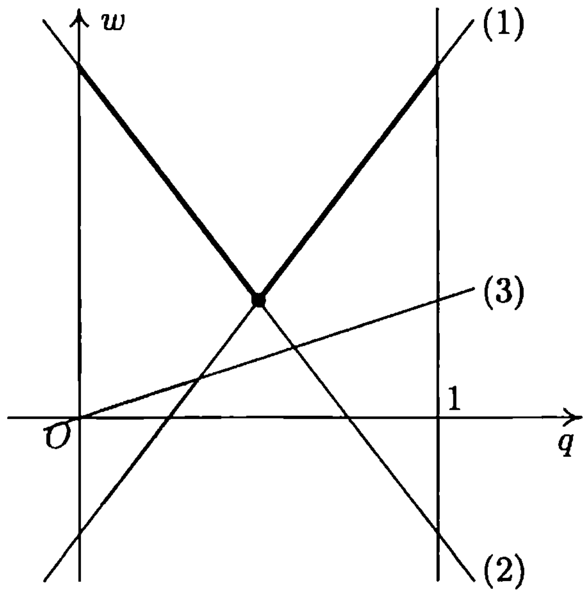


Рис. 12

Рассмотрим конкретный пример.

Пример 5. 3×2-игра задана матрицей

$$\begin{pmatrix} 3 & -1 \\ -1 & 3 \\ 1 & 0 \end{pmatrix}.$$

Решение.

1-й шаг. Анализ игры на наличие седловой точки.

Нижняя цена игры равна 0, верхняя — равна 3. Седловой точки нет. Решение игры нужно искать в смешанных стратегиях.

2-й шаг. Вычисление средних выигрышей игрока B (проводится при условии, что игрок A выбирает только чистые стратегии).

Из таблицы

q	$1-q$
3	−1
−1	3
1	0

получаем:

$$\begin{aligned}&(1): w = 3q - (1-q),\\&(2): w = -q + 3(1-q),\\&(3): w = q.\end{aligned}$$

3-й шаг. Построение верхней огибающей.

Построим на координатной плоскости (q, w) все три прямые, а затем и их верхнюю огибающую (рис. 12).

4-й шаг. Отыскание цены игры и оптимальной смешанной стратегии игрока B.

Нижняя точка верхней огибающей является точкой пересечения прямых (1) и (2). Решая систему уравнений

$$\begin{aligned}w &= 3q - (1-q),\\w &= -q + 3(1-q),\end{aligned}$$

получаем

$$q^0 = \frac{1}{2}, \quad w^0 = 1.$$

5-й шаг. Отыскание оптимальной смешанной стратегии игрока A.

Полагая

$$p_1^0 = p, \quad p_2^0 = 1 - p, \quad p_3 = 0,$$

приравниваем средние выигрыши игрока A, соответствующие чистым стратегиям игрока B:

$$3p - (1-p) = -p + 3(1-p),$$

и находим $p^0 = 1/2$.

Таким образом, цена игры и оптимальные смешанные стратегии игроков A и B соответственно равны:

$$\nu = 1, \quad \mathrm{P}^0 = \left\{\frac{1}{2}, \frac{1}{2}, 0\right\}, \quad \mathrm{Q}^0 = \left\{\frac{1}{2}, \frac{1}{2}\right\}.$$

17.3.3. $m \times n$-игры

В принципе решение любой матричной игры сводится к решению стандартной задачи линейного программирования и тем самым может быть найдено методами линейного программирования. При этом требуемый объем вычислений напрямую зависит от числа чистых стратегий игроков (растет с его увеличением и, значит, с увеличением размеров матрицы игры). Поэтому любые приемы предварительного анализа игры, позволяющие уменьшать размеры ее платежной матрицы или еще как-то упрощать эту матрицу, не нанося ущерба решению, играют на практике весьма важную роль.

Правило доминирования. В целом ряде случаев анализ платежной матрицы обнаруживает, что некоторые чистые стратегии не могут внести никакого вклада в искомые оптимальные смешанные стратегии. Отбрасывание подобных стратегий позволяет заменить первоначальную матрицу на матрицу выигрышей меньших размеров.

Опишем одну из таких возможностей более подробно.

Сравнение строк и столбцов матрицы

Будем говорить, что i-я строка матрицы $\mathbf{A}$

$$a_{i1} \quad a_{i2} \quad \ldots \quad a_{in}$$

не больше j-й строки этой матрицы

$$a_{j1} \quad a_{j2} \quad \ldots \quad a_{jn},$$

если одновременно выполнены следующие n неравенств:

$$a_{i1} \leqslant a_{j1}, \; a_{i2} \leqslant a_{j2}, \; \ldots, \; a_{in} \leqslant a_{jn}.$$

При этом говорят также, что j-я строка *доминирует* i-ю строку или что стратегия A_j игрока A *доминирует* стратегию A_i.

Замечание. Игрок A поступит разумно, если будет избегать стратегий, которым в матрице игры отвечают доминируемые строки.

Если в матрице $\mathbf{A}$ одна из строк (j-я) доминирует другую строку (i-ю), то число строк в матрице $\mathbf{A}$ можно уменьшить путем отбрасывания доминируемой строки (i-й).

Далее, будем говорить, что k-й столбец матрицы $\mathbf{A}$

$$\begin{matrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{matrix}$$

не меньше l-го столбца этой матрицы

$$\begin{matrix} a_{1l} \\ a_{2l} \\ \vdots \\ a_{ml}, \end{matrix}$$

если одновременно выполнены следующие m неравенств:

$$a_{1k} \geqslant a_{1l}, \quad a_{2k} \geqslant a_{2l}, \quad \ldots, \quad a_{mk} \geqslant a_{ml}.$$

При этом говорят также, что l-й столбец *доминирует* k-й столбец или что стратегия B_l игрока B *доминирует* стратегию B_k.

Замечание. Игрок B поступит разумно, если будет избегать стратегий, которым в матрице игры отвечают доминируемые столбцы.

Если в матрице $\mathbf{A}$ один из столбцов (l-й) доминирует другой столбец (k-й), то число столбцов в матрице $\mathbf{A}$ можно уменьшить путем отбрасывания доминируемого столбца (k-го).

Важное замечание. Оптимальные смешанные стратегии в игре с матрицей, полученной усечением исходной за счет доминируемых строк и столбцов, дадут оптимальное решение в исходной игре: доминируемые чистые стратегии игроков в смешении не участвуют — соответствующие им вероятности следует взять равными нулю.

Пример 6. Рассмотрим игру с матрицей

$$\begin{pmatrix} -1 & 0 & 2 & 1 \\ -2 & 0 & 1 & 0 \\ 2 & 1 & -1 & -2 \\ -1 & 0 & 2 & 1 \end{pmatrix}.$$

Сравнивая строки матрицы, видим, что 1-я строка совпадает с 4-й строкой, или, что то же, стратегия A_4 дублирует стратегию A_1.

Тем самым одну из этих строк можно вычеркнуть, не нанося ущерба решению:

$$\begin{pmatrix} -1 & 0 & 2 & 1 \\ -2 & 0 & 1 & 0 \\ 2 & 1 & -1 & -2 \end{pmatrix}.$$

Сравнивая поэлементно 1-ю и 2-ю строки, замечаем, что 1-я строка доминирует 2-ю строку, или, что то же, стратегия A_1 доминирует стратегию A_2. Это вновь позволяет уменьшить число строк матрицы:

$$\begin{pmatrix} -1 & 0 & 2 & 1 \\ 2 & 1 & -1 & -2 \end{pmatrix}.$$

Замечая, что 4-й столбец полученной матрицы доминирует ее 3-й столбец, приходим к игре с 2×3-матрицей:

$$\begin{pmatrix} -1 & 0 & 1 \\ 2 & 1 & -2 \end{pmatrix}.$$

Решая эту 2×3-игру графическим методом, находим ее решение — цену игры и оптимальные смешанные стратегии игроков A и B:

$$\nu = 0, \quad \left\{\frac{2}{3}, \frac{1}{3}\right\}, \quad \left\{\frac{1}{2}, 0, \frac{1}{2}\right\}.$$

Возвращаясь к исходной 4×4-игре, получаем окончательный ответ:

$$\nu = 0, \quad \mathrm{P}^0 = \left\{\frac{2}{3}, 0, \frac{1}{3}, 0\right\}, \quad \mathrm{Q}^0 = \left\{\frac{1}{2}, 0, 0, \frac{1}{2}\right\}.$$

Замечание. При отбрасывании доминируемых строк и столбцов некоторые из оптимальных стратегий могут быть потеряны. Однако цена игры не изменится, и по усеченной матрице может быть найдена хотя бы одна пара оптимальных смешанных стратегий.

Аффинное правило. При поиске решения матричных игр часто оказывается полезным следующее свойство.

Оптимальные стратегии у матричных игр, элементы матриц **A** *и* **C** *которых связаны равенствами*

$$c_{ik} = \lambda a_{ik} + \mu, \quad i = 1, 2, \ldots, m; \quad k = 1, 2, \ldots, n,$$

где $\lambda > 0$, а μ произвольное, имеют одинаковые равновесные ситуации (либо в чистых, либо в смешанных стратегиях), а их цены удовлетворяют следующему условию:

$$\nu_{\mathbf{C}} = \lambda \nu_{\mathbf{A}} + \mu.$$

Элементы матриц

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 1 \\ 2 & 1 & -2 \end{pmatrix} \quad \text{и} \quad \mathbf{C} = \begin{pmatrix} 2 & 5 & 8 \\ 11 & 8 & -1 \end{pmatrix}$$

связаны равенством

$$c_{ik} = 3 \cdot a_{ik} + 5, \quad i = 1, 2; \quad k = 1, 2, 3.$$

Поэтому цена игры с матрицей **C** легко вычисляется:

$$\nu_{\mathrm{C}} = 3 \cdot \nu_{\mathrm{A}} + 5 = 3 \cdot 0 + 5 = 5$$

(см. пример 6).

Основные этапы поиска решения матричной игры. *1-й этап* — проверка наличия (или отсутствия) равновесия в чистых стратегиях (при наличии равновесной ситуации указываются соответствующие оптимальные стратегии игроков и цена игры).

2-й этап — поиск доминирующих стратегий (в случае успеха этого поиска — отбрасывание доминируемых строк и столбцов в исходной матрице игры).

3-й этап — замена игры на ее смешанное расширение и отыскание оптимальных смешанных стратегий и цены игры.

17.3.4. Итерационный метод решения матричных игр

Опишем метод отыскания решения матричной игры — цены игры и оптимальных смешанных стратегий, в известной степени верно отражающий некоторую реальную ситуацию накопления опыта постепенной выработки игроками *хороших* стратегий в результате многих повторений конфликтных ситуаций. Основная идея этого метода заключается в том, чтобы мысленно как бы смоделировать реальное практическое "обучение" игроков в ходе самой игры, когда каждый из них на опыте прощупывает способ поведения противника и старается отвечать на него наиболее выгодным для себя образом. Иными словами, всякий раз при возобновлении игры игрок выбирает наиболее выгодную для себя стратегию, опираясь на предыдущий выбор противника.

Проиллюстрируем этот метод на примере игры, заданной матрицей

$$\begin{pmatrix} 2 & 0 & 3 \\ 1 & 3 & -3 \end{pmatrix}$$

(здесь $maxmin = 0$, $minmax = 2$, следовательно, седловой точки нет).

Опишем правила выбора ходов игроками, предположив для определенности, что начинает игрок A:

ход игрока A — стратегия A_1 – $(2 \quad \mathbf{0} \quad 3)$;

игрок B выбирает свою стратегию так, чтобы выигрыш игрока A был минимален (отмечен выше полужирным шрифтом):

ход игрока B — стратегия B_2 – $(0 \quad \mathbf{3})$;

игрок A выбирает свою стратегию так, чтобы его выигрыш при стратегии B_2 игрока B был максимален (отмечен выше полужирным шрифтом):

ход игрока A — стратегия A_2 – $(1 \quad 3 \quad -3)$;

игрок B выбирает свою стратегию так, чтобы “накопленный” выигрыш игрока A при стратегиях A_1 и A_2,

$$(2 \quad 0 \quad 3) + (1 \quad 3 \quad -3) = (3 \quad 3 \quad \mathbf{0}),$$

был минимален:

ход игрока B — стратегия B_3 – $(3 \quad -3)$;

игрок A выбирает свою стратегию так, чтобы его “накопленный” выигрыш при стратегиях B_2 и B_3 игрока B,

$$(0 \quad 3) + (3 \quad -3) = (\mathbf{3} \quad 0),$$

был максимален:

ход игрока A — стратегия A_1 – $(2 \quad 0 \quad 3)$;

игрок B выбирает свою стратегию так, чтобы “накопленный” выигрыш игрока A при стратегиях A_1, A_2 и A_1,

$$(3 \quad 3 \quad 0) + (2 \quad 0 \quad 3) = (5 \quad \mathbf{3} \quad 3),$$

был минимален:

ход игрока B — стратегия B_2 – $(0 \quad 3)$;

и т. д.

Разобьем последовательные ходы игроков A и B на пары

$$(\text{ход игрока } A, \text{ход игрока } B)$$

и запишем результаты в таблице:

n	i	B_1	B_2	B_3	$\nu_*(n)$	k	A_1	A_2	$\nu^*(n)$	$\nu(n)$
1	1	2	**0**	3	0,00	2	0	**3**	3,00	1,50
2	2	3	3	**0**	0,00	3	**3**	0	1,50	0,75
3	1	5	**3**	3	1,00	2	**3**	3	1,00	1,00
4	1	7	**3**	6	0,75	2	3	**6**	1,50	1,12
5	2	8	6	**3**	0,60	3	**6**	3	1,20	0,90
6	1	10	**6**	6	1,00	2	**6**	6	1,00	1,00
7	1	12	**6**	9	0,86	2	6	**9**	1,44	1,15
8	2	13	9	**6**	0,75	3	**9**	6	1,13	0,93
9	1	15	**9**	9	1,00	2	**9**	9	1,00	1,00
10	1	17	**9**	12	0,90	2	9	**12**	1,20	1,05
...	...	...	...	...	...	...	...	...	...	...

требующей некоторых пояснений.

Описание таблицы

1-й столбец — номер n-го шага (пары последовательных ходов игроков A и B),

2-й столбец — номер i-й стратегии, выбранной игроком A,

3-й столбец — "накопленный" суммарный выигрыш игрока A за первые n шагов при стратегии B_1 игрока B,

4-й столбец — "накопленный" суммарный выигрыш игрока A за первые n шагов при стратегии B_2 игрока B,

5-й столбец — "накопленный" суммарный выигрыш игрока A за первые n шагов при стратегии B_3 игрока B

(минимальный из этих выигрышей выделяется полужирным шрифтом),

6-й столбец — минимальный средний выигрыш игрока A, равный минимальному накопленному им выигрышу за первые n шагов, деленному на число этих шагов,

7-й столбец — номер k стратегии, выбранной игроком B,

8-й столбец — "накопленный" суммарный выигрыш игрока A за первые n шагов при стратегии A_1,

9-й столбец — "накопленный" суммарный выигрыш игрока A за первые n шагов при стратегии A_2

(максимальный из этих выигрышей выделяется полужирным шрифтом),

10-й столбец — максимальный средний выигрыш игрока A, равный максимальному накопленному им выигрышу за первые n шагов, деленному на число этих шагов,

11-й столбец — среднее арифметическое минимального среднего выигрыша и максимального среднего выигрыша игрока A.

Цена игры определяется приближенно по окончании любого из шагов. Например, за приближенную цену игры можно взять среднее арифметическое $\nu(n)$, полученное на n-м шаге.

Смешанные стратегии противников определяются частотами появления чистых стратегий. Например, после 9-го шага имеем:

$$\nu(9) = 1,00, \quad \mathrm{P}_9 = \left\{\frac{6}{9}, \frac{3}{9}\right\}, \quad \mathrm{Q}_9 = \left\{0, \frac{6}{9}, \frac{3}{9}\right\},$$

а после 10-го:

$$\nu(10) = 1,05, \quad \mathrm{P}_{10} = \left\{\frac{7}{10}, \frac{3}{10}\right\}, \quad \mathrm{Q}_{10} = \left\{0, \frac{7}{10}, \frac{3}{10}\right\}.$$

Так как эта игра легко решается графически, полезно сравнить полученные результаты с точными:

$$\nu = 1, \quad \mathrm{P} = \left\{\frac{2}{3}, \frac{1}{3}\right\}, \quad \mathrm{Q} = \left\{0, \frac{2}{3}, \frac{1}{3}\right\}.$$

Сделаем несколько замечаний.

Замечание 1. При увеличении числа шагов все три величины $\nu_*(n)$, $\nu^*(n)$ и $\nu(n)$ будут приближаться к цене игры ν, но среднее арифметическое $\nu(n)$ будет приближаться к ν сравнительно быстрее.

Замечание 2. Хотя сходимость итераций весьма медленна, тем не менее даже такой небольшой расчет всегда дает возможность находить ориентировочное значение цены игры и доли чистых стратегий.

Замечание 3. Сравнительно медленную скорость сходимости можно объяснить целым рядом причин. Укажем одну из них, психологически наиболее интересную. Если, к примеру, игрок A уже нашел оптимальную смешанную стратегию, то он не склонен останавливаться на ней — он продолжит попытки выиграть у противника B побольше, особенно если последний еще не достиг оптимальной смешанной стратегии. Тем самым игрок A может невольно ухудшить свое положение.

Замечание 4. Отметим два основных преимущества описанного метода:

1) итерационный метод прост и одновременно универсален (при его помощи можно легко найти приближенное решение любой матричной игры),

2) объем и сложность вычислений сравнительно слабо растут по мере увеличения числа стратегий игроков (размеров матрицы игры).

17.4. Некоторые задачи, сводимые к матричным играм

В чистом виде антагонистические конфликты встречаются редко (разве только в боевых действиях и в спортивных состязаниях). Однако довольно часто конфликты, в которых интересы сторон противоположны, при допущении, что множество способов действия сторон конечно, можно моделировать матричными играми.

Рассмотрим несколько конкретных ситуаций.

Пример 7 (планирование посева). Сельскохозяйственное предприятие имеет возможность выращивать две культуры — A_1 и A_2. Необходимо определить, как сеять эти культуры, если при прочих равных условиях их урожаи зависят от погоды, а план посева должен обеспечить наибольший доход (прибыль от реализации выращенной культуры определяется полученным объемом). В зоне рискованного земледелия (а таковой является бóльшая часть России) планирование посева должно осуществляться с учетом наименее благоприятного состояния погоды.

Таким образом, одной из сторон выступает сельскохозяйственное предприятие, заинтересованное в том, чтобы получить наибольший доход (игрок A), а другой стороной — природа, способная навредить сельскохозяйственному предприятию в максимальной степени (от нее зависят погодные условия) и как бы преследующая тем самым прямо противоположные цели (игрок B).

Принятие природы за противника равносильно планированию посева с учетом наиболее неблагоприятных условий; если же погодные условия окажутся благоприятными, то выбранный план даст возможность увеличить доход.

Налицо антагонистический конфликт, в котором у игрока A две стратегии — A_1 и A_2, а у игрока B три — B_1 (засушливое лето), B_2 (нормальное лето) и B_3 (дождливое лето).

В качестве выигрыша игрока A возьмем прибыль от реализации и будем считать, что расчеты прибыли сельскохозяйственного предприятия (в млн. руб.) в зависимости от состояний погоды сведены в следующую матрицу:

$$\begin{pmatrix} 8 & 5 & 3 \\ 2 & 3 & 6 \end{pmatrix}.$$

Нетрудно заметить, что седловой точки у этой матрицы нет. Поэтому оптимальная стратегия игрока A будет смешанной. Применяя графический метод, получаем

$$\mathrm{P}^0 = \left\{\frac{3}{5}, \frac{2}{5}\right\}, \qquad \nu = \frac{21}{5}.$$

Замечание. Здесь мы столкнулись со сравнительно редкой ситуацией, когда оптимальная смешанная стратегия одного из игроков допускает так называемую физическую реализацию. Полученное решение сельскохозяйственное предприятие может использовать так:

на $3/5$ всех площадей выращивать культуру A_1,

на $2/5$ всех площадей выращивать культуру A_2

и получать прибыль в размере, не меньшем 4,2 млн. руб.

Пример 8(переговоры о заключении контракта между профсоюзом и администрацией). Рассмотрим фирму, администрация которой ведет переговоры с профсоюзом рабочих и служащих о заключении контракта.

Предположим, что платежная матрица, отражающая интересы договаривающихся сторон, имеет следующий вид:

$$\begin{pmatrix} 75 & 105 & 65 & 45 \\ 70 & 60 & 55 & 40 \\ 80 & 90 & 35 & 50 \\ 95 & 100 & 50 & 55 \end{pmatrix}.$$

Выплаты указаны в центах в час и представляют собой среднюю зарплату служащего фирмы вместе со всеми добавками. Тем самым заданная матрица описывает прибыль профсоюза (игрок A) и затраты администрации (игрок B).

Ясно, что профсоюз стремится максимизировать доходы рабочих и служащих, в то время как администрации хотелось бы минимизировать собственные потери.

Нетрудно заметить, что седловой точки у платежной матрицы нет. Кроме того, для дальнейшего анализа существенными являются лишь стратегии A_1 и A_4 игрока A и стратегии B_3 и B_4 игрока B (в этом легко убедиться, воспользовавшись правилом доминирования стратегий). В результате соответствующего усечения получим матрицу

$$\begin{pmatrix} 65 & 45 \\ 50 & 55 \end{pmatrix}.$$

Элементы матрицы

$$\begin{pmatrix} 4 & 0 \\ 1 & 2 \end{pmatrix}$$

связаны с элементами предыдущей матрицы соотношениями

$$65 = 5 \cdot 4 + 45, \quad 45 = 5 \cdot 0 + 45, \quad 50 = 5 \cdot 1 + 45, \quad 55 = 5 \cdot 2 + 45.$$

Применяя графический метод, в итоге получим

$$\mathrm{P} = \left\{\frac{1}{5}, 0, 0, \frac{4}{5}\right\}, \quad \mathrm{Q} = \left\{0, 0, \frac{2}{5}, \frac{3}{5}\right\}, \quad \nu = 53.$$

Тем самым профсоюзу следует выбирать стратегию A_1 в 20% случаев и стратегию A_4 в 80%. Что касается администрации, то ей следует выбирать стратегию B_3 с вероятностью 0,4 и стратегию B_4 с вероятностью 0,6. При этом ожидаемая цена игры равна 53.

Замечание. Следует отметить, что если процесс переговоров будет повторяться много раз, то среднее должно сходиться к ожидаемому значению 53. Если же переговоры пройдут лишь единожды, то реальный результат получится при выборе каждым игроком некоторой своей чистой стратегии. Поэтому один из игроков, профсоюз или администрация, будет неудовлетворен.

Пример 9 (локальный конфликт). Рассмотрим войну между двумя небольшими государствами A и B, которая ведется в течение 30 дней.

Для бомбардировки небольшого моста — важного военного объекта страны B — страна A использует оба имеющихся у нее самолета. Разрушенный мост восстанавливается в течение суток, а каждый самолет совершает один полет в день по одному из двух воздушных маршрутов, соединяющих эти страны. У страны B есть два зенитных орудия, при помощи которых можно сбивать самолеты страны A. Если самолет собьют, то некая третья страна в течение суток поставит стране A новый самолет.

Страна A может послать самолеты либо по одному маршруту, либо по разным.

Страна B может поместить либо обе зенитки на одном маршруте, либо по одной зенитке на каждом маршруте.

Если один самолет летит по маршруту, на котором расположена одна зенитка, то этот самолет будет сбит.

Если два самолета летят по маршруту, на котором расположены две зенитки, то оба самолета будут сбиты.

Если два самолета летят по маршруту, на котором расположена одна зенитка, то сбит будет только один самолет.

Если самолет доберется до цели, то мост будет уничтожен.

У страны A есть две стратегии:
послать самолеты по разным маршрутам — A_1,
послать самолеты по одному маршруту — A_2.
У страны B также две стратегии:
поместить зенитки на разных маршрутах — B_1,
поместить зенитки на одном маршруте — B_2.

Если страна A выберет стратегию A_1, а страна B — стратегию B_1, то страна A получит нулевой выигрыш, так как ни один из самолетов не достигнет цели.

Если страна A выберет стратегию A_2, а страна B — стратегию B_1, то хотя бы один самолет достигнет цели и вероятность разрушения моста будет равна 1.

Если страна A выберет стратегию A_1, а страна B — стратегию B_2, то вновь хотя бы один самолет достигнет цели и вероятность разрушения моста будет равна 1.

Если страна A выберет стратегию A_2, а страна B — стратегию B_2, то страна A с вероятностью 1/2 выберет маршрут, на котором установлены зенитки, и, следовательно, цель будет уничтожена с вероятностью 1/2.

Запишем результаты проведенного анализа в стандартной игровой форме:

$$\begin{pmatrix} 0 & 1 \\ 1 & 1/2 \end{pmatrix}.$$

При помощи графического метода получаем оптимальные смешанные стратегии игроков и цену игры:

$$\mathrm{P} = \left\{\frac{1}{3}, \frac{2}{3}\right\}, \quad \mathrm{Q} = \left\{\frac{1}{3}, \frac{2}{3}\right\}, \quad \nu = \frac{2}{3}.$$

Это означает, что если страна A будет посылать самолеты по разным маршрутам в течение десяти дней из тридцати, отпущенных на войну (и, значит, по одному маршруту в течение двадцати дней), то в среднем она будет иметь 66,7% удачных случаев (мост будет находиться в нерабочем состоянии). Воспользовавшись для своих зениток предложенным выбором, страна B не позволит бомбить мост чаще, чем в 66,7% случаев.

Подведем некоторые итоги.

Матричные игры моделируют конфликтные ситуации, в которых каждая из сторон-участниц делает свой ход одновременно со второй стороной. При этом наибольший интерес представляет случай, когда игра не заканчивается сразу же после совершения игроками одной такой пары одновременных ходов, а повторяется многократно. Причем считается, что перед каждым возобновлением игры игроки не получают никаких *новых* сведений ни о конфликте, ни о возможных действиях противной стороны. Иными словами, при многократном повторении матричной игры каждая из сторон всякий раз оказывается перед выбором некоторой стратегии из одного и того же множества стратегий, неизменного у каждого из игроков.

Тем не менее в таких многократно повторяющихся обстоятельствах большую роль играет анализ игры, как предварительный, так и промежуточный.

В результате разумно проведенного предварительного анализа матричной игры заинтересованная в нем сторона может определить свою линию поведения (правило выбора стратегий) на всю серию игр. Разумеется, описанный нами выше максиминный подход является далеко не единственным средством. Однако не следует забывать, что принципиальной особенностью этого подхода является то обстоятельство, что игрок, придерживающийся выводимого на его основе правила выбора стратегий, заранее может довольно точно оценить нетривиальные размеры своего гарантированного выигрыша. Кроме того, максиминный подход позволяет сводить задачу поиска решения игры к рассмотрению сравнительно несложных задач линейного программирования и тем самым получать эффективные рекомендации, как лучше выбирать стратегии в конкретной игре при многократном ее повторении.

Если игра повторяется много раз, то некоторые дополнительные сведения — какие именно стратегии выбирает противная сторона и какими правилами выбора стратегий она руководствуется — игрок все же получает. На основании этих сведений и результатов предварительного анализа игры он может довольно точно оценить противника, и если тот не придерживается компромиссного минимаксного подхода, внести соответствующие изменения в собственную линию поведения и увеличить выигрыш.

17.5. Задания и ответы

1. Найдите нижнюю цену игры, верхнюю цену игры, определите седловые точки, оптимальные чистые стратегии и цену игры (если они существуют):

$$\text{а)}\ \begin{pmatrix} -1 & 3 & -2 & 1 \\ 2 & 3 & 4 & -1 \\ 7 & 1 & -5 & 2 \\ -8 & 4 & 3 & -1 \end{pmatrix}, \quad \text{б)}\ \begin{pmatrix} 1 & 2 & 1 & 3 & 7 \\ 3 & -5 & -1 & 6 & -2 \\ 3 & 2 & 1 & 4 & 2 \\ -1 & 7 & -2 & -3 & 4 \end{pmatrix}.$$

Ответы: а) $\alpha = -2$, $\beta = 2$; б) $\nu = 1$, $\{A_1, B_3\}$ и $\{A_3, B_3\}$.

2. Найдите решения следующих матричных игр:

$$\text{а)}\ \begin{pmatrix} -5 & 8 \\ 4 & -7 \end{pmatrix},\quad \text{б)}\ \begin{pmatrix} -1 & 1 & -1 & 2 \\ 0 & -1 & 2 & -2 \end{pmatrix},\quad \text{в)}\ \begin{pmatrix} 4 & 2 & 3 & -1 \\ -4 & 0 & -2 & 2 \end{pmatrix}$$

$$\text{г)}\ \begin{pmatrix} 1 & 4 \\ 3 & -2 \\ 0 & 5 \end{pmatrix},\quad \text{д)}\ \begin{pmatrix} 4 & 2 & 0 & 1 & 0 & 2 \\ 4 & 2 & 0 & 2 & 1 & 1 \\ 4 & 3 & 1 & 3 & 1 & 2 \\ 4 & 3 & 4 & -1 & 2 & 2 \\ 4 & 3 & 3 & -2 & 2 & 2 \end{pmatrix}.$$

Ответы: а) $\mathrm{P} = \left\{\frac{11}{24}, \frac{13}{24}\right\}$, $\mathrm{Q} = \left\{\frac{5}{8}, \frac{3}{8}\right\}$, $\nu = -\frac{1}{8}$;

б) $\mathrm{P} = \left\{\frac{2}{5}, \frac{3}{5}\right\}$, $\mathrm{Q} = \left\{\frac{4}{5}, 0, 0, \frac{1}{5}\right\}$, $\nu = -\frac{2}{5}$;

в) $\mathrm{P} = \left\{\frac{6}{11}, \frac{5}{11}\right\}$, $\mathrm{Q} = \left\{\frac{3}{11}, 0, 0, \frac{8}{11}\right\}$, $\nu = \frac{4}{11}$;

г) $\mathrm{P} = \left\{\frac{5}{8}, \frac{3}{8}, 0\right\}$, $\mathrm{Q} = \left\{\frac{3}{4}, \frac{1}{4}\right\}$, $\nu = \frac{7}{4}$;

д) $\mathrm{P} = \left\{0, 0, \frac{3}{5}, \frac{2}{5}, 0\right\}$, $\mathrm{Q} = \left\{0, 0, 0, \frac{1}{5}, \frac{4}{5}, 0\right\}$, $\nu = \frac{7}{5}$.

3. Найдите приближенные решения следующих матричных игр:

$$\text{а)}\ \begin{pmatrix} 4 & -2 & 0 \\ 0 & 1 & 2 \\ -3 & 3 & -1 \end{pmatrix},\quad \text{б)}\ \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 2 \\ 0 & 2 & -4 \end{pmatrix}.$$

Ответы: а) $\mathrm{P} \approx \{0{,}14; 0{,}86; 0\}$, $\mathrm{Q} \approx \{0{,}43; 0{,}57; 0\}$, $\nu \approx 0{,}46$.

б) $\mathrm{P} \approx \{0{,}05; 0{,}62; 0{,}33\}$, $\mathrm{Q} \approx \{0; 0{,}67; 0{,}33\}$, $\nu \approx 1{,}00$.

Глава 18
ПОЗИЦИОННЫЕ ИГРЫ

Во многих практически важных конфликтных ситуациях, располагая той или иной информацией об их прошлом развитии, стороны-участницы совершают свой выбор не раз и навсегда, а последовательно во времени, шаг за шагом. Тем самым они используют стратегии, отражающие как динамику конфликта, так и степень собственной информированности о фактически складывающейся обстановке в развитии этого конфликта.

18.1. Структура позиционной игры

Одним из классов игр, описывающих конфликты, динамика которых оказывает влияние на поведение участников, являются так называемые позиционные игры.

Позиционная игра — это бескоалиционная игра, моделирующая процессы последовательного принятия решений игроками в условиях меняющейся во времени и, вообще говоря, неполной информации.

Процесс самой игры состоит в последовательном переходе от одного состояния игры к другому, который осуществляется либо путем выбора игроками одного из возможных действий в соответствии с правилами игры, либо случайным образом (*случайный ход*).

В качестве примеров позиционных игр можно привести крестики-нолики, шашки, шахматы, карточные игры, домино и др. Интересно, что право выбора первого хода в этих играх часто определяется случайным образом.

Состояния игры принято называть *позициями* (отсюда и название — позиционные игры), а возможные выборы в каждой позиции — *альтернативами.*

Характерной особенностью позиционной игры является возможность представления множества позиций в виде древовидного упорядоченного множества, которое называется *деревом игры* (рис. 1).

Для определенности мы будем рассматривать позиционные игры, в каждой позиции которых кроме окончательных ровно две альтернативы, первая и вторая.

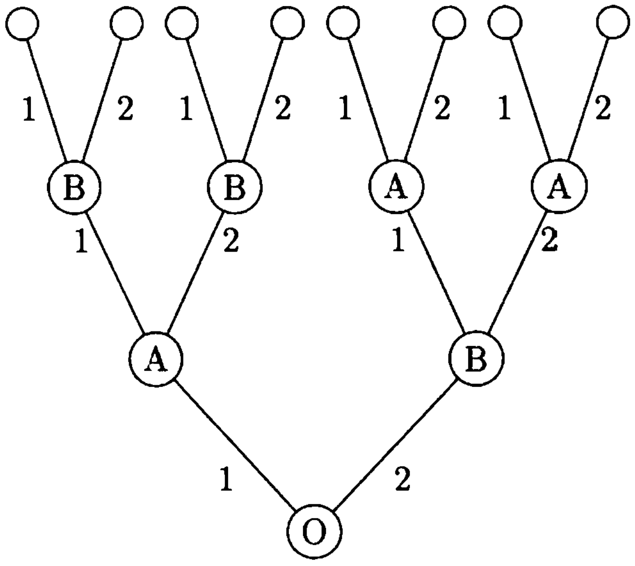


Рис. 1

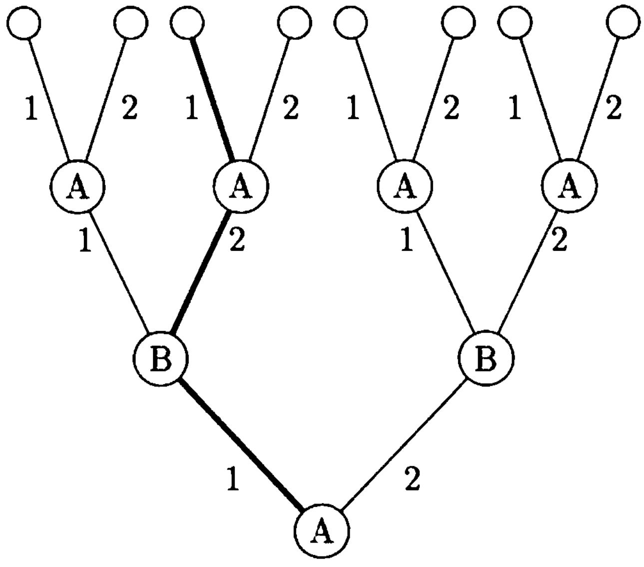


Рис. 2

Замечание. Символ O, A или B в кружке указывает, кто из игроков, O, A или B, делает очередной ход. При этом символом O обычно обозначается ход в игре, осуществляемый не игроком, а каким-нибудь случайным механизмом (иногда его называют *природой*). Например, в позиционной игре, представленной на рис. 1 своим деревом, первый ход производится случайно.

Пользуясь графическим описанием игры, можно сказать, что процесс игры состоит в переходе от начальной позиции к окончательной через непосредственно следующие одна за другой промежуточные позиции.

Каждая окончательная вершина определяет единственную цепь (последовательность идущих друг за другом звеньев), связывающую начальную вершину с данной (рис. 2). Такая цепь называется *партией.* Число различных партий равно числу окончательных вершин (позиций).

В каждой окончательной позиции задан числовой выигрыш игрока A.

Замечание. Мы будем рассматривать здесь только антагонистические позиционные игры.

В шахматах функция выигрышей игрока A (белых) определяется так:

$+1$ — в выигрываемых партиях,

0 — в ничейных партиях,

-1 — в проигрываемых партиях.

Функция выигрышей игрока B (черных) отличается от функции выигрышей белых только знаком.

Различают позиционные игры *с полной информацией* и позиционные игры *с неполной информацией.*

В позиционных играх с полной информацией (пример — шашки, шахматы) каждый игрок при своем ходе знает ту позицию дерева игры, в которой он находится.

В позиционных играх с неполной информацией (пример — домино) игроку при своем ходе позиция дерева игры, в которой он фактически находится, точно не известна. Этот игрок знает лишь некоторое множество позиций, включающее в себя его фактическую позицию. Такое множество позиций называется *информационным множеством.*

Таким образом, в игре с неполной информацией игрок при своем ходе знает, в каком информационном множестве он находится, но ему неизвестно, в какой именно позиции этого множества.

Позиции, принадлежащие одному и тому же информационному множеству, объединяются пунктирными линиями.

Рассмотрим примеры двух игр, состоящих из двух ходов, которые последовательно делают участвующие в ней игроки A и B. Начинает игрок A: он выбирает одну из двух возможных альтернатив — число x, равное либо 1 (первая альтернатива), либо 2 (вторая альтернатива). На ход игрока A игрок B отвечает своим ходом, выбирая одну из двух возможных альтернатив — число y, равное либо 1 (первая альтернатива), либо 2 (вторая альтернатива).

И в результате игрок A получает вознаграждение или вынужден платить штраф.

Пример 1. *1-й ход.* Игрок A выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход. Игрок B выбирает число y из множества двух чисел $\{1, 2\}$, зная выбор числа x игроком A.

Функция $W(x, y)$ выплат игроку A за счет игрока B задается так:

$$W(1,1) = 1, \qquad W(2,1) = -2,$$
$$W(1,2) = -1, \qquad W(2,2) = 2.$$

На рис. 3 показаны дерево игры и информационные множества.

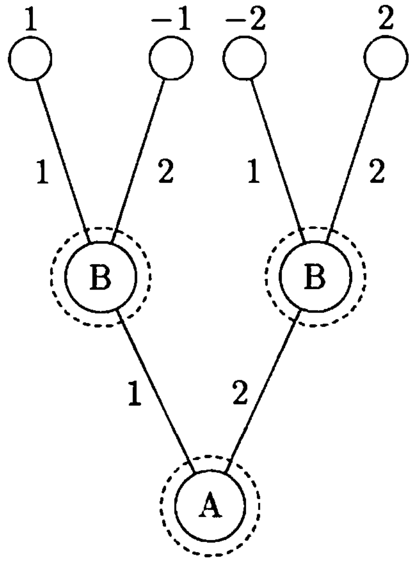


Рис. 3

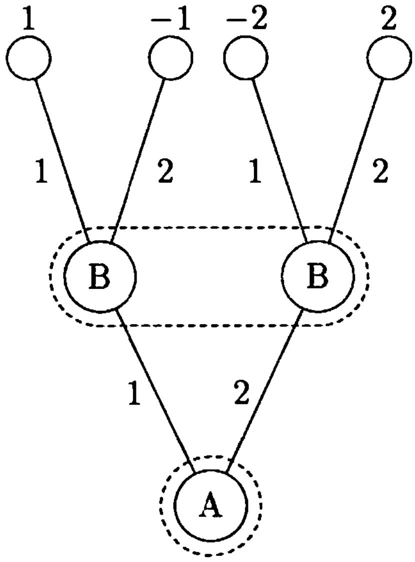


Рис. 4

Пример 2. В случае, если выполнены все условия предыдущего примера, кроме одного — хода игрока B,

2-й ход — игрок B выбирает число y из множества двух чисел $\{1, 2\}$, не зная выбора числа x игроком A.

Информационные множества выглядят так, как показано на рис. 4.

18.2. Нормализация позиционной игры

Заранее определенную последовательность ходов игрока, выбранную им в зависимости от информации о ходах другого игрока и ходах игрока O (природы), будем называть *чистой стратегией* этого игрока.

В том случае, если в игре нет случайных ходов (игрок O в игре не участвует), выбор игроком A и игроком B чистых стратегий однозначно определяет исход игры — приводит к окончательной позиции, где игрок A и получает свой выигрыш. Это обстоятельство позволяет сводить позиционную игру к матричной игре.

Процесс сведения позиционной игры к матричной называется *нормализацией позиционной игры.*

Покажем на нескольких примерах, как это делается.

Пример 3. Опишем стратегии игроков из примера 1.

Стратегию игрока A можно задать одним числом x, показывающим, какую альтернативу, первую или вторую, выбрал игрок.

Тем самым у игрока A две чистые стратегии:

$$A_1 \text{ — выбрать } x = 1, \quad A_2 \text{ — выбрать } x = 2.$$

Стратегию игрока B, принимая во внимание, что выбор игрока A на 1-м ходе ему известен, удобно описывать упорядоченной парой

$$[y_1, y_2].$$

Здесь y_1 $(y_1 = 1, 2)$ — альтернатива, выбираемая игроком B при условии, что игрок A выбрал первую альтернативу, $x = 1$, а y_2 $(y_2 = 1, 2)$ — альтернатива, выбираемая игроком B при условии, что игрок A выбрал вторую альтернативу, $x = 2$.

Например, выбор игроком B стратегии $[2, 1]$ означает, что если на 1-м ходе игрок A выбрал $x = 1$, то игрок B на своем ходе должен выбрать $y = 2$. Если же на 1-м ходе игрок A выбрал $x = 2$, то, согласно этой стратегии, игрок B на своем ходе должен выбрать $y = 1$.

Таким образом, у игрока B четыре чистые стратегии:

B_1 - $[1, 1]$, $y = 1$ при любом выборе x;

B_2 - $[1, 2]$, $y = x$ при любом выборе x;

B_3 - $[2, 1]$, $y \neq x$ при любом выборе x;

B_4 - $[2, 2]$, $y = 2$ при любом выборе x.

Покажем теперь, как рассчитываются выигрыши игрока A в зависимости от примененных стратегий.

Пусть, например, игрок A выбрал стратегию A_1 - (1), а игрок B — стратегию B_2 - $[1, 2]$. Тогда $x = 1$, а из стратегии $[1, 2]$ вытекает, что $y = 1$. Отсюда

$$W(x, y) = W(1, 1) = 1.$$

Остальные выигрыши рассчитываются совершенно аналогично.

Результаты расчетов записываются обычно или в виде таблицы выигрышей игрока A:

		B_1	B_2	B_3	B_4
		$[1,1]$	$[1,2]$	$[2,1]$	$[2,2]$
A_1	$x=1$	$W(1,1)$	$W(1,1)$	$W(1,2)$	$W(1,2)$
A_2	$x=2$	$W(2,1)$	$W(2,2)$	$W(2,1)$	$W(2,2)$

или в виде матрицы игры:

$$\begin{pmatrix} 1 & 1 & -1 & -1 \\ -2 & 2 & -2 & 2 \end{pmatrix},$$

где, как обычно, строки соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B.

Полученная матрица имеет седловую точку. Оптимальные стратегии игроков: A_1 – (1) и B_3 – $[2,1]$. Тем самым игрок A на 1-м ходе выбирает $x = 1$, а игрок B на 2-м ходе выбирает $y = 2$. Цена игры $\nu = -1$.

Пример 4. Опишем стратегии игроков из примера 2.

У игрока A они те же, что и в предыдущем примере:

$$A_1 \text{ — выбрать } x = 1, \quad A_2 \text{ — выбрать } x = 2.$$

Так как игроку B выбор игрока A неизвестен, т. е. игрок B не знает, в какой именно из двух позиций он находится (см. рис. 4), то у него те же две стратегии:

$$B_1 \text{ — выбрать } y = 1, \quad B_2 \text{ — выбрать } y = 2.$$

Соответствующие таблица выигрышей игрока A и матрица игры имеют следующий вид:

		B_1	B_2
		$y=1$	$y=2$
A_1	$x=1$	$W(1,1)$	$W(1,2)$
A_2	$x=2$	$W(2,1)$	$W(2,2)$

$$\begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix}.$$

Полученная матрица седловой точки не имеет. Оптимальные смешанные стратегии игроков: $P = \{2/3, 1/3\}$ и $Q = \{1/2, 1/2\}$. Цена игры $\nu = 0$.

Замечание 1. На этих двух примерах хорошо видно, что результат сведения позиционной игры к матричной напрямую зависит от степени информированности игроков. В частности, отсутствие у игрока B сведений о выборе, сделанном игроком A, приводит к уменьшению количества его возможных стратегий. Сравнивая ответы, полученные в примерах 3 и 4, замечаем, что снижение уровня информированности игрока (в данном случае игрока B) делает для него исход игры менее благоприятным.

Замечание 2. Приведенные выше примеры всех возможных вариантов не исчерпывают даже в этом самом простом случае двухходовых позиционных игр.

Рассмотрим теперь несколько примеров сведения к матричным играм позиционных игр, состоящих из трех ходов, сосредоточив при этом основное внимание на одном из наиболее ответственных шагов нормализации — описании стратегий игроков.

Пример 5. *1-й ход* делает игрок A: он выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход делает игрок B: зная выбранное игроком A число x, он выбирает число y из множества двух чисел $\{1, 2\}$.

3-й ход делает игрок A: не зная о выбранном игроком B числе y на 2-м ходе и забыв выбранное им самим на 1-м ходе число x, он выбирает число z из множества двух чисел $\{1, 2\}$.

После этого игрок A получает вознаграждение $W(x, y, z)$ за счет игрока B, например, такое:

$$\begin{aligned}
W(1,1,1) &= -2, & \qquad W(2,1,1) &= 3,\\
W(1,1,2) &= 4, & W(2,1,2) &= 0,\\
W(1,2,1) &= 1, & W(2,2,1) &= -3,\\
W(1,2,2) &= -4, & W(2,2,2) &= 5.
\end{aligned}$$

На рис. 5 показаны дерево игры и информационные множества.

Нормализуем эту игру.

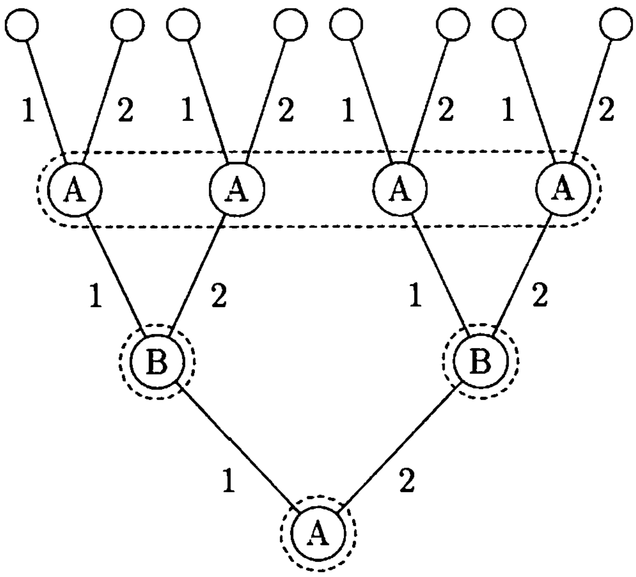


Рис. 5

Поскольку игроку B выбор игрока A на 1-м ходе известен, то у игрока B те же четыре стратегии, что и в примере 3:

$$B_1 - [1,1],\ B_2 - [1,2],\ B_3 - [2,1],\ B_4 - [2,2].$$

Игрок A на 3-м ходе не знает предыдущих выборов — ни значения x, ни значения y. Поэтому каждая его стратегия состоит просто из пары чисел (x, z), где x $(x = 1, 2)$ — альтернатива, выбираемая игроком A на 1-м ходе, а z $(z = 1, 2)$ — альтернатива, выбираемая игроком A на 3-м ходе.

Например, выбор игроком A стратегии $(2, 1)$ означает, что на 1-м ходе он выбирает $x = 2$, а на 3-м ходе — $z = 1$.

Таким образом, у игрока A четыре стратегии:

$$A_1 - (1,1),\ A_2 - (1,2),\ A_3 - (2,1),\ A_4 - (2,2).$$

Покажем теперь, как рассчитываются выигрыши игрока A в зависимости от стратегий, применяемых игроками в данной игре. Пусть, например, игрок A выбрал стратегию $A_2 - (1,2)$, а игрок B — стратегию $B_3 - [2,1]$. Тогда $x = 1$, откуда вытекает, что $y = 2$. Значение $z = 2$ выбрано игроком A независимо от выбора игрока B. Вычисляя значение функции выигрышей для этого набора, получаем

$$W(x, y, z) = W(1, 2, 2) = -4.$$

В результате подобных рассуждений получаются и остальные пятнадцать выигрышей. Это позволяет построить таблицу выигрышей игрока A. Имеем:

		B_1	B_2	B_3	B_4
		$[1,1]$	$[1,2]$	$[2,1]$	$[2,2]$
A_1	$(1,1)$	$W(1,1,1)$	$W(1,1,1)$	$W(1,2,1)$	$W(1,2,1)$
A_2	$(1,2)$	$W(1,1,2)$	$W(1,1,2)$	$W(1,2,2)$	$W(1,2,2)$
A_3	$(2,1)$	$W(2,1,1)$	$W(2,2,1)$	$W(2,1,1)$	$W(2,2,1)$
A_4	$(2,2)$	$W(2,1,2)$	$W(2,2,2)$	$W(2,1,2)$	$W(2,2,2)$

или

$$\begin{pmatrix} -2 & -2 & 1 & 1 \\ 4 & 4 & -4 & -4 \\ 3 & -3 & 3 & -3 \\ 0 & 5 & 0 & 5 \end{pmatrix}.$$

Пример 6. *1-й ход* делает игрок A: он выбирает число x из множества двух чисел $\{1,2\}$.

2-й ход делает игрок B: не зная о выборе игрока A на 1-м ходе, он выбирает число y из множества двух чисел $\{1,2\}$.

3-й ход делает игрок A: он выбирает число z из множества двух чисел $\{1,2\}$, не зная ни значения x, ни значения y.

После этого игроки расплачиваются по правилу, указанному в примере 5.

Графическое представление этой игры показано на рис. 6.

Ясно, что у игрока A те же четыре стратегии, что и в примере 5:

A_1 – $(1,1)$, A_2 – $(1,2)$, A_3 – $(2,1)$, A_4 – $(2,2)$.

У игрока B всего две стратегии:

B_1 — выбрать $y=1$, B_2 — выбрать $y=2$.

В этом случае (весьма слабой информированности игроков) таблица выигрышей игрока A и соответствующая матрица строятся совсем просто. Имеем:

		B_1	B_2
		$y=1$	$y=2$
A_1	$(1,1)$	$W(1,1,1)$	$W(1,2,1)$
A_2	$(1,2)$	$W(1,1,2)$	$W(1,2,2)$
A_3	$(2,1)$	$W(2,1,1)$	$W(2,2,1)$
A_4	$(2,2)$	$W(2,1,2)$	$W(2,2,2)$

$$\begin{pmatrix} -2 & 1 \\ 4 & -4 \\ 3 & -3 \\ 0 & 5 \end{pmatrix}.$$

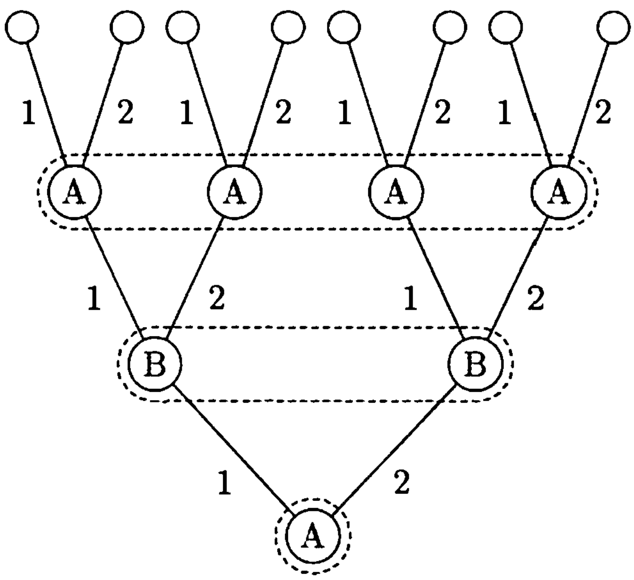


Рис. 6

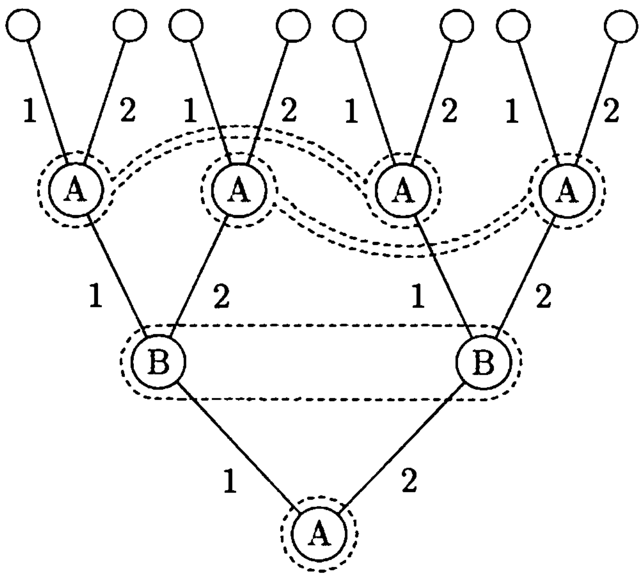


Рис. 7

Оптимальные смешанные стратегии игроков и цена игры соответственно равны:

$$\mathrm{P} = \left\{0, \frac{9}{13}, 0, \frac{4}{13}\right\}, \quad \mathrm{Q} = \left\{\frac{5}{13}, \frac{8}{13}\right\}, \quad \nu = \frac{20}{13}.$$

В следующем примере информационые множества выглядят немного иначе.

Пример 7. *1-й ход* делает игрок A: он выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход делает игрок B: не зная о выборе игрока A на 1-м ходе, он выбирает число y из множества двух чисел $\{1, 2\}$.

3-й ход делает игрок A: он выбирает число z из множества двух чисел $\{1, 2\}$, зная выбор y игрока B на 2-м ходе, но не помня собственного выбора x на 1-м ходе.

После этого игроки расплачиваются по правилу, указанному в примере 5.

Графическое представление этой игры показано на рис. 7.

Поскольку игроку B неизвестен выбор игрока A на 1-м ходе, то, выполняя свой ход, он не знает, в какой именно из двух возможных позиций он находится. Поэтому у игрока B всего две стратегии:

B_1 — выбрать $y = 1$, B_2 — выбрать $y = 2$.

При описании стратегий игрока A нужно исходить из того, что к 3-му ходу игрок A утратил сведения о собственном выборе на 1-м ходе, но ему известен выбор игрока B на 2-м ходе. Поэтому выбор числа z игроку A следует связать с известным ему к 3-му ходу значением y. Удобнее всего это сделать подобно тому, как были рассчитаны стратегии игрока B в примерах 3 и 5, т. е. при помощи упорядоченной пары

$$[z_1, z_2].$$

Здесь z_1 ($z_1 = 1, 2$) — альтернатива, выбираемая игроком A при условии, что игрок B выбрал первую альтернативу, $y = 1$, а z_2 ($z_2 = 1, 2$) — альтернатива, выбираемая игроком A при условии, что игрок B выбрал вторую альтернативу, $y = 2$.

Чистую стратегию игрока A в данной игре можно записать так:

$$(x, [z_1, z_2]).$$

Здесь x ($x = 1, 2$) — альтернатива, которую игрок A выбирает на 1-м ходе, z_1 ($z_1 = 1, 2$) — альтернатива, которую игрок A выбирает на 3-м ходе, если на 2-м ходе игрок B выбрал первую альтернативу ($y = 1$) и z_2 ($z_2 = 1, 2$) — альтернатива, которую игрок A выбирает на 3-м ходе, если на 2-м ходе игрок B выбрал вторую альтернативу ($y = 2$).

Например, выбор игроком A стратегии $(2, [2, 1])$ означает, что на 1-м ходе игрок A выбирает $x = 2$, а на 3-м — $z = 2$, если игрок B выбрал $y = 1$, и $z = 1$, если игрок B выбрал $y = 2$.

Тем самым у игрока A восемь чистых стратегий:

$$A_1 - (1, [1, 1]),\ A_2 - (1, [1, 2]),\ A_3 - (1, [2, 1]),\ A_4 - (1, [2, 2]),$$

$$A_5 - (2, [1, 1]),\ A_6 - (2, [1, 2]),\ A_7 - (2, [2, 1]),\ A_8 - (2, [2, 2]).$$

Покажем теперь, как в зависимости от применяемых стратегий определяются элементы таблицы выигрышей игрока A.

Пусть, например, игрок A выбрал стратегию $A_3 - (1, [2, 1])$, а игрок B — стратегию $B_2 - (2)$. Тогда $x = 1$, $y = 2$, а из $[2, 1]$ вытекает, что $z = 1$. Отсюда

$$W(x, y, z) = W(1, 2, 1) = 1.$$

По этой же схеме вычисляются и остальные элементы таблицы.

В результате получаем:

		B_1	B_2
		(1)	(2)
A_1	(1, [1, 1])	$W(1,1,1)$	$W(1,2,1)$
A_2	(1, [1, 2])	$W(1,1,1)$	$W(1,2,2)$
A_3	(1, [2, 1])	$W(1,1,2)$	$W(1,2,1)$
A_4	(1, [2, 2])	$W(1,1,2)$	$W(1,2,2)$
A_5	(2, [1, 1])	$W(2,1,1)$	$W(2,2,1)$
A_6	(2, [1, 2])	$W(2,1,1)$	$W(2,2,2)$
A_7	(2, [2, 1])	$W(2,1,2)$	$W(2,2,1)$
A_8	(2, [2, 2])	$W(2,1,2)$	$W(2,2,2)$

$$\begin{pmatrix} -2 & 1 \\ -2 & -4 \\ 4 & 1 \\ 4 & -4 \\ 3 & -3 \\ 3 & 5 \\ 0 & -3 \\ 0 & 5 \end{pmatrix}.$$

Оптимальные смешанные стратегии игроков и цена игры соответственно равны:

$$\mathrm{P} = \left\{0, 0, \frac{2}{5}, 0, 0, \frac{3}{5}, 0, 0\right\}, \quad \mathrm{Q} = \left\{\frac{4}{5}, \frac{1}{5}, \right\}, \quad \nu = \frac{17}{5}.$$

Рассмотрим позиционную игру со случайным ходом.

Пример 8. *1-й ход* производится случайно: игрок O выбирает число x, равное 1 с вероятностью 0,5 и равное 2 с такой же вероятностью.

2-й ход делает игрок A: он выбирает число y из множества двух чисел $\{1, 2\}$, не зная результатов случайного выбора на 1-м ходе.

3-й ход делает игрок B: он выбирает число z из множества двух чисел $\{1, 2\}$, зная о том, какое именно число x случайно выбрано игроком O на 1-м ходе, и не зная выбора y игрока A на 2-м ходе.

После этого игроки расплачиваются, используя функцию $W(x, y, z)$, ту же, что и в предыдущих примерах.

Графическое представление этой игры показано на рис. 8.

Опишем стратегии игроков.

Поскольку игроку A исход случайного испытания неизвестен, то он имеет всего две стратегии:

A_1 – (1), A_2 – (2).

При построении своих стратегий игроку B естественно воспользоваться имеющейся у него информацией о результатах 1-го хода. Это позволит ему описать свою стратегию упорядоченной парой

$$[z_1, z_2].$$

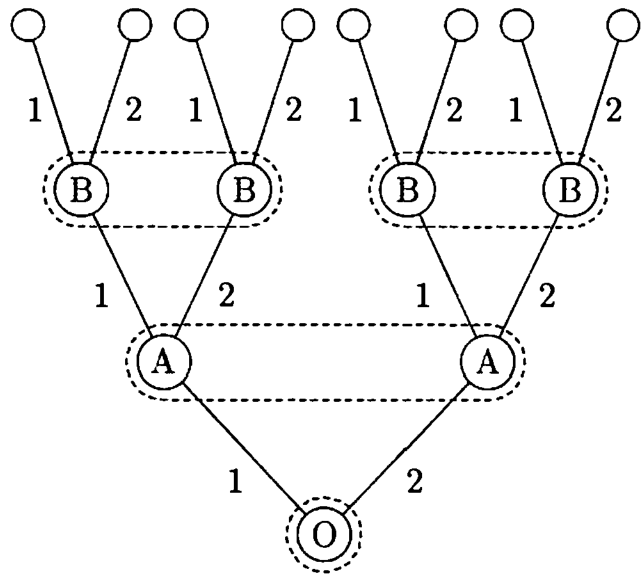


Рис. 8

Здесь z_1 ($z_1 = 1, 2$) — альтернатива, выбираемая игроком B при условии, что $x = 1$, а z_2 ($z_2 = 1, 2$) — альтернатива, выбираемая игроком B при условии, что $x = 2$.

Тем самым у игрока B четыре стратегии:

B_1 – $[1, 1]$, B_2 – $[1, 2]$, B_3 – $[2, 1]$, B_4 – $[2, 2]$.

Покажем теперь, как определяются элементы таблицы выигрышей игрока A.

Пусть, например, игрок A выбрал стратегию A_1 – (1), а игрок B — стратегию B_3 – $[2, 1]$.

Различаются два случая:

1) $x = 1$,

2) $x = 2$.

Если $x = 1$, то стратегия B_3 указывает игроку B его выбор $z = 2$. А так как $y = 1$, то в результате имеем

$$W(x, y, z) = W(1, 1, 2) = 4.$$

Если $x = 2$, то стратегия B_3 указывает игроку B его выбор $z = 1$. А так как $y = 1$, то в результате имеем

$$W(x, y, z) = W(2, 1, 1) = 3.$$

Поскольку первая и вторая альтернативы на 1-м ходе выбираются с вероятностями 0,5 и 0,5, то и вышеуказанные выигрыши появляются с теми же вероятностями и, следовательно, средний выигрыш (математическое ожидание) игрока A при этих стратегиях определяется так:

$$4 \cdot 0,5 + 3 \cdot 0,5 = 3,5.$$

Аналогичным образом рассчитывая остальные средние выигрыши, получим

при $x = 1$:

		B_1	B_2	B_3	B_4
		$[1,1]$	$[1,2]$	$[2,1]$	$[2,2]$
A_1	(1)	$W(1,1,1)$	$W(1,1,1)$	$W(1,1,2)$	$W(1,1,2)$
A_2	(2)	$W(1,2,1)$	$W(1,2,1)$	$W(1,2,2)$	$W(1,2,2)$

или

$$\begin{pmatrix} -2 & -2 & 4 & 4 \\ 1 & 1 & -4 & -4 \end{pmatrix},$$

при $x = 2$:

		B_1	B_2	B_3	B_4
		$[1,1]$	$[1,2]$	$[2,1]$	$[2,2]$
A_1	(1)	$W(2,1,1)$	$W(2,1,2)$	$W(2,1,1)$	$W(2,1,2)$
A_2	(2)	$W(2,2,1)$	$W(2,2,2)$	$W(2,2,1)$	$W(2,2,2)$

или

$$\begin{pmatrix} 3 & 0 & 3 & 0 \\ -3 & 5 & -3 & 5 \end{pmatrix}.$$

Отсюда получаем искомую матрицу игры:

$$\begin{pmatrix} 0{,}5 & -1 & 3{,}5 & 2 \\ -1 & 3 & -3{,}5 & 0{,}5 \end{pmatrix}.$$

Наконец, рассмотрим пример позиционной игры со случайным разыгрыванием права первого хода.

Пример 9. *1-й ход* делает игрок O, выбирая число x, равное 1 с вероятностью $2/3$ и равное 2 с вероятностью $1/3$.

Если $x = 1$, то на *2-м ходе* игрок A выбирает число y из множества двух чисел $\{1, 2\}$, зная результат случайного выбора на 1-м ходе, а на *3-м ходе* игрок B выбирает число z из множества двух чисел $\{1, 2\}$, зная x, но не зная y.

Если $x = 2$, то на *2-м ходе* игрок B выбирает число y из множества двух чисел $\{1, 2\}$, зная результат случайного выбора на 1-м ходе, а на *3-м ходе* игрок A выбирает число z из множества двух чисел $\{1, 2\}$, зная x, но не зная y.

После этого игроки расплачиваются, используя функцию $W(x, y, z)$, ту же, что и в предыдущих примерах.

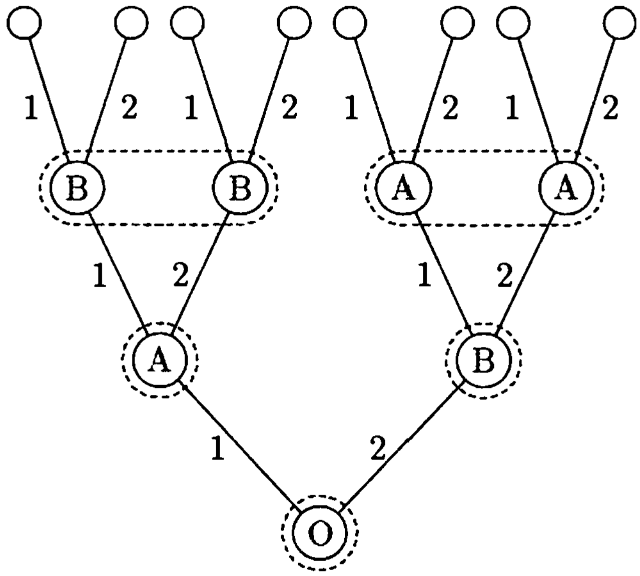


Рис. 9

Графическое представление этой игры показано на рис. 9.

Чистую стратегию игрока A в данной игре можно описать упорядоченной парой

$$|y, z|,$$

где y $(y = 1, 2)$ — выбор игрока A на 2-м ходе, если на 1-м ходе выбрано $x = 1$, а z $(z = 1, 2)$ — выбор игрока A на 3-м ходе, если на 1-м ходе выбрано $x = 2$.

Например, стратегия $|1, 2|$ означает, что на 2-м ходе игрок A выбирает $y = 1$, а на 3-м ходе — $z = 2$.

Тем самым у игрока A четыре стратегии:

$$A_1 - |1, 1|,\ A_2 - |1, 2|,\ A_3 - |2, 1|,\ A_4 - |2, 2|.$$

У игрока B те же четыре стратегии:

$$B_1 - |1, 1|,\ B_2 - |1, 2|,\ B_3 - |2, 1|,\ B_4 - |2, 2|.$$

Покажем теперь, как находятся элементы матрицы выигрышей игрока A.

Пусть, например, игрок A применяет стратегию $A_2 - |1,2|$, а игрок B — стратегию $B_3 - |2,1|$.

Различаются два случая:

1) $x = 1$,

2) $x = 2$.

По условию при $x = 1$ игрок A имеет возможность сделать только 2-й ход (выбрать y), а игрок B — только 3-й (выбрать z). При $x = 2$ их возможности меняются местами: игроку B предоставлено право 2-го хода (выбор y), а игроку A — 3-го (выбор z).

Если $x = 1$, то стратегия A_2 указывает игроку A при 2-м ходе выбор $y = 1$, а стратегия B_3 указывает игроку B при 3-м ходе выбор $z = 1$. В результате

$$W(x, y, z) = W(1, 1, 1) = -2.$$

Если $x = 2$, то стратегия B_3 указывает игроку B при 2-м ходе выбор $y = 2$, а стратегия A_2 указывает игроку A при 3-м ходе выбор $z = 2$. В результате

$$W(x, y, z) = W(2, 2, 2) = 5.$$

Поскольку первая и вторая альтернативы на 1-м ходе выбираются соответственно с вероятностями $2/3$ и $1/3$, то и найденные выигрыши появляются с теми же вероятностями. Следоватеьно, математическое ожидание выигрыша игрока A при таких стратегиях рассчитывается так:

$$-2 \cdot \frac{2}{3} + 5 \cdot \frac{1}{3} = \frac{1}{3}.$$

Итак,

при $x = 1$:

		B_1	B_2	B_3	B_4
		$\|1,1\|$	$\|1,2\|$	$\|2,1\|$	$\|2,2\|$
A_1	$\|1,1\|$	$W(1,1,1)$	$W(1,1,2)$	$W(1,1,1)$	$W(1,1,2)$
A_2	$\|1,2\|$	$W(1,1,1)$	$W(1,1,2)$	$W(1,1,1)$	$W(1,1,2)$
A_3	$\|2,1\|$	$W(1,2,1)$	$W(1,2,2)$	$W(1,2,1)$	$W(1,2,2)$
A_4	$\|2,2\|$	$W(1,2,1)$	$W(1,2,2)$	$W(1,2,1)$	$W(1,2,2)$

или

$$\begin{pmatrix} -2 & 4 & -2 & 4 \\ -2 & 4 & -2 & 4 \\ 1 & -4 & 1 & -4 \\ 1 & -4 & 1 & -4 \end{pmatrix},$$

при $x = 2$:

		B_1	B_2	B_3	B_4
		$\|1,1\|$	$\|1,2\|$	$\|2,1\|$	$\|2,2\|$
A_1	$\|1,1\|$	$W(2,1,1)$	$W(2,1,1)$	$W(2,2,1)$	$W(2,2,1)$
A_2	$\|1,2\|$	$W(2,1,2)$	$W(2,1,2)$	$W(2,2,2)$	$W(2,2,2)$
A_3	$\|2,1\|$	$W(2,1,1)$	$W(2,1,1)$	$W(2,2,1)$	$W(2,2,1)$
A_4	$\|2,2\|$	$W(2,1,2)$	$W(2,1,2)$	$W(2,2,2)$	$W(2,2,2)$

или

$$\begin{pmatrix} 3 & 3 & -3 & -3 \\ 0 & 0 & 5 & 5 \\ 3 & 3 & -3 & -3 \\ 0 & 0 & 5 & 5 \end{pmatrix}.$$

Отсюда получаем искомую матрицу игры:

$$\frac{1}{3}\begin{pmatrix} -1 & 11 & -7 & 5 \\ -4 & 8 & 1 & 13 \\ 5 & -5 & -1 & -11 \\ 2 & -8 & 7 & -3 \end{pmatrix}.$$

Замечание. Графическое представление и функция выигрышей полностью определяют позиционную игру. В рассмотренных выше примерах 6–9 мы пользовались одной и той же функцией и одним и тем же деревом. Отличие было только в маркировке вершин дерева и информационных множествах. При построении последних необходимо соблюдать два правила:

1) в одно информационное множество могут входить позиции только одного игрока,

2) цепь, определяющая партию игры, может иметь с информационным множеством не более одной общей позиции.

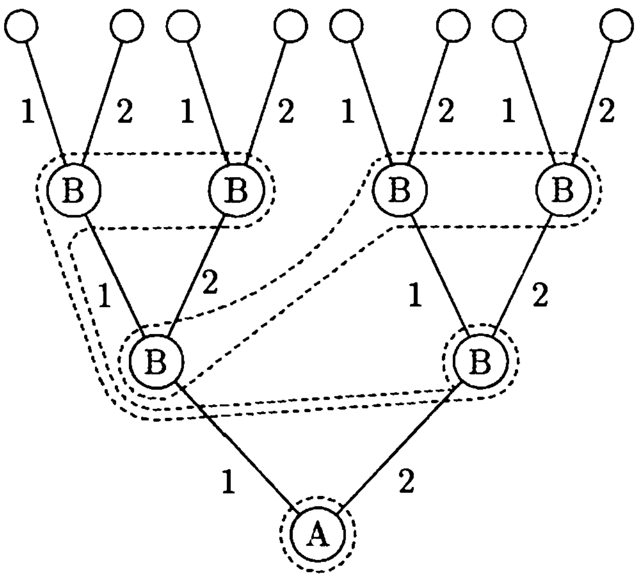


Рис. 10

Как показывает рис. 10, и при таких ограничениях информационные множества могут выглядеть довольно необычно.

18.3. Позиционные игры с полной информацией

Позиционная игра называется *игрой с полной информацией*, если в каждой позиции любой ее партии игрок, делающий ход, знает, какие альтернативы были выбраны на предыдущих ходах. В графическом описании каждая вершина дерева такой игры представляет собой отдельное информационное множество.

Примерами позиционных игр с полной информацией могут служить крестики-нолики, шашки и шахматы.

Основная особенность позиционной игры с полной информацией состоит в том, что соответствующая ей матрица выигрышей всегда имеет седловую точку, т. е. в игре с полной информацией существуют оптимальные чистые стратегии и, значит, равновесная ситуация.

Сказанное означает, что в шахматах (крестиках-ноликах, шашках) уже в начальной позиции имеется способ выигрыша либо у белых, либо у черных, либо как та, так и другая сторона способна форсировать ничью.

Однако известное доказательство существования равновесной ситуации неконструктивно и не дает эффективных приемов фактического нахождения решения игры.

И такие способы (стратегии) в шахматах не найдены до сих пор, и даже неизвестно, какая из перечисленных возможностей имеет место на самом деле.

Иное дело с игрой крестики-нолики: стратегий в ней немного и она разобрана до самого конца — существуют оптимальные чистые стратегии, ведущие игроков к ничьей.

Рассмотрим несколько примеров.

1. Как нетрудно заметить, двухходовая игра из примера 1 является игрой с полной информацией. Ее нормализация приводит к матрице с седловой точкой (см. пример 3).

2. *Выкладывание монет на стол.* Два игрока поочередно кладут монеты одинаковых размеров на обыкновенный стол, всякий раз выбирая произвольное доступное место для монеты (взаимное накрывание монет не допускается). Тот из игроков, кто положит монету, не оставляющую места для новых монет, выигрывает.

Это игра с полной информацией. Существует вполне определенная стратегия, обеспечивающая выигрыш тому из игроков, кто начинает игру. А именно, начинающий игру должен положить первую монету точно в центр стола и на каждый ход противника отвечать симметричным ходом. Исход игры от стратегии второго игрока не зависит.

3. *Переговоры.* В переговорах участвуют две стороны: A и B. В слегка идеализированном варианте это может выглядеть, например, так.

Сначала сторона A высказывает одно из нескольких предложений, способных заинтересовать сторону B. Затем сторона B, ознакомившись с предложением стороны A, высказывает одно из нескольких встречных предложений, способных, по ее мнению, заинтересовать сторону A. В свою очередь, сторона A, ознакомившись с реакцией стороны B на сделанные предложения, высказывает ей новое предложение, внеся одну из нескольких возможных корректировок в свое первоначальное предложение с учетом мнения стороны B, и т. д.

Если предмет переговоров сложен, то подобный обмен ходами может затянуться. Однако любые переговоры непременно заканчиваются. И там, на финише, ждет функция выигрышей.

Попробуем смоделировать короткий переговорный процесс трехходовой позиционной игрой.

Предположим, что переговоры заканчиваются через три хода, на каждом из которых соответствующая сторона имеет возможность выбора из двух альтернатив, и опишем соответствующую позиционную игру.

1-й ход делает сторона A: она выбирает одно из двух возможных предложений — число x из множества двух чисел $\{1, 2\}$.

2-й ход делает сторона B: она выбирает число y из множества двух чисел $\{1, 2\}$, зная число x, предложенное стороной A.

3-й ход делает сторона A: она выбирает число z из множества двух чисел $\{1, 2\}$, зная о предложении стороны B на 2-м ходе и помня собственное предложение на 1-м ходе.

После этого сторона A либо получает вознаграждение (например, в виде кредита от стороны B), либо выплачивает стороне B штраф.

Все эти возможности описываются функцией выигрышей $W(x, y, z)$, заданной следующей таблицей:

$$
\begin{array}{ll}
W(1,1,1) = a, & W(2,1,1) = e, \\
W(1,1,2) = b, & W(2,1,2) = f, \\
W(1,2,1) = c, & W(2,2,1) = g, \\
W(1,2,2) = d, & W(2,2,2) = h.
\end{array}
$$

Графическое представление этой игры показано на рис. 11.

Ясно, что описанная позиционная игра является игрой с полной информацией.

Начнем с описания возможных стратегий игрока B.

Поскольку игроку B выбор игрока A на 1-м ходе известен, то у игрока B те же четыре стратегии, что и в примерах 3 и 5:

B_1 - [1, 1], B_2 - [1, 2], B_3 - [2, 1], B_4 - [2, 2].

С описанием возможных стратегий игрока A дело обстоит немного посложнее — их восемь.

Чистая стратегия игрока A в данной игре описывается упорядоченной тройкой

$$(x, [z_1, z_2]).$$

Здесь x ($x = 1, 2$) — альтернатива, которую игрок A выбирает на 1-м ходе, z_1 ($z_1 = 1, 2$) — альтернатива, которую игрок A выбирает на 3-м ходе, если на 2-м ходе игрок B выбрал первую альтернативу ($y = 1$), и z_2 ($z_2 = 1, 2$) — альтернатива, которую игрок A выбирает на 3-м ходе, если на 2-м ходе игрок B выбрал вторую альтернативу ($y = 2$).

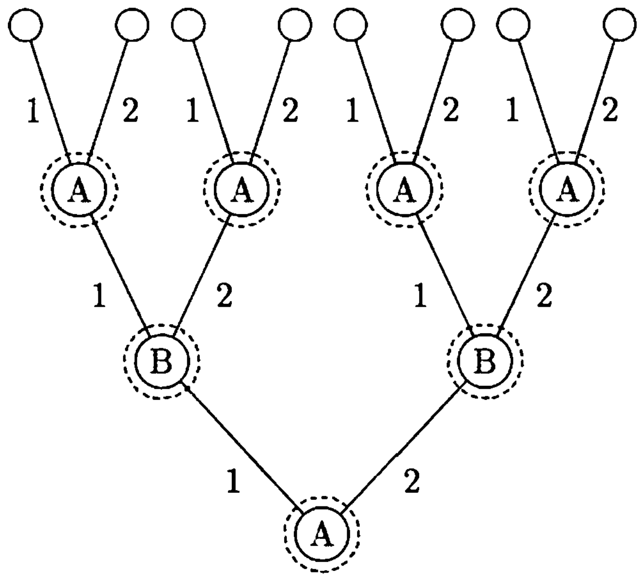


Рис. 11

Например, выбор игроком A стратегии $(1, [2, 1])$ означает, что на 1-м ходе игрок A выбирает $x = 1$, а на 3-м — $z = 2$, если игрок B выбрал $y = 1$, и $z = 1$, если игрок B выбрал $y = 2$.

Тем самым у игрока A восемь чистых стратегий:

$$A_1 - (1, [1, 1]),\ A_2 - (1, [1, 2]),\ A_3 - (1, [2, 1]),\ A_4 - (1, [2, 2]),$$

$$A_5 - (2, [1, 1]),\ A_6 - (2, [1, 2]),\ A_7 - (2, [2, 1]),\ A_8 - (2, [2, 2]).$$

Покажем теперь, как в зависимости от применяемых стратегий определяются элементы таблицы выигрышей игрока A.

Пусть, например, игрок A выбрал стратегию $A_6 - (2, [1, 2])$, а игрок B — стратегию $B_3 - [2, 1]$. Тогда $x = 2$. Из $[2, 1]$ вытекает, что $y = 1$, а из $(2, [1, 2])$ — что $z = 1$. Отсюда

$$W(x, y, z) = W(2, 1, 1) = e.$$

Рассчитывая по этой же схеме все остальные элементы таблицы выигрышей, в итоге получим таблицу

		B_1	B_2	B_3	B_4
		$[1,1]$	$[1,2]$	$[2,1]$	$[2,2]$
A_1	$(1,[1,1])$	$W(1,1,1)$	$W(1,1,1)$	$W(1,2,1)$	$W(1,2,1)$
A_2	$(1,[1,2])$	$W(1,1,1)$	$W(1,1,1)$	$W(1,2,2)$	$W(1,2,2)$
A_3	$(1,[2,1])$	$W(1,1,2)$	$W(1,1,2)$	$W(1,2,1)$	$W(1,2,1)$
A_4	$(1,[2,2])$	$W(1,1,2)$	$W(1,1,2)$	$W(1,2,2)$	$W(1,2,2)$
A_5	$(2,[1,1])$	$W(2,1,1)$	$W(2,2,1)$	$W(2,1,1)$	$W(2,2,1)$
A_6	$(2,[1,2])$	$W(2,1,1)$	$W(2,2,2)$	$W(2,1,1)$	$W(2,2,2)$
A_7	$(2,[2,1])$	$W(2,1,2)$	$W(2,2,1)$	$W(2,1,2)$	$W(2,2,1)$
A_8	$(2,[2,2])$	$W(2,1,2)$	$W(2,2,2)$	$W(2,1,2)$	$W(2,2,2)$

и соответствующую матрицу

$$\begin{pmatrix} a & a & c & c \\ a & a & d & d \\ b & b & c & c \\ b & b & d & d \\ e & g & e & g \\ e & h & e & h \\ f & g & f & g \\ f & h & f & h \end{pmatrix}$$

Вследствие того что рассматриваемая позиционная игра является игрой с полной информацией, полученная матрица имеет седловую точку при любой функции выигрышей. В этом легко убедиться, произвольно выбирая значения параметров a, b, c, d, e, f, g и h.

Несколько слов в заключение.

В рассмотренных примерах основное внимание было уделено описанию процесса нормализации позиционной игры — построению дерева игры и информационных множеств, выработке стратегий игроков и вычислению элементов платежной матрицы.

Следующий естественный шаг — отыскание цены игры и оптимальных стратегий игроков — проводится методами, о которых рассказывается в главе “Матричные игры”.

18.4. Задания

Дайте графическое представление и приведите к нормальной форме позиционную игру с функцией выигрышей $W(x, y, z)$:

$$\begin{array}{ll} W(1,1,1) = 2, & W(2,1,1) = -1, \\ W(1,1,2) = -2, & W(2,1,2) = 3, \\ W(1,2,1) = 1, & W(2,2,1) = 0, \\ W(1,2,2) = 0, & W(2,2,2) = -3. \end{array}$$

а)

1-й ход делает игрок A: он выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход делает игрок B: не зная о выборе игрока A на 1-м ходе, он выбирает число y из множества двух чисел $\{1, 2\}$.

3-й ход делает игрок A: он выбирает число z из множества двух чисел $\{1, 2\}$, зная значение y, выбранное игроком B на 2-м ходе, но не помня собственного выбора x на 1-м ходе.

б)

1-й ход делает игрок A: он выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход делает игрок B: зная выбор игрока A на 1-м ходе, он выбирает число y из множества двух чисел $\{1, 2\}$.

3-й ход делает игрок A: он выбирает число z из множества двух чисел $\{1, 2\}$, не зная значения y, выбранного игроком B на 2-м ходе, но помня собственный выбор x на 1-м ходе.

в)

1-й ход производится случайно: игрок O выбирает число x, равное 1 с вероятностью 0,3 и равное 2 с вероятностью 0,7.

2-й ход делает игрок A: он выбирает число y из множества двух чисел $\{1, 2\}$, зная результат случайного выбора на 1-м ходе.

3-й ход делает игрок B: он выбирает число z из множества двух чисел $\{1, 2\}$, зная выбор y игрока A на 2-м ходе, но не зная случайного выбора x на 1-м ходе.

Глава 19
БИМАТРИЧНЫЕ ИГРЫ

Предыдущие рассмотрения касались игр двух лиц, в которых интересы игроков были прямо противоположны (антагонистические, или матричные, игры), а также позиционных игр, сводимых к матричным. Однако ситуации, в которых интересы игроков хотя и не совпадают, но уже необязательно являются противоположными, встречаются значительно чаще.

Рассмотрим, например, конфликтную ситуацию, в которой каждый из двух участников имеет следующие возможности для выбора своей линии поведения:

игрок A — может выбрать любую из стратегий $A_1, \ldots, A_m$,
игрок B — любую из стратегий $B_1, \ldots, B_n$.

При этом всякий раз их совместный выбор оценивается вполне определенно:

если игрок A выбрал i-ю стратегию A_i, а игрок B — k-ю стратегию B_k, то в итоге выигрыш игрока A будет равен некоторому числу a_{ik}, а выигрыш игрока B — некоторому, вообще говоря, другому числу b_{ik}.

Иными словами, всякий раз каждый из игроков получает свой приз.

Последовательно перебирая все стратегии игрока A и все стратегии игрока B, мы можем заполнить их выигрышами две таблицы:

	B_1	$\ldots$	B_k	$\ldots$	B_n
A_1	a_{11}	$\ldots$	a_{1k}	$\ldots$	a_{1n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
A_i	a_{i1}	$\ldots$	a_{ik}	$\ldots$	a_{in}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
A_m	a_{m1}	$\ldots$	a_{mk}	$\ldots$	a_{mn}

	B_1	$\ldots$	B_k	$\ldots$	B_n
A_1	b_{11}	$\ldots$	b_{1k}	$\ldots$	b_{1n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
A_i	b_{i1}	$\ldots$	b_{ik}	$\ldots$	b_{in}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
A_m	b_{m1}	$\ldots$	b_{mk}	$\ldots$	b_{mn}

Первая из таблиц описывает выигрыш игрока A, а вторая — выигрыш игрока B. Обычно эти таблицы записывают в виде матриц:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{1k} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{i1} & \ldots & a_{ik} & \ldots & a_{in} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mk} & \ldots & a_{mn} \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} b_{11} & \ldots & b_{1k} & \ldots & b_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ b_{i1} & \ldots & b_{ik} & \ldots & b_{in} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ b_{m1} & \ldots & b_{mk} & \ldots & b_{mn} \end{pmatrix}.$$

Здесь $\mathbf{A}$ — *платежная матрица игрока* A, а $\mathbf{B}$ — *платежная матрица игрока* B.

При выборе игроком A i-й стратегии, а игроком B — k-й стратегии их выигрыши находятся в матрицах выплат на пересечении i-х строк и k-х столбцов:

в матрице $\mathbf{A}$ это элемент a_{ik}, а в матрице $\mathbf{B}$ — элемент b_{ik}.

Таким образом, в случае, когда интересы игроков различны (но необязательно противоположны), получаются две платежные матрицы: одна — матрица выплат игроку A, другая — матрица выплат игроку B. Поэтому совершенно естественно звучит название, которое обычно присваивается подобной игре, — *биматричная.*

Замечание. Рассмотренные ранее матричные игры, разумеется, можно отнести и к биматричным, где матрица выплат игроку B *противоположна* матрице выплат игроку A:

$$b_{ik} = -a_{ik}$$

или

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = -\mathbf{A} = \begin{pmatrix} -a_{11} & \ldots & -a_{1n} \\ \ldots & \ldots & \ldots \\ -a_{m1} & \ldots & -a_{mn} \end{pmatrix}.$$

Тем не менее в общем случае биматричная игра — это игра с ненулевой суммой.

Нам кажется вполне естественным время от времени сопоставлять наши рассмотрения с рассуждениями, приведенными ранее для матричных игр (особенно при попытках разрешения сходных проблем). Подобные сопоставления часто оказываются одновременно и удобными и полезными. Конечно, класс биматричных игр значительно шире класса матричных (разнообразие новых моделируемых конфликтных ситуаций весьма заметно), а значит, неизбежно увеличиваются и трудности, встающие на пути их успешного разрешения. Впрочем, мы надеемся, что часть этих трудностей мы сумеем преодолеть уже в настоящем издании.

19.1. Примеры биматричных игр

Примеры этого раздела описывают некоторые типические конфликтные ситуации, приводящие к биматричным играм. Сначала мы обсудим вопросы, связанные с формализацией рассматриваемых конфликтов (построение платежных матриц), а позднее — с рекомендациями по их разрешению.

19.1.1. Борьба за рынки

Небольшая фирма (игрок A) намерена сбыть партию товара на одном из двух рынков, контролируемых другой, более крупной фирмой (игрок B). Для этого фирма A готова сделать на одном из рынков соответствующие приготовления (например, развернуть рекламную кампанию). Господствующая на рынках фирма B может попытаться воспрепятствовать этому, приняв на одном из рынков предупредительные меры (разумеется, в рамках закона). Не встречая противодействия на рынке, фирма A захватывает его; при наличии препятствий — терпит поражение.

Будем считать для определенности, что проникновение фирмы A на первый рынок более выгодно для нее, нежели на второй. Естественно также считать, что и борьба за первый рынок потребует вложения бо́льших средств. Например, победа фирмы A на первом рынке принесет ей вдвое больший выигрыш, чем победа на втором, но зато и поражение при попытке освоиться на первом рынке полностью ее разорит, а фирму B избавит от конкурента.

Что же касается второго рынка, то при поражении фирмы A ее потери будут не столь разорительны, но и победа принесет не много.

Таким образом, у фирмы A две стратегии:

A_1 — выбор первого рынка, A_2 — выбор второго рынка.

Такие же стратегии и у фирмы B:

B_1 — выбор первого рынка, B_2 — выбор второго рынка.

Для того чтобы составить платежные матрицы игроков, нужны расчетные количественные показатели, которые мы приведем здесь в условных единицах:

$$\mathbf{A} = \begin{pmatrix} -10 & 2 \\ 1 & -1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 5 & -2 \\ -1 & 1 \end{pmatrix}.$$

Посмотрим на выписанные матрицы выплат. Из сказанного выше ясно, что если оба игрока выберут один и тот же рынок, то победа останется за более сильной фирмой B.

То, что в ситуации (A_1, B_1) выигрыш игрока B равен 5, а в ситуации (A_2, B_2) — 1, подчеркивает, что первый рынок более выгоден (удобно расположен, хорошо посещаем и т. п.), чем второй. Выигрыш (-10) игрока A в ситуации (A_1, B_1) (а точнее, проигрыш) в сопоставлении с его выигрышем (-1) в ситуации (A_2, B_2) выглядит, разумеется, вполне сокрушительно. Что же касается ситуации, когда фирмы уделяют основное внимание разным рынкам — (A_1, B_2) и (A_2, B_1), то здесь фирму A ждет настоящий выигрыш, бóльший на более выгодном рынке. Потери, которые при этом несет фирма B, оказываются прямо противоположными.

Замечание. Ясно, что точно рассчитать выгоду и ущерб сторон в этом конфликте заранее довольно трудно. Зато в следующей конфликтной ситуации размеры выигрышей игроков известны им со всей определенностью.

19.1.2. Дилемма узников

Игроками являются два узника, находящиеся в предварительном заключении по подозрению в совершении преступления. При отсутствии прямых улик возможность их осуждения в большой степени зависит от того, заговорят они или будут молчать.

Если оба будут молчать, то наказанием будет лишь срок предварительного заключения (потери каждого из узников составят (-1)). Если сознаются, то получат срок, учитывающий признание как смягчающее обстоятельство (потери каждого из узников составят в этом случае (−6)). Если же заговорит только один из узников, а другой будет молчать, то в этом случае заговоривший будет выпущен на свободу (его потери равны 0), а сохраняющий молчание получит максимально возможное наказание (его потери будут равны (−9)).

Эта конфликтная ситуация приводит к биматричной игре, в которой каждый из игроков имеет по две стратегии — молчать (М) или говорить (Г).

Выигрыши игроков A и B соответственно описываются так:

	(М)	(Г)
(М)	−1	−9
(Г)	0	−6

	(М)	(Г)
(М)	−1	0
(Г)	−9	−6

19.1.3. Семейный спор

Два партнера договариваются о проведении одного из двух действий, (1) и (2), каждое из которых требует их совместного участия.

В случае осуществления первого из этих двух действий выигрыш первого партнера (игрок A) будет вдвое выше выигрыша второго партнера (игрок B). Напротив, в случае осуществления второго из этих двух действий выигрыш игрока A будет вдвое меньше выигрыша игрока B. Если же партнеры выполнят различные действия, то выигрыш каждого из них будет равен нулю.

Эта конфликтная ситуация приводит к биматричной игре, в которой каждый из игроков имеет по две стратегии. Выигрыши игроков A и B соответственно описываются таблицами следующего вида:

	(1)	(2)
(1)	2	0
(2)	0	1

	(1)	(2)
(1)	1	0
(2)	0	2

Пояснение. Понятно, что различные конфликтные ситуации могут иметь одну и ту же формализацию. В частности, рассмотренная биматричная игра часто интерпретируется как одновременный выбор супругами совместного развлечения: посещение оперного спектакля или хоккейного матча. При этом в посещении оперного театра жена заинтересована в большей степени, чем муж, а при посещении стадиона наблюдается обратная картина. В случае же непреодолимости разногласий, возникших при выборе, день оказывается вообще испорченным.

Отсюда и название, вынесенное в заголовок.

19.1.4. Студент – преподаватель

Рассмотрим следующую ситуацию. Студент (игрок A) готовится к зачету, который принимает преподаватель (игрок B). Можно считать, что у студента две стратегии — подготовиться к сдаче зачета (+) и не подготовиться (−). У преподавателя также две стратегии — поставить зачет [+] и не поставить зачета [−].

В основу значений функций выигрыша игроков положим следующие соображения:

Выигрыш студента

	[+]	[−]
(+)	Оценка заслужена	Очень обидно
(−)	Удалось обмануть	Оценка заслужена

Выигрыш преподавателя

	[+]	[−]
(+)	Все нормально	Был неправ
(−)	Дал себя обмануть	Опять придет

Количественно это можно выразить, например, так:

$$\begin{array}{c|cc} & [+] & [-] \\ \hline (+) & 2 & -1 \\ (-) & 1 & 0 \end{array} \qquad \begin{array}{c|cc} & [+] & [-] \\ \hline (+) & 1 & -3 \\ (-) & -2 & -1 \end{array}$$

19.2. Смешанные стратегии

В приведенных примерах (позже мы вернемся к подробному рассмотрению каждого) описаны ситуации, в которых интересы игроков не совпадают. Естественно встает вопрос о том, какие рекомендации необходимо дать игрокам для того, чтобы моделируемая конфликтная ситуация разрешилась.

Иными словами, что мы будем понимать под решением биматричной игры?

Попробуем ответить на этот вопрос так: вследствие того что интересы игроков не совпадают, нам нужно построить такое (компромиссное) решение, которое бы в том или ином, но в одинаковом смысле удовлетворяло обоих игроков.

Иначе говоря, попробуем найти некую равновесную ситуацию, явное отклонение от которой уменьшает выигрыш игрока.

Подобный вопрос мы ставили и при рассмотрении матричных игр. Напомним, что возникавшее при разработке минимаксного подхода понятие равновесной ситуации приводило нас к поиску седловой точки, которая, как оказалось, существует далеко не всегда, конечно, если ограничиваться только чистыми стратегиями игроков A и B, т. е. стратегиями

$$A_1, \ldots, A_m, \ B_1, \ldots, B_n.$$

Естественно ожидать, что в более сложном случае биматричной игры дело вряд ли будет обстоять проще.

В матричных играх эта трудность была преодолена путем перехода к смешанным стратегиям, т. е. к такому поведению игроков, при котором они чередуют (чистые) стратегии $A_1, \ldots, A_m$, $B_1, \ldots, B_n$ (конечно, свои собственные):

игрок A — стратегии A_1, ..., A_m, игрок B — стратегии B_1, ..., B_n

с определенными частотами:

игрок A — стратегии A_1, ..., A_m с частотами p_1, ..., p_m, где

$$p_1 \geqslant 0, \ \ldots, \ p_m \geqslant 0, \qquad \sum_{i=1}^{m} p_i = 1,$$

а игрок B — стратегии B_1, ..., B_n с частотами q_1, ..., q_n, где

$$q_1 \geqslant 0, \ \ldots, \ q_n \geqslant 0, \qquad \sum_{k=1}^{n} q_k = 1,$$

и оказалось, что в смешанных стратегиях равновесная ситуация существует всегда.

Иными словами, любая матричная игра в смешанных стратегиях разрешима.

Поэтому, рассматривая здесь биматричные игры, разумно попробовать сразу же перейти к смешанным стратегиям игроков. Тем самым мы предполагаем, что каждая игра может быть повторена в неизменных обстоятельствах многократно.

В матричном случае смешивание стратегий приводило к расширению возможности выплат в том смысле, что расчет строился на вычислении средних выигрышей игроков A и B, которые определялись по элементам платежной матрицы A и вероятностям p_i и q_k:

$$H_A = \sum_{i,k} a_{ik} p_i q_k, \qquad H_B = -\sum_{i,k} a_{ik} p_i q_k.$$

При смешанных стратегиях в биматричных играх также естественно возникают средние выигрыши игроков A и B, вычисляемые по правилам, в которых уже нет никакой дискриминации игрока B:

$$H_A = \sum_{i,k} a_{ik} p_i q_k, \qquad H_B = \sum_{i,k} b_{ik} p_i q_k.$$

19.3. 2×2-биматричные игры. Ситуация равновесия

Мы предполагаем уделить основное внимание случаю, когда у каждого из игроков имеется ровно две стратегии, т. е. случаю $m = n = 2$. Поэтому нам кажется уместным выписать приведенные выше формулы именно для такого случая.

В 2×2-биматричной игре платежные матрицы игроков имеют следующий вид:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

вероятности

$$p_1 = p, \quad p_2 = 1 - p, \quad q_1 = q, \quad q_2 = 1 - q,$$

а средние выигрыши вычисляются по формулам

$$H_A(p,q) = a_{11}pq + a_{12}p(1-q) + a_{21}(1-p)q + a_{22}(1-p)(1-q),$$
$$H_B(p,q) = b_{11}pq + b_{12}p(1-q) + b_{21}(1-p)q + b_{22}(1-p)(1-q),$$

где

$$0 \leqslant p \leqslant 1, \qquad 0 \leqslant q \leqslant 1.$$

Сформулируем основное определение.

Определение. Будем говорить, что пара чисел

$$(p^*, q^*), \qquad 0 \leqslant p^* \leqslant 1, \quad 0 \leqslant q^* \leqslant 1,$$

определяет *равновесную ситуацию*, если для любых p и q, подчиненных условиям $0 \leqslant p \leqslant 1$, $0 \leqslant q \leqslant 1$, одновременно выполнены следующие неравенства:

$$H_A(p, q^*) \leqslant H_A(p^*, q^*), \qquad H_B(p^*, q) \leqslant H_B(p^*, q^*). \tag{1}$$

Пояснение. Выписанные неравенства (1) означают следующее: ситуация, определяемая смешанной стратегией (p^*, q^*), является равновесной, если отклонение от нее одного из игроков при условии, что другой сохраняет свой выбор, приводит к уменьшению выигрыша первого. Тем самым получается, что если равновесная ситуация существует, то отклонение от нее невыгодно самому игроку.

Но может ли быть подобная ситуация равновесия в биматричной игре?

Ответ на поставленный вопрос дает следующее утверждение.

ТЕОРЕМА (Дж. Нэш). *Всякая биматричная игра имеет хотя бы одну равновесную ситуацию (точку равновесия) в смешанных стратегиях.*

Заметим, что весьма похожее утверждение мы уже встречали при изучении вопроса разрешимости матричных игр.

Напомним также, что в соответствующем разделе после формулировки аналогичного утверждения о существовании равновесной ситуации описывался и способ отыскания точки равновесия.

Та же проблема встает перед нами и здесь.

Итак, равновесная ситуация существует. Но как ее найти?

Если предполагается, что некоторая пара чисел (p^*, q^*) определяет ситуацию равновесия, то для того, чтобы подтвердить или опровергнуть это предположение, необходимо проверить справедливость неравенств (1) для любого p в пределах от 0 до 1 и для любого q в пределах от 0 до 1.

В общем случае число таких проверок бесконечно. И следовательно, действенный способ определения равновесной ситуации должен быть иным.

Для его обоснования сошлемся на следующий теоретический результат.

ТЕОРЕМА. *Выполнение неравенств*

$$H_A(p, q^*) \leqslant H_A(p^*, q^*), \qquad H_B(p^*, q) \leqslant H_B(p^*, q^*)$$

равносильно выполнению неравенств

$$\begin{aligned} H_A(0, q^*) \leqslant H_A(p^*, q^*), \quad H_B(p^*, 0) \leqslant H_B(p^*, q^*), \\ H_A(1, q^*) \leqslant H_A(p^*, q^*), \quad H_B(p^*, 1) \leqslant H_B(p^*, q^*). \end{aligned} \tag{2}$$

Иными словами, для того чтобы убедиться, что пара (p^*, q^*) определяет равновесную ситуацию, достаточно проверить справедливость неравенства

$$H_A(p, q^*) \leqslant H_A(p^*, q^*)$$

только для двух чистых стратегий игрока A ($p = 0$ и $p = 1$) и неравенства

$$H_B(p^*, q) \leqslant H_B(p^*, q^*)$$

только для двух чистых стратегий игрока B ($q = 0$ и $q = 1$).

Четыре неравенства (2) позволяют провести поиск точки равновесия уже вполне конструктивно.

Запишем средние выигрыши игроков A и B в более удобной форме. Имеем:

$$H_A(p, q) = (a_{11} - a_{12} - a_{21} + a_{22})pq + (a_{12} - a_{22})p + (a_{21} - a_{22})q + a_{22},$$
$$H_B(p, q) = (b_{11} - b_{12} - b_{21} + b_{22})pq + (b_{12} - b_{22})p + (b_{21} - b_{22})q + b_{22}.$$

Обратимся к первой из полученных формул. Полагая в ней сначала $p = 1$, а потом $p = 0$, получаем, что

$$H_A(1, q) = (a_{11} - a_{12} - a_{21} + a_{22})q + a_{12} + (a_{21} - a_{22})q,$$
$$H_A(0, q) = (a_{21} - a_{22})q + a_{22}.$$

Рассмотрим разности

$$\begin{aligned} H_A(p, q) - H_A(1, q) ={}& (a_{11} - a_{12} - a_{21} + a_{22})pq + (a_{12} - a_{22})p - \\ & - (a_{11} - a_{12} - a_{21} + a_{22})q + a_{22} - a_{12}, \\ H_A(p, q) - H_A(0, q) ={}& (a_{11} - a_{12} - a_{21} + a_{22})pq + (a_{12} - a_{22})p. \end{aligned}$$

Полагая

$$C = a_{11} - a_{12} - a_{21} + a_{22}, \qquad \alpha = a_{22} - a_{12},$$

получим для них следующие выражения:

$$\begin{aligned} H_A(p, q) - H_A(1, q) &= Cpq - \alpha p - Cq + \alpha = \\ &= Cq(p - 1) - \alpha(p - 1) = (p - 1)(Cq - \alpha), \\ H_A(p, q) - H_A(0, q) &= Cpq - \alpha p = p(Cq - \alpha). \end{aligned}$$

В случае если пара (p, q) определяет точку равновесия, эти разности неотрицательны:

$$H_A(p, q) - H_A(1, q) \geqslant 0, \qquad H_A(p, q) - H_A(0, q) \geqslant 0.$$

Поэтому окончательно получаем

$$\begin{aligned}(p-1)(Cq-\alpha) &\geqslant 0,\\ p(Cq-\alpha) &\geqslant 0.\end{aligned}$$

Из формул для функции $H_B(p,q)$ при $q=1$ и $q=0$ соответственно имеем:

$$\begin{aligned}&H_B(p,1)=(b_{11}-b_{12}-b_{21}+b_{22})p+(b_{12}-b_{22})p+b_{21},\\ &H_B(p,0)=(b_{12}-b_{22})p+b_{22}.\end{aligned}$$

Разности

$$H_B(p,q)-H_B(p,1) \quad \text{и} \quad H_B(p,q)-H_B(p,0)$$

с учетом обозначений

$$D=b_{11}-b_{12}-b_{21}+b_{22}, \qquad \beta=b_{22}-b_{21}$$

приводятся к виду

$$\begin{aligned}&H_B(p,q)-H_B(p,1)=(q-1)(Dp-\beta),\\ &H_B(p,q)-H_B(p,0)=q(Dp-\beta)\end{aligned}$$

совершенно так же, как соответствующие разности для функции H_A.

Если пара (p,q) определяет точку равновесия, то эти разности неотрицательны:

$$H_B(p,q)-H_B(p,1)\geqslant 0, \qquad H_B(p,q)-H_B(p,0)\geqslant 0.$$

Поэтому

$$\begin{aligned}(q-1)(Dp-\beta) &\geqslant 0,\\ q(Dp-\beta) &\geqslant 0.\end{aligned}$$

Прежде чем приступать к последующим шагам, подведем некоторые итоги.

Для того чтобы в биматричной игре

$$\mathbf{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}, \qquad \mathbf{B}=\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}$$

пара (p, q) определяла равновесную ситуацию, необходимо и достаточно одновременное выполнение следующих неравенств:

$$\begin{aligned} (p-1)(Cq-\alpha) &\geqslant 0, \\ p(Cq-\alpha) &\geqslant 0, \\ (q-1)(Dp-\beta) &\geqslant 0, \\ q(Dp-\beta) &\geqslant 0, \\ 0 \leqslant p &\leqslant 1, \\ 0 \leqslant q &\leqslant 1, \end{aligned} \tag{3}$$

где

$$\begin{aligned} C &= a_{11} - a_{12} - a_{21} + a_{22}, & \alpha &= a_{22} - a_{12}, \\ D &= b_{11} - b_{12} - b_{21} + b_{22}, & \beta &= b_{22} - b_{21}. \end{aligned}$$

19.4. Поиск равновесных ситуаций

Геометрический смысл условий (3) рассмотрим на примерах описанных выше биматричных игр.

19.4.1. Борьба за рынки

Напомним, что ситуация, сложившаяся в этой задаче, задается платежными матрицами следующего вида:

$$\mathbf{A} = \begin{pmatrix} -10 & 2 \\ 1 & -1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 5 & -2 \\ -1 & 1 \end{pmatrix}.$$

Заменяя в неравенствах (3) величины C, α, D и β их конкретными значениями

$$\begin{aligned} C &= -10 - 2 - 1 - 1 = -14, & \alpha &= -1 - 2 = -3, \\ D &= \quad 5 + 2 + 1 + 1 = \quad 9, & \beta &= \quad 1 + 1 = \quad 2, \end{aligned}$$

получаем

$$(l) \quad \begin{aligned} (p-1)(-14q-(-3)) &\geqslant 0, \\ p(-14q-(-3)) &\geqslant 0, \end{aligned} \qquad (r) \quad \begin{aligned} (q-1)(9p-2) &\geqslant 0, \\ q(9p-2) &\geqslant 0. \end{aligned}$$

Рассмотрим сначала левую пару неравенств (l):

$$(p-1)(-14q+3) \geqslant 0,$$
$$p(-14q+3) \geqslant 0.$$

Возможны следующие три случая:

$$1)\ p=1, \quad 2)\ p=0, \quad 3)\ 0<p<1.$$

Рассмотрим каждый из этих случаев подробно.

1. Полагая $p=1$, получаем

$$0 \geqslant 0, \quad -14q+3 \geqslant 0.$$

Отсюда

$$14q-3 \leqslant 0$$

и, значит,

$$q \leqslant \frac{3}{14}.$$

2. Полагая $p=0$, получаем

$$-(-14q+3) \geqslant 0, \quad 0 \geqslant 0,$$

откуда

$$14q-3 \geqslant 0$$

и, значит,

$$q \geqslant \frac{3}{14}.$$

3. Наконец, положив $0<p<1$, получим

$$-14q+3 \geqslant 0,$$
$$-14q+3 \leqslant 0,$$

что возможно лишь в случае, если

$$-14q+3=0,$$

т. е.

$$q=\frac{3}{14}.$$

Сформулируем полученный результат:

1°. $p = 1$, $q \leqslant \dfrac{3}{14}$,

2°. $p = 0$, $q \geqslant \dfrac{3}{14}$,

3°. $0 < p < 1$, $q = \dfrac{3}{14}$.

Перенесем теперь полученные результаты на чертеж.

Введем на плоскости прямоугольную систему координат (p, q) и выделим на ней единичный квадрат, соответствующий неравенствам

$$0 \leqslant p \leqslant 1, \quad 0 \leqslant q \leqslant 1$$

(рис. 1).

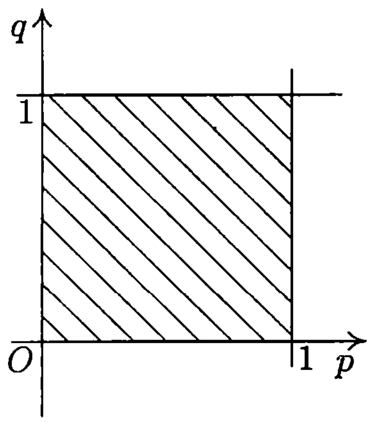


Рис. 1

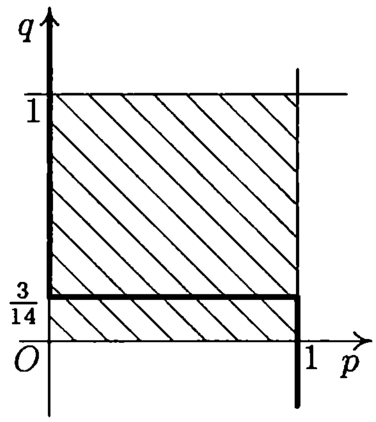


Рис. 2

Нанесем на этот чертеж то множество точек, которое описывается условиями 1°, 2° и 3°. Это множество (на рис. 2 его точки выделены жирной линией) состоит из трех прямолинейных участков — двух вертикальных лучей и одного горизонтального отрезка — и представляет собой "зигзаг". Нас будет интересовать только та его часть, которая попала в заштрихованный на рис. 2 единичный квадрат.

Теперь обратимся к правой части неравенств (r):

$$\begin{aligned} (q-1)(9p-2) &\geqslant 0, \\ q(9p-2) &\geqslant 0. \end{aligned}$$

Три интересных для нас случая:

$$1)\ q = 1, \qquad 2)\ q = 0, \qquad 3)\ 0 < q < 1$$

приводят к следующему результату:

$$1^\circ.\ q=1,\ p \geqslant \frac{2}{9},$$

$$2^\circ.\ q=0,\ p \leqslant \frac{2}{9},$$

$$3^\circ.\ 0<q<1,\ p=\frac{2}{9}.$$

Перенося его на чертеж, получим второй "зигзаг", но уже горизонтальный (рис. 3).

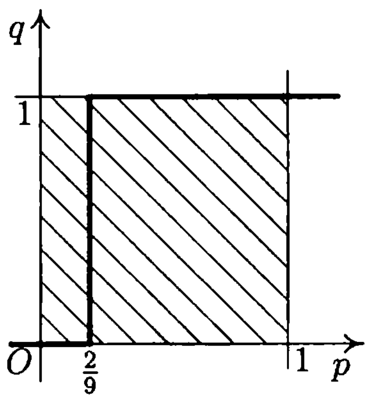


Рис. 3

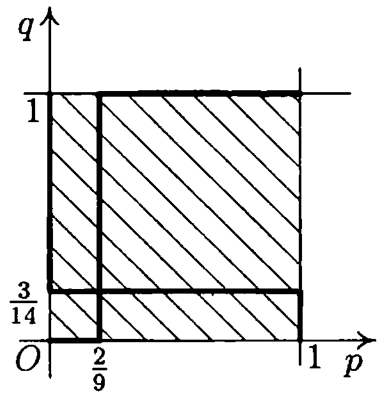


Рис. 4

Теперь остается только объединить полученное на рис. 4.

Общая точка построенных зигзагов — точка равновесия — имеет координаты

$$\left(\frac{2}{9}, \frac{3}{14}\right).$$

Соответствующие смешанные стратегии игроков имеют следующий вид:

$$\mathrm{P}=\left\{\frac{2}{9}, \frac{7}{9}\right\}, \quad \mathrm{Q}=\left\{\frac{3}{14}, \frac{11}{14}\right\},$$

а средние выигрыши игроков таковы:

$$H_A\left(\frac{2}{9}, \frac{3}{14}\right)=-\frac{4}{7}, \quad H_B\left(\frac{2}{9}, \frac{3}{14}\right)=\frac{1}{3}.$$

Замечание. Попробуем разбить рассмотренную биматричную игру на две матричные игры с нулевой суммой.

1. *Игра с матрицей* **A**.

$$\mathbf{A} = \begin{pmatrix} -10 & 2 \\ 1 & -1 \end{pmatrix}.$$

Решая эту игру графическим методом, найдем оптимальную смешанную стратегию для игрока A:

$$\left\{\frac{1}{7}, \frac{6}{7}\right\},$$

цену этой игры:

$$\nu_A^0 = -\frac{4}{7},$$

а затем и оптимальную смешанную стратегию для игрока B:

$$\left\{\frac{3}{14}, \frac{11}{14}\right\}.$$

2. *Игра с матрицей* **B**.

$$\mathbf{B} = \begin{pmatrix} 5 & -2 \\ -1 & 1 \end{pmatrix}.$$

Решая эту игру графическим методом, найдем оптимальную смешанную стратегию для игрока B:

$$\left\{\frac{1}{3}, \frac{2}{3}\right\},$$

цену этой игры:

$$\nu_B^0 = \frac{1}{3},$$

а затем и оптимальную смешанную стратегию для игрока A:

$$\left\{\frac{2}{9}, \frac{7}{9}\right\}.$$

Сравнивая полученные результаты с решением биматричной игры, можно заметить следующее:

если каждый игрок будет применять свои стратегии в этой игре исходя только из матрицы своих выигрышей, то его оптимальный средний выигрыш совпадет с его выигрышем при равновесной ситуации; кстати, по своей матрице игрок может найти и оптимальную смешанную стратегию другого игрока (но не свою!).

19.4.2. Дилемма узников

Выигрыши игроков A и B описываются соответствующими матрицами выплат:

$$\mathbf{A} = \begin{pmatrix} -1 & -9 \\ 0 & -6 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -1 & 0 \\ -9 & -6 \end{pmatrix}.$$

Проведем необходимые вычисления. Имеем:

$$C = -1 - (-9) - 0 + (-6) = 2, \quad \alpha = -6 - (-9) = 3,$$
$$D = -1 - 0 - (-9) + (-6) = 2, \quad \beta = -6 - (-9) = 3.$$

Отсюда

$$(l) \quad \begin{aligned} (p-1)(2q-3) &\geqslant 0, \\ p(2q-3) &\geqslant 0, \end{aligned} \qquad (r) \quad \begin{aligned} (q-1)(2p-3) &\geqslant 0, \\ q(2p-3) &\geqslant 0, \end{aligned}$$

и тогда получаем, что

$$1^l. \quad p = 1, \ q \geqslant \frac{3}{2}, \qquad 2^l. \quad p = 0, \ q \leqslant \frac{3}{2}, \qquad 3^l. \quad 0 < p < 1, \ q = \frac{3}{2};$$

$$1^r. \quad q = 1, \ p \geqslant \frac{3}{2}, \qquad 2^r. \quad q = 0, \ p \leqslant \frac{3}{2}, \qquad 3^r. \quad 0 < q < 1, \ p = \frac{3}{2}.$$

Полученные зигзаги изображены на рис. 5.

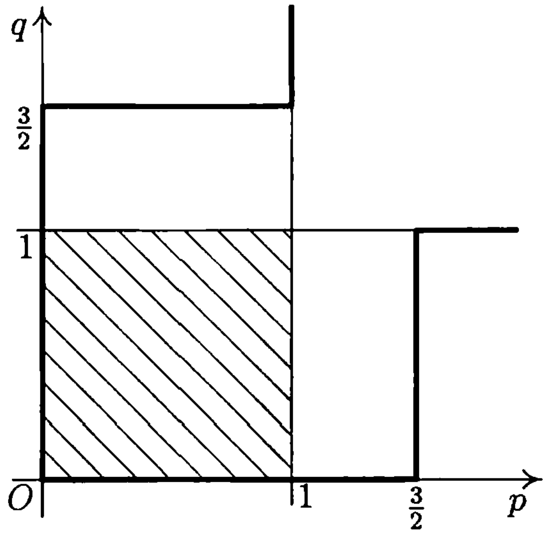


Рис. 5

Единственная равновесная ситуация — (0,0). Это ситуация, в которой каждый из игроков выбирает вторую чистую стратегию — сознаться — и его потери составляют 6.

Как мы уже отмечали ранее, отклонение от ситуации равновесия одного из игроков не дает ему никаких преимуществ. Однако при одновременном отклонении обоих каждый из них может получить больший выигрыш, нежели в равновесной ситуации. Например, в ситуации (1,1), когда оба игрока выбирают первую чистую стратегию — молчать, каждый из них теряет лишь 1.

Напомним, что по условию задачи сговор (создание коалиции) между игроками недопустим.

Совершенно ясно, однако, что в рассматриваемых обстоятельствах ситуация (1,1) неустойчива — любой из узников, изменяя свою стратегию, увеличивает свой выигрыш (избегает наказания).

19.4.3. Семейный спор

Выигрыши игроков A и B в этой биматричной игре задаются так:

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Проводя необходимые вычисления:

$$C = 2 - 0 - 0 + 1 = 3, \qquad \alpha = 1 - 0 = 1,$$
$$D = 1 - 0 - 0 + 2 = 3, \qquad \beta = 2 - 0 = 2$$

и рассуждения:

$$(l)\quad \begin{aligned} (p-1)(3q-1) &\geqslant 0, \\ p(3q-1) &\geqslant 0, \end{aligned} \qquad (r)\quad \begin{aligned} (q-1)(3p-2) &\geqslant 0, \\ q(3p-2) &\geqslant 0, \end{aligned}$$

получаем, что

$$1^l.\quad p = 1,\ q \geqslant \frac{1}{3}, \quad 2^l.\quad p = 0,\ q \leqslant \frac{1}{3}, \quad 3^l.\quad 0 < p < 1,\ q = \frac{1}{3};$$

$$1^r.\quad q = 1,\ p \geqslant \frac{2}{3}, \quad 2^r.\quad q = 0,\ p \leqslant \frac{2}{3}, \quad 3^r.\quad 0 < q < 1,\ p = \frac{2}{3}.$$

Геометрически полученный результат изображен на рис. 6.

Данная игра имеет три точки равновесия.

Две из них отвечают чистым стратегиям игроков:

$$p = 1, \qquad q = 1: \qquad H_A(1,1) = 2, \qquad H_B(1,1) = 1,$$
$$p = 0, \qquad q = 0: \qquad H_A(0,0) = 1, \qquad H_B(0,0) = 2,$$

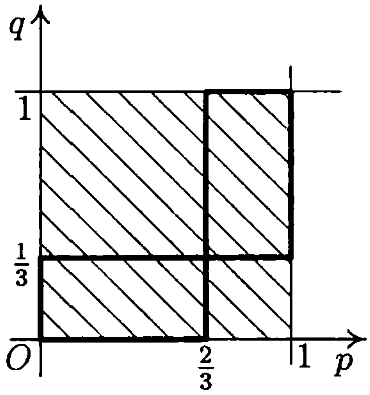


Рис. 6

одна — смешанной:

$$p = \frac{2}{3}, \quad q = \frac{1}{3}: \qquad H_A\left(\frac{2}{3}, \frac{1}{3}\right) = \frac{2}{3}, \quad H_B\left(\frac{2}{3}, \frac{1}{3}\right) = \frac{2}{3}.$$

В полученных результатах больше вопросов, чем ответов.

Ситуации (1,1) и (0,0) означают одновременный выбор игроками первых или соответственно вторых стратегий, т. е. определенную договоренность о совместных действиях.

Однако в данном случае есть еще одна ситуация равновесия, состоящая в выборе игроками вполне определенных смешанных стратегий. В ней оба игрока получают одинаковые выигрыши, правда, меньшие тех, которые давали две другие равновесные ситуации.

Какой же из этих трех ситуаций равновесия следует отдать предпочтение? Какую выбрать игрокам?

Если бы игроки договорились выбрать одновременно, скажем, первую чистую стратегию, причем игрок A за получение большего выигрыша, чем игрок B, заплатил бы ему $1/2$, то выигрыш каждым полутора единиц можно было бы считать и выгодным, и справедливым. Однако в рамках теории бескоалиционных игр такого рода дележи не рассматриваются.

Вполне ясно, что для выбора каждым из игроков своей линии поведения (напомним, что подобная ситуация может повторяться и повторяется многократно) необходимы либо расширение возможностей, имеющихся у игроков, либо иные, измененные критерии.

19.4.4. Студент – преподаватель

Наконец, обратимся к последнему из приведенных выше примеров биматричных игр — студент – преподаватель. Впечатления у каждого из них относительно результатов общения в матричном виде выглядят следующим образом:

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & -3 \\ -2 & -1 \end{pmatrix}.$$

Проводя необходимые вычисления:

$$C = 2 + 1 - 1 + 0 = 2, \qquad \alpha = 0 + 1 = 1,$$
$$D = 1 + 3 + 2 - 1 = 5, \qquad \beta = -1 + 2 = 1$$

и рассуждения:

$$(l) \quad \begin{aligned} (p-1)(2q-1) &\geqslant 0, \\ p(2q-1) &\geqslant 0, \end{aligned} \qquad (r) \quad \begin{aligned} (q-1)(5p-1) &\geqslant 0, \\ q(5p-1) &\geqslant 0, \end{aligned}$$

получаем, что

$$1^l.\quad p = 1,\ q \geqslant \frac{1}{2}, \qquad 2^l.\quad p = 0,\ q \leqslant \frac{1}{2}, \qquad 3^l.\quad 0 < p < 1,\ q = \frac{1}{2},$$

$$1^r.\quad q = 1,\ p \geqslant \frac{1}{5}, \qquad 2^r.\quad q = 0,\ p \leqslant \frac{1}{5}, \qquad 3^r.\quad 0 < q < 1,\ p = \frac{1}{5}$$

(рис. 7).

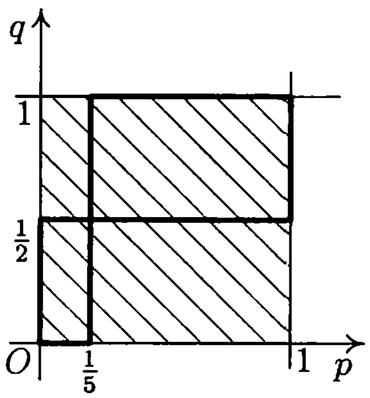


Рис. 7

Число точек пересечения у зигзагов (равновесных ситуаций) равно трем.

Две из них отвечают чистым стратегиям игроков:

$$p = 1, \qquad q = 1: \qquad H_A(1,1) = 2, \qquad H_B(1,1) = 1,$$
$$p = 0, \qquad q = 0: \qquad H_A(0,0) = 0, \qquad H_B(0,0) = -1,$$

одна — смешанной:

$$p = \frac{1}{5}, \quad q = \frac{1}{2}: \qquad H_A\left(\frac{1}{5}, \frac{1}{2}\right) = \frac{1}{2}, \quad H_B\left(\frac{1}{5}, \frac{1}{2}\right) = -\frac{7}{5}.$$

В данной задаче в отличие от предыдущей все довольно ясно: наилучшим является выбор каждым из игроков первой чистой стратегии — хорошо подготовиться к зачету и поставить зачет.

Как нетрудно заметить, тем самым в этой задаче реализуется весьма редкая возможность, когда функции выигрыша каждого из игроков достигают своих максимумов одновременно.

Выгодность такой ситуации совершенно ясна. Ее устойчивость также вполне очевидна: любое отклонение от ситуации (1,1) одного из игроков или обоих игроков может привести разве что к уменьшению их выигрышей.

19.5. Некоторые итоги

На анализе полученных результатов стоит остановиться чуть подробнее.

Из приведенных примеров видно, что числа C и D могут быть как положительными, так и отрицательными. Они могут, в частности, даже обращаться в нуль.

Рассмотрим, однако, наиболее интересный в приложениях случай, когда ни C ни D нулю не равны, т. е.

$$CD \neq 0.$$

Тогда, как нетрудно видеть, точка равновесия определяется парой

$$p = \frac{\beta}{D}, \quad q = \frac{\alpha}{C}.$$

Эти формулы являются весьма примечательными: в равновесной ситуации выбор игрока A полностью определяется элементами платежной матрицы игрока B,

$$p = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}}$$

(и не зависит от элементов его собственной платежной матрицы), а выбор игрока B в равновесной ситуации полностью определяется элементами платежной матрицы игрока A,

$$q = \frac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}$$

(и не зависит от элементов его собственной платежной матрицы).

Иными словами, равновесная ситуация обоих игроков определяется не столько стремлением увеличить собственный выигрыш, сколько желанием держать под контролем выигрыш другого игрока (минимизировать этот выигрыш). И если, например, заменить в биматричной игре матрицу выплат игроку A, а матрицу выплат игроку B оставить прежней, то игрок A никак не изменит своего "равновесного" поведения (просто не обратит внимания на эту замену), в то время как игрок B изменит свою стратегию на новую, равновесную.

Таким образом, в биматричной (неантагонистической) игре мы вновь встречаемся с антагонизмом. Правда, теперь это уже не *антагонизм интересов* (как было в антагонистической, матричной игре), а *антагонизм поведения.*

Отметим, что в биматричными играх (в отличие от матричных) при наличии нескольких ситуаций равновесия средний выигрыш игрока в разных равновесных ситуациях различен (напомним, что в матричной игре выигрыш игрока один и тот же вне зависимости от количества точек равновесия).

Но если средние выигрыши разнятся, то какую равновесную ситуацию следует считать оптимальной?

Наконец, еще одно, не менее интересное обстоятельство. Вспомним, с какими трудностями мы столкнулись, пытаясь перевести эмоциональные оценки результатов общения студент – преподаватель в количественные показатели. В целом сохраняя основные соотношения, эти количественные оценки могут, конечно, изменяться как от студента к студенту, так и от преподавателя к преподавателю. Однако если эти изменения будут не слишком значительными — элементы платежной матрицы "пошевельнутся" слегка — то слегка "пошевельнутся" и зигзаги, не изменяя ни своей общей формы, ни взаимного расположения, а значит, число равновесных ситуаций не изменится. Впрочем, сказанное относится лишь к случаю, когда множество ситуаций равновесия конечно и состоит из нечетного числа точек (одной или трех).

Как принято говорить в подобных случаях, это число *устойчиво относительно малых шевелений.*

Конечно, в некоторых биматричных играх равновесные ситуации случаются и в чистых стратегиях (в последнем из разобранных примеров таких ситуаций даже две). И (в принципе, это совсем нетрудно) можно дать определение ситуации равновесия в чистых стратегиях. Найти ее (если она, конечно, существует) — дело довольно простое. Но, как показывают разобранные примеры, во-первых, чистой ситуации равновесия может вовсе не быть, и, во-вторых, даже при ее наличии не исключено существование равновесных ситуаций в смешанных стратегиях. И, чтобы найти их все, неизбежно приходится обращаться к описанному выше подходу.

19.6. Задания и ответы

Найдите решение биматричной игры:

а) $\mathbf{A} = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 2 & 6 \\ 8 & 2 \end{pmatrix}$,

б) $\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$,

в) $\mathbf{A} = \begin{pmatrix} 5 & 0 \\ 10 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 5 & 10 \\ 0 & 1 \end{pmatrix}$.

Ответы:

а) $p = \frac{3}{5}$, $q = \frac{1}{3}$: $H_A\left(\frac{3}{5}, \frac{1}{3}\right) = \frac{10}{3}$, $H_B\left(\frac{3}{5}, \frac{1}{3}\right) = \frac{22}{5}$.

б) $p = 1$, $q = 1$: $H_A(1,1) = 4$, $H_B(1,1) = 1$;
$p = 0$, $q = 0$: $H_A(0,0) = 1$, $H_B(0,0) = 4$;
$p = \frac{4}{5}$, $q = \frac{1}{5}$: $H_A\left(\frac{4}{5}, \frac{1}{5}\right) = \frac{4}{5}$, $H_B\left(\frac{4}{5}, \frac{1}{5}\right) = \frac{4}{5}$.

в) $p = 0$, $q = 0$: $H_A(0,0) = 1$, $H_B(0,0) = 1$.

Глава 20
НЕКОТОРЫЕ ДРУГИЕ ИГРЫ

Реальные конфликтные ситуации приводят к различным видам игр. К настоящему времени общепризнанной классификации игр не сложилось. Тем не менее легко заметить, что игры можно различать по целому ряду признаков: по количеству участвующих в них игроков, возможных стратегий, ходов, по характеру взаимоотношений между игроками, выигрышей, информационной обеспеченности игроков, по виду функций выигрышей и т. д.

В зависимости от вида игры разрабатывается и метод ее решения.

В этой главе мы приведем примеры игр, отличных от рассмотренных ранее. Но начнем с обсуждения (на примере биматричной игры) несколько иного подхода к вопросу устойчивости искомой ситуации.

20.1. Ситуации, оптимальные по Парето

Содержательные представления об устойчивости, выгодности и справедливости многообразны. Выше мы рассматривали проявление устойчивости ситуации через равновесие. Существует и иной вариант устойчивости ситуации, в большей степени, чем равновесность, отражающий черты ее выгодности, — это оптимальность по Парето.

Поговорим об этом на примере биматричной игры с 2×2-матрицами $\mathbf{A}$ и $\mathbf{B}$. Пусть $H_A(p, q)$ и $H_B(p, q)$ — средние выигрыши игроков A и B.

Ситуация (p_*, q_*) в биматричной игре

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

называется *оптимальной по Парето*, если из того, что

$$H_A(p_*, q_*) \leqslant H_A(p, q) \quad \text{и} \quad H_B(p_*, q_*) \leqslant H_B(p, q)$$

вытекает, что

$$p = p_*, \qquad q = q_*.$$

Иными словами, в оптимальной по Парето ситуации игроки не могут совместными усилиями увеличить выигрыш одного из них, не уменьшив при этом выигрыш другого.

Отличие ситуации равновесия от ситуации, оптимальной по Парето, состоит в следующем:

в ситуации равновесия ни один из игроков, действуя в одиночку, не может увеличить свой собственный выигрыш;

в ситуации, оптимальной по Парето, игроки, действуя совместно, не могут увеличить выигрыш хотя бы одного из них, не уменьшая выигрыша другого.

Покажем, как практически отыскиваются оптимальные по Парето ситуации.

Обратимся к игре “Дилемма узников”.

Напомним, что соответствующие платежные матрицы в этой игре имеют следующий вид:

$$\mathbf{A} = \begin{pmatrix} -1 & -9 \\ 0 & -6 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -1 & 0 \\ -9 & -6 \end{pmatrix}.$$

Тем самым на единичном квадрате

$$0 \leqslant p \leqslant 1, \qquad 0 \leqslant q \leqslant 1$$

(рис. 1) возможных значений вероятностей p и q заданы две функции:

$$U = H_A(p, q) = -pq - 9p(1 - q) - 6(1 - p)(1 - q),$$
$$V = H_B(p, q) = -pq - 9(1 - p)q - 6(1 - p)(1 - q).$$

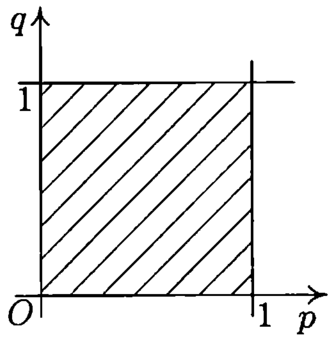


Рис. 1

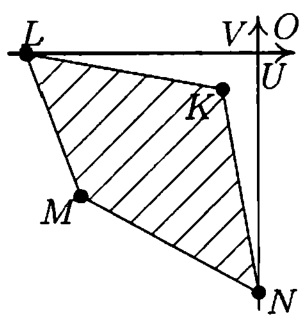


Рис. 2

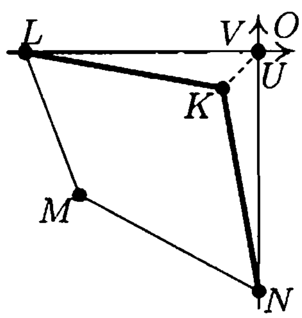


Рис. 3

Точки с координатами (U, V), вычисленными по приведенным формулам, на плоскости (U, V) заполняют четырехугольник с вершинами $K(-1, -1)$, $L(-9, 0)$, $M(-6, -6)$ и $N(0, -9)$ (рис. 2). Граница Парето этого множества — ломаная NKL.

В качестве точки утопии здесь естественно рассматривать начальную точку $O(0, 0)$. Идеальная точка $K(-1, -1)$ — точка с наибольшими выигрышами для каждого из игроков — оказывается лучше, чем равновесная (рис. 3). Ей соответствуют чистые стратегии обоих игроков

$$p = 1, \qquad q = 1.$$

20.2. Неантагонистические позиционные игры

Мы достаточно подробно остановились на позиционных играх двух лиц, где были явно выражены интересы одного из игроков (игрока A). Следует, однако, иметь в виду, что в одних случаях интересы игрока B могут быть полностью противоположными интересам игрока A, в то время как в других вполне может оказаться, что то, что хорошо для одного игрока, необязательно плохо для другого.

Приведем два простых примера.

Пример 1. *1-й ход.* Игрок A выбирает число x из множества двух чисел $\{1, 2\}$.

2-й ход. Игрок B выбирает число y из множества двух чисел $\{1, 2\}$, зная выбор числа x игроком A.

Функции выплат игрокам A и B — $W_A(x, y)$ и $W_B(x, y)$ соответственно — задаются так:

$$\begin{array}{llll} W_A(1,1) = 1, & W_A(1,2) = -1, & W_A(2,1) = -2, & W_A(2,2) = 2, \\ W_B(1,1) = 2, & W_B(1,2) = 1, & W_B(2,1) = 1, & W_B(2,2) = 2. \end{array}$$

Дерево игры показано на рис. 4.

Исход игры зависит от того, каковы намерения игрока B — *максимизировать свой выигрыш*:

$$W_B(x, y) \to \max$$

или

максимизировать свой относительный выигрыш:

$$W_B(x, y) - W_A(x, y) \to \max.$$

В первом случае это достигается так:
при $x = 1$: $y = 1$ и $W_B(1,1) = 2$ ($W_A(1,1) = 1$);
при $x = 2$: $y = 2$ и $W_B(2,2) = 2$ ($W_A(2,2) = 2$).

Во втором случае:
при $x = 1$: $y = 2$ и $W_B(1,2) - W_A(1,2) = 1 - (-1) = 2$;
при $x = 2$: $y = 1$ и $W_B(2,1) - W_A(2,1) = 1 - (-2) = 3$.

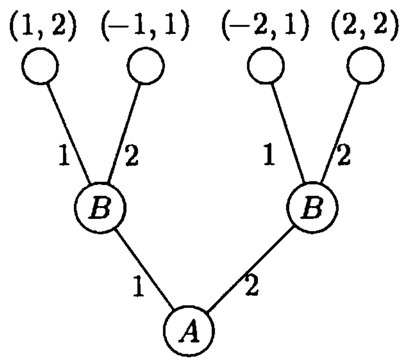


Рис. 4

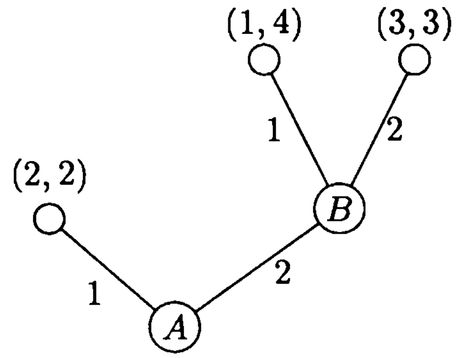


Рис. 5

Пример 2. Игра задается деревом (рис. 5).

1-й ход. Игрок A выбирает число x из множества двух чисел $\{1, 2\}$.

Если $x = 1$, то каждый из игроков получает свой выигрыш, равный 2.

Если $x = 2$, то игрок B получает право 2-го хода, где он и выбирает число y из множества двух чисел $\{1, 2\}$.

При $y = 1$ выигрыш игрока A равен 1, а игрока B — 4. При $y = 2$ оба игрока получают поровну — по 3.

В случае когда каждый из игроков стремится к получению максимального выигрыша и любые виды кооперации запрещены, исход игры ясен — игрок A выбирает $x = 1$ и игра заканчивается. Но при $x = 2$ и $y = 2$ каждый из игроков получает по 3 (такой исход предпочтительнее простейшего (2,2)), и, если допустить соглашение между игроками, это обстоятельство вполне может изменить исход игры.

20.3. Бесконечные игры

Выше мы остановились достаточно подробно лишь на трех видах игр — матричных, позиционных и биматричных. Сделанный выбор обусловлен тем, что уже здесь можно наглядно показать, какой смысл вкладывается в термин *игра* и чем именно занимается *теория игр*, а также познакомить с относительно несложным математическим *инструментарием*, опирающимся на ключевые понятия вероятности, матрицы и координаты и позволяющим разрешать простейшие из этих видов игр.

Вместе с тем нам не хотелось бы, чтобы у читателя сложилось впечатление, что доступными содержательному анализу могут быть только игры, описанные выше.

Существует интересный, привлекающий неослабевающее внимание исследователей класс игр, в которых хотя бы один из игроков имеет бесконечное множество возможных стратегий, — *бесконечные* игры.

В этом заключительном разделе мы приведем примеры бесконечных игр двух лиц трех видов — непрерывной игры на единичном квадрате (*непрерывными* называются бесконечные игры, в которых функции выигрышей непрерывно зависят от стратегий, выбираемых игроками), дуэли (играми *с выбором момента времени*, или играми *типа дуэли*, называются игры, характеризующиеся моментом выбора хода и вероятностями получения выигрыша в зависимости от времени, прошедшего от начала игры до момента выбора) и дифференциальной игры поиска (в *дифференциальных* играх допускается делать ходы непрерывно и связывать поведение игроков условиями, описываемыми дифференциальными уравнениями).

Мы ограничимся здесь лишь постановками задач — описанием возможностей в поведении игроков и построением функций выигрышей, хотя для каждой из приводимых игр разработаны достаточно эффективные подходы к построению их решения.

20.3.1. Борьба за рынки (игра на единичном квадрате)

Одна из конкурирующих фирм (игрок A) пытается вытеснить другую фирму (игрок B) с одного из двух рынков сбыта. Предположим, что общая сумма средств, выделенная на это игроком A, равна 1. Типичной стратегией игрока A является разделение выделенной суммы на две части: x $(0 \leqslant x \leqslant 1)$ для первого рынка и $1 - x$ для второго. Подобным образом выглядят и стратегии игрока B: выделение им части y $(0 \leqslant y \leqslant 1)$ своей суммы на первый рынок и $1 - y$ на второй.

Будем считать, что если игрок A добился превосходства на одном из рынков (на другом превосходства автоматически добивается игрок B), то он вытесняет противника с этого рынка и получает выигрыш, пропорциональный избытку вложенных средств с коэффициентом, характеризующим важность рынка (этот коэффициент равен k_1 для первого рынка и k_2 для второго). Тогда функция выигрыша $H(x,y)$ игрока A определяется формулой

$$H(x,y) = \begin{cases} k_1(x-y), & \text{если } 0 \leqslant y < x \leqslant 1, \\ 0, & \text{если } x = y, \\ -k_2(y-x), & \text{если } 0 \leqslant x < y \leqslant 1. \end{cases}$$

Ясно, что функция выигрыша игрока B равна $-H(x,y)$.

20.3.2. Игра типа дуэли

Два дуэлянта (игроки A и B) начинают сходиться в момент времени $t = 0$. У каждого пистолет заряжен одной пулей. Они встретятся в момент времени $t = 1$ (если только ни один из них не застрелит другого раньше). Каждый из дуэлянтов может выстрелить, когда пожелает. Если при этом одному из них удастся поразить противника, а самому остаться невредимым, то он становится победителем (его выигрыш равен 1) и дуэль тут же прекращается. Если оба промахнутся, дуэль закончится вничью (выигрыш каждого из игроков равен 0). Если оба выстрелят одновременно и каждый поразит противника, то дуэль также считается окончившейся вничью.

Если игрок A произведет выстрел в момент времени x $(0 \leqslant x \leqslant 1)$, то его выстрел будет успешным с вероятностью $p(x)$. Подобным же образом выстрел игрока B в момент времени y $(0 \leqslant y \leqslant 1)$ будет успешным с вероятностью $q(y)$. При условии $x < y$ игрок A выиграет с вероятностью $p(x)$, а проиграет с вероятностью $(1 - p(x))q(y)$. Его средний выигрыш при $x < y$ будет равен

$$H(x,y) = 1 \cdot p(x) + (-1) \cdot (1 - p(x))q(y) = p(x) - q(y) + p(x)q(y).$$

С другой стороны, если $x > y$, его средний выигрыш будет равен

$$H(x,y) = 1 \cdot (1 - q(y))p(x) + (-1) \cdot q(y) = p(x) - q(y) - p(x)q(y).$$

При $x = y$ средний выигрыш

$$H(x,x) = 1 \cdot p(x) + (-1) \cdot q(x) = p(x) - q(x).$$

Таким образом, функция выигрыша $H(x,y)$ игрока A имеет вид

$$H(x,y) = \begin{cases} p(x) - q(y) + p(x)q(y), & \text{если } 0 \leqslant x < y \leqslant 1, \\ p(x) - q(x), & \text{если } x = y, \\ p(x) - q(y) - p(x)q(y), & \text{если } 0 \leqslant y < x \leqslant 1 \end{cases}$$

и антагонистическая игра задана.

В частности, если игроки стреляют без промаха, $p(x) = q(y) \equiv 1$,

$$H(x,y) = \begin{cases} 1, & \text{если } 0 \leqslant x < y \leqslant 1, \\ 0, & \text{если } x = y, \\ -1, & \text{если } 0 \leqslant y < x \leqslant 1. \end{cases}$$

20.3.3. Дифференциальная игра поиска

Ищущий (игрок A) стремится обнаружить уклоняющегося (игрок B). Оба игрока перемещаются с постоянными скалярными скоростями (α и β соответственно) по плоскости внутри некоторой поисковой области Ω. В любой момент каждый из игроков управляет своим перемещением, задавая направление вектора скорости. Пусть (x_A, y_A) и (x_B, y_B) — координаты игроков. Тогда имеем

$$\frac{dx_A}{dt} = \alpha\cos\varphi, \qquad \frac{dx_B}{dt} = \beta\cos\psi, \qquad \frac{dy_A}{dt} = \alpha\sin\varphi, \qquad \frac{dy_B}{dt} = \beta\sin\psi.$$

Игра поиска заканчивается в тот момент, когда игроки сблизятся на расстояние $l > 0$, иными словами, когда будет выполнено неравенство

$$(x_A - x_B)^2 + (y_A - y_B)^2 \leqslant l^2.$$

В случае успешного обнаружения выигрыш игрока A считается равным 1.

Построение решения в этой игре существенно зависит от характера и степени информированности игроков.

20.4. Несколько слов в заключение

Все это были примеры *бескоалиционных* игр, когда любые соглашения, обмен информацией, побочные платежи, совместный выбор стратегий запрещены. Другой важный класс составляют *кооперативные* игры, в которых разрешены самые разнообразные формы сотрудничества (что особенно важно при наличии числа игроков, большего двух). Возможность соглашений между игроками оказывает существенное влияние на исход игры. Если допустить, например, в игре “Дилемма узников” совместный выбор стратегий, то исход игры может оказаться совсем иным. При наличии побочных платежей по-иному окончится и “Семейный спор”.

Глава 21
УПРАВЛЕНИЕ ОРГАНИЗАЦИОННЫМИ СИСТЕМАМИ

21.1. Распределение ресурсов

21.1.1. Постановка задачи распределения ресурсов

Организационная система (оргсистема, организация) — это система, включающая технику и коллективы людей, интересы которых существенно связаны с ее функционированием. Примерами здесь могут служить семья, фирма, университет, город, страна. Каждая оргсистема состоит из элементов (которые в свою очередь тоже могут представлять собой системы).

Для нас существенными являются следующие два обстоятельства. С одной стороны, система существует для достижения каких-либо определенных целей, т. е. можно говорить об интересах системы в целом. С другой стороны, элементы системы зачастую преследуют собственные интересы, вообще говоря не совпадающие с интересами системы в целом. Все это дает основание формализовать некоторые аспекты функционирования оргсистем в терминах теории игр.

В данном разделе мы будем рассматривать простейшую двухуровневую модельную оргсистему, состоящую из *Центра* и некоторого числа однотипных *Элементов*. Управление такой системой мы рассмотрим на примере задачи распределения ресурсов. Суть этой задачи состоит в следующем. Элементы (в дальнейшем мы будем называть их *Потребителями*) представляют Центру заявки на получение некоторого ресурса (для простоты рассматривается один вид ресурса). Центр на основании этих заявок распределяет имеющийся в его распоряжении ресурс (который предполагается делимым).

Если все заявки могут быть полностью удовлетворены, то Центру, по-видимому, так и следует поступить — выделить каждому Потребителю столько, сколько он просит.

Существенно сложнее ситуация *дефицита*, когда суммарный объем заявок превосходит имеющийся в распоряжении Центра ресурс. В этом случае задача распределения ресурса становится нетривиальной. Универсальных рекомендаций здесь не существует. Ниже будут рассмотрены некоторые способы, или *механизмы*, распределения ресурсов, каждый из которых обладает определенными достоинствами и недостатками.

Проведем формализацию вышеописанной задачи. Имеется n Потребителей, каждый из которых сообщает Центру число s_i ($i = 1, 2, \ldots, n$) — заявку (рис. 1), а также, быть может, еще некоторую информацию (на рис. 1 обозначено пунктирной стрелкой). Далее Центр на основании заявок Потребителей, имеющегося в его распоряжении ресурса R и дополнительной информации о Потребителях вычисляет по некоторому правилу числа x_i ($i = 1, 2, \ldots, n$) — объем ресурса, выделяемый i-му Потребителю.

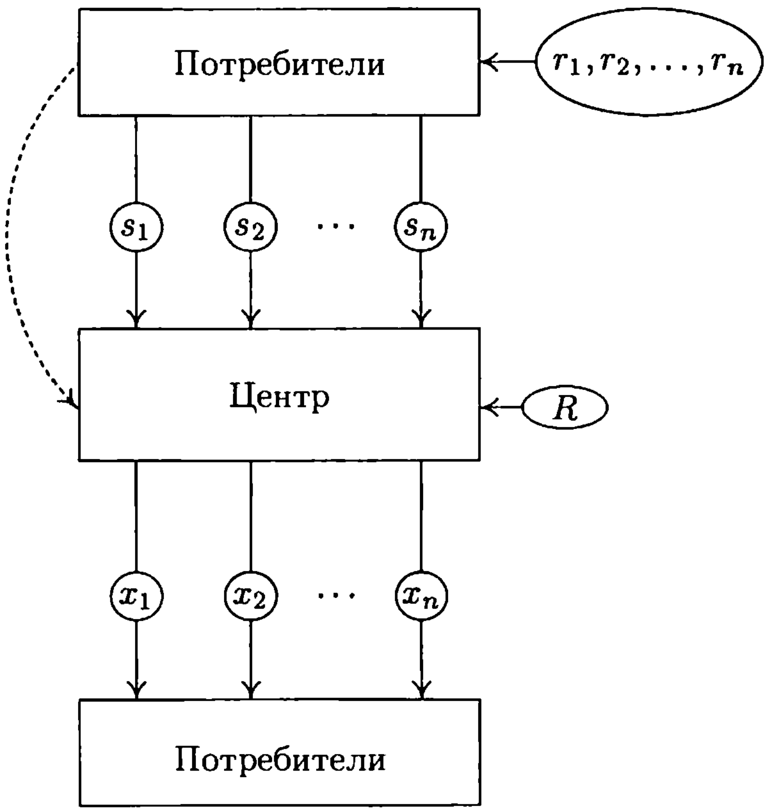


Рис. 1

В случае

$$\sum_{i=1}^{n} s_i \leqslant R$$

(отсутствие дефицита) естественным решением Центра является следующее:

$$x_1 = s_1, \quad x_2 = s_2, \quad \ldots, \quad x_n = s_n$$

(каждый Потребитель получает столько, сколько просил). В дальнейшем мы будем считать выполненным неравенство

$$\sum_{i=1}^{n} s_i > R$$

(суммарная заявка Потребителей превосходит ресурс Центра).

Отметим следующее важное обстоятельство. Потребители формируют свои заявки на основании собственных реальных потребностей r_i, которые им известны, но неизвестны Центру. Можно сказать, что числа s_i являются *стратегиями* Потребителей как участников иерархической игры. В свою очередь, стратегией Центра являются числа x_i.

21.1.2. Механизм прямых приоритетов

Механизм прямых приоритетов относится к числу так называемых приоритетных механизмов, отличительной чертой которых является приписывание каждому Потребителю некоторого приоритета. Итак, наряду с размерами заявок s_i $(i = 1, 2, \ldots, n)$ Центр учитывает приоритет каждого Потребителя, который определяется числом A_i $(i = 1, 2, \ldots, n)$.

В соответствии с механизмом прямых приоритетов распределение ресурса осуществляется по правилу

$$x_i = \max\{s_i, \gamma A_i s_i\}, \quad i = 1, 2, \ldots, n, \tag{1}$$

где γ — общий для всех Потребителей параметр — определяется из условия

$$\sum_{i=1}^{n} x_i = R \tag{2}$$

(весь ресурс распределяется без остатка).

Особенно простой вид формула (1) приобретает в случае "равенства" Потребителей с точки зрения Центра, т. е. при

$$A_1 = A_2 = \cdots = A_n = 1$$

(это условие не ограничивает общности, но упрощает дальнейшие выкладки). Тогда

$$x_i = \min\{s_i, \gamma s_i\} = \gamma s_i, \quad i = 1, 2, \ldots, n$$

(случай $x_i = s_i$ невозможен, поскольку при этом каждый Потребитель получает столько, сколько просил, а это противоречит предположению о наличии дефицита). Из условия (2) получаем

$$\sum_{i=1}^{n} \gamma s_i = R,$$

откуда

$$\gamma = \frac{R}{\sum\limits_{i=1}^{n} s_i}.$$

Описанный механизм распределения ресурсов является, пожалуй, самым простым. Смысл его состоит в том, что все заявки пропорционально "урезаются" путем умножения на число γ.

Пример 1. Пусть пять Потребителей подали заявки в размере 5, 8, 12, 7 и 8. Имеющийся в распоряжении Центра ресурс составляет 32. Как должен быть распределен этот ресурс в соответствии с механизмом прямых приоритетов?

Решение. Имеем:

$$s_1 = 5,\ s_2 = 8,\ s_3 = 12,\ s_4 = 7,\ s_5 = 8; \quad R = 32.$$

Поскольку

$$\sum_{i=1}^{5} s_i = 5 + 8 + 12 + 7 + 8 = 40 > 32 = R,$$

налицо дефицит. Определяем коэффициент γ:

$$\gamma = \frac{32}{40} = 0{,}8.$$

На это число и умножаются заявки. В итоге получаем

$$\begin{aligned} x_1 &= 0{,}8 \cdot \ \ 5 = 4; \\ x_2 &= 0{,}8 \cdot \ \ 8 = 6{,}4; \\ x_3 &= 0{,}8 \cdot 12 = 9{,}6; \\ x_4 &= 0{,}8 \cdot \ \ 7 = 5{,}6; \\ x_5 &= 0{,}8 \cdot \ \ 8 = 6{,}4. \end{aligned}$$

Ответ: $x_1 = 4$; $x_2 = 6{,}4$; $x_3 = 9{,}6$; $x_4 = 5{,}6$; $x_5 = 6{,}4$.

Достоинства механизма прямых приоритетов очевидны. Отметим два недостатка.

Во-первых, каждый Потребитель получает меньше, чем просит. Между тем нетрудно представить себе ситуацию, когда Потребителю требуется на осуществление какого-либо проекта именно s_i единиц ресурса, а γs_i уже не хватает.

Во-вторых, данный механизм "толкает" Потребителей к завышению заявок в условиях дефицита. Действительно, поскольку чем больше Потребитель просит, тем больше получает, он может, завышая свои потребности, попытаться приблизить итоговое решение Центра x_i к своим реальным потребностям r_i. Тем самым дефицит еще более возрастает, причем Центр даже не имеет возможности узнать реальные запросы Потребителей r_i, поскольку они сообщают заявки $s_i > r_i$.

21.1.3. Механизм обратных приоритетов

Механизм обратных приоритетов основывается на предположении, что, чем меньше требуется Потребителю ресурса, тем больше эффективность его использования. В соответствии с этим распределение ресурса осуществляется по правилу

$$x_i = \min \left\{ s_i, \gamma \frac{A_i}{s_i} \right\}, \quad i = 1, 2, \ldots, n, \tag{3}$$

где число γ определяется, как и в механизме прямых приоритетов, из условия

$$\sum_{i=1}^{n} x_i = R.$$

Из формулы (3) видно, что, подавая очень малую либо очень большую заявку s_i, Потребитель получает малый ресурс x_i.

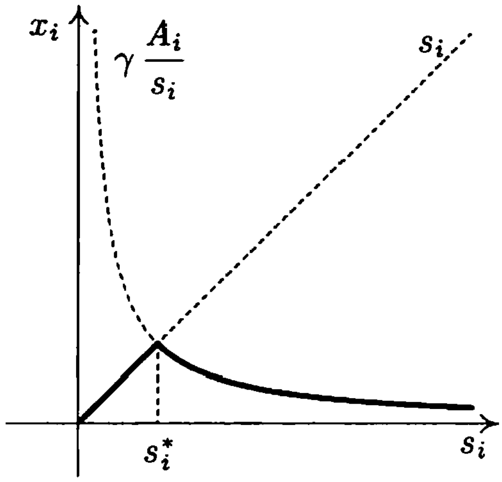


Рис. 2

Найдем, какую же заявку s_i должен подавать i-й Потребитель, чтобы получить максимальный ресурс x_i (в условиях дефицита такая цель Потребителя представляется вполне понятной). На рис. 2 изображен график функции $x_i = x_i(s_i)$. Видно, что максимум достигается в точке s_i^*, являющейся решением уравнения

$$s_i^* = \gamma \frac{A_i}{s_i^*}.$$

Преобразуя последнее равенство, получаем

$$s_i^* = \sqrt{\gamma A_i}.$$

Таким образом, *равновесным* является набор стратегий Потребителей

$$s_1^* = \sqrt{\gamma A_1},\ s_2^* = \sqrt{\gamma A_2},\ \ldots,\ s_n^* = \sqrt{\gamma A_n},$$

при этом

$$x_1 = s_1^*,\ x_2 = s_2^*,\ \ldots,\ x_n = s_n^*.$$

Выбирая вместо s_i^* любую другую стратегию s_i, i-й Потребитель лишь уменьшает выделяемый ему ресурс x_i.

Осталось вычислить константу γ. Имеем:

$$R = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} s_i^* = \sum_{i=1}^{n} \sqrt{\gamma A_i} = \sqrt{\gamma} \sum_{i=1}^{n} \sqrt{A_i},$$

откуда

$$\sqrt{\gamma} = \frac{R}{\sum\limits_{i=1}^{n} \sqrt{A_i}}.$$

Замечание. Еще раз отметим, что набор стратегий s_i^* $(i = 1, 2, \ldots, n)$ является равновесным, т. е., подавая любую заявку $s_i \neq s_i^*$, i-й Потребитель лишь уменьшает выделяемый ему ресурс x_i. Можно доказать, что каждая из стратегий s_i^* является также гарантирующей, т. е. в случае применения i-м Потребителем этой стратегии он в любом случае (т. е. при любых заявках остальных Потребителей) получает не меньше, чем $x_i = s_i^*$.

Пример 2. Пусть имеется пять Потребителей, приоритеты которых определяются числами 8, 6, 12, 15, 11. Ресурс Центра составляет 60. Определить равновесные стратегии (заявки) Потребителей, если ресурс распределяется в соответствии с механизмом обратных приоритетов.

Решение. Имеем:

$$A_1 = 8, \; A_2 = 6, \; A_3 = 12, \; A_4 = 15, \; A_5 = 11; \quad R = 60.$$

Вычислим константу $\sqrt{\gamma}$:

$$\sqrt{\gamma} = \frac{60}{\sqrt{8} + \sqrt{6} + \sqrt{12} + \sqrt{15} + \sqrt{11}} \approx 3{,}77.$$

Определять γ необязательно, поскольку в формулы для s_i^* можно подставить сразу $\sqrt{\gamma}$:

$$\begin{aligned}
s_1^* &= 3{,}77 \cdot \sqrt{8} \approx 10{,}7,\\
s_2^* &= 3{,}77 \cdot \sqrt{6} \approx 9{,}2,\\
s_3^* &= 3{,}77 \cdot \sqrt{12} \approx 13{,}1,\\
s_4^* &= 3{,}77 \cdot \sqrt{15} \approx 14{,}6,\\
s_5^* &= 3{,}77 \cdot \sqrt{11} \approx 12{,}5.
\end{aligned}$$

Ответ: $s_1^* = 10{,}7$; $s_2^* = 9{,}2$; $s_3^* = 13{,}1$; $s_4^* = 14{,}6$; $s_4^* = 12{,}5$.

Замечание 1. Из-за ошибок округления сумма заявок немного отличается от $R = 60$.

Замечание 2. На самом деле мы рассмотрели случай, когда $s_i^* < r_i$ для всех i, т. е. когда каждый из Потребителей вынужден, подавая заявку, занижать свою реальную потребность. Может быть и так, что для некоторых Потребителей $s_i^* \geqslant r_i$. Тогда эти Потребители подают заявку на ресурс $s_i = r_i$ и столько же получают.

Механизм обратных приоритетов обладает рядом достоинств. В частности, не происходит неоправданного завышения заявок, т. е. не возникает ситуации $s_i > r_i$. Кроме того, при условии разумного поведения Потребителей (т. е. при использовании каждым из них равновесной стратегии s_i^*) они получают столько, сколько просят.

Недостатком является то, что числа s_i^* скорее всего оказываются меньше реальных потребностей r_i. Вследствие этого Центр не получает достоверной информации о реальном дефиците

$$\left(\sum_{i=1}^{n} r_i\right) - R.$$

21.1.4. Конкурсный механизм

Конкурсный механизм применяется в тех случаях, когда нецелесообразно "урезать" заявки, поскольку Потребителям ресурс нужен на реализацию каких-либо конкретных проектов, на которые меньшего ресурса не хватит. В этих условиях Центр проводит конкурс заявок. Те, кто побеждают в конкурсе, полностью получают требуемый ресурс, а проигравшие не получают ничего.

Реализация этого происходит следующим образом. Потребители сообщают Центру свои заявки s_i, а также величины w_i, характеризующие *эффект*, который они намереваются получить. На основании этих данных Центр вычисляет для каждого Потребителя показатель *эффективности*:

$$e_i = \frac{w_i}{s_i}, \; i = 1, 2, \ldots, n.$$

После этого ресурс распределяется следующим образом. Сначала рассматривается Потребитель с наибольшей эффективностью. Ему выделяется столько, сколько он просит (если у Центра хватает ресурса). Затем берется второй по эффективности и т. д. В какой-то момент оказывается, что на удовлетворение очередной заявки оставшегося у Центра ресурса не хватает. Тогда этот потребитель, равно как и все оставшиеся, ничего не получает.

Пример 3. Имеется шесть Потребителей, подавших заявки в размере 14, 18, 10, 15, 8, 14 и сообщивших Центру соответственно следующие показатели эффекта: 36, 38, 25, 42, 28, 29. Каким должно быть распределение ресурса объемом 60 в соответствии с конкурсным механизмом?

Решение. По условию имеем

$$s_1 = 14, \quad s_2 = 18, \quad s_3 = 10, \quad s_4 = 15, \quad s_5 = 8, \quad s_6 = 14;$$
$$w_1 = 36, \quad w_2 = 38, \quad w_3 = 25, \quad w_4 = 42, \quad w_5 = 28, \quad w_6 = 29.$$

Вычислим показатели эффективности для каждого Потребителя:

$$e_1 = \frac{36}{14} \approx 2{,}57, \qquad e_4 = \frac{42}{15} = 2{,}8,$$
$$e_2 = \frac{38}{18} \approx 2{,}11, \qquad e_5 = \frac{28}{8} = 3{,}5,$$
$$e_3 = \frac{25}{10} = 2{,}5, \qquad e_6 = \frac{29}{14} \approx 2{,}07.$$

Расположим эти числа в порядке убывания:

$$e_5 > e_4 > e_1 > e_3 > e_2 > e_6.$$

Распределение ресурса начинаем с 5-го Потребителя:

$$x_5 = 8.$$

Ресурса осталось $60 - 8 = 52$. Дальше в порядке убывания показателей эффективности следует 4-й Потребитель:

$$x_4 = 15.$$

Ресурса осталось $52 - 15 = 37$. Далее:

$$x_1 = 14.$$

Ресурса осталось $37 - 14 = 23$. Далее,

$$x_3 = 10.$$

Ресурса осталось $23 - 10 = 13$.

Следующему, 2-му Потребителю требуется 18 единиц ресурса, а у Центра осталось лишь 15. Поэтому 2-й, а также 6-й Потребители ничего не получают:

$$x_2 = x_6 = 0.$$

Ответ: $x_1 = 14$, $x_2 = 0$, $x_3 = 10$, $x_4 = 15$, $x_5 = 8$, $x_6 = 0$.

Замечание. В эффективности описанного механизма могут возникнуть сомнения. Ведь Потребители могут пообещать большой эффект, получить ресурс, а затем не выполнить обещанного. Поэтому при реальном применении конкурсного механизма необходима действенная система контроля (возможно, поэтапный контроль для проектов с длительным временем реализации).

21.1.5. Механизм открытого управления

Во всех рассмотренных выше механизмах распределения ресурсов Потребители могут добиться лучшего для себя решения Центра путем искажения информации. Таким образом, Центр не получает достоверных данных о запросах Потребителей.

Возможность эффективно управлять на основании недостоверной информации представляется, вообще говоря, сомнительной. Поэтому интересны механизмы открытого управления, идея которых заключается в создании для Потребителей стимулов к сообщению в заявке своих реальных потребностей.

Опишем один из возможных механизмов открытого управления. Распределение ресурсов проводится в несколько этапов. На первом этапе ресурс разделяется поровну между всеми Потребителями, т. е. по R/n каждому. Если заявки каких-либо Потребителей оказались не больше чем R/n, то они полностью удовлетворяются. Тем самым число Потребителей уменьшается до n_1, уменьшается и ресурс Центра — до R_1. На втором этапе ресурс разделяется поровну между оставшимися n_1 Потребителями и т. д.

На каком-то этапе оказывается, что, разделив ресурс поровну между оставшимися Потребителями, не удается удовлетворить ни одной заявки. Тогда все эти Потребители получают поровну.

Пример 4. Восемь Потребителей подали Центру свои заявки. Они таковы: 12, 3, 6, 1, 5, 7, 10, 2. Центр обладает ресурсом $R = 40$. Требуется распределить этот ресурс в соответствии с вышеописанным механизмом.

Решение. В данном случае на первом этапе получается следующее ($R/n = 5$):

$$\begin{array}{cccccccccc} s_1 = 12; & s_2 = 3; & s_3 = 6; & s_4 = 1; & s_5 = 5; & s_6 = 7; & s_7 = 10; & s_8 = 2; & R = 40. \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & \end{array}$$

Видно, что можно удовлетворить заявки второго, четвертого, пятого и восьмого Потребителей:

$$\boxed{x_2 = 3} \quad \boxed{x_4 = 1} \quad \boxed{x_5 = 5} \quad \boxed{x_8 = 2}\,.$$

При этом $R_1 = 40 - 3 - 1 - 5 - 2 = 29$, $n_1 = 4$.

На втором этапе имеем ($R_1/n_1 = 7\frac{1}{4}$):

$$\begin{array}{ccccc} s_1 = 12; & s_3 = 6; & s_6 = 7; & s_7 = 10; & R = 29. \\ 7\frac{1}{4} & 7\frac{1}{4} & 7\frac{1}{4} & 7\frac{1}{4} & \end{array}$$

Можно удовлетворить заявки третьего и шестого Потребителей:

$$\boxed{x_3 = 6} \quad \boxed{x_6 = 7}\,.$$

При этом $R_2 = 29 - 6 - 7 = 16$, $n_2 = 2$.

На третьем этапе имеем ($R_2/n_2 = 8$):

$$\begin{array}{ccc} s_1 = 12; & s_7 = 10; & R = 16. \\ 8 & 8 & \end{array}$$

Обе оставшиеся заявки превышают 8, поэтому первый и седьмой Потребители получают по 8 единиц ресурса:

$$\boxed{x_1 = 8} \quad \boxed{x_7 = 8}\,.$$

Ответ: $x_1 = 8$, $x_2 = 3$, $x_3 = 6$, $x_4 = 1$, $x_5 = 5$, $x_6 = 7$, $x_7 = 8$, $x_8 = 2$.

Описанный механизм является механизмом открытого управления. Действительно, в конечном счете все Потребители делятся на приоритетных (которые получили столько, сколько просили) и неприоритетных (к последним в приведенном примере относятся первый и восьмой Потребители). Приоритетные получают столько, сколько просят, поэтому им не имеет смысла искажать свои реальные потребности. Неприоритетные же, как нетрудно видеть, не могут увеличить выделенный им ресурс ни повышая, ни понижая свою заявку.

Таким образом, при распределении ресурсов в соответствии с описанным механизмом Центр получает достоверную информацию о реальных запросах Потребителей.

21.2. Открытое управление и экспертный опрос

Если требуется определить объем финансирования крупного проекта, то часто прибегают к проведению экспертного опроса. Мы рассмотрим следующую процедуру опроса. Каждому из n экспертов предлагается сообщить число s из отрезка $[d; D]$, после чего на основании экспертных оценок определяется итоговое решение x. Задача состоит как раз в том, чтобы определить число x, исходя из заданных s_i $(i = 1, 2, \ldots, n)$.

На первый взгляд кажется, что наилучшее решение здесь — взять в качестве итогового решения среднее арифметическое мнений экспертов

$$x = \frac{1}{n}\sum_{i=1}^{n} s_i. \tag{4}$$

Однако у такого решения есть существенный недостаток.

Дело состоит в следующем. У каждого эксперта есть мнение r_i относительно объема финансирования. И если эксперт каким-либо образом заинтересован в том, чтобы итоговая оценка x совпала с его мнением r_i, то он может попытаться добиться этого совпадения, сообщая оценку $s_i \neq r_i$.

Пример 5. Пусть три эксперта имеют следующие мнения:

$$r_1 = 10, \quad r_2 = 10, \quad r_3 = 40.$$

Если каждый из них сообщит свое мнение без искажений, то при принятии решения по способу (4) результат будет таким:

$$x = \frac{10 + 10 + 40}{3} = 20.$$

Однако третий эксперт может (имея представление о мнениях остальных двух экспертов) сообщить оценку $s_3 = 100$. Тогда итоговый результат

$$x = \frac{10 + 10 + 100}{3} = 40.$$

как раз совпадет с его истинным мнением r_3.

Замечание. В теории коллективного принятия решений такой способ действий называется *манипулированием.* В свою очередь, если механизм коллективного принятия решений допускает манипулирование с чьей-либо стороны, то он называется *манипулируемым.* Рассмотренный только что пример показал, что механизм (4) является манипулируемым: искажая свои истинные предпочтения, можно приблизить итоговое коллективное решение к собственному истинному предпочтению.

Пример манипулирования со стороны избирателей можно было наблюдать в первом туре президентских выборов в России в 1996 г. Определенное число избирателей, считавших лучшей кандидатурой Явлинского, голосовали за Ельцина (с целью предотвратить победу Зюганова).

Вернемся к экспертному опросу. Говоря более строго, i-й эксперт решает задачу

$$|x - r_i| \longrightarrow \min_{s_i},$$

т. е. пытается минимизировать разность между итоговым решением x и своим истинным мнением r_i путем надлежащего выбора сообщаемой оценки s_i.

Опишем механизм выработки решения x^*, являющийся механизмом открытого управления (т. е. неманипулируемым механизмом).

Напомним, что эксперты сообщают свои оценки

$$s_i \in [d, D], \quad i = 1, 2, \ldots, n.$$

Будем считать, не ограничивая общности, что оценки экспертов расположены по неубыванию:

$$s_1 \leqslant s_2 \leqslant \cdots \leqslant s_n$$

(этого всегда можно добиться перенумерацией экспертов).

Вычисляются n вспомогательных чисел

$$v_i = D - (i-1)\frac{D-d}{n}, \quad i = 1, 2, \ldots, n$$

(эти числа делят отрезок $[d, D]$ на n равных частей). После этого для каждого i берется меньшее из двух чисел s_i и v_i:

$$\min\{s_i, v_i\}.$$

И наконец, из всех этих минимумов выбирается наибольший, который и является итоговым решением:

$$x^* = \max_{1 \leqslant i \leqslant n} \min\{s_i, v_i\}.$$

Можно доказать, что описанный механизм является механизмом открытого управления.

Пример 6. Пусть 6 экспертов сообщили следующие оценки из промежутка $[30, 90]$: 65, 90, 45, 80, 75, 90. Определить итоговое решение в соответствии с описанным механизмом.

Решение. Выпишем числа v_i (здесь $\frac{D-d}{n} = \frac{90-30}{6} = 10$):

$$v_1 = 90, \qquad v_4 = 90 - 30 = 60,$$
$$v_2 = 90 - 10 = 80, \qquad v_4 = 90 - 40 = 50,$$
$$v_3 = 90 - 20 = 70, \qquad v_4 = 90 - 50 = 40.$$

Дальнейшее удобно изобразить в виде таблицы, в первой строке которой записаны упорядоченные по неубыванию оценки экспертов:

$$\begin{array}{r|cccccc} s_i: & 45 & 65 & 75 & 80 & 90 & 90 \\ v_i: & 90 & 80 & 70 & 60 & 50 & 40 \\ \hline \min\{s_i, v_i\}: & 45 & 65 & \boxed{70} & 60 & 50 & 40 \end{array}$$

В качестве итогового решения берется максимальное число в последней строке:

$$x^* = 70.$$

Замечание. Во всех предыдущих рассуждениях квалификация экспертов предполагается одинаковой. Можно в случае необходимости вводить коэффициенты, позволяющие учитывать мнение разных экспертов различным образом — принципиально это ничего не меняет, лишь несколько усложняется вычисление итогового результата x^*.

21.3. Задания и ответы

1. Восемь Потребителей подали Центру заявки в размере 9, 18, 15, 14, 10, 13, 7, 14. Имеющийся в распоряжении Центра ресурс составляет 70. Как должен быть распределен этот ресурс в соответствии с механизмом прямых приоритетов?

Ответ: 6,3; 12,6; 10,5; 9,8; 7; 9,1; 4,9; 9,8.

2. Распределение ресурса производится в соответствии с механизмом обратных приоритетов. Приоритеты четырех Потребителей определяются числами 26, 18, 24, 20. Какими являются равновесные стратегии (заявки) Потребителей, если имеющийся в распоряжении Центра ресурс составляет 50?

Ответ: 13,6; 11,3; 13,1; 11,9.

3. Распределение ресурса осуществляется в соответствии с конкурсным механизмом. Пять Потребителей сообщили Центру свои заявки: 5, 8, 6, 9, 8 и показатели эффекта: 12, 21, 18, 23, 23 соответственно. Как должен быть распределен между Потребителями ресурс объемом 25?

Ответ: 0; 8; 6; 0; 8.

4. Восемь Потребителей подали Центру заявки 13, 10, 16, 19, 9, 12, 14, 11. Центр располагает ресурсом объемом 100. Как должен быть распределен этот ресурс в соответствии с механизмом открытого управления?

Ответ: 13; 10; 15,5; 15,5; 9; 12; 14; 11.

5. Восьми экспертам было предложено сообщить оценку объема финансирования из промежутка $[0, 80]$. Эксперты сообщили следующие оценки: 45, 10, 35, 80, 65, 35, 60, 55. Определите итоговое решение при помощи механизма открытого управления.

Ответ: 45.

Глава 22
ДИНАМИЧЕСКИЕ МОДЕЛИ

Модель — это представление объекта, системы или идеи в некоторой форме, отличной от самой целостности.

Р.Шеннон

22.1. Коротко о типах моделей

Не ставя перед собой задачи дать сколько-нибудь полную классификацию существующих моделей, коротко опишем некоторые их типы.

22.1.1. Физические модели

Так называют увеличенное или уменьшенное описание объекта или системы. Отличительная характеристика физической модели состоит в том, что в некотором смысле она выглядит как моделируемая целостность.

Наиболее известным примером физической модели является копия конструируемого самолета, выполненная с полным соблюдением пропорций, скажем 1 : 50. На одном из этапов разработки самолета новой конструкции возникает необходимость проверить его основные аэродинамические параметры. С этой целью подготовленную копию продувают в специальной (аэродинамической) трубе, а полученные показания затем тщательно исследуют. Выгодность такого подхода совершенно очевидна. И потому все ведущие самолетостроительные компании используют физические модели подобного рода при разработке каждого нового летательного аппарата.

Часто в аэродинамическую трубу помещают уменьшенные копии многоэтажных зданий, имитируя при этом розу ветров, характерную для той местности, где предполагается их строительство. Пользуются физическими моделями и в кораблестроении.

22.1.2. Аналоговые модели

Так называют модели, представляющие исследуемый объект аналогом, который ведет себя как реальный объект, но не выглядит как таковой.

Приведем два достаточно характерных примера.

Пример 1. График, иллюстрирующий соотношения между затраченными усилиями и результатами, является аналоговой моделью. График на рис. 1 показывает, как количество времени, отведенное студентом на подготовку к экзамену, влияет на его результат.

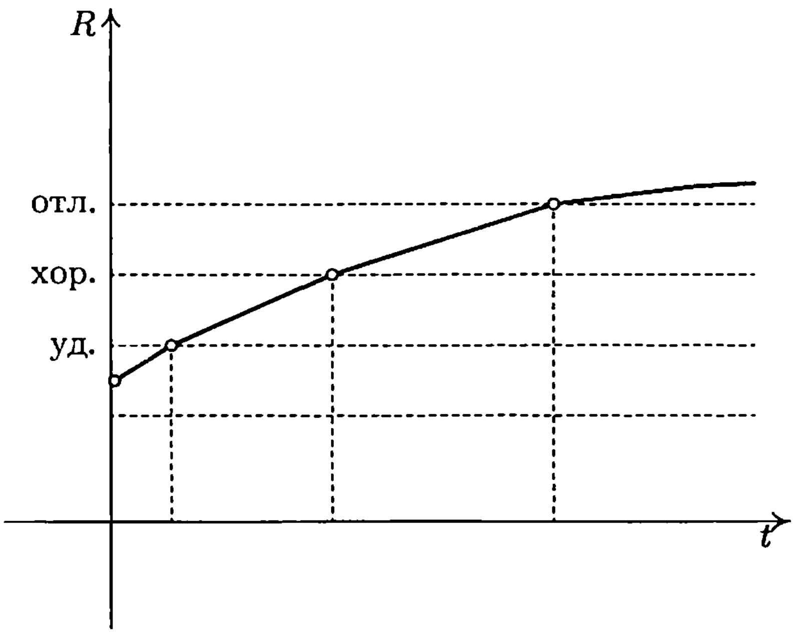


Рис. 1

Пример 2. Предположим, что нужно найти наиболее экономичный способ для регулярных известных поставок товаров в три города, построив для этого только один склад. Основное требование: место для склада должно быть таким, чтобы полные транспортные расходы были наименьшими (считается, что стоимость каждой перевозки равна произведению расстояния от склада до пункта назначения на общий вес перевозимых товаров и измеряется в тонна-километрах).

Наклеим карту местности на лист фанеры. Затем в месте нахождения каждого города пропилим сквозные отверстия, пропустим через них нити и привяжем к ним грузики, пропорциональные запросам товаров в этот город (рис. 2). Свяжем свободные концы нитей в один узел и отпустим. Под действием силы тяжести система придет

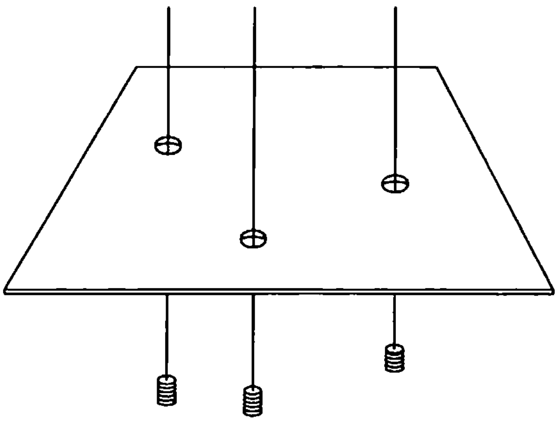


Рис. 2

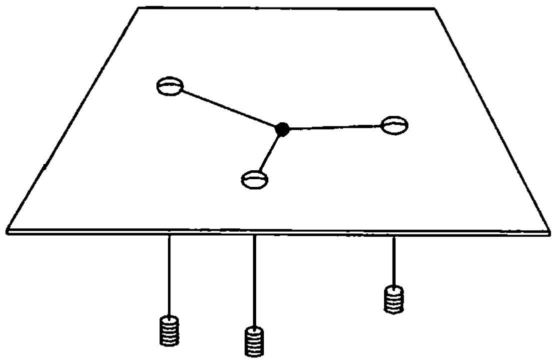


Рис. 3

в состояние равновесия. То место на листе фанеры, которое при этом займет узел, и будет соответствовать оптимальному расположению склада (рис. 3).

Замечание. Стоимость дорог, которые придется построить заново, мы для простоты рассуждений в расчет не принимаем.

22.1.3. Математические модели

Так называют модели, использующие для описания свойств и характеристик объекта или события математические символы и методы.

Если некоторую проблему удается перенести на язык формул, то она сильно упрощается. Математический подход прост еще и потому, что он подчиняется вполне определенным жестким правилам, которые нельзя отменить указом или иным способом. Сложность нашей жизни как раз и состоит в том, что многое, что в ней случается, нередко свободно от пут условностей.

Математика имеет дело с упрощенным описанием явлений. По существу, любая формула (или совокупность формул) представляет собой определенный этап в построении математической модели. Опыт показывает, что построить модель (написать уравнение) довольно легко. Трудно в этой модельной и, следовательно, упрощенной форме суметь передать суть изучаемого явления.

Для нахождения приемлемого или оптимального решения задачи полезно знать, в чем она состоит. Как ни просто и прозрачно данное утверждение, чересчур многие < ... > игнорируют очевидное (Р. Шеннон).

В предыдущих главах мы рассмотрели достаточное число разнообразных математических моделей, детерминированных, стохастических и игровых. В этой главе мы приведем примеры динамических моделей, на основании которых можно делать прогнозы на будущее и по-новому заглядывать в прошлое.

Итак, мы рассматриваем модели, в которые входят изменяющиеся во времени величины, уделяя основное внимание простейшим из них. Дело в том, что сами модельные уравнения (модели) строятся на основе простых и зачастую почти очевидных соображений. Именно анализ предлагаемых уравнений позволяет как-то оценить степень их адекватности описываемым ими обстоятельствам.

22.2. Модель народонаселения

Интересно, что построить математическую модель часто совсем нетрудно. Нередко для этого используются самые простые и легкообъяснимые предположения.

Опишем, как это можно сделать, на одном почти реальном примере.

Представим себе следующую картину. Середина XVIII в. Центральная Европа. Приход в глубинке. Церковь. Прихожане — жители окрестных деревень. Приходский священник замечает, что храм стал тесноват для богослужений: возросло число прихожан. Священник размышляет: если число прихожан будет увеличиваться и в будущем, то придется строить новую церковь, для чего понадобятся средства, и немалые.

Священик понимает, что срок, за который должен быть построен храм, и его размеры во многом зависят от того, как именно будет изменяться число окрестных жителей. И он решает попытаться рассчитать это.

Попробуем и мы изложить возможный ход его рассуждений, пользуясь современными обозначениями и языком.

Обозначим через x_n количество прихожан к концу n-го года. Их численность через год, т. е. к концу $(n+1)$-го года, естественно обозначить через x_{n+1}. Тогда изменение численности за этот год можно описать разностью

$$\Delta x_n = x_{n+1} - x_n.$$

Оно происходит по двум естественным причинам — люди рождаются и умирают (для простоты будем считать, что вирус миграций эту местность тогда еще не поразил). Определить число родившихся и число умерших за год по приходским книгам особого труда не составляет. Подсчитывая число родившихся и умерших в разные годы, священник решает сопоставить полученные числа

$$b_1, \dots, b_k$$

и

$$d_1, \dots, d_k$$

с общим числом прихожан за эти годы

$$x_1, \dots, x_k$$

и замечает, что отношения

$$\frac{b_1}{x_1}, \dots, \frac{b_k}{x_k}$$

год от года различаются весьма мало. То же касается и отношений

$$\frac{d_1}{x_1}, \dots, \frac{d_k}{x_k}.$$

Для простоты расчетов будем считать эти отношения постоянными и обозначим их через α и β соответственно.

Тем самым число родившихся в n-м году оказывается равным

$$\alpha x_n,$$

число умерших —

$$\beta x_n,$$

а изменение численности по естественным причинам составляет

$$+\alpha x_n - \beta x_n.$$

В результате мы приходим к соотношению

$$\Delta x_n = \alpha x_n - \beta x_n,$$

или подробнее:

$$x_{n+1} = x_n + \alpha x_n - \beta x_n.$$

Положим

$$\gamma = 1 + \alpha - \beta.$$

Тогда интересующая нас формула примет вид

$$x_{n+1} = \gamma x_n. \qquad (1)$$

Модель построена.

Попробуем теперь разобраться с тем, что же получилось, т. е. проанализировать построенную модель.

Возможны три случая:

1) $\gamma > 1$ ($\delta = \alpha - \beta > 0$ — рождается больше, чем умирает) и численность прихожан растет год от года,

2) $\gamma = 1$ ($\delta = \alpha - \beta = 0$ — умирает столько же, сколько рождается) и численность прихожан год от года остается неизменной,

3) $\gamma < 1$ ($\delta = \alpha - \beta < 0$ — умирает больше, чем рождается) и численность прихожан неуклонно снижается.

Так как побудительным мотивом для построения модели было желание узнать, как быстро будет расти число прихожан, начнем с рассмотрения случая 1.

Случай 1. Итак, численность прихожан растет. Но как, насколько быстро?

Здесь самое время кратко вспомнить поучительную историю (печальную притчу) о безвестном изобретателе шахмат.

Говорят, что игра очень понравилась богатому и всесильному магарадже, который тут же решил наградить изобретателя и щедро предложил выбрать вознаграждение ему самому. Тот, как рассказывают, смахнув фигуры с шахматной доски, положил на 1-ю клетку одно пшеничное зернышко, на 2-ю — два зернышка, на 3-ю — четыре зернышка, на 4-ю — восемь зернышек (рис. 4) и предложил магарадже, чтобы он отдал распоряжение слугам выкладывать зерна пшеницы на другие клетки шахматной доски по предложенному

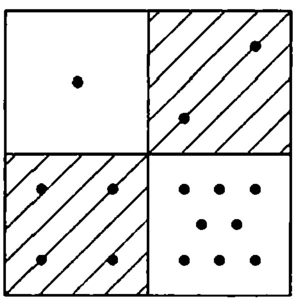


Рис. 4

закону, т. е. так:

$$1, 2, 4, 8, 16, \ldots, 2^{63}.$$

Магараджу эта простая просьба почти обидела, и он согласился выполнить ее далеко не сразу. Но изобретатель настаивал. Магараджа приказал. И слуги тут же кинулись исполнять это "легкое" задание. Нужно ли говорить, что выполнить распоряжение магараджи им не удалось. Дело в том, что общее количество зерен пшеницы на шахматной доске должно было быть равным

$$2^{64} - 1,$$

что намного превышает выращиваемое сейчас во всем мире за год.

Закончим притчу совсем коротко: магараджа оказался в непривычном для себя положении — он прилюдно дал обещание и не смог его выполнить. Виновного, впрочем, тут же и нашли. Возможно, именно поэтому история и не сохранила имени изобретателя шахмат.

Попробуем, однако, изобразить на графике, как быстро растет число зерен в каждой следующей клетке, для большей наглядности соединяя соседние точки (рис. 5).

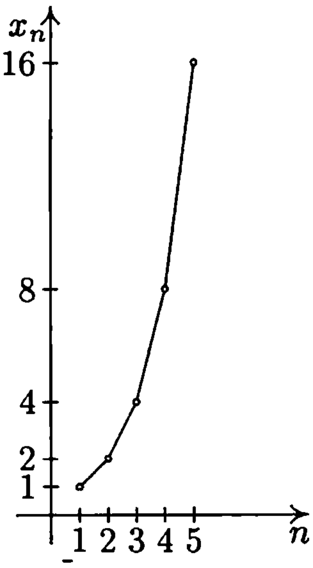


Рис. 5

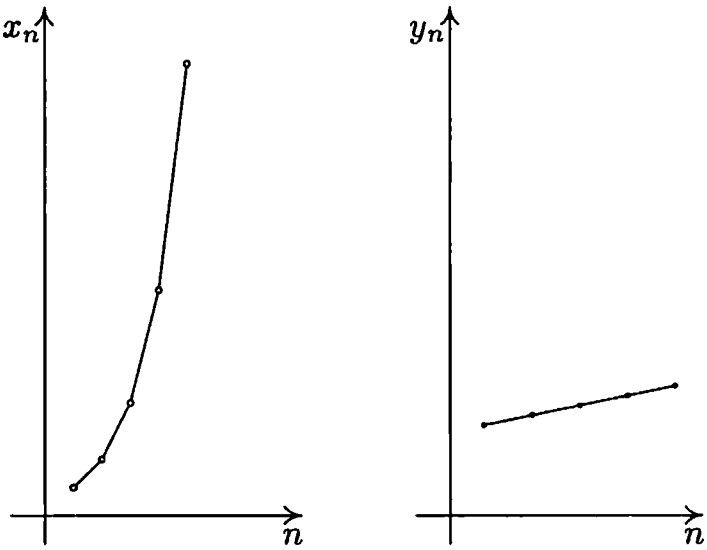


Рис. 6

Правило, предложенное изобретателем шахмат,

$$x_{n+1} = 2x_n,$$

является частным случаем формулы (1) при $\gamma = 2$ и, так же как и она, описывает закон, следуя которому мы получаем последовательность чисел, образующих геометрическую прогрессию.

При любом $\gamma > 1$ картинка, иллюстрирующая изменение x_n, имеет похожий вид — x_n будет расти *экспоненциально.*

В 1820 г. в Лондоне Т. Р. Мальтусом была опубликована работа "Principles of political economy considered with a view to their practical application" (в русском переводе — "Опыт о законе народонаселения ..." Т. 1–2. СПб., 1868), в которой, в частности, говорилось о том, что в силу биологических особенностей людей население имеет тенденцию размножаться по закону геометрической прогрессии:

$$x_{n+1} = \gamma x_n, \qquad \gamma > 1,$$

в то время как средства существования могут увеличиваться лишь по закону арифметической прогрессии:

$$y_{n+1} = y_n + d, \qquad d > 0.$$

Такое различие в скорости изменения величин, непосредственно связанных с проблемами выживаемости популяции (рис. 6), не могло остаться незамеченным и вызвало довольно жесткую критику и сильно политизированную полемику в соответствующих кругах.

Попробуем извлечь из самого факта критики полезный для нас вывод об адекватности построенной модели (1).

Разумеется, при попытке упрощенного описания ситуации некоторыми обстоятельствами приходится пренебрегать, считая их несущественными. Однако единого взгляда на то, что́ именно существенно, а что́ не очень, по-видимому, нет. Можно, например, не обращать внимания на то, что начался дождик. Но согласитесь, что одно дело пробежать под накрапывающим дождем сотню метров, и совсем другое — часовая прогулка под таким дождем без зонта.

Нечто аналогичное мы наблюдаем и здесь: при расчете на 3–4 года вперед формула (1) работает достаточно хорошо, но долгосрочный прогноз, основанный на ней, оказывается ошибочным.

Важный вывод. Предлагая построенную или выбранную вами модель, вы непременно должны указать пределы, в которых ею можно пользоваться, и предупредить о том, что нарушение этих ограничений может привести (и, скорее всего, приведет) к серьезным ошибкам. Коротко говоря, у каждой модели есть свой ресурс.

Покупая блузку или рубашку, мы привыкли к наличию меток, на которых указаны максимально допустимая температура глажения, дозволенные виды стирки и т. п. Это, конечно, ни в коей мере не означает, что вам запрещается, взяв докрасна раскаленный утюг, пройтись им раз-другой по ткани. Такое вы сделать можете. Но вот захотите ли вы носить блузку или рубашку после такого глажения?

Случай 2. Численность населения не изменяется (рис. 7).

Случай 3. Население вымирает (рис. 8).

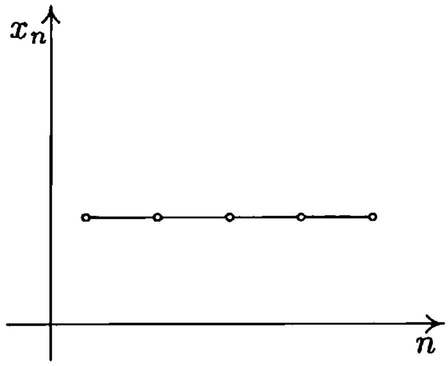


Рис. 7

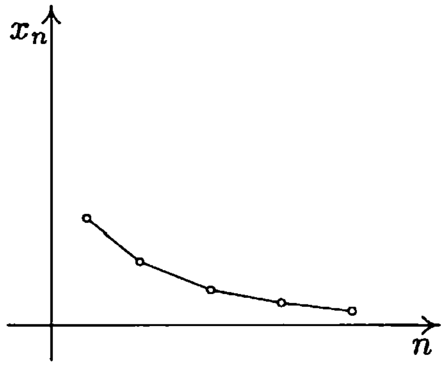


Рис. 8

Мы умышленно весьма подробно остановились на описании модели народонаселения, во-первых, потому, что она является одной из первых моделей подобного рода, и, во-вторых, чтобы на ее примере показать, через какие основные этапы проходит решение задачи построения математической модели.

Замечание 1. Очень часто, описывая эту модель народонаселения, привлекают ее дифференциальный вариант:

$$x' = \delta x$$

(здесь $x = x(t)$ — зависящая от времени численность популяции, x' — производная по времени, δ — постоянная величина).

Замечание 2. При больших значениях x конкурентная борьба за средства существования приводит к уменьшению δ, и эта жесткая модель должна быть заменена более мягкой моделью:

$$x' = \delta(x)x,$$

в которой коэффициент δ зависит от численности населения. В простейшем случае эта зависимость описывается так:

$$\delta(x) = a - bx,$$

где a и b — постоянные числа, а соответствующее уравнение принимает вид

$$x' = ax - bx^2.$$

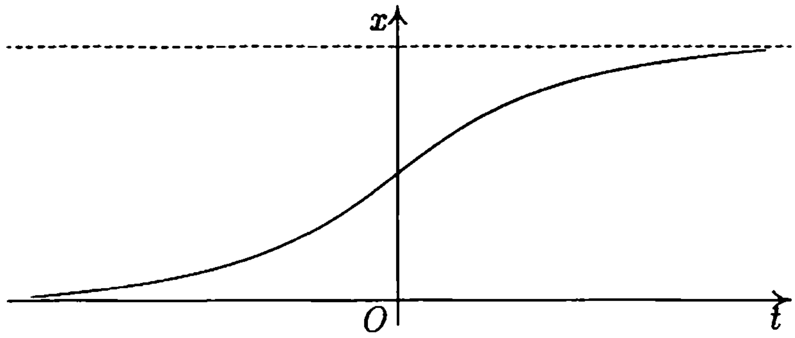


Рис. 9

И мы приходим к более сложной, так называемой *логистической* модели, которая описывает динамику популяции уже достаточно хорошо. Анализ логистической кривой (рис. 9) весьма поучителен, и его проведение может быть любопытно читателю.

Логистическая модель хорошо описывает и другие процессы, например эффективность рекламы.

22.3. Модель мобилизации

Под термином *политическая*, или *социальная, мобилизация* понимается вовлечение людей в партию или в число ее сторонников, в какое-либо общественное движение и т. п.

Вследствие того что текущий уровень мобилизации тесно связан с прошлым ее уровнем, а будущая мобилизация зависит от сегодняшних успехов пропагандистской кампании, ясно, что при построении соответствующей модели необходимо учитывать временной фактор. Иными словами, нужно понимать, что искомая модель должна быть *динамической.*

Постановка задачи

Отразить логику изменения уровня мобилизации в данном регионе между двумя соседними моментами времени, скажем за месяц (за год, неделю, день и т. п.).

Построение модели

Примем за единицу ту часть населения, для которой мобилизация данного типа имеет смысл. Пусть M_n — доля мобилизованного населения в момент времени $t_n = n$. Тогда доля немобилизованного населения будет равна $1 - M_n$ (рис. 10).

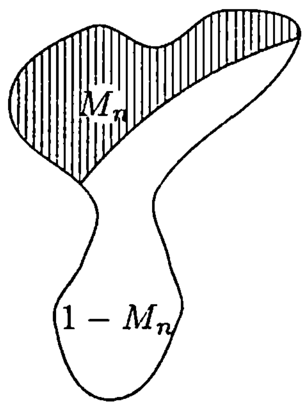


Рис. 10

За месяц уровень мобилизации может измениться по двум основным причинам:

1) часть населения удалось привлечь дополнительно; ясно, что эта величина тем больше, чем выше доля еще несагитированного населения на момент $t_n = n$, и поэтому можно считать ее равной

$$\alpha(1 - M_n)$$

(здесь $\alpha > 0$ — коэффициент агитируемости, постоянный для данного региона);

2) часть населения убыла (по разным причинам); ясно, что это уменьшает долю сагитированного населения тем больше, чем выше была эта доля на момент $t_n = n$, и поэтому потери, связанные с выбытием, можно считать равными

$$\beta M_n$$

(здесь $\beta > 0$ — постоянный коэффициент выбытия).

Подчеркнем, что числовые параметры α и β отражают пропорциональное изменение интересов, взглядов и намерений соответствующих частей населения рассматриваемого региона.

Таким образом, изменение уровня мобилизации за единицу времени

$$\Delta M_n = M_{n+1} - M_n$$

равно разности между долей населения, привлеченного дополнительно, и долей выбывшего сагитированного населения:

$$M_{n+1} - M_n = \alpha(1 - M_n) - \beta M_n.$$

Это и есть *уравнение процесса мобилизации.* Модель мобилизации построена.

Последнее соотношение легко преобразуется к следующему виду:

$$M_{n+1} = \alpha + \gamma M_n, \tag{2}$$

где

$$\gamma = 1 - \alpha - \beta.$$

Замечание. Вспомогательный параметр γ не может быть больше 1 вследствие того, что исходные параметры α и β положительны.

Полученное уравнение (2) называется *линейным разностным уравнением с постоянными коэффициентами.*

С уравнениями подобного рода читатель, несомненно, уже не раз сталкивался, правда, по большей части в простейших вариантах.

Один из них (при $\gamma = 1$) описывает правило, по которому каждый член последовательности, начиная со второго, получается из предыдущего путем сложения с некоторым постоянным числом:

$$M_{n+1} = \alpha + M_n,$$

т. е. *арифметическую прогрессию.*

Второй (при $\alpha = 0$) описывает правило, по которому каждый член последовательности, начиная со второго, получается из предыдущего путем умножения на некоторое постоянное число:

$$M_{n+1} = \gamma M_n,$$

т. е. *геометрическую прогрессию.*

Предположим, что начальная доля привлеченного населения M_0 известна. Тогда уравнение (2) легко решается (для определенности считаем, что $\gamma \neq 1$). Имеем:

$$M_n = \alpha \frac{1-\gamma^n}{1-\gamma} + \gamma^n M_0.$$

Применение модели

Попробуем проанализировать возможности этой (построенной на основании простейших соображений) модели.

Начнем со случая $|\gamma| < 1$.

Для этого перепишем последнее соотношение в виде

$$M_n = M^* + \gamma^n (M_0 - M^*),$$

где через M^* обозначена следующая величина:

$$M^* = \frac{\alpha}{1-\gamma}.$$

Замечание. Тот же результат получается, если в уравнении (2) положить

$$M_{n+1} = M_n = M^*.$$

В самом деле, тогда получим

$$M^* = \alpha + \gamma M^*,$$

откуда

$$M^* = \frac{\alpha}{1-\gamma}.$$

Найденная величина M^* не зависит от начального значения M_0, выражается через исходные параметры α и β по формуле

$$M^* = \frac{\alpha}{\alpha+\beta}$$

и, следовательно, подчинена условию

$$0 < M^* < 1.$$

Для придания полученной формуле большей наглядности вновь воспользуемся методом координат.

На рис. 11 показаны области возможных значений вспомогательного параметра γ, на рис. 12 — исходных параметров α и β, а на рис. 13–15 — соответствующие им наборы значений M_n при разных n, M_0 и M^* (для удобства восприятия соседние точки (n, M_n) и $(n+1, M_{n+1})$ соединены прямолинейными отрезками).

Случай $\gamma < -1$ проиллюстрирован на рис. 16.

Конечно, на этих рисунках представлена качественная картина. Но ничто не мешает взять вполне конкретные значения величин M_0, α и β и подробно рассчитать соответствующую ситуацию.

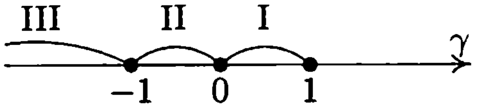


Рис. 11

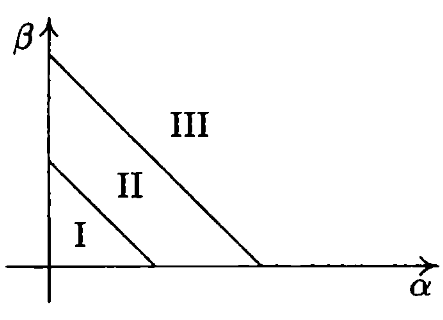


Рис. 12

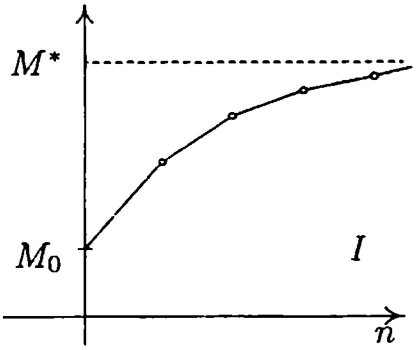


Рис. 13

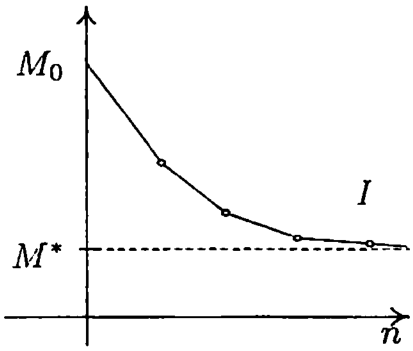


Рис. 14

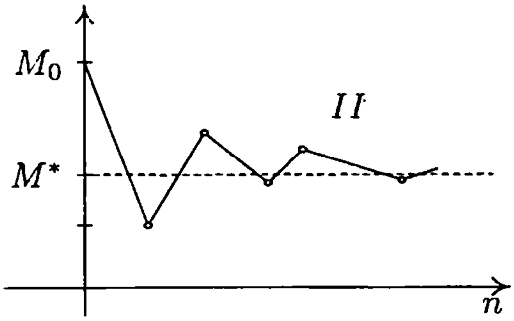


Рис. 15

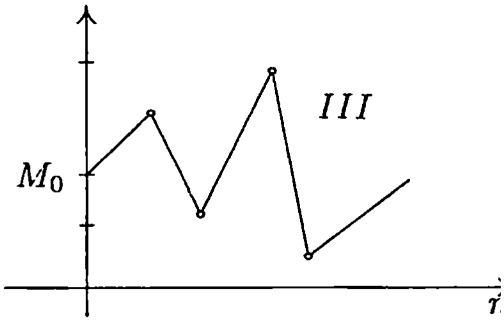


Рис. 16

Например, для

$$M_0 = \frac{1}{4}, \quad \alpha = \frac{1}{4}, \quad \beta = \frac{1}{4}$$

имеем

$$M_1 = \frac{3}{8}, \ M_2 = \frac{7}{16}, \ M_3 = \frac{15}{32}, \ M_4 = \frac{31}{64}, \ \ldots$$

(рис. 17).

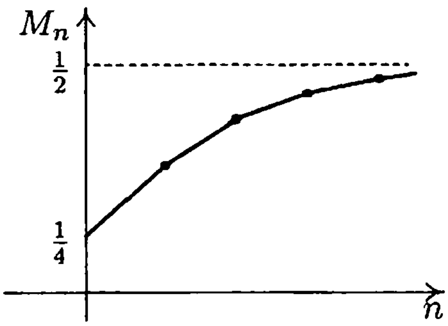


Рис. 17

Интересно отметить, что построенная модель, несмотря на простоту подходов и рассуждений, довольно хорошо отражает реальные процессы. Так, предложенная модель мобилизации использовалась для изучении динамики числа голосов, поданных за демократическую партию в Лейк Кантри (США) в 1920–1968 гг., и оказалось, что она достаточно хорошо описывает качественные характеристики процесса мобилизации.

22.4. Модель гонки вооружений

Рассмотрим конфликтную ситуацию, в которой могут оказаться две соседние страны, для определенности названные странами X и Y.

Обозначим через $x=x(t)$ расходы на вооружение страны X и через $y=y(t)$ расходы на вооружение страны Y в момент времени $t\geqslant 0$.

Предположение 1

Страна X вооружается, опасаясь потенциальной угрозы войны со стороны страны Y, которая в свою очередь, зная о росте затрат на вооружение страны X, также увеличивает свои расходы на вооружение. Каждая страна изменяет скорость роста (или сокращения) вооружений пропорционально уровню затрат другой. В простейшем случае это можно описать так:

$$\begin{cases} x' &= \alpha y, \\ y' &= \beta x, \end{cases}$$

где α и β — положительные постоянные.

Однако написанные уравнения имеют очевидный недостаток — уровень вооружения ничем не лимитируется. Поэтому правые части этих уравнений нуждаются в естественной корректировке.

Предположение 2

Чем больше текущий уровень расходов страны на оборону, тем меньше скорость его роста. Это позволяет внести в предыдущую систему следующие изменения:

$$\begin{cases} x' = \alpha y - \gamma x, \\ y' = \beta x - \delta y, \end{cases}$$

где γ и δ — положительные постоянные.

Предположение 3

Каждая страна наращивает вооружение, руководствуясь своими державными притязаниями и враждебностью к соседней стране, даже если эта страна не угрожает существованию данной. Обозначим соответствующие претензии через a и b (a и b — положительные постоянные). В случае если постоянные a и b отрицательны, их можно назвать коэффициентами доброй воли.

Основываясь на всех трех предположениях, в результате получаем следующую систему уравнений:

$$\begin{cases} x' = \alpha y - \gamma x + a, \\ y' = \beta x - \delta y + b. \end{cases}$$

Модель гонки вооружений построена.

Решением полученной системы являются функции $x(t)$ и $y(t)$, определяемые для данных начальных условий $x_0 \geqslant 0$ и $y_0 \geqslant 0$ (начального состояния гонки вооружений).

Проанализируем полученную систему, предполагая, что уровни затрат обеих стран на вооружение не зависят от времени (являются стационарными). Это означает, что

$$x' = 0, \quad y' = 0,$$

или по-иному:

$$\begin{cases} \alpha y - \gamma x + a = 0, \\ \beta x - \delta y + b = 0. \end{cases}$$

Рассмотрим конкретный пример.

Пример 3. Пусть система уравнений гонки вооружений имеет следующий вид:

$$\begin{cases} x' = 3y - 5x + 15, \\ y' = 3x - 4y + 12. \end{cases}$$

Если скорости изменения величин x и y равны нулю, то эти величины с необходимостью связаны условиями

$$\begin{aligned} &(\text{а}) : 3y - 5x + 15 = 0, \\ &(\text{б}) : 3x - 4y + 12 = 0. \end{aligned}$$

Каждое из этих уравнений описывает прямую на плоскости (x, y), и точка пересечения этих прямых

$$M^* \left(\frac{96}{11}, \frac{105}{11}\right)$$

лежит в первой четверти (рис. 18).

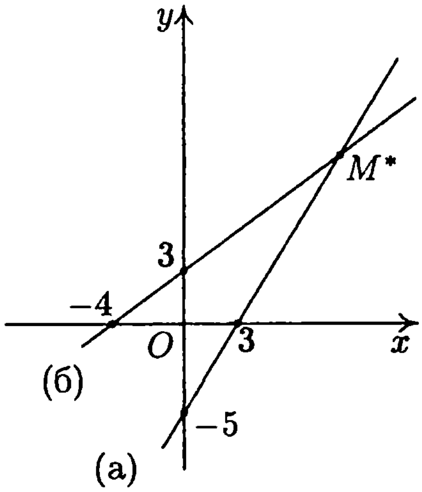


Рис. 18

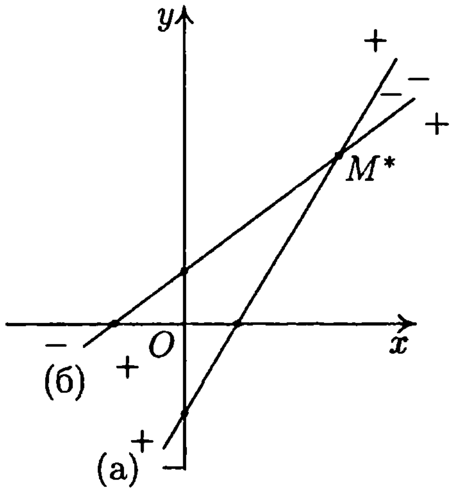


Рис. 19

Прямая, заданная уравнением (а), разбивает плоскость, и начальная точка $O(0,0)$ лежит в положительной полуплоскости. В рассматриваемом случае то же справедливо и для прямой, заданной уравнением (б) (рис. 19).

Тем самым первая четверть (а нас интересует только она, так как всегда $x \geqslant 0$ и $y \geqslant 0$) разбивается на четыре области, которые удобно обозначить так:

$$\mathrm{I}-(+,+),\quad \mathrm{II}-(-,+),\quad \mathrm{III}-(-,-),\quad \mathrm{IV}-(+,-).$$

Пусть начальное состояние (x_0, y_0) находится в области I. Тогда выполнены неравенства

$$\begin{aligned}
&(\text{а}):\quad 3y_0 - 5x_0 + 15 > 0,\\
&(\text{б}):\quad 3x_0 - 4y_0 + 12 > 0,
\end{aligned}$$

из которых следует, что скорости x' и y' в этой точке положительны:

$$x' > 0, \qquad y' > 0$$

и, значит, обе величины (x и y) должны возрастать (рис. 20).

Таким образом, с течением времени в области I решение приходит в точку равновесия.

Подобным же образом анализируя возможные расположения начального состояния в областях II, III и IV, получим в итоге, что стабильное состояние (баланс сил) достигается независимо от начальных уровней вооружения стран X и Y. Отличие состоит лишь в том, что если переход к стационарному состоянию из области I сопровождается одновременным увеличением уровней вооруженности,

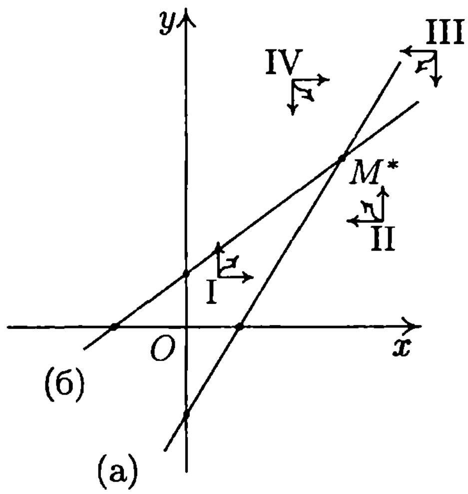


Рис. 20

то из области III — их одновременным снижением; для областей II и IV иная ситуация — одна из сторон наращивает свое вооружение, в то время как другая разоружается.

Возможны и другие случаи (рис. 21).

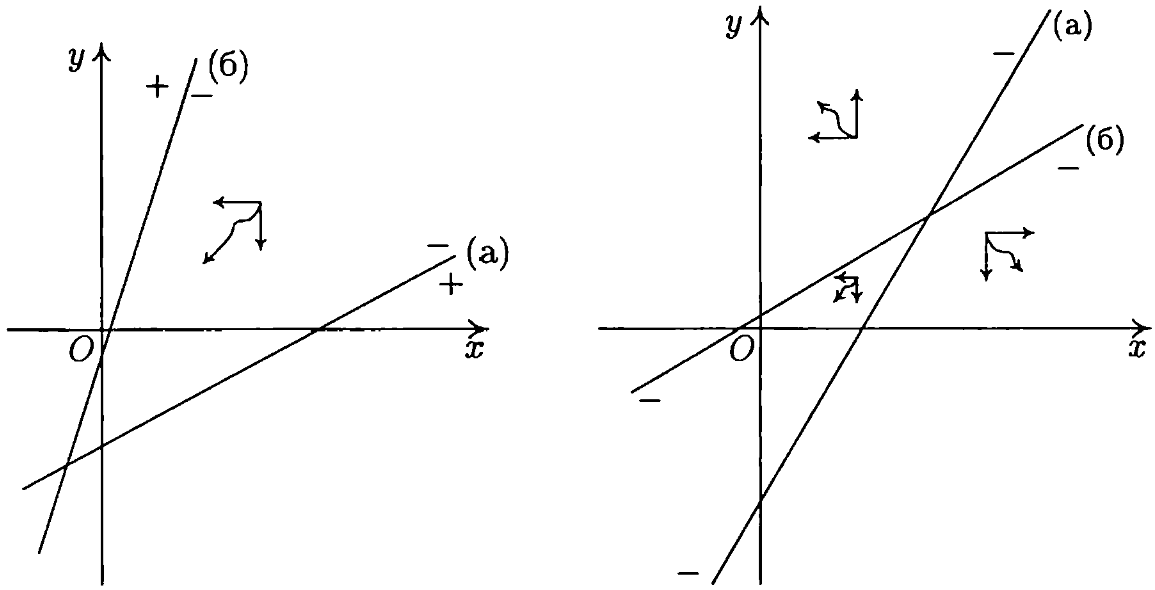


Рис. 21

Интересно отметить, что возможности построенной модели проверялись на реальной ситуации — гонке вооружений перед первой мировой войной. Проведенные исследования показали, что, несмотря на свою простоту, эта модель достаточно достоверно описывает положение дел в Европе в 1909–1913 гг.

В завершение этого раздела процитируем высказывание Т. Саати об этой модели: “Модель представляется гораздо более убедительной, если вместо вооружений провести на ней изучение проблем угрозы, поскольку люди реагируют на абсолютный уровень враждебности, проявляемый по отношению к ним другими, и испытывают чувство тревоги в степени, пропорциональной уровню враждебности, которую они испытывают сами”.

22.5. Модель хищник – жертва

Выше рассказывалось о беспрепятственном размножении популяции. Однако в реальных обстоятельствах популяция сосуществует с другими популяциями, находясь с ними в самых разных взаимоотношениях.

Здесь мы коротко рассмотрим антагонистическую пару *хищник – жертва* (это может быть и пара рысь – заяц и пара божья коровка – тля) и попытаемся проследить, как может изменяться со временем численность обеих взаимодействующих сторон.

Популяция жертвы может существовать сама по себе, в то время как популяция хищника — только за счет жертвы.

Обозначим численность популяции жертвы через x, а численность популяции хищника через y.

В отсутствие хищника жертва размножается согласно уравнению

$$x' = \alpha x, \qquad \alpha > 0,$$

а хищник в отсутствие жертвы вымирает по закону

$$y' = -\beta y, \qquad \beta > 0.$$

Хищник съедает тем больше жертвы, чем ее больше и чем более многочислен он сам. Поэтому при наличии хищника численность жертвы меняется по закону

$$x' = \alpha x - \gamma xy, \qquad \gamma > 0.$$

Съеденное количество жертвы способствует размножению хищника, что можно записать так:

$$y' = -\beta y + \delta xy, \qquad \delta > 0.$$

Таким образом, мы получаем систему уравнений

$$\begin{cases} x' = \quad \alpha x - \gamma xy, \\ y' = -\beta y + \delta xy, \end{cases}$$

причем

$$x \geqslant 0, \qquad y \geqslant 0.$$

Модель *хищник – жертва* построена.

Как и в предыдущей модели, наибольший интерес для нас представляет точка равновесия (x^*, y^*), где x^* и y^* — отличное от нуля решение системы уравнений

$$\begin{cases} \quad \alpha x - \gamma xy = 0, \\ -\beta y + \delta xy = 0, \end{cases}$$

или

$$x(\alpha - \gamma y) = 0, \qquad y(-\beta + \delta x) = 0.$$

Эта система получается из условия стабильности численности обеих популяций

$$x' = 0, \qquad y' = 0.$$

Координаты точки равновесия — она является точкой пересечения прямых

$$\alpha - \gamma y = 0, \tag{3}$$

$$-\beta + \delta x = 0 \tag{4}$$

— легко вычисляются:

$$x^* = \frac{\beta}{\delta}, \qquad y^* = \frac{\alpha}{\gamma}$$

(рис. 22).

Начало координат $O(0,0)$ лежит в положительной полуплоскости относительно горизонтальной прямой, задаваемой уравнением (3), а относительно вертикальной прямой, задаваемой уравнением (4), — в отрицательной полуплоскости (рис. 23).

Тем самым первая четверть (а нас интересует только она, так как $x > 0$ и $y > 0$) разбивается на четыре области, которые удобно обозначить так:

$$\mathrm{I} - (+,+), \quad \mathrm{II} - (-,+), \quad \mathrm{III} - (-,-), \quad \mathrm{IV} - (+,-).$$

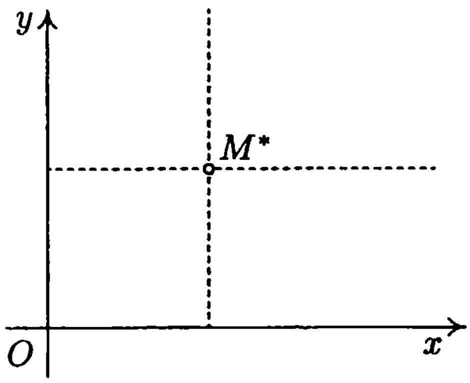


Рис. 22

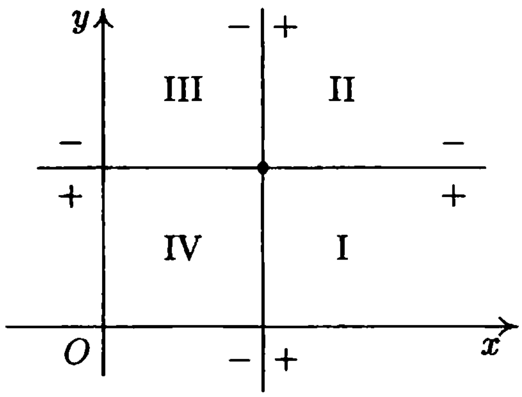


Рис. 23

Пусть начальное состояние $Q(x_0, y_0)$ находится в области IV. Тогда выполнены неравенства

$$\alpha - \gamma y_0 > 0, \qquad -\beta + \delta x_0 < 0,$$

из которых следует, что скорости x' и y' в этой точке должны быть разных знаков,

$$x' > 0, \qquad y' < 0,$$

и, значит, величина x должна возрастать, а величина y убывать.

Подобным же образом анализируя поведение x и y в областях II, III и IV, получим в итоге картину, изображенную на рис. 24.

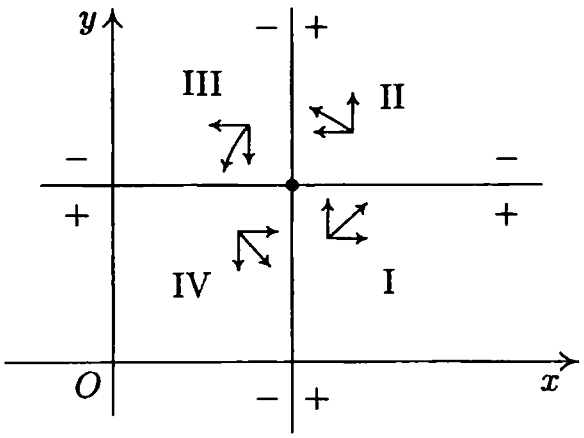


Рис. 24

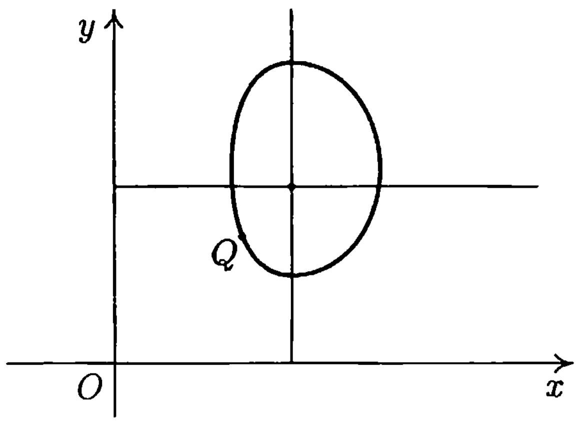


Рис. 25

Тем самым начальное состояние Q приводит к периодическому колебанию численности как жертвы, так и хищника, так что по прошествии какого-то времени система вновь возвращается в состояние Q (рис. 25).

Как показывают наблюдения, несмотря на свою простоту, предложенная модель качественно верно отражает колебательный характер численности в системе хищник – жертва (рис. 26).

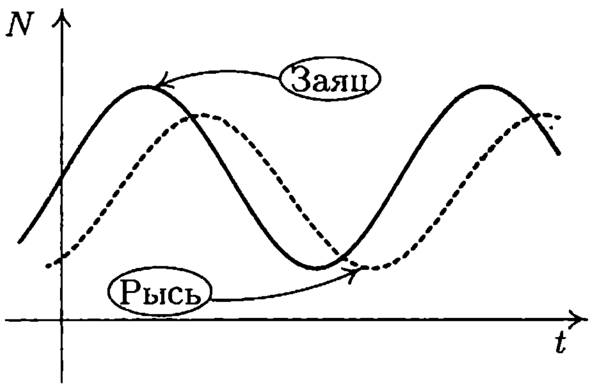


Рис. 26

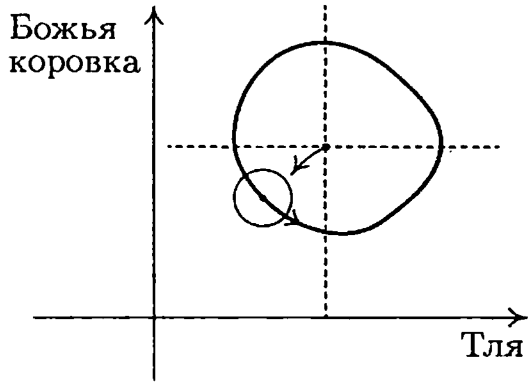


Рис. 27

Реальные наблюдения. Вмешиваться в действия непонятных нам законов природы иногда довольно опасно — применение инсектицидов (если только они не уничтожают насекомых практически полностью) в конечном счете приводит к увеличению популяции тех насекомых, численность которых находится под контролем других насекомых-хищников.

Случайно попавшая в Америку тля поставила под угрозу все производство цитрусовых. Вскоре туда же был завезен ее естественный враг — божья коровка, которая немедленно принялась за дело и сильно сократила популяцию тли. Чтобы ускорить процесс уничтожения, фермеры применили ДДТ, но в результате количество тли увеличилось, что, глядя на рис. 27, нетрудно предугадать.

22.6. Заключение

Построение модели опирается на значительное упрощение изучаемой ситуации, и, следовательно, к получаемым на ее основе выводам нужно относиться достаточно осторожно — модель может не все. Вместе с тем даже весьма грубая на вид идеализация нередко позволяет глубже вникнуть в суть проблемы. Пробуя как-то влиять на параметры модели (выбирать их, управлять ими), мы получаем возможность подвергнуть исследуемое явление качественному анализу и сделать выводы общего характера.

Глава 23
О ТОМ, ЧТО НЕ ВОШЛО В ЭТУ КНИГУ

Предмет данной книги — применение количественных методов в управлении — не имеет четко очерченных границ. Возникают новые практические задачи, для их решения разрабатываются адекватные методы. Развитие компьютерных технологий также играет важную роль в изменении облика науки управления.

В рамках одной книги трудно осветить даже и некоторые устоявшиеся разделы. С целью как-то заполнить вынужденные пробелы мы кратко обозначим разделы, которые вполне могли бы быть включены как в книгу, так и в учебный курс.

Среди детерминированных методов отметим задачи так называемого МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ, когда требуется максимизировать (либо минимизировать) заданную целевую функцию при некоторых заданных ограничениях. Частными случаями являются рассмотренные в этой книге задача линейного программирования и детерминированные модели управления запасами.

В зависимости от вида ограничений и целевой функции выделяют задачи линейного, квадратичного, выпуклого программирования и др. Решение задач математического программирования обычно требует применения довольно сложного математического аппарата.

Несколько особняком стоит задача ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ, основные идеи которого доступно изложены в кн.: *Вентцель Е. С.* Исследование операций — задачи, принципы, методология. М.: Наука, 1980.

Принятие решений при помощи приоритетов и иерархий подробно описывается в кн.: *Саати Т.* Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1993.

Из стохастических методов отметим весьма широко применяемую ТЕОРИЮ МАССОВОГО ОБСЛУЖИВАНИЯ (в зарубежной литературе обычно называемую ТЕОРИЕЙ ОЧЕРЕДЕЙ, queue theory). Ситуации, допускающие применение теории массового обслуживания, разнообразны: покупатели в магазине, пациенты в приемной врача, автомобили у автозаправочной станции, телефонные звонки на АТС и т. п. См.: *Вентцель Е. С.* Указ. соч.; *Тернер Д.* Вероятность, статистика, исследование операций. М.: Статистика, 1976.

ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ исходит из того, что уже в процессе построения модели необходимо учитывать опыт и предпочтения лица, принимающего решение (ЛПР). В связи с этим весьма важными являются понятия *субъективной вероятности* и *полезности.* О теории принятия решений можно прочитать в кн.: Исследование операций: В 2 т. / Под ред. Дж. Моудера, С. Элмаграби. М.: Мир, 1988; *Райфа Г.* Анализ решений. М.: Наука, 1977.

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ — универсальный метод исследования систем, функционирование которых зависит от тех или иных случайных факторов (в частности, от реализации случайных величин). Имитационная модель последовательно, шаг за шагом воспроизводит процесс функционирования системы. Исследователь при этом имеет возможность наблюдать, какие значения принимают те или иные значимые параметры. Об имитационном моделировании см.: *Вентцель Е. С.* Указ. соч.; Исследование операций. Указ. соч.; *Эддоус М., Стэнсфилд Р.* Методы принятия решений. М.: Аудит, ЮНИТИ, 1997.

Из многочисленных СТАТИСТИЧЕСКИХ методов обработки информации в данной книге затронуты лишь некоторые. Более подробный обзор можно найти в кн.: *Тюрин Ю. Н., Макаров А. А.* Анализ данных на компьютере / Под ред. В. Э. Фигурнова. М.: Инфра-М, Финансы и статистика, 1995 (ее второе издание носит название “Статистический анализ данных на компьютере”).

Управлению организационными системами посвящена кн.: *Бурков В. Н., Ириков В. А.* Модели и методы управления организационными системами. М.: Наука, 1994.

ПРИЛОЖЕНИЕ

Таблица 1

Значения функции $\Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,1	0,040	1,1	0,364	2,1	0,482
0,2	0,079	1,2	0,385	2,2	0,486
0,3	0,118	1,3	0,403	2,3	0,489
0,4	0,155	1,4	0,419	2,4	0,492
0,5	0,192	1,5	0,433	2,5	0,494
0,6	0,226	1,6	0,445	2,6	0,495
0,7	0,258	1,7	0,455	2,7	0,497
0,8	0,288	1,8	0,464	2,8	0,497
0,9	0,316	1,9	0,471	2,9	0,498
1,0	0,341	2,0	0,477	3,0	0,499

Таблица 2

Критические значения $\chi^2_{\text{кр}}$

α \ k	1	2	3	4	5	6	7	8	9	10
0,01	6,8	9,2	11,3	13,3	15,1	16,8	18,5	20,1	21,7	23,2
0,05	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9	18,3