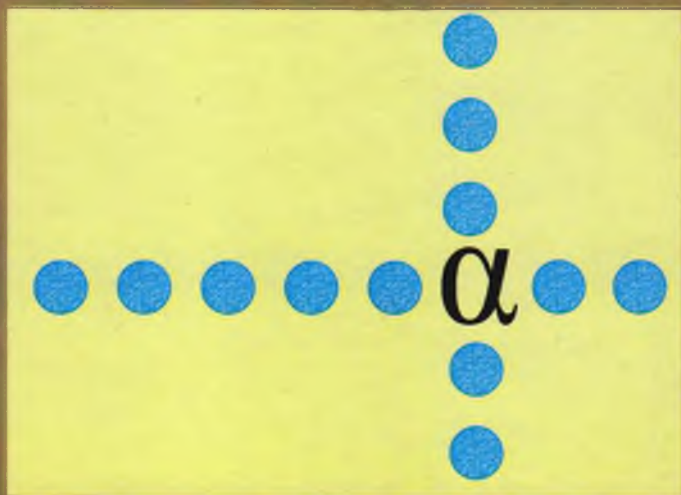


Ш.Ш. БАБАДЖАНОВ



МАТЕМАТИКА

$$\int_{\alpha}^{\beta} f(x) dx$$

ЭКОНОМИЧЕСКАЯ

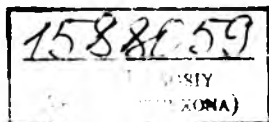
**МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО
СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ УЗБЕКИСТАН**

ТАШКЕНТСКИЙ ФИНАНСОВЫЙ ИНСТИТУТ

Ш.Ш. БАБАДЖАНОВ

ЭКОНОМИЧЕСКАЯ МАТЕМАТИКА

***Рекомендовано Министерством высшего и среднего
специального образования Республики
Узбекистан в качестве учебного пособия***



**Ташкент
«IQTISOD-MOLIYA»
2018**

УДК: 519.1(075)

ББК: 22.171

Рецензенты: *канд. физ.-мат. наук, доц. ТУИТ Х. Абдуванитов;*

канд. физ.-мат. наук, доц. А. Хашимов

Б-4 Экономическая математика: Учебное пособие / Ш.Ш. Бабаджанов; Т.: «Iqtisod-Moliya», 2018. — 564 с.

Учебное пособие содержит 27 параграфов по теории вероятностей, математической статистике, математическому программированию и их экономическим приложениям. Материал, изложенный в книге, является математической базой для многих (но не всех) математических методов, применяемых в экономических моделях.

Пособие предназначено бакалаврам всех направлений области знания «Экономика», и написано в соответствии с требованиями государственного стандарта.

УДК: 519.1(075)

ББК: 22.171

ISBN 978-9943- 13-734-9

© Ш.Ш. Бабаджанов, 2018

© «Iqtisod-Moliya», 2018

ПРЕДИСЛОВИЕ

Учебное пособие написано с учетом требований по внедрению зарубежного опыта преподавания математических дисциплин бакалаврам всех направлений области знания «Экономика». Пособие содержит 27 параграфов по теории вероятностей, математической статистике, математическому программированию и их экономическим приложениям.

Пособие отражает математическое содержание курсов «Теория вероятностей и математическая статистика», «Математическое программирование» или родственных им по названию, которые традиционно читаются на экономических специальностях вузов.

Отметим, что математический материал, изложенный в книге, является математической базой далеко не всех математических методов, применяемых в экономических моделях. Например, не представлена математическая основа математических методов, применяемых в экономических моделях, которые описываются дифференциальными и разностными уравнениями, прикладной статистики и т.д.п. В данной книге не ставилась цель охватить в полном объеме математическую основу всех методов, применяемых в экономических моделях.

Автор

§1. Пространство элементарных исходов (событий). Определения вероятности

Ключевые слова: событие, случайное событие, достоверное и невозможное события, несовместные события, равновозможные события, массовые или статистические случайные события, опыт (эксперимент, испытание), воспроизводимый опыт, наблюдаемый результат, пространство (множество) элементарных исходов, поле событий, элементарный исход, операция суммы, произведение событий, противоположное событие, алгебра событий, классическое определение, классическая схема или схема урн, факториал натурального числа, упорядоченное множество, размещения, перестановки, сочетания, статистическое определение вероятности, геометрическая вероятность.

1. Случайные события и предмет теории вероятностей

Каждая математическая дисциплина, в том числе и теория вероятностей, отражает свойства реального мира. В каждой математической дисциплине есть свои основные понятия, связанные с предметом изучения. Эти понятия происходят из явлений и фактов окружающей нас действительности.

Начнем с описания тех явлений и фактов, которые лежат в основе понятия «случайное событие». Это описание еще не является математическим определением, однако оно включает в себе некоторые свойства реального мира, изучение которых математическими методами и является содержанием теории вероятностей.

При неоднократном воспроизведении одного и того же опыта результаты его могут изменяться от случая к случаю.

Так, в результате подбрасывания монеты может оказаться, что она падет либо гербом вверх («герб»), либо гербом вниз («цифра»).

При бросании игральной кости на верхней ее грани может появиться одно из чисел 1, 2, 3, 4, 5, 6.

Если из большого количества произведенных изделий выбрать наудачу одно, то такое изделие может оказаться либо стандартным, либо бракованным.

Всякой электролампе можно поставить в соответствие некоторое число количество часов работы до момента перегорания.

В каждом из перечисленных выше примеров рассмотрен некоторый опыт (эксперимент, испытание), исход которого заранее предвидеть невозможно.

Определение 1. Некоторое событие называют *случайным* по отношению к данному опыту, если в его результате оно может появиться или не появиться.

Примеры. Следующие события являются случайными:

1. Событие, состоящее в том, что при бросании монеты выпадет герб (решетка).
2. Событие, состоящее в том, что при бросании игральной кости выпадет четное число очков.
3. Событие, состоящее в том, что наудачу выбранное изделие окажется бракованным.
4. Событие, состоящее в том, что электролампа будет гореть не менее трех часов.

Можно говорить о таких случайных событиях, как поражение мишени, выигрыш в лотерею, наличие сильных помех при радиоприеме и др.

Случайные события обозначаются в дальнейшем латинскими прописными буквами A , B , C и т.д.

Виды случайных событий. Сделаем сразу же одно замечание. Согласно данному выше определению, событие считается случайным, если его наступление в результате опыта представляет собой лишь одну из возможностей. Под это определение формально подходят и такие события, которые в результате данного опыта обязательно наступают; эти события называют

достоверными. Например, достоверным является событие, состоящее в том, что при бросании игральной кости выпадает целое число очков или, что выбранное наугад слово из данной лекции содержит не более 50 букв. Итак, достоверное событие можно рассматривать как одну из разновидностей случайного события. Оно обозначается символом Ω (омега).

Аналогичное замечание относится и к *невозможным* событиям, т.е. таким, которые никогда не наступят при осуществлении данного опыта. Невозможное событие тоже можно рассматривать как случайное. Примером невозможного события может служить получение двух выигрышей по одному билету. Невозможное событие обозначается символом $\emptyset$ пустое множество.

Определение 2. События называют *несовместными*, если появление одного из них исключает появление других событий в одном и том же испытании.

Комментарий к определению 2. Несовместные события не могут появиться в результате одного опыта.

Можно говорить о нескольких несовместных событий $A_1, A_2, \dots, A_n$: в этом случае имеют в виду, что события $A_1, A_2, \dots, A_n$, не могут наступить все сразу в результате одного испытания. Если в множестве $A_1, A_2, \dots, A_n$ событий каждые два несовместны, то говорят, что эти события *попарно несовместны*.

Например, при бросании игральной кости события A «выпадет количество очков, равное 1 или 2» и B «выпадет количество очков, равное 4 или 5» несовместны.

Определение 3. События называют *равновозможными*, если есть основания считать, что ни одно из этих событий не является более возможными, чем другие.

В дальнейшем нас будут интересовать только такие опыты, которые можно повторить (в принципе) неограниченное число раз; именно такой характер носит опыт с бросанием монеты, покупкой лотерейного билета, обследованием изделия на годность или брак. Любое случайное событие, наступ-

ление которого возможно в такого рода опытах, называется *массовым* или *статистическим*.

Массовые случайные события следует отличать от единичных, исключительных, обладающих той особенностью, что опыт, с которым связаны эти события, принципиально невоспроизводим. Например «1 сентября 2018 года в Ташкенте шел дождь» является исключительным, так как воспроизвести наступление указанного дня невозможно. В то же время события «1 сентября в Ташкенте шел дождь» (без упоминания о годе) является, несомненно, *массовым* ведь наблюдают погоду в Ташкенте 1 сентября в течение многих лет.

Теперь мы в состоянии ответить на вопрос, с которого, собственно, и должно начинаться знакомство с теорией вероятностей: чем занимается и какие задачи ставит перед собой эта дисциплина? В самых общих словах предмет теории вероятностей может быть определен следующим образом.

Теория вероятностей занимается изучением закономерностей, присущих массовым случайным событиям.

Простейший пример закономерности такого рода дает опыт с бросанием монеты. Предположим, что бросание производится много раз подряд. Исход каждого отдельного бросания является случайным, неопределенным. Однако *средний результат* большого числа бросаний утрачивает случайный характер, становится закономерным. А именно: «доля» тех бросаний, при которых выпадает герб (т.е. отношение числа таких бросаний к числу всех бросаний) с увеличением числа бросаний приближается к $1/2$.

Это некоторый предварительный, интуитивный подход к понятию «случайное событие». Он базируется на совершенно естественных, но вместе с тем не вполне строгих рассуждениях. Точный смысл основных понятий теории вероятностей связано с понятием пространства элементарных исходов (событий).

Строгое математическое понятие «случайного события» должно обобщить соответствующее обиходное понятие, т.е. отражать его существенные черты и в то же время быть безупречно точным. Точный смысл основного

понятия теории вероятностей, в том числе и понятия «случайного события» связано с понятием пространства элементарных исходов (событий).

2. Элементарные события (исходы)

Следует подчеркнуть, что всякий случайный опыт (эксперимент, испытание) состоит в осуществлении некоторого вполне определенного комплекса условий и наблюдений результата. *Рассматриваются только такие опыты, которые можно повторять (воспроизводить) при неизменном комплексе условий произвольное число раз (по крайней мере, теоретически).*

Предметом наблюдения в том или ином опыте может быть некоторый процесс, физическое явление или действующая система. Для реально воспроизводимого опыта понятие «наблюдаемый результат» означает, что существует принципиальная возможность зарегистрировать данный результат опыта с помощью того или иного прибора (в простейшем случае, например, визуально). Любой наблюдаемый результат интерпретируется как случайный исход опыта (*случайное событие*). Событие может произойти, а может и не произойти в результате опыта.

При математической формализации модели опыта отправным пунктом является понятие *пространства (множества) элементарных исходов* (обозначается Ω), связанного с данным опытом. Под этим понимают множество взаимоисключающих исходов такое, что результатом опыта всегда является один и только один исход. Любое подмножество данного множества Ω интерпретируется как событие (возможно, и ненаблюдаемое). Совокупность всех наблюдаемых событий составляет *поле событий* для данного опыта.

Говорят, что событие A появилось (наступило, произошло, осуществилось, реализовалось), если результатом опыта является элементарный исход ω , принадлежащий A ($\omega \in A$). Событие, совпадающее с пустым множеством $\emptyset$, называется - *невозможным* событием, а событие, совпадающее со всем множеством Ω - *достоверным*.

Два события A и B называются *совместными (несовместными)*, если в результате опыта возможно (невозможно) их совместное осуществление. Другими словами, события A и B совместны, если соответствующие множества A и B имеют общие элементы, и несовместны - в противном случае.

Множество Ω для данного опыта может быть дискретным, непрерывным или иметь более сложную структуру. К дискретным относятся конечные или счетные множества элементарных исходов. К непрерывным относятся множества типа континуума. Любой конечный или бесконечный промежуток числовой прямой является примером множества типа континуума. В дальнейшем мы рассматриваем только такие модели опытов, для которых множество элементарных исходов Ω либо дискретно, либо непрерывно.

Построение множества Ω (если оно не задано при описании опыта) осуществляется на практике, исходя из требования, чтобы все интересующие нас результаты данного опыта могли быть однозначно описаны на основе построенного множества Ω . Другими словами, если нас интересуют события A , B , C и т.д., являющиеся наблюдаемыми событиями в данном опыте, то множество Ω должно состоять из таких исходов, чтобы существовали подмножества данного множества, равносильные событиям A , B , C и т.д.

Так как понятие «элементарный исход» строго не определяемо, то указанная задача допускает не единственное решение и зависит от набора интересующих нас событий. Если потребовать, чтобы правило однозначного описания выполнялось для всего поля событий, то в этом случае понятие элементарного исхода становится более определенным. Именно в совокупности всех подмножеств множества Ω , составляющих поле событий, элементарные исходы являются одноэлементными подмножествами.

Пример 1. Опыт состоит в подбрасывании один раз правильной шестигранной игральной кости. Обозначим через X число очков, выпавших на верхней грани кости. Описать множество элементарных исходов Ω и указать состав подмножеств, соответствующих следующим событиям:

$$A = \{X - \text{кратно трем}\}, \quad B = \{X - \text{нечетно}\}, \quad C = \{X > 3\},$$

$$D = \{X < 7\}, \quad E = \{X - \text{дробно}\}, \quad F = \{0,5 < X < 1,5\}.$$

Выявить пары совместных событий.

Решение

Введем обозначения для следующих наблюдаемых в данном опыте событий:

$$\omega_k = \{X = k\}, \quad k = 1, 2, \dots, 6;$$

$$\omega^{(1)} = \{X - \text{нечетное число}\}; \quad \omega^{(2)} = \{X - \text{четное число}\}.$$

На базе данных исходов можно сконструировать два достоверных события:

$$\Omega_1 = \{\omega_1, \omega_2, \dots, \omega_6\} \quad \text{и} \quad \Omega_2 = \{\omega^{(1)}, \omega^{(2)}\}.$$

Какое из них больше подходит в качестве множества элементарных исходов?

Ясно, что Ω_2 следует «забраковать», поскольку, например, наблюдаемые события $\omega_1, \omega_2, \dots, \omega_6$, A , B , D , E не являются подмножествами множества Ω_2 . С другой стороны, все перечисленные события могут быть описаны как подмножества множества Ω_1 . Действительно,

$$A = \{\omega_3, \omega_6\}, \quad B = \omega^{(1)} = \{\omega_1, \omega_3, \omega_5\}, \quad C = \{\omega_4, \omega_5, \omega_6\},$$

$$D = \Omega_1 = \{\omega_1, \omega_2, \dots, \omega_6\}, \quad E = \emptyset, \quad F = \{\omega_1\}.$$

Из написанных равенств, в частности, усматриваем, что исходы $\omega^{(1)}$ и $\omega^{(2)}$ разложимы на элементы, которые сами являются исходами данного опыта. Таким образом, исходы $\omega_1, \omega_2, \dots, \omega_6$ более «элементарны», чем исходы $\omega^{(1)}$ и $\omega^{(2)}$.

Сопоставляя попарно события и проверяя наличие общих элементов, находим пары совместных событий: A и B , A и C , A и D , B и C , B и D , B и F , C и D , D и F .

Пример 2. Опыт состоит в радиолокационном обнаружении воздушной цели. Наблюдаемый результат - положение светящегося пятна (отраженного

импульса от цели) на экране индикатора цели, имеющего форму круга радиуса 10 см, в системе декартовых координат с началом, совпадающим с центром экрана. Описать множество элементарных исходов и состав подмножеств, соответствующих следующим событиям:

$$A = \{\text{цель находится в первом квадранте}\};$$

$$B = \left\{ \begin{array}{l} \text{цель находится в круге радиуса 5 см, центр} \\ \text{которого совпадает с центром экрана} \end{array} \right\};$$

$$C = \left\{ \begin{array}{l} \text{цель находится в круге радиуса 2,5 см, центр которого сдвинут} \\ \text{на 5 см вдоль оси } O_x \text{ в отрицательном направлении} \end{array} \right\}.$$

Совместны ли пары событий A и B , A и C , B и C ?

Решение

Все интересующие нас в данном опыте наблюдаемые события связаны с регистрацией положения светящегося пятна на экране индикатора. Удобной формой математического описания элементарного исхода являются в данном случае координаты случайной точки на плоскости, соответствующей, например, центру пятна (предполагается, что пятно представляет собой круг достаточно малого радиуса). Хотя очевидно, что точку (как абстрактное математическое понятие) наблюдать физически на экране индикатора невозможно, тем не менее такой идеализированный способ описания элементарного исхода упрощает математическую формализацию модели данного опыта.

Таким образом, множество Ω непрерывно и может быть записано в виде

$$\Omega = \{(x, y): x^2 + y^2 \leq 100\}.$$

Подмножества, равносильные указанным событиям, имеют вид:

$$A = \{(x, y): x^2 + y^2 \leq 100, x > 0, y > 0\};$$

$$B = \{(x, y): x^2 + y^2 \leq 25\};$$

$$C = \{(x, y): (x + 5)^2 + y^2 \leq 6,25\}.$$

По определению события совместны, если соответствующие им подмножества имеют общие элементы (пересекаются), и несовместны - в про-

тивном случае. Поэтому события A и B , B и C совместны, а события A и C несовместны.

Пример 3. Рассмотрим следующий случайный опыт: матч на первенство страны по футболу между командами «Пахтакор» и «Бунёдкор». Интересующие нас события:

$$A = \{\text{выиграла команда "Пахтакор"}\};$$

$$B = \{\text{игра закончилась победой одной из команд}\};$$

$$C = \{\text{игра закончилась со счетом } 3:1 \text{ в пользу "Бунёдкор"}\};$$

$$D = \{\text{в игре забито не меньше трех голов}\}.$$

Требуется описать множество элементарных исходов и указать состав подмножеств, соответствующих указанным событиям.

Решение

Очевидно, что описанный опыт не удовлетворяет требованию воспроизводимости при неизменном комплексе условий, поскольку условия проведения матча меняются от игры к игре. Поэтому построение вероятностной модели исходов футбольного матча возможно лишь при определенной идеализации реальных условий. Предположим, что такая модель существует. Спрашивается, что представляет собой результат (элементарный исход) данного опыта? Ответ зависит от того, какой круг событий мы собираемся наблюдать (регистрировать) в данном опыте. События, перечисленные в условии задачи, определяют круг интересов обычного «болельщика» за ту или иную команду. Для полного и однозначного описания всех указанных событий достаточно принять в качестве элементарного исхода конечный счет в матче.

Запишем множество Ω следующим образом:

$$\Omega = \{\omega = (x, y): x \in \mathbb{Z}_0, y \in \mathbb{Z}_0\},$$

где x - количество голов, забитых командой «Пахтакор», y - количество голов, забитых командой «Бунёдкор», $\mathbb{Z}_0$ - множество неотрицательных целых чисел.

Мы вынуждены считать множества возможных значений x и y , по крайней мере, теоретически - неограниченными, поскольку до игры нет никаких оснований для того, чтобы установить явную точную границу возможного счета. Практически, конечно, бесконечно большой счет ни в какой игре не осуществим.

Подмножества, соответствующие интересующим нас событиям, имеют следующий вид:

$$A = \{(x, y): x \in \mathbb{Z}_0, y \in \mathbb{Z}_0, x > y\}, \quad B = \{(x, y): x \in \mathbb{Z}_0, y \in \mathbb{Z}_0, x \neq y\},$$

$$C = \{(x, y): x = 1, y = 3\} = \{(1, 3)\}, \quad D = \{(x, y): x \in \mathbb{Z}_0, y \in \mathbb{Z}_0, x + y \geq 3\}.$$

Заметим, что для другого «экспериментатора», например, для тренера команды «Пахтакор» - может оказаться более важным предметом наблюдения не конечный счет матча, а количество травмированных или оштрафованных игроков. В этом случае придется строить другое множество элементарных исходов, а следовательно, и другую вероятностную модель футбольного матча.

Упражнения

В задачах 1- 6 построить множество элементарных исходов Ω по описанию опыта и указанных подмножеств, соответствующие указанным событиям

1. Игральная кость подбрасывается дважды. Наблюдаемый результат - пара чисел, соответствующих числам очков, выпавших в первый и второй раз. События:

$A = \{\text{оба раза выпало число очков, кратное трем}\};$

$B = \{\text{ни разу не выпало число шесть}\};$

$C = \{\text{оба раза выпало число очков, большее трех}\};$

$D = \{\text{оба раза выпало одинаковое число очков}\}.$

2. Монета подбрасывается три раза. Наблюдаемый результат - появление герба (г) или цифры (ц) на верхней стороне - монеты. События:

$A = \{\text{герб выпал ровно один раз}\};$

$B = \{\text{ни разу не выпала цифра}\};$

$C = \{\text{выпало больше гербов, чем цифр}\};$

$D = \{\text{герб выпал не менее, чем два раза подряд}\}.$

3. Монета подбрасывается до первого появления герба. Наблюдаемый результат - общее число подбрасываний. События:

$A = \{\text{герб выпал при третьем подбрасывании}\};$

$B = \{\text{герб выпал не ранее, чем при третьем подбрасывании}\}.$

4. Производится стрельба по плоской прямоугольной мишени:

$$-2 \leq x \leq 2, \quad -1 \leq y \leq 1.$$

Наблюдаемый результат - координаты точки попадания в декартовой системе координат. По условиям стрельбы непопадание в указанный прямоугольник исключено. События:

$A = \{\text{абсцисса точки попадания не меньше ординаты}\};$

$B = \{\text{произведение координат точки неотрицательно}\};$

$C = \{\text{сумма абсолютных величин координат точки превышает единицу}\}.$

Выявить пары совместных событий.

5. На отрезке $[a, b]$ наудачу ставится точка. Пусть x координата этой точки. Затем на отрезке $[a, x]$ наудачу ставится еще одна точка с координатой y . Наблюдаемый результат - пара чисел (x, y) . События:

$A = \{\text{вторая точка ближе к правому концу отрезка } [a, b], \text{ чем к левому}\};$

$B = \{\text{расстояние между двумя точками меньше половины длины отрезка}\};$

$C = \{\text{первая точка ближе к левому концу отрезка } [a, b], \text{ чем к правому}\};$

$D = \{\text{первая точка ближе ко второй, чем к правому концу отрезка } [a, b]\}.$

Выявить пары несовместных событий.

6. Иван и Петр договорились о встрече в определенном месте между одиннадцатью и двенадцатью часами. Каждый приходит в случайный момент указанного промежутка и ждет появления другого до истечения часа, но не более 15 минут, после чего уходит. Наблюдаемый результат - пара чисел (x, y) , где x - время прихода Петра, y - время прихода Ивана (время исчисляется в минутах, начиная от 11 часов). События:

$A = \{\text{Петр пришел после 11 ч 45 мин}\};$

$B = \{\text{Петр пришел после Ивана}\};$

$C = \{\text{Иван пришел до 11 ч 45 мин}\};$

$D = \{\text{встреча не состоялась}\};$

$E = \{\text{Петр ждал Ивана все обусловленное время и не дождался}\};$

$F = \{\text{Ивану не пришлось ждать Петра}\};$

$K = \{\text{встреча состоялась когда до истечения часа оставалось меньше пяти минут}\}.$

3. Операции над событиями

В этом пункте мы ознакомимся с тремя основными видами комбинации событий: суммой событий, произведением события, противоположным событием.

В п. 2 было отмечено, что любое подмножество множества элементарных исходов Ω интерпретируется как событие (возможно, и ненаблюдаемое). Для любых двух подмножеств A, B пространства Ω определены подмножества $A \cup B$ (объединение A и B) и $A \cap B$ (пересечение A и B): $A \cup B$ состоит из тех элементов, которые принадлежат подмножеству A или под-

множеству B , или обоим подмножествам вместе; $A \cap B$ состоит из тех элементов, которые принадлежат и множеству A , и множеству B .

Рассматривают также дополнение подмножества A в Ω , обозначаемое $\overline{A}$: дополнение $\overline{A}$ состоит из тех точек, которые не принадлежат множеству A .

Операциям $\cup$ и $\cap$ соответствуют операции над событиями - сумма и произведение событий.

Определение 4. Суммой событий A и B называется событие, обозначаемое $A + B$ и состоящее в том, что в результате опыта наступит или событие A , или событие B , или оба вместе.

Определение 5. Произведением двух событий A и B называется событие, обозначаемое AB и состоящее в том, что в результате опыта наступит и событие A , и событие B .

Операции дополнения множества A соответствует операция перехода от заданного события к противоположному.

Определение 6. Событием, противоположным событию A , называется событие, обозначаемое $\overline{A}$ и состоящее в том, что в результате опыта событие A не наступит.

Комментарии к определениям 4-6

1. Определения 4-6 описывают подмножества элементарных событий, благоприятствующих сумме событий A и B , произведению событий A и B , событию, противоположному к A .

2. Операции суммы событий, произведения событий и дополнения обладают такими свойствами, как

а) $A + B = B + A$, $AB = BA$, (коммутативность);

б) $(A + B) + C = A + (B + C)$, $(AB)C = A(BC)$ (ассоциативность);

в) $A(B + C) = AB + AC$ (дистрибутивность умножения относительно сложения).

множеству B , или обоим подмножествам вместе; $A \cap B$ состоит из тех элементов, которые принадлежат и множеству A , и множеству B .

Рассматривают также дополнение подмножества A в Ω , обозначаемое $\overline{A}$: дополнение $\overline{A}$ состоит из тех точек, которые не принадлежат множеству A .

Операциям $\cup$ и $\cap$ соответствуют операции над событиями - сумма и произведение событий.

Определение 4. Суммой событий A и B называется событие, обозначаемое $A + B$ и состоящее в том, что в результате опыта наступит или событие A , или событие B , или оба вместе.

Определение 5. Произведением двух событий A и B называется событие, обозначаемое AB и состоящее в том, что в результате опыта наступит и событие A , и событие B .

Операции дополнения множества A соответствует операция перехода от заданного события к противоположному.

Определение 6. Событием, противоположным событию A , называется событие, обозначаемое $\overline{A}$ и состоящее в том, что в результате опыта событие A не наступит.

Комментарии к определениям 4-6

1. Определения 4-6 описывают подмножества элементарных событий, благоприятствующих сумме событий A и B , произведению событий A и B , событию, противоположному к A .

2. Операции суммы событий, произведения событий и дополнения обладают такими свойствами, как

а) $A + B = B + A$, $AB = BA$, (коммутативность);

б) $(A + B) + C = A + (B + C)$, $(AB)C = A(BC)$ (ассоциативность);

в) $A(B + C) = AB + AC$ (дистрибутивность умножения относительно сложения).

$$A_1 \cdot A_2 \cdots A_n \cdots = \prod_{i=1}^{\infty} A_i.$$

Событие $\prod A_i$ состоит в том, что в результате опыта наступит каждое из событий $A_1, A_2, \dots, A_n, \dots$.

Примеры

1. Пусть при бросании игральной кости событие A означает, что выпадет четное число очков, а событие B - что количество очков не превзойдет четырех. Ясно, что

$$A = \{2, 4, 6\}; B = \{1, 2, 3, 4\}; (A + B) = \{1, 2, 3, 4, 6\}; (AB) = \{2, 4\}.$$

Иными словами, событие $A + B$ наступает при выпадении чисел 1, 2, 3, 4, 6, а событие AB - при выпадении чисел 2, 4.

Для событий $\overline{A}$, $\overline{B}$, $\overline{A+B}$, $\overline{AB}$, противоположным, соответственно, к событиям A , B , $A+B$, AB , получим

$$\overline{A} = \{1, 3, 5\}; \overline{B} = \{5, 6\}; (\overline{A+B}) = \{5\}; (\overline{AB}) = \{1, 3, 5, 6\}.$$

2. Пусть при игре в спортлото события A , B , C заключаются в том, что выбранная в результате опыта шестерка чисел содержит соответственно числа 7, 12, 20. Результатом испытания оказалась шестерка чисел 7, 8, 9, 20, 21, 42. Какие из следующих событий наступили при этом:

$$A, B, C, \overline{A}, \overline{B}, \overline{C}, ABC, A+B, A+C, \overline{A+C}?$$

Решение

Обозначим множество $\{7, 8, 9, 20, 21, 42\}$ через M . Тогда получим, что из перечисленных в условии событий наступили следующие: A , поскольку $7 \in M$; C , так как $20 \in M$; $\overline{B}$, поскольку $12 \notin M$; $A+B$, так как $7 \in M$; $A+C$, поскольку $7 \in M$ и $20 \in M$. Остальные из перечисленных событий не наступили.

В алгебре событий справедливы формулы

$$\overline{A_1 + A_2 + \dots + A_n} = \overline{A_1} \cdot \overline{A_2} \cdots \overline{A_n}, \quad (*)$$

$$\overline{A_1 \cdot A_2 \cdots A_n} = \overline{A_1} + \overline{A_2} + \dots + \overline{A_n}. \quad (**)$$

Строгий вывод этих формул мы опускаем. Здесь мы ограничимся лишь некоторым пояснением к ним. Левая часть формулы (*) есть событие, противоположное к $A_1 + A_2 + \dots + A_n$. Последнее событие наступает тогда и только тогда, когда наступает хотя бы одно из A_i ($i=1, 2, \dots, n$). Событием, противоположным тому, что наступит хотя бы одно из A_i , является событие, состоящее в том, что не наступит ни одно из A_i , т.е. наступит каждое из $\overline{A_i}$; это означает, что наступит событие $\overline{A_1} \cdot \overline{A_2} \cdots \overline{A_n}$.

Левая часть формулы (**) есть событие, противоположное событию $A_1 \cdot A_2 \cdots A_n$. Последнее событие состоит в том, что наступит каждое из событий $A_1, A_2, \dots, A_n$. Событием, противоположным тому, что наступит каждое из A_i ($i=1, 2, \dots, n$), является событие, состоящее в том, что не наступит хотя бы одно из $A_1 \cdot A_2 \cdots A_n$, т.е. что наступит хотя бы одно из $\overline{A_1}, \overline{A_2}, \dots, \overline{A_n}$; это означает, что наступит событие $\overline{A_1} + \overline{A_2} + \dots + \overline{A_n}$.

Пример 3. Каждый из стрелков делает по одному выстрелу в мишень.

- а) Какое событие является противоположным к событию A : «хотя бы один стрелок попал в цель»?
- б) Какое событие противоположно событию «каждый из стрелков попал в цель?»

Решение

а) Таким событием является «каждый из стрелков промахнулся» или, что то же самое, «ни один не попал в цель». Справедливость ответа вытекает из того, что событие A означает поражение мишени, а событие $\overline{A}$ - непоражение мишени. Этот пример иллюстрирует формулу (*).

б) Таким событием является «хотя бы один из стрелков промахнулся». Этот пример иллюстрирует формулу (**).

Пример 4. Игральная кость подбрасывается один раз. Наблюдаемый результат - число очков на верхней грани. События A, B, C, D, E, F описаны в примере 1. из п. 2 данного параграфа. Описать состав и выяснить смысл следующих событий:

$$E_1 = \overline{B}, \quad E_2 = \overline{C}, \quad E_3 = AB, \quad E_4 = A + B, \quad E_5 = A - B, \quad E_6 = E + D, \quad E_7 = EF.$$

Решение

В обозначениях примера 1 напомним состав указанных событий, используя определение соответствующей алгебраической операции:

$$E_1 = \{\omega_2, \omega_4, \omega_6\} = \{\text{выпало четное число очков}\};$$

$$E_2 = \{\omega_1, \omega_2, \omega_3\} = \{\text{выпало число очков, не больше трех}\};$$

$$E_3 = \{\omega_3\} = \{\text{выпавшее число очков, нечетно и кратно трем}\};$$

$$E_4 = \{\omega_1, \omega_3, \omega_5, \omega_6\} = \{\text{выпавшее число очков или нечетно, или кратно трем}\};$$

$$E_5 = \overline{AB} = AB - E_3; \quad E_6 = \emptyset + D = D = \Omega.$$

Пример 5. Рассмотрим снова случайный эксперимент, описанный в примере 2. Как уже говорилось, множество идеализированных элементарных исходов

$$\Omega = \{(x, y): x^2 + y^2 \leq 100\}$$

является удобной формой математического описания наблюдаемых событий, связанных с положением на экране индикатора светящегося пятна - отраженного импульса от цели. Пусть наблюдения в данном опыте состоят в регистрации факта принадлежности светящегося пятна произвольно выбранному участку круга, имеющему площадь. Например, проверяется результат: попадает ли отраженный импульс от цели в некоторую часть кольца, ограниченную радиусами r_1 и r_2 и полярными углами φ_1 и φ_2 . Составляет ли множество всех quadriруемых подмножеств множества Ω алгебру событий?

Решение

Из интегрального исчисления известно, что объединение, пересечение и дополнение конечного или счетного числа квадрируемых, подмножеств не-которого квадрируемого множества на плоскости являются квадрируемыми множествами. Отсюда следует, что система квадрируемых подмножеств множества Ω образует алгебру для данного опыта.

Упражнения

В задачах 1-2 требуется по описанию опыта построить множество элементарных исходов и выявить состав подмножеств, соответствующих указанным событиям.

1. Пусть A, B, C - три события, наблюдаемые в данном эксперименте. Выразить в алгебре событий следующие события:

$E_1 = \{\text{из трех событий } A, B, C \text{ произойдет ровно одно}\};$

$E_2 = \{\text{из трех событий } A, B, C \text{ произойдет хотя бы одно}\};$

$E_3 = \{\text{из трех событий } A, B, C \text{ произойдет ровно два}\};$

$E_4 = \{\text{из трех событий } A, B, C \text{ произойдет не меньше двух}\};$

$E_5 = \{\text{из трех событий } A, B, C \text{ не произойдет ни одного}\};$

$E_6 = \{\text{из трех событий } A, B, C \text{ произойдет хотя бы два}\};$

$E_7 = \{\text{из трех событий } A, B, C \text{ не произойдет хотя бы одно}\}.$

2. Произведено три выстрела из орудия по цели. Событие $A_k = \{\text{попадание при } k\text{-м выстреле}\}$ ($k = 1, 2, 3$).

а) Выяснить состав множества Ω , выразив каждый элементарный исход ω_i , через события A_k .

б) Записать в алгебре событий следующие события:

$A = \{\text{ровно одно попадание}\};$

$B = \{\text{хотя бы одно попадание}\};$

$C = \{\text{хотя бы один промах}\};$

$D = \{\text{не меньше двух попаданий}\};$

$E = \{\text{попадание не раньше, чем при третьем выстреле}\}.$

4. Классическое определение вероятности

Опыт, связанный с пространством элементарных событий Ω , позволяет каждому событию A , сопоставить число $p(A)$, где $0 \leq p(A) \leq 1$, измеряющее «степень вероятности» наступления события A , причем для соответствия $A \rightarrow p(A)$ выполняются все требования здравого смысла, перечисленные выше.

Рассмотрим опыт, удовлетворяющий тому условию, что соответствующее ему множество элементарных исходов $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ представляет собой конечное множество.

Соответствие $A \rightarrow p(A)$ строится следующим образом. Пусть дано событие $A = \{\omega_{k_1}, \omega_{k_2}, \dots, \omega_{k_m}\} \subset \Omega$: это значит, что заданы все элементарные исходы (события), благоприятствующие событию A . Общее количество таких элементарных исходов обозначим через m , а число всех исходов опыта - число элементов множества Ω через n .

Положим

$$p(A) = \frac{m}{n}. \quad (1)$$

Формула (1) сопоставляет каждому событию A в рассматриваемом испытании число $p(A)$, называемое *вероятностью* события A .

Определение 8. Вероятностью $p(A)$ события A называют отношение числа исходов благоприятствующих этому событию к общему числу всех элементарных исходов испытания: $p(A) = \frac{m}{n}$.

Легко убедиться в том, что более вероятным событиям формула (1) сопоставляет большие вероятности, равновероятным событиям - равные вероятности, невозможному событию - нуль, достоверному событию - единицу. В частности, вероятность каждого элементарного исхода (события) ω_i равна

$$\frac{1}{n} \quad (i = 1, 2, \dots, n).$$

Формула (1) задает так называемое «классическое определение» вероятности события.

Определение 9. Всякий опыт, удовлетворяющий тому условию, что соответствующее ему множество элементарных исходов $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ представляет собой конечное множество равновероятных исходов, называется **классической схемой или схемой урн**.

Классическая схема является математической формализацией опытов, в которых элементарные исходы обладают определенной симметрией по отношению к условиям опыта, так что нет оснований считать какой-либо из исходов более вероятным, чем другой. Таким свойством, например, обладают опыты по извлечению наудачу определенного числа шаров из урны, содержащей заданное количество неразличимых на ощупь шаров. Отсюда и название - схема урн.

Этим способом пользовались французские математики XVII в. Блез Паскаль и Пьер Ферма, рассматривавшие формулу (1) как определение вероятности. В далекое от нашей эпохи время трудно было предвидеть роль понятия вероятности, разнообразие и серьезность будущих приложений теории вероятностей к естествознанию, технике и экономике. Первоначальным материалом, на котором «отрабатывались» простейшие факты теории, были азартные игры. С тех пор задачи о бросании игральной кости, об извлечении карт из колоды, шаров из урны и т. п. стали традиционными для теории вероятностей. Заметим, что и по сей день, эти задачи сохраняют свою роль как тренировочные упражнения, а в некоторых случаях – как наглядные модели для более серьезных вероятностных схем.

Оказывается, что рассмотренный выше опыт, связанный с пространством элементарных исходов (событий), носит довольно общий характер: многие задачи теории вероятностей (в том числе далеко не простые) представляют собой частный случай этой общей схемы.

Отметим, однако, что пространство элементарных исходов может быть

бесконечным, этот случай рассмотрен в п. 6.

4. 1. Элементы комбинаторики

Для подсчета m - общего количества благоприятствующих элементарных исходов некоторого наблюдаемого события и n - числа всех исходов опыта, связанного с этим событием (число элементов множества Ω), необходимо подсчитать число возможных способов совершения каких-либо действий. Задачи такого типа называются *комбинаторными*, а раздел математики, занимающийся решением таких задач, - *комбинаторикой*. Сформулируем два универсальных правила, применяемых при решении комбинаторных задач.

Правило произведения. Пусть требуется выполнить одно за другим какие-либо r действий. Если первое действие можно выполнить l_1 способами, второе действие - l_2 способами и так до r -го действия, которое можно выполнить l_r способами, то все r действий могут быть выполнены $l_1 l_2 \dots l_r$ способами.

Правило суммы. Пусть требуется выполнить одно из каких-либо r действий, взаимно исключающих друг друга. Если первое действие можно выполнить l_1 способами, второе действие - l_2 способами и так до r -го действия, которое можно выполнить l_r способами, то выполнить одно из этих r действий можно $l_1 + l_2 + \dots + l_r$ способами.

Напомним понятие факториала, активно используемое в комбинаторике. **Факториалом натурального числа l** называется число

$$l! = l(l-1)(l-2) \dots 3 \cdot 2 \cdot 1. \quad (2)$$

По определению, факториалом нуля является единица: $0! = 1$.

Рассмотрим некоторое множество S , состоящее из l различных элементов. Пусть $1 \leq k \leq l$. Назовём множество, состоящее из k элементов, упорядоченным, если каждому элементу этого множества поставлено в соответствие число от 1 до k , причём различным элементам множества соответствуют разные числа.

Размещениями из l элементов по k называются упорядоченные подмножества множества S , состоящие из k различных элементов и отличающиеся друг от друга составом элементов или порядком их расположения.

Число размещений из l элементов по k равно

$$A_l^k = \frac{l!}{(l-k)!} = l(l-1)(l-2)\cdots(l-k+1). \quad (3)$$

Перестановками из l элементов называются размещения из l элементов по l , т.е. упорядоченные подмножества множества S , состоящие из всех элементов данного множества и отличающиеся друг от друга только порядком их расположения.

Число перестановок из l элементов равно

$$P_l = l! = l(l-1)(l-2)\cdots 3 \cdot 2 \cdot 1. \quad (4)$$

Сочетаниями из l элементов по k называются подмножества множества S , состоящие из k различных элементов и отличающиеся друг от друга только составом элементов.

Число сочетаний из l элементов по k равно

$$C_l^k = \frac{A_l^k}{P_l} = \frac{l!}{k!(l-k)!}. \quad (5)$$

Размещениями с повторениями из l элементов по k называются упорядоченные подмножества множества S , состоящие из k элементов, среди которых могут оказаться одинаковые, и отличающиеся друг от друга составом элементов или порядком их расположения.

Число размещений с повторениями из l элементов по k равно

$$\tilde{A}_l^k = l^k. \quad (6)$$

Сочетаниями с повторениями из l элементов по k называются подмножества множества S , состоящие из k элементов, среди которых могут оказаться одинаковые, и отличающиеся друг от друга только составом элементов.

Число сочетаний с повторениями из l элементов по k равно

$$\tilde{C}_l^k = C_{l-k-1}^k \frac{A_l^k}{P_l} = \frac{(l+k-1)!}{k!(l-1)!}. \quad (7)$$

Отметим, что формулы (4) - (7) сохраняют смысл и остаются справедливыми и при $k = 0$.

Если во множестве S , состоящим из l элементов, есть только r различных элементов, то **перестановками с повторениями из l элементов** называются упорядоченные подмножества множества S , в которые первый элемент множества S входит l_1 раз, второй элемент - l_2 раз и так до r -го элемента, который входит l_r раз ($l_1 + l_2 + \dots + l_r = l$).

Число перестановок с повторениями из l элементов, в которые первый элемент множества S входит l_1 раз, второй элемент - l_2 раз и так до r -го элемента, который входит l_r раз ($l_1 + l_2 + \dots + l_r = l$), равно

$$\tilde{P}_l(l_1, l_2, \dots, l_r) = \frac{l!}{l_1! l_2! \dots l_r!}. \quad (8)$$

Примеры

1. Маша поссорилась с Петей и не хочет ехать с ним в одном автобусе. От общежития до института с 7 до 8 ч отправляется пять автобусов. Не успевший на последний из этих автобусов опаздывает на лекцию. Сколькими способами Маша и Петя могут доехать до института в разных автобусах и не опоздать на лекцию?

Решение

Петя может доехать до института $l_1 = 5$ различными способами (на одном из пяти автобусов), при этом Маше остаётся только $l_2 = 4$ способа (так как один из автобусов занят Петей). Таким образом, по правилу произведения у Пети и Маши есть $l_1 l_2 = 5 \cdot 4 = 20$ различных способов добраться до института в разных автобусах и не опоздать на лекцию.

2. В информационно-технологическом управлении банка работают три аналитика, десять программистов и 20 инженеров. Для сверхурочной работы

в праздничный день начальник управления должен выделить одного сотрудника. Сколько способов существует у начальника управления?

Решение

Начальник управления может отобрать одного аналитика $l_1 = 3$ способами, одного программиста - $l_2 = 10$ способами, а одного инженера - $l_3 = 20$ способами. Поскольку по условию задачи начальник управления может выделить любого из своих сотрудников, согласно правилу суммы, у него существует $l_1 + l_2 + l_3 = 3 + 10 + 20 = 33$ различных способа выбрать сотрудника для сверхурочной работы.

3. Начальник службы безопасности банка должен ежедневно расставлять десять охранников по десяти постам. В целях усиления безопасности одна и та же комбинация расстановки охранников по постам не может повторяться чаще одного раза в месяц. Чтобы оценить, возможно ли это, найти число различных комбинаций расстановки охранников.

Решение

Первый способ. На первый пост начальник службы безопасности может назначить любого из $l_1 = 10$ охранников, на второй пост - любого из оставшихся $l_2 = 9$ охранников и так до девятого поста, на который можно назначить любого из оставшихся $l_9 = 2$ охранников, при этом оставшийся $l_{10} = 1$ охранник будет назначен на десятый пост. Поэтому, согласно правилу произведения, у начальника службы безопасности есть

$$l_1 l_2 \dots l_{10} = 10 \cdot 9 \cdot \dots \cdot 2 \cdot 1 = 10! = 3628800$$

способов расстановки охранников по постам. Поскольку количество дней в месяце не превышает 31, у начальника службы безопасности заведомо существует достаточное число способов расстановки своих подчинённых по постам.

Второй способ. Число способов расстановки десяти охранников по десяти постам, существующих у начальника службы безопасности, описывается числом перестановок из 10 элементов, т. е. $P_{10} = 10! = 3628800$.

Упражнение

Определить, сколькими способами можно разместить на шахматной доске восемь ладей так, чтобы они не били друг друга.

Пример 4. Новый президент банка должен назначить двух новых вице-президентов из числа десяти директоров. Сколько способов существует у президента, если: а) один из вице-президентов (первый) выше другого по должности; б) вице-президенты по должности равны между собой.

Решение

Первый способ. а) Первого вице-президента можно выбрать из $l_1 = 10$ претендентов, при этом на пост второго вице-президента будут претендовать $l_2 = 9$ оставшихся директоров. Поэтому, согласно правилу произведения, у нового президента банка есть $l_1 l_2 = 10 \cdot 9 = 90$ способов назначения двух вице-президентов, один из которых подчиняется другому, из числа десяти директоров. б) Пусть первое действие заключается в том, что президент отбирает двух человек на должность вице-президентов, а второе действие - в том, что президент говорит отобранным людям, кто из них является первым вице-президентом, а кто - вторым. Пусть первое действие можно выполнить l_1 способами, второе действие, очевидно, можно выполнить $l_2 = 2$ способами, и по правилу произведения число способов назначения двух вице-президентов, один из которых подчиняется другому, из числа десяти директоров составляет $l_1 l_2 = 2l_1$. С другой стороны, в пункте а) мы нашли это число, и оно оказалось равным 90, поэтому $l_1 = \frac{90}{2} = 45$.

Второй способ. а) Число способов выбора двух кандидатов на две различные должности из десяти претендентов описывается числом размещений из 10 элементов по 2, т.е. $A_{10}^2 = 90$. б) Число способов выбора двух кандидатов на две одинаковые должности из десяти претендентов описывается числом сочетаний из 10 элементов по 2, т.е. $C_{10}^2 = \frac{10!}{2!8!} = 45$.

Упражнения

1. В кредитном отделе банка работают восемь человек. Сколько существует способов распределить между ними три премии: а) одинакового размера; б) разных размеров, известных заранее?

2. Одна из воюющих сторон захватила в плен 12 солдат, а другая 15. Определить, сколькими способами стороны могут обменять семерых военнопленных.

3. Петя и Маша коллекционируют видеокассеты. У Пети есть 30 комедий, 80 боевиков и 7 мелодрам, у Маши - 20 комедий, 5 боевиков и 90 мелодрам. Сколькими способами Петя и Маша могут обменяться тремя комедиями, двумя боевиками и одной мелодрамой?

4. В период сдачи итоговых контрольных работ в течение 20 дней студенты одной группы должны сдать итоговые контрольные работы по пяти дисциплинам. Сколькими способами можно составить расписание сдачи итоговых контрольных работ, если: а) запрещается сдавать две сдачи итоговых контрольных работ в один день; б) между двумя сдачами итоговых контрольных работ должно пройти хотя бы один день, для подготовки?

5. В банке девять учредителей. Регистрационные документы хранятся в сейфе. Сколько замков должен иметь сейф, и сколько ключей к ним нужно изготовить, чтобы доступ к содержимому сейфа был возможен только тогда, когда соберётся не менее шести учредителей?

6. Маша решила помириться с Петей и позвонить ему, но забыла две последних цифры его телефона и набирает их наудачу. Найти наибольшее возможное число неудачных попыток, которые сделает Маша, прежде чем дозвонится до Пети.

Пример 5. Маша очень любит пирожные и ежедневно в булочной, рядом с институтом покупает шесть пирожных (одинаковых или разных). Всего в булочной продаётся 11 сортов пирожных. Сколькими способами Маша может выбрать из них шесть штук?

Решение

Каждому набору пирожных, которые выберет Маша, будем ставить в соответствие последовательность нулей и единиц, определяемую по следующему правилу. Напишем подряд столько единиц, сколько пирожных первого вида выбрала Маша, далее поставим нуль и после него запишем количество отобранных пирожных второго вида и т.д. Например, комбинации «одно пирожное второго вида, три пирожных пятого вида и одно пирожное восьмого вида» соответствует такая последовательность:

«010001110001000»

(нули отделяют виды пирожных друг от друга, поэтому нуль после одиннадцатого вида не нужен). При этом каждому набору пирожных взаимно однозначным образом соответствует последовательность, построенная по описанному правилу. Все такие последовательности состоят, очевидно, из 16 знаков, причём 10 из них нули, которые могут занимать любое место. Поэтому количество способов выбора пирожных равно количеству всех таких последовательностей, т.е. числу размещений десяти нулей по 16 местам:

$$C_{16}^{10} = 8008.$$

Упражнение

В конкурсе по трём номинациям участвуют десять кинофильмов. Вычислить число вариантов распределения призов, если по каждой номинации установлены: а) различные призы; б) одинаковые призы.

Пример 6. Сколько различных слов можно составить, переставляя буквы в слове «мама»? Выписать все эти слова.

Решение

Число различных слов, которые можно составить, переставляя буквы в слове «мама», описывается числом перестановок с повторениями из $l = 4$ элементов (букв в слове «мама»), в которые первый элемент (буква «м») входит $l_1 = 2$ раза, а второй элемент (буква «а») - $l_2 = 2$ раза

$$(l_1 + l_2 = 4 = l).$$

Это число равно $\bar{P}_4(2,2) = \frac{4!}{2!2!} = 6$. Шесть различных слов, получающихся перестановками букв в слове «мама», таковы: «*mmaa*», «*мама*», «*маам*», «*амма*», «*амам*», «*аамм*».

Упражнение

Сколько различных слов можно составить, переставляя буквы в слове «*математика*»?

4. 2. Примеры вычисления вероятности по формуле (1)

1. В испытании, связанном с бросанием игральной кости, пространство элементарных исходов состоит из шести элементов: 1, 2, 3, 4, 5, 6. Вероятность того, что в результате опыта выпадет количество очков, большее двух, равна $4/6$ четыре благоприятствующих элементарных исходов: 3, 4, 5, 6 ($m = 4$) и общее количество элементарных исходов $n = 6$.

2. При игре в спортлото элементарным исходом является выбор шести различных чисел (шести пронумерованных шаров) из множества 1, 2, 3, ..., 43, 44, 45. Так как шары, на которых обозначены цифры, одинаковы по физическим параметрам, то естественно считать, что все шестерки шаров равновероятны (и, значит, требование, предъявляемое к пространству элементарных исходов, выполнено). Пространством элементарных исходов является множество неупорядоченных наборов из 6 чисел $\{1, 2, \dots, 45\}$, состоящее из

$$C_{45}^6 = \frac{45!}{(45-6)!6!} = 8145060$$

элементов (общее количество различных шестерок равно 8145060).

Вероятности событий A , связанных с рассматриваемым испытанием, можно вычислить по формуле $p(A) = \frac{m}{n}$, где m - количество элементарных исходов, благоприятствующих событию A , т.е. количеству шестерок, означающих наступление события A . Тогда вероятность того, что в результате опыта появится заранее заданная шестерка чисел (т.е. вероятность угадать все шесть чисел), равна

$$p(A) = \frac{m}{n} = \frac{1}{8145060} = 0,000000123.$$

3. Брошено две игральные кости. Предполагая, что элементарные события равновероятны, найти вероятность события $A = \{\text{сумма выпавших очков делится на } 6\}$.

Решение

Исход эксперимента (опыта) можно описать парой чисел $\omega_{ij} = (i, j)$, где i – число очков, выпавших на первой кости, а j – на второй ($i, j = 1, 2, \dots, 6$). Поэтому множество элементарных исходов

$$\Omega = \{\omega_{ij} = (i, j) : i = 1, 2, \dots, 6; \quad j = 1, 2, \dots, 6\}.$$

Нетрудно проверить, что число элементов множества Ω (число всех исходов эксперимента) - $n = 36$. Другими словами, общее число элементарных событий $n = 36$. Событие A соответствует подмножеству

$$\{(1,5), (5,1), (2,4), (4,2), (3,3), (6,6)\} \subset \Omega.$$

Другими словами

$$A = \{(1,5), (5,1), (2,4), (4,2), (3,3), (6,6)\} \subset \Omega.$$

Так как число элементов множества A (число всех благоприятствующих событию A исходов) - $m = 6$, то по формуле классической вероятности получаем

$$p(A) = \frac{m}{n} = \frac{6}{36} = \frac{1}{6}.$$

В этом примере помимо нахождения n - число элементов множества Ω (число всех исходов эксперимента) и m - число элементов множества A (число всех благоприятствующих событию A исходов) нам удалось, полностью описать то из чего состоят множества Ω и A . Однако во многих случаях описания содержания этих множеств является практически невозможным. Следует отметить, что это и не является обязательным. Обязательным

зается нахождение чисел m и n . Нахождение этих чисел, следовательно, вычисление вероятностей в классической схеме облегчается, если применить элементы комбинаторики см. п.п. 4.1.

Еще заметим, что подсчет числа элементов тех или иных подмножеств множества Ω часто облегчается благодаря следующей формуле. Число элементов прямого произведения множеств равно произведению числа элементов составляющих множеств, т. е.

$$N(\Omega_1 \times \Omega_2 \times \dots \times \Omega_l) = N(\Omega_1)N(\Omega_2) \dots N(\Omega_l).$$

При решении вероятностных задач важно выделять опыты, где можно использовать те или иные комбинаторные формулы. Каждая из комбинаторных формул определяет общее число элементарных исходов в некотором опыте, состоящем в выборе наудачу k элементов из l различных элементов исходного множества $E = \{e_1, e_2, \dots, e_l\}$. При этом, в постановке каждого такого опыта строго оговорено, каким способом производится выбор и, что понимается под различными выборками.

Существуют две принципиально различные схемы выбора. В первой схеме выбор осуществляется без возвращения элементов (это значит, что отбираются либо сразу все k элементов, либо последовательно по одному элементу, причем каждый отобранный элемент исключается из исходного множества). Во второй схеме выбор осуществляется поэлементно с обязательным возвращением отобранного элемента на каждом шаге и тщательным перемешиванием исходного множества перед следующим выбором. После того как выбор тем или иным способом осуществлен, отобранные элементы (или их номера) могут быть либо упорядочены (т.е. выложены в последовательную цепочку), либо нет. В результате получаются следующие четыре различные постановки опыта по выбору наудачу k элементов из общего числа l различных элементов множества E .

Первая схема

а) Схема выбора, приводящая к сочетаниям

Пример 4. Из партии, содержащей 10 изделий, среди которых 3 бракованных, наудачу извлекают три изделия для контроля. Найти вероятности следующих событий:

$A = \{\text{в полученной выборке ровно одно изделие бракованное}\};$

$B = \{\text{в полученной выборке нет ни одного бракованного изделия}\}.$

Решение

Занумеруем изделия числами от 1 до 10, и пусть множество номеров $E_1 = \{1, 2, \dots, 7\}$ соответствует годным изделиям, а множество номеров $E_2 = \{8, 9, 10\}$ - бракованным.

Согласно описанию эксперимента производится выбор без возвращения и без упорядочивания трех элементов из множества $E = E_1 \cup E_2 = \{1, 2, \dots, 10\}$. Поэтому $n = C_{10}^3 = 120$.

Событию A благоприятствуют только такие исходы, когда один элемент выборки принадлежит E_2 , а остальные два элемента - множеству E_1 . По формуле прямого произведения множеств получаем, что число всех таких исходов $m = C_3^2 \cdot C_7^1 = 63$, поэтому

$$p(A) = \frac{m}{n} = \frac{63}{120} = \frac{21}{40}.$$

Событию B благоприятствуют только такие исходы, когда все три отобранных элемента принадлежат множеству E_1 , поэтому $m = C_7^3 = 35$. Отсюда следует, что

$$p(B) = \frac{m}{n} = \frac{35}{120} = \frac{7}{24}.$$

Упражнения

1. Из десяти первых букв русского алфавита наудачу составляется новый алфавит, состоящий из трех букв. Какова вероятность того, что случайно выбранный алфавит будет содержать букву a ?

2. Из полного набора домино (28 штук) наудачу выбирают 7 костей. Какова вероятность, что среди них окажется по крайней мере одна кость с шестью очками?

3. Из десяти первых букв русского алфавита наудачу составляется новый алфавит, состоящий из пяти букв. Найти вероятности следующих событий:

$A = \{\text{в состав нового алфавита входит буква } o\};$

$B = \{\text{в состав нового алфавита входят только согласные буквы}\}.$

4. Среди кандидатов в студенческий совет факультета 3 первокурсника, 5 второкурсников и 7 третьекурсников. Из этого состава наудачу выбирают пять человек на предстоящую конференцию. Найти вероятности следующих событий:

$A = \{\text{будут выбраны одни третьекурсники}\};$

$B = \{\text{все первокурсники попадут на конференцию}\};$

$C = \{\text{не будет выбрано ни одного второкурсника}\}.$

5. Из колоды в 52 карты извлекаются наудачу 4 карты. Найти вероятности следующих событий:

$A = \{\text{в полученной выборке все карты бубновой масти}\},$

$B = \{\text{окажется хотя бы один туз}\}.$

б) Схема выбора, приводящая к размещению

Пример 5. Множество E состоит из 10 первых букв русского алфавита. Опыт состоит в выборе без возвращения 4 букв и записи слова в порядке поступления букв. Сколько 4-буквенных слов может быть получено в данном опыте? Какова вероятность того, что наудачу составленное слово будет оканчиваться буквой a ?

Решение

n число всех 4-буквенных слов в данном опыте - равно числу 4-элементных упорядоченных подмножеств из 10 элементов, т.е.

$$n = A_{10}^4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040.$$

Пусть событие $A = \{\text{наудачу составленное слово из 4 букв множества } E \text{ оканчивается буквой } a\}$. Число элементов множества A равно числу способов разместить на три оставшиеся места по одному символу из 9 (символ a исключен из рассмотрения, поскольку его место уже определено); таким образом,

$$m = A_3^9 = 9 \cdot 8 \cdot 7 = 504$$

и

$$p(A) = \frac{m}{n} = \frac{504}{5040} = \frac{1}{10}.$$

Упражнения

1. Числа 1, 2, ..., 9 записываются в случайном порядке. Найти вероятности следующих событий:

$A = \{\text{числа будут записаны в порядке возрастания}\};$

$B = \{\text{числа 1 и 2 будут стоять рядом и в порядке возрастания}\};$

$C = \{\text{числа 3, 6 и 9 будут следовать друг за другом и в порядке возрастания}\};$

$D = \{\text{на четных местах будут стоять четные числа}\};$

$E = \{\text{сумма каждых двух чисел, стоящих на одинаковом расстоянии от концов, равна 10}\}.$

2. Группа, состоящая из 8 человек, занимает места с одной стороны прямоугольного стола. Найти вероятность того, что два определенных лица окажутся рядом, если а) число мест равно 8; б) число мест равно 12.

3. На пяти карточках написаны цифры от 1 до 5. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятности следующих событий:

$A = \{\text{появится число 123}\};$

$B = \{\text{появится число, не содержащее цифры 3}\},$

$C = \{\text{появится число, состоящее из последовательных цифр}\};$

$D = \{\text{появится четное число}\};$

$E = \{\text{появится число, содержащее хотя бы одну из цифр 2 или 3}\}.$

4. 10 вариантов контрольной работы, написанные каждый на отдельной карточке, перемешиваются и распределяются случайным образом среди восьми студентов, сидящих в одном ряду, причем, каждый получает по одному варианту. Найти вероятности следующих событий:

$A = \{\text{варианты с номерами 1 и 2 останутся неиспользованными}\};$

$B = \{\text{варианты 1 и 2 достанутся рядом сидящим студентам}\};$

$C = \{\text{будут распределены последовательные номера вариантов}\}.$

5. 12 студентов, среди которых Иванов и Петров, случайным образом занимают очередь за учебниками в библиотеку. Какова вероятность, что между Ивановым и Петровым, в образовавшейся очереди, окажутся ровно 5 человек?

Вторая схема

а) Схема выбора, приводящая к сочетаниям с повторениями

Пример 6. В технической библиотеке имеются книги по математике, экономике, статистике и т.д., всего по 16 разделам науки. Поступили очередные четыре заказа на литературу. Считая, что любой состав заказанной литературы равновозможен, найти вероятности следующих событий:

$A = \{\text{заказаны книги из различных разделов науки}\},$

$B = \{\text{заказаны книги из одного и того же раздела науки}\}.$

Решение

Число всех равновероятных исходов данного эксперимента равно, очевидно, числу сочетаний с повторениями из 16 элементов по 4, т.е.

$$n = C_{16+4-1}^4 = C_{19}^4.$$

Число исходов, благоприятствующих событию A , равно числу способов отобрать без возвращения четыре элемента из 16, т.е. $m = C_{16}^4$, поэтому

$$p(A) = \frac{m}{n} = \frac{C_{16}^4}{C_{19}^4} = \frac{455}{969} \approx 0,47.$$

Число исходов, благоприятствующих событию B , равно числу способов выбрать один элемент из шестнадцати, поэтому

$$p(A) = \frac{m}{n} = \frac{C_{16}^1}{C_{19}^4} = \frac{4}{969} \approx 0.004.$$

Упражнения

1. В кондитерской имеется 7 видов пирожных. Очередной покупатель выбил чек на 4 пирожных. Считая, что любой заказываемый набор пирожных равновероятен, вычислить вероятность того, что покупатель заказал: а) пирожные одного вида; б) пирожные разных видов; в) по два пирожных различных видов.

2. Из общего количества костей домино, содержащих числа $0, 1, 2, \dots, n$, наудачу извлекли одну кость. Оказалось, что это не дубль. Какова вероятность p_n того, что вторую, извлеченную также наудачу кость домино, можно будет приставить к первой. Найти числовые значения вероятности для $n = 6$ (обычный набор домино) и $n = 9$ (расширенный набор).

б) Схема выбора, приводящая к размещениям с повторениями

Пример 7. 7 одинаковых шариков случайным образом рассыпаются по 4 лункам (в одну лунку может поместиться любое число шариков). Сколько существует различных способов распределения 7 шариков по 4 лункам? Какова вероятность того, что в результате данного опыта первая лунка окажется пустой (при этом может оказаться пустой и еще какая-либо лунка)?

Решение

Занумеруем лунки и шарики. Можно считать, что опыт состоит в 7-кратном выборе с возвращением номера лунки и записи 7-буквенного слова. При этом, каждому порядковому номеру буквы (номеру шарика) будет поставлена в соответствие одна из четырех букв алфавита (номер лунки).

Так, например, слово

1	1	3	1	4	4	2
1	2	3	4	5	6	7

значает, что в первую лунку попали шары № 1, № 2 и № 4, во вторую лунку - шар № 7, в третью - шар № 3, в четвертую - шары № 5 и № 6. Таким образом, число всех способов распределить 7 шариков по 4 лункам равно числу различных 7-буквенных слов из алфавита в 4 буквы, т.е. $n = 4^7$.

Событие $A = \{\text{первая лунка окажется пустой}\}$ соответствует такому выбору, когда символ 1 (номер первой лунки) удален из алфавита. Поэтому $m = 3^7$ и

$$p(A) = \frac{m}{n} = \frac{3^7}{4^7} = \left(\frac{3}{4}\right)^7 \approx 0,133.$$

Упражнения

1. Бросается 10 одинаковых игральных костей. Вычислить вероятности следующих событий:

$A = \{\text{ни на одной кости не выпадет 6 очков}\};$

$B = \{\text{хотя бы на одной кости выпадет 6 очков}\};$

$C = \{\text{ровно на 3 костях выпадет 6 очков}\}.$

2. Опыт состоит в четырехкратном выборе с возвращением одной из букв алфавита $E = \{a, б, к, о, м\}$ и выкладывании слова в порядке поступления букв. Какова вероятность того, что в результате будет выложено слово «мама»?

3. В подъезде дома установлен замок с кодом. Дверь автоматически отпирается, если в определенной последовательности набрать три цифры из имеющихся десяти. Некто вошел в подъезд и, не зная кода, стал наудачу пробовать различные комбинации из трех цифр. На каждую попытку он тратит 20 секунд. Какова вероятность события $A = \{\text{вошедшему удастся открыть дверь за один час}\}?$

4. Телефонная книга раскрывается наудачу и выбирается случайный номер телефона. Считая, что телефонные номера состоят из 7 цифр, причем все комбинации цифр равновероятны, найти вероятности следующих событий:

$A = \{\text{четыре последние цифры телефонного номера одинаковы}\};$

$B = \{\text{все цифры различны}\};$

$C = \{\text{номер начинается с цифры 5}\};$

$D = \{\text{номер содержит три цифры 5, две цифры 1 и две цифры 2}\}.$

5. Шесть человек вошли в лифт на первом этаже семиэтажного дома. Считая, что любой пассажир может с равной вероятностью выйти на 2-м, 3-м, ..., 7-м этажах, найти вероятности следующих событий:

$A = \{\text{на втором, третьем и четвертом этажах не выйдет ни один из пассажиров}\};$

$B = \{\text{трое пассажиров выйдут на седьмом этаже}\};$

$C = \{\text{на каждом этаже выйдет по одному пассажиру}\};$

$D = \{\text{все пассажиры выйдут на одном этаже}\}.$

6. К четырехстороннему перекрестку с каждой стороны подъехало по одному автомобилю. Каждый автомобиль может с равной вероятностью совершить один из четырех маневров на перекрестке: развернуться и поехать обратно, поехать прямо, налево или направо. Через некоторое время все автомобили покинули перекресток. Найти вероятности следующих событий:

$A = \{\text{все автомобили поедут по одной и той же улице}\};$

$B = \{\text{по определенной улице поедут ровно три автомобиля}\};$

$C = \{\text{по крайней мере по одной из улиц не поедет ни один автомобиль}\}.$

7. Из разрезной азбуки выкладывается слово математика. Затем все буквы этого слова тщательно перемешиваются и снова выкладываются в случайном порядке. Какова вероятность того, что снова получится слово «математика»?

8. 52 карты раздаются четырем игрокам (каждому по 13 карт). Найти вероятности следующих событий:

$A = \{\text{каждый игрок получит туза}\};$

$B = \{\text{один из игроков получит все 13 карт одной масти}\};$

$C = \{\text{все тузы попадут к одному из игроков}\};$

$D = \{\text{двое определенных игроков не получают ни одного туза}\}.$

5. Статистический подход к понятию вероятности

Существует еще один подход к понятию вероятности, не связанный с пространством элементарных событий. В основе этого подхода лежит явление устойчивости частоты наступления события при многократном повторении испытаний. Пусть в результате некоторого опыта возможно появление события A . Повторим опыт n раз и подсчитаем количество m наступлений событий A . Предположим, что в данной серии опытов результаты предшествующих испытаний не влияют на последующие. Как известно из практики, отношение $\frac{m}{n}$ мало изменяется при больших n несмотря на то, что событие A случайное и величина m также зависит от случая. В частности, если монету подбрасывать многократно, то отношение количества выпавших гербов к общему числу бросаний монеты приближенно будет равно $1/2$.

Отношение $\frac{m}{n}$ называется *частотой* появления события A в n испытаниях.

При условии устойчивости частоты появления события A число $\frac{m}{n}$ удовлетворяет требованиям, предъявляемым к вероятности как к характеристике того, насколько вероятно изучаемое событие с точки зрения практики и здравого смысла. Действительно, число $\frac{m}{n}$ удовлетворяет неравенству $0 \leq \frac{m}{n} \leq 1$, более вероятным с точки зрения здравого смысла событиям соответствует большее значение $\frac{m}{n}$, для невозможных событий $\frac{m}{n} = 0$, для досто-

верных $\frac{m}{n} = 1$; если событие A является суммой двух несовместных событий

B и C , т. е. $A = B + C$, то $\frac{m_A}{n} = \frac{m_B}{n} + \frac{m_C}{n}$, где m_A , m_B , m_C - количество на-

ступлений событий в n опытах. В силу сказанного (и при том условии, что

возможно многократное повторение опыта с устойчивостью частоты) отно-

шение $\frac{m}{n}$ называют *статистической вероятностью* события A .

В дальнейшем будет доказано (закон больших чисел Бернулли), что если для события A определена его классическая вероятность p , то частное

$\frac{m_n}{n}$, где m_n случайная величина, выражающая количество наступлений со-

бытия A в n независимых друг от друга испытаниях, при больших n «почти всегда» мало отличается от p . Таким образом, известное из практики явление оказывается строго доказанным математическим утверждением.

Следует заметить, что статистическое определение вероятности в некоторых случаях единственно доступно для измерения и исследования: действительно, далеко не всегда можно изучать случайные события с помощью пространства элементарных событий. Например, при изучении стрельбы по мишени невозможно указать разумную модель пространства элементарных событий, отражающую существенные стороны исследуемых испытаний.

6. Геометрические вероятности

Рассмотрим, наконец, вопрос о том, что называется вероятностью события в строго математическом смысле. Строгое определение понятия вероятности развивает идею конечного пространства элементарных событий и приводится в солидных учебниках по теории вероятностей. Здесь будут указаны лишь главные черты такого определения.

Пусть дано некоторое множество X (пространство элементарных событий, описанное в п. 2). Изучаемые события представляют собой подмно-

жества в X . Задана также функция p , которая подмножествам в X сопоставляет число из числового отрезка $[0,1]$. Требуется, чтобы эта функция обладала свойством аддитивности, т.е., чтобы для всяких двух непересекающихся подмножеств $A, B \subset X$, для которых p определена, было справедливо равенство

$$p(A \cup B) = p(A) + p(B); \quad (9)$$

кроме того, $p(\emptyset) = 0$, $p(X) = 1$.

Определение 10. Значение функции p на множестве A называется **вероятностью события A** .

Следует отметить, что множество X в отличие от п. 4¹ может быть бесконечным; например, X - отрезок числовой прямой, вся прямая, квадрат на плоскости, вся плоскость и т. п.

Часто функцию p бывает невозможно определить для всех подмножеств множества X ; в этом случае ее определяют только для некоторых (так называемых измеримых) подмножеств (подробнее см. в солидных учебниках по теории вероятностей).

Наглядным примером вероятностной модели может служить *геометрическая вероятность*.

Пусть X - множество точек квадрата на плоскости, т.е.

$$a \leq x \leq b, \quad a \leq y \leq b.$$

Для подмножества $A \subset X$ определим значение f равенством

$$p(A) = \frac{S(A)}{(b-a)^2}, \quad (10)$$

¹ Классическое определение вероятности предполагает, что число элементарных исходов испытания – конечно. На практике же весьма часто встречаются испытания, число возможных исходов которых – бесконечно. В таких случаях классическое определение неприменимо.

где $S(A)$ - площадь фигуры A . Ясно, что $p(X)=1$. Свойство (9) для функции (10) выполняется, поскольку площадь объединения двух непересекающихся фигур равна сумме площадей этих фигур.

Пример. Два приятеля договорились встретиться в установленном месте в промежутке времени от 6 до 7 ч. По взаимному соглашению каждый приходит на место встречи в случайный наугад выбранный момент и ждет другого ровно 10 мин. Какова вероятность того, что приятели встретятся?

Решение. Пусть x и y означают моменты прихода на место встречи первого и второго приятеля, соответственно. Такое событие удобно отметить точкой квадрата (рис. 3). Условие встречи заключается в том, что

$$|x - y| < \frac{1}{6} \quad (11)$$

(сторона квадрата соответствует часу времени, 10 мин составляет $1/6$ ч).

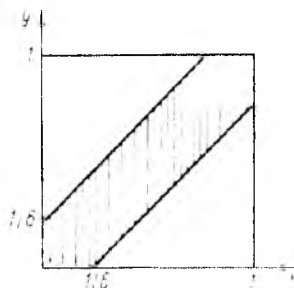


Рис. 3

Множество точек, удовлетворяющих условию (11), отмечено на рисунке штриховкой. Площадь этого множества равна $11/36$. Согласно формуле (10), вероятность встречи

$$p = \frac{11/36}{(7-6)^2} = \frac{11}{36}.$$

Здесь применение формулы (10) обосновано следующими соображениями. По условию каждый из приятелей приходит на место встречи, выбирая момент прихода наугад.

Формула (10) естественным образом обобщается на случай пространства любой размерности.

Пусть $\Omega \subset R^n$. Для подмножества $A \subset \Omega$ определим значение p равенством

$$P(A) = \frac{\text{mes}(A)}{\text{mes}(\Omega)},$$

где $\text{mes}(A)$ – мера множества A (длина, площадь, объем и т.д. в зависимости от размерности того пространства, в котором рассматриваются данные множества).

Упражнения

1. Внутри квадрата с вершинами $(0,0)$, $(1,0)$, $(1,1)$ и $(0,1)$ наудачу выбирается точка $M(x, y)$. Найти вероятности следующих событий:

$$A = \{(x, y) : \max(x, y) < a, a > 0\};$$

$$B = \{(x, y) : \min(x, y) < a, 0 \leq a \leq 1\};$$

$$C = \{(x, y) : x^2 + y^2 < a, a > 0\};$$

$$D = \{(x, y) : xy < a, a > 0\}.$$

2. Какова вероятность того, что сумма трех наудачу взятых отрезков, длина каждого из которых не превосходит l , будет больше l ?

3. Луч локатора перемещается в горизонтальной плоскости с постоянной угловой скоростью. Какова вероятность того, что цель будет обнаружена в угловом секторе α радиан, если появление цели по любому направлению одинаково возможно?

4. Значения a и b равновозможны в квадрате $|a| \leq 1, |b| \leq 1$. Найти вероятности следующих событий:

$$A = \{\text{корни квадратного трехчлена } x^2 + 2ax + b \text{ действительны}\};$$

$$B = \{\text{корни квадратного трехчлена } x^2 + 2ax + b \text{ положительные}\}.$$

5. Какова вероятность, не целясь, попасть бесконечно малой пулей в прутья квадратной решетки, если толщина прутьев равна a , а расстояние между их осями равно l ($l > a$)?

6. Иван и Петр договорились о встрече в определенном месте между одиннадцатью и двенадцатью часами. Каждый приходит в случайный момент указанного промежутка и ждет появления другого до истечения часа, но не более 15 минут, после чего уходит. Наблюдаемый результат - пара чисел (x, y) , где x - время прихода Петра, y - время прихода Ивана (время исчисляется в минутах, начиная от 11 часов). Найти вероятности событий:

$B = \{\text{Петр пришел после Ивана}\};$

$D = \{\text{встреча не состоялась}\};$

$F = \{\text{Ивану не пришлось ждать Петра}\};$

$K = \{\text{встреча состоялась, когда до истечения часа оставалось меньше пяти минут}\}.$

7. Из отрезка $[-1, 2]$ наудачу взяты два числа. Какова вероятность того, что их сумма больше единицы, а произведение меньше единицы?

8. (**Задача Бюффона**). На плоскость, разграфленную параллельными прямыми линиями, отстоящими друг от друга на расстоянии $2a$, наудачу бросается игла длиной $2l$. Какова вероятность того, что игла пересечет одну из параллельных прямых, если $l < a$.

Вопросы для самопроверки

1. Приведите интуитивное определение случайного события. Приведите примеры случайных событий.

2. Дайте определения достоверного и невозможного событий. Какими символами они обозначаются?

3. Когда события называют несовместными?

4. Какие события называют равновозможными?

5. Какое случайное событие называется массовым или статистическим?

Чем отличаются массовые случайные события от единичных?

6. Чем занимается теория вероятностей?
7. С каким понятием связан точный смысл основных понятий теории вероятностей, в том числе, понятия «случайного события»?
8. Каково должно быть строгое математическое понятие «случайного события»?
9. Из чего состоит всякий случайный опыт (эксперимент, испытание)?
10. Что означает воспроизводимый опыт?
11. Что может быть предметом наблюдения в том или ином опыте?
12. Что означает понятие «наблюдаемый результат» для реально воспроизводимого опыта?
13. Как интерпретируется любой наблюдаемый результат?
14. Что понимают под понятием пространства (множества) элементарных исходов?
15. Как интерпретируется любое подмножество пространства (множества) элементарных исходов?
16. Из чего состоит поле событий для данного опыта?
17. Исходя из теоретико-множественного подхода разъясните понятие события; достоверного, невозможного событий и двух совместных (несовместных) событий.
18. Какую структуру может иметь множество Ω - пространство элементарных исходов? Прокомментируйте возможные варианты этого множества.
19. Исходя из какого требования на практике осуществляется построение множества Ω ?
20. Когда строго не определяемое понятие «элементарный исход» становится более определенным?
21. Введите операции суммы, произведения событий и понятия противоположного события исходя из теоретико-множественного подхода к понятию события.
22. Что называется алгеброй событий?
23. Дайте классическое определение вероятности события.
24. Что называется классической схемой или схемой урн?

25. Сформулируйте два универсальных правила, применяемых при решении комбинаторных задач.

26. Что называется факториалом натурального числа l ?

27. Когда множество, состоящее из k элементов называется упорядоченным?

28. Что называются размещениями из l элементов по k ? Чему равно число размещений из l элементов по k ?

29. Что называются перестановками из l элементов размещения из l элементов по l ? Чему равно число перестановок из l элементов?

30. Что называются сочетаниями из l элементов по k ?

31. Что называются размещениями с повторениями из l элементов по k ? Чему равно число размещений с повторениями из l элементов по k ?

32. Что называются сочетаниями с повторениями из l элементов по k ? Чему равно число сочетаний с повторениями из l элементов по k ?

33. Если во множестве S , состоящем из l элементов, есть только r различных элементов, то что называются перестановками с повторениями из l элементов? Чему равно число перестановок с повторениями из l элементов в данном случае?

34. Существуют две принципиально различные схемы выбора. Разъясните эти схемы.

35. Прокомментируйте четыре различные постановки опыта по выбору наудачу k элементов из общего числа l различных элементов множества.

36. Разъясните статистический подход к понятию вероятности. Что лежит в основе этого подхода?

37. В каких случаях целесообразно применить статистическое определение вероятности?

38. Что называется вероятностью события в строго математическом смысле? Укажите главные черты строгого определения понятия вероятности, которая развивает идею конечного пространства элементарных событий.

39. Разъясните понятие «геометрическая вероятность». Почему геометрическую вероятность считают наглядным примером вероятностной модели?

§2. Теоремы сложения и умножения вероятностей. Формула полной вероятности и формулы Байеса

Ключевые слова: аксиомы теории вероятностей, независимость (зависимость) одного события от другого, теорема сложения вероятностей, теорема умножения вероятностей, попарно независимость событий, независимость событий в совокупности, полная группа событий, теорема о полной вероятности, теорема гипотез или Байеса.

1. Теорема сложения вероятностей

В п. 6 §1 в попытке определения вероятности события в строго математическом смысле на множестве событий - системе подмножеств множества Ω - пространства элементарных событий рассмотрели функцию $p(A)$ с требованием выполнения условия аддитивности и $p(\emptyset)=0$, $p(X)=1$. Затем вероятность события A определили как значение функции p на множестве A .

Ниже эту идею, следуя А.Н. Колмогорову, сформулируем в виде аксиом.

Вероятностью $p(A)$ называется числовая функция, определенная на¹ множестве событий - системе подмножеств множества Ω - пространства элементарных событий и удовлетворяющая трем условиям (аксиомам вероятностей):

1⁰. Каждому событию A поставлено в соответствие неотрицательное число $p(A)$.

2⁰. Если события $A_1, A_2, \dots$ попарно несовместны, то

$$p(A_1 + A_2 + \dots) = p(A_1) + p(A_2) + \dots$$

¹ поле событий (см. п.2 §1) для данного опыта, и удовлетворяющая трем условиям (аксиомам вероятностей): Далее по тексту.

Заметим, что при бесконечном числе событий $A_1, A_2, \dots$ в правой части написанного равенства стоит *сумма ряда*.

Очевидно, если множество Ω является конечным, то любая совокупность попарно не пересекающихся подмножеств состоит лишь из конечного числа подмножеств. Отсюда ясно, что для случая конечного Ω аксиома 2^0 равнозначна такому (в общем случае более слабому) требованию:

$$2^{*0}. \quad p(A + B) = p(A) + p(B), \text{ если } A \text{ и } B \text{ несовместны.}$$

Чтобы подчеркнуть существенное различие между аксиомами 2^0 и 2^{*0} , часто аксиому 2^{*0} называют аксиомой аддитивности, а 2^0 - аксиомой счетной аддитивности.

$$3^0. \quad p(\Omega) = 1.$$

Аксиомы 1-3 составляют основу всей теории вероятностей. Все теоремы этой теории, включая самые сложные, выводятся из них формально-логическим путем.

Укажем несколько примеров такого вывода.

Следствие. Исходя из очевидного соотношения между подмножествами

$$A + \bar{A} = \Omega,$$

и применяя аксиомы 2^0 и 3^0 , получаем, что

$$p(A) + p(\bar{A}) = 1,$$

т.е. сумма вероятностей противоположных событий равна единице.

В качестве другого следствия аксиом выведем следующее утверждение.

Теорема 1. Вероятность суммы двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:

$$p(A + B) = p(A) + p(B) - p(AB). \quad (*)$$

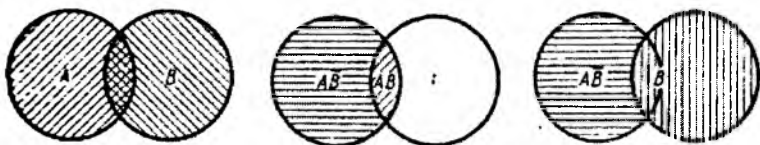
Доказательство. Рассмотрим очевидные соотношения между событиями (как между подмножествами множества Ω):

$$A = AB + A\bar{B}, \quad A + B = B + A\bar{B}.$$

Применяя к обоим равенствам аксиому сложения, получим два число-
~~вые~~ равенства:

$$p(A) = p(AB) + p(A\bar{B}), \quad p(A+B) = p(B) + p(A\bar{B}).$$

Вычитая из последнего равенства предыдущее, приходим к формуле



Теорема 1 допускает следующее обобщение.

Теорема 2 (обобщенная теорема сложения вероятностей). Если $A_1, A_2, \dots, A_n$ совместные события, то справедливо

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + \dots + P(A_n) - P(A_1 A_2) - \dots - P(A_{n-1} A_n) + \\ + p(A_1 A_2 A_3) + p(A_1 A_2 A_4) + \dots + p(A_{n-2} A_{n-1} A_n) - \dots$$

Доказательство опускаем.

2. Условная вероятность. Теорема умножения вероятностей.

Определение 1. Событие A называется *независимым* от события B , если вероятность события A не зависит от того, наступило событие B или нет. В противном случае событие A называется *зависимым* от события B .

Комментарий к определению 1. В дальнейшем будет доказано, что если случайное событие A не зависит от B , то и B не зависит от A .

Примеры

1. Монету подбрасывают два раза. Событие A состоит в том, что в первый раз выпадет герб, а событие B - в том, что во второй раз выпадет решка. Эти события, очевидно, независимы, поскольку второе бросание никак не может повлиять на первое.

2. Из деревьев леса наудачу выбирают одно. Событие A состоит в том, что дерево высокое (выше некоторого уровня, равного, например, 15 м), а

событие B - в том, что дерево толстое (скажем, диаметр ствола больше 1 м). Эти события зависимы, поскольку вероятность выбрать высокое дерево увеличивается, если выбирать его среди толстых деревьев.

Определение 2. Вероятность наступления события A при условии, что событие B наступило, называется *условной вероятностью* и обозначается $p_B(A)$.

Теорема 1 (теорема умножения вероятностей). Вероятность произведения событий A и B равна произведению вероятности события A на условную вероятность события B при условии, что A произошло, т.е.

$$p(AB) = p(A)p_A(B). \quad (1)$$

Для независимых событий A и B справедливо равенство

$$p(AB) = p(A)p(B). \quad (2)$$

Доказательство. Убедимся в справедливости теоремы для случая конечного пространства элементарных исходов.

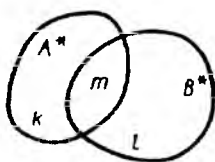


Рис. 1

Пусть событиям A и B соответствуют множества A^* и B^* , изображенные на рис. 1. Далее, пусть область A^* содержит k точек, область B^* - l точек и в пересечении $A^* \cap B^*$ содержится m точек; всего же в пространстве элементарных исходов имеется n точек. По определению, $p(A) = \frac{k}{n}$, $p(AB) = \frac{m}{n}$ (произведение AB изображается множеством $A^* \cap B^*$).

Заметим, что условная вероятность $p_A(B) = \frac{m}{k}$.

Действительно, событию B при условии, что A произошло, благоприятствуют m элементарных исходов, а наступление события A означает наступление одного из k элементарных исходов. Так как $\frac{m}{n} = \frac{k}{n} \cdot \frac{m}{k}$, то

$$p(AB) = p(A)p_A(B),$$

т. е. имеет место равенство (1).

В случае, когда событие B не зависит от события A , имеем $p(B) = p_A(B)$ и равенство (1) примет вид $p(AB) = p(A)p(B)$.

Теперь введем более широкое понятие независимости, а именно:

Определение 3. Говорят, что событие A не зависит от B если выполняется равенство (2).

В дальнейшем независимость A от B будет пониматься как выполнение равенства (2).

Следствие 1. Если случайное событие B не зависит от A , то и A не зависит от B .

Действительно, по условию $p(AB) = p(A)p(B) = p(B)p(A)$; кроме того, по теореме умножения вероятностей $p(AB) = p(B)p_B(A)$. Поэтому в случае $p(B) \neq 0$ имеем $p(A) = p_B(A)$, т.е. событие A не зависит от B .

Иначе говоря, отношение независимости является симметричным. Поэтому в дальнейшем мы можем говорить просто о независимых событиях A и B .

Следствие 2. Нетрудно видеть, что если A и B независимы, то независимы также события $\bar{A}$ и B , A и $\bar{B}$, $\bar{A}$ и $\bar{B}$.

Примеры

1. В урне находятся 3 черных и 2 белых шара. Из урны извлекают последовательно (без возвращения) два шара. Событие A состоит в том, что первым будет взят белый шар, а событие B - в том, что второй

шар окажется черным. Найти вероятность произведения (т.е. совместного наступления) событий A и B .

Решение

В силу теоремы умножения вероятностей имеем

$$p(AB) = p(A)p_A(B).$$

Очевидно, что $p(A) = \frac{2}{5}$. Так как после извлечения белого шара в урне осталось 4 шара - 1 белый и 3 черных, то при этих условиях вероятность извлечения черного шара $p_A(B) = \frac{3}{4}$. Итак, $p(AB) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} = \frac{3}{10}$.

2. Какова вероятность выпадения двух гербов при двукратном бросании монеты?

Решение

Пусть A - выпадение герба при первом бросании, а B - выпадение герба при втором бросании; тогда $p(AB) = p(A)p(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Аналогично, применяя теорему умножения несколько раз, получаем: вероятность того, что при n бросаниях монеты герб выпадает n раз, равна $\left(\frac{1}{2}\right)^n$.

Замечание. Выразим условную вероятность из соотношения (1), считая $p(A) \neq 0$:

$$p_A(B) = \frac{p(AB)}{p(A)}. \quad (3)$$

Пример 1. Все грани игральной кости заклеены непрозрачной бумагой: грани 1, 2, 3 – красной, грани 4, 5, 6 – черной. При бросании кости выпала черная грань. Какова вероятность того, что на этой грани стоит четное число?

Решение

Очевидно, мы должны найти условную вероятность $p_A(B)$, где событие B есть выпадение четного числа очков, а событие A – выпадение числа большего 3. Имеем:

$$p_A(B) = \frac{p(AB)}{p(A)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}.$$

Для сравнения отметим, что безусловная вероятность события B (просте A B) равна $\frac{1}{2}$.

Пример 2. Из колод игральных карт наугад выбирают одну карту. Пусть событие A заключается в том, что вынутая карта является «тузом», а события B – в том, что карта красной масти («бубновая» или «червовая»). Интуитивно ясно, что A не зависит от B (цена карты не зависит от масти). Проверим это подсчетом. Так как

$$p(A) = \frac{4}{36}, \quad p(B) = \frac{18}{36}, \quad p(AB) = \frac{2}{36},$$

то равенство (2) выполняется. Следовательно, события A и B независимы.

В практических вопросах для установления независимости одного события от другого редко прибегают к проверке равенства (2). Обычно при этом довольствуются интуитивными соображениями. Так, например, если бросают подряд две монеты, то ясно, что выпадение той или другой стороны на одной монете не оказывает никакого влияния на условия бросания другой, и, значит, следующие два события: выпадение герба на одной монете (событие A) и выпадение герба на другой (событие B) – являются независимыми.

Определение 3. Несколько событий называют *попарно независимыми*, если каждые два из них независимы.

Пример 3. Монета брошена 3 раза. Пусть A, B, C – события, состоящие в появлении герба, соответственно, в первом, втором и третьем испытаниях.

Ясно, что каждые два из рассматриваемых событий (т.е. A и B , A и C , B и C) - независимы.

Таким образом, события A, B и C - попарно независимые.

Определение 4. Несколько событий называют *независимыми в совокупности*, если каждое из них и любая комбинация остальных событий (содержащая либо все остальные события, либо часть из них) есть события независимые.

Например, если события A_1, A_2 и A_3 независимые в совокупности, то независимыми являются события: A_1 и A_2 , A_1 и A_3 , A_2 и A_3 , $A_1 A_2$ и A_3 , $A_1 A_3$ и A_2 , $A_2 A_3$ и A_1 .

Комментарий к определению 4. Если несколько событий независимы попарно, то отсюда еще не следует их независимость в совокупности. В этом смысле требование независимости событий в совокупности сильнее требования их попарной независимости.

Привести пример, показывающий, что из попарной независимости событий A, B, C не следует их независимость в совокупности.

Теорема 2 (обобщенная теорема умножения вероятностей). Вероятность произведения нескольких событий равна произведению вероятности одного из них на условные вероятности всех остальных, причем вероятность каждого последующего события вычисляется в предположении, что все предыдущие события уже появились:

$$p(A_1 A_2 \dots A_n) = p(A_1) \cdot p_{A_1}(A_2) \cdot p_{A_1 A_2}(A_3) \dots p_{A_1 A_2 \dots A_{n-1}}(A_n), \quad (4)$$

где $p_{A_1 A_2 \dots A_{n-1}}(A_n)$ - вероятность события, вычисленная в предположении, что события $A_1, A_2, \dots, A_{n-1}$ наступили.

Для нескольких событий, независимых в совокупности, вероятность произведения нескольких событий равна произведению вероятностей этих событий:

$$p(A_1 A_2 \dots A_n) = p(A_1) p(A_2) \dots p(A_n). \quad (5)$$

Комментарий к теореме 2. Порядок, в котором расположены события, может быть выбран любой, т.е. безразлично, какое событие считать первым, вторым и т.д.

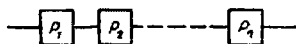
Доказательство производится методом математической индукции.

Теперь введем более широкое понятие независимости в совокупности, а именно:

Определение 5. События $A_1, A_2, \dots, A_n$ называются *независимыми в совокупности*, если выполняется следующее условие: каково бы ни было подмножество $\{i_1, i_2, \dots, i_k\}$ множеств $\{1, 2, \dots, n\}$ справедливо равенство

$$p(A_{i_1} A_{i_2} \dots A_{i_k}) = p(A_{i_1}) p(A_{i_2}) \dots p(A_{i_k}). \quad (6)$$

Пример 4. Электрическая схема состоит из n последовательно соединенных блоков.



Надежность (т.е. вероятность безотказной работы) каждого блока равна, соответственно $p_1, p_2, \dots, p_n$. Считая выходы из строя различных блоков независимыми событиями, найти надежность всей схеме в целом.

Решение

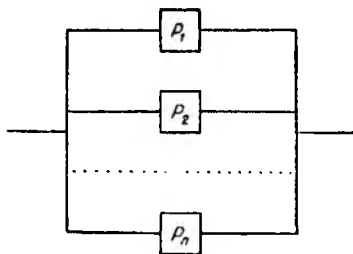
Событие, заключающееся в исправной работе i -го блока, обозначим A_i ; исправность схемы в целом обозначим A . Так как блоки собраны последовательно, то A имеет место в том и только в том случае, когда имеют место все A_i . Поэтому

$$A = A_1 A_2 \dots A_n,$$

откуда в силу независимости событий $A_1, A_2, \dots, A_n$ следует

$$p(A) = p(A_1) p(A_2) \dots p(A_n) = p_1 p_2 \dots p_n.$$

Та же самая задача для схемы из параллельно соединенных блоков приводит к другому результату.



В этом случае выход схемы из строя происходит лишь в том случае, когда выходят из строя все блоки. Это значит, что

$$\bar{A} = \bar{A}_1 \bar{A}_2 \dots \bar{A}_n$$

и, следовательно,

$$p(\bar{A}) = p(\bar{A}_1) p(\bar{A}_2) \dots p(\bar{A}_n) = (1 - p_1)(1 - p_2) \dots (1 - p_n).$$

Таким образом, надежность всей схемы оказывается равной

$$p(A) = 1 - (1 - p_1)(1 - p_2) \dots (1 - p_n).$$

Пример 5. Слово «лotos», составленное из букв-кубиков, рассыпано на отдельные буквы, которые затем сложены в коробке. Из коробки наугад извлекаются одна за другой три буквы. Какова вероятность того, что при этом появится слово «сто».

Решение. Введем обозначение для событий:

$$A_1 = \{\text{первой извлечена буква «с»}\};$$

$$A_2 = \{\text{второй извлечена буква «т»}\};$$

$$A_3 = \{\text{третьей извлечена буква «о»}\};$$

$$A = \{\text{получилось слово «сто»}\}.$$

Очевидно, $A = A_1 A_2 A_3$. Имеем последовательно:

$$p(A_1) = \frac{1}{5}; \quad p(A_1 A_2) = p(A_1) p_{A_1}(A_2) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20};$$

$$p(A_1 A_2 A_3) = p(A_1 A_2) p_{A_1 A_2}(A_3) = \frac{1}{20} \cdot \frac{2}{3} = \frac{1}{30}.$$

Итак, $p(A) = \frac{1}{30}.$

3. Полная группа событий

Определение. Полной группой называют систему $A_1, A_2, \dots, A_n$ попарно несовместных событий, если появления в результате опыта одного и только одного из них является достоверным событием.

Пример 6. Стрелок производит по мишени 2 выстрела. Следующие события A_1 – одно попадание; A_2 – два попадания; A_3 – промах образуют полную группу.

Для событий, образующих полную группу, имеет место следующая теорема.

Теорема 3. Сумма вероятностей событий, образующих полную группу, равна единице:

$$p(A_1) + p(A_2) + \dots + p(A_n) = 1.$$

Доказательство: Так как появление одного из событий полной группы достоверно, а вероятность достоверного события равна единице, то

$$p(A_1 + A_2 + \dots + A_n) = 1.$$

Так как любые два события полной группы несовместны, то по теореме сложения несовместных событий

$$p(A_1 + A_2 + \dots + A_n) = p(A_1) + p(A_2) + \dots + p(A_n) = 1.$$

С помощью понятия полной группы событий можно дать другое определение противоположных событий.

Противоположными событиями называют два события, образующие полную группу.

На основании приведенной выше теоремы, сумма вероятностей противоположных событий равна единице:

$$p(A) + p(\bar{A}) = 1.$$

Как правило, при рассмотрении противоположных событий вероятность одного из них обозначают p , а вероятность другого – q .

Таким образом, $p + q = 1$.

Пример 7. Вероятность того, что день будет ясный равна 0,7. Найти вероятность того, что день будет дождливый.

Решение

По условию $p = 0,7$. Тогда как $q = 1 - p$, т.е. $q = 1 - 0,7 = 0,3$.

Очень часто при решении задач на нахождение вероятности события A бывает удобно найти прежде вероятность противоположного события $\bar{A}$, а затем вероятность искомого события по формуле:

$$p(A) = 1 - p(\bar{A}).$$

Пример 8. Студент из 50 экзаменационных вопросов знает ответ на 30. Какова вероятность того, что из заданных ему наудачу 5 вопросов он знает ответ хотя бы на один?

Решение

Обозначим через A событие, заключающееся в том, что студент знает ответ хотя бы на один вопрос. Тогда событие $\bar{A}$ - студент не знает ответа ни на один вопрос - противоположное событие. Вероятность события $\bar{A}$ легко находится по классическому определению вероятности:

$$p(\bar{A}) = \frac{m}{n} = \frac{C_{20}^5}{C_{50}^5}.$$

Тогда искомая вероятность

$$p(A) = 1 - p(\bar{A}) = 1 - \frac{C_{20}^5}{C_{50}^5}.$$

Пример 9. В продукции завода брак составляет 5 % от общего количества выпускаемых деталей. Для контроля отобрано 20 деталей. Какова вероятность того, что среди них имеется хотя бы одна бракованная?

Решение

Для любой детали из продукции завода вероятность быть бракованной равна по условию $p = 0,05 = p(A_k)$, $k = 1, 2, \dots, 20$, где событие

$$A_k = \{k\text{-я по счету извлеченная деталь бракованная}\}.$$

Очевидно, нас интересует событие $A_1 + A_2 + \dots + A_{20}$. В условиях отлаженного технологического процесса можно считать, что события

$$A_1, A_2, \dots, A_{20}$$

в совокупности. Тогда очевидно, что вероятность осуществления **любого** из $A_1, A_2, \dots, A_{20}$ проще вычисляется не по формуле сложения, а с помощью формулы умножения:

$$\begin{aligned} p(A_1 + A_2 + \dots + A_{20}) &= 1 - p(\overline{A_1 + A_2 + \dots + A_{20}}) = \\ &= 1 - p(\overline{A_1}) p(\overline{A_2}) \dots p(\overline{A_{20}}) = 1 - 0,95^{20} \approx 0,64. \end{aligned}$$

Упражнения

1. Из 100 студентов, находящихся в аудитории, 50 человек знают английский язык, 40 - французский и 35 - немецкий. Английский и французский языки знают 20 студентов, английский и немецкий - 8, французский и немецкий - 10. Все три языка знают 5 человек. Один из студентов вышел из аудитории. Вычислить вероятности следующих событий:

$A = \{\text{вышедший знает или английский или французский язык}\};$

$B = \{\text{вышедший знает только английский язык}\};$

$C = \{\text{вышедший не знает ни одного языка}\}.$

2. Статистика, собранная среди студентов кредитно-экономического факультета Ташкентского финансового института, обнаружила следующие факты: 60 % всех студентов занимаются спортом, 30 % участвуют в художественной самодеятельности, 50 % работают в банке, 20 % занимаются спортом и участвуют в художественной самодеятельности, 10 % занимаются спортом и работают в банке, 5 % участвуют в самодеятельности и работают в банке, наконец, 5 % участвуют во всех трех видах деятельности. Корреспондент местной газеты подошел к наудачу выбранному студенту. Найти вероятности следующих событий:

$A = \{\text{студент занимается по крайней мере одним из двух видов деятельности: занимается спортом или участвует в художественной самодеятельности}\};$

$B = \{\text{студент занимается одним только спортом}\};$

$C = \{\text{студент занимается только одним видом деятельности}\};$

$D = \{\text{студент занимается двумя и только двумя видами деятельности}\}.$

3. (Задача де Мере). Сколько раз нужно бросить пару игральных костей, чтобы с вероятностью, не меньшей 0,5 хотя бы один раз появилась сумма очков, равная 12?

4. Формула полной вероятности. Формулы Байеса

Следствием основных теорем – теоремы сложения вероятностей и теоремы умножения вероятностей – является так называемая формула полной вероятности. А следствием теоремы умножения и формулы полной вероятности является так называемая теорема гипотез или формула Байеса. Этот пункт посвящается этим формулам.

4.1. Одним из эффективных методов подсчета вероятностей является формула полной вероятности, с помощью которой решается широкий круг задач.

Теорема 4 (теорема о полной вероятности). Пусть $B_1, B_2, \dots, B_n$ – попарно несовместные события, имеющие, соответственно вероятности $p(B_1), p(B_2), \dots, p(B_n)$. Пусть событие A может наступить только вместе с одним из событий $B_1, B_2, \dots, B_n$, и $p_{B_1}(A), p_{B_2}(A), \dots, p_{B_n}(A)$ – условные вероятности события A при условии, что $B_1, B_2, \dots, B_n$ наступили. Тогда вероятность $p(A)$ события A равна сумме произведений вероятностей событий B_n на условные вероятности $p_{B_n}(A)$:

$$p(A) = p(B_1) \cdot p_{B_1}(A) + p(B_2) \cdot p_{B_2}(A) + \dots + p(B_n) \cdot p_{B_n}(A). \quad (7)$$

Комментарий к теореме 4

1. Вероятности $p_A(B_i)$ называются послеопытными (апостериорными) вероятностями событий B_i , а вероятности $p(B_i)$ – доопытными (априорными) вероятностями событий B_i). Эти вероятности различаются, что будет видно на примерах.

2) События $B_1, B_2, \dots, B_n$ часто называются *гипотезами*.

Доказательство. По условию,

$$A(B_1 + B_2 + \dots + B_n) = A \text{ и } AB_1 + AB_2 + \dots + AB_n = A.$$

Применяя сначала теорему сложения, а затем теорему умножения вероятностей, получим

$$\begin{aligned} p(A) &= p(AB_1) + p(AB_2) + \dots + p(AB_n) = \\ &= p(B_1) \cdot p_{B_1}(A) + p(B_2) \cdot p_{B_2}(A) + \dots + p(B_n) \cdot p_{B_n}(A). \end{aligned}$$

Формула (1) называется *формулой полной вероятности*.

Пример 10. Производится серия из четырех выстрелов по некоторому объекту. Вероятности попадания в цель одного, двух, трех и четырех снарядов заданы таблицей

1	2	3	4
0,4	0,26	0,22	0,03

Вероятности разрушения объекта при условии попадания одного, двух, трех и четырех снарядов даны в таблице

1	2	3	4
0,5	0,7	0,8	0,99

Найти вероятность разрушения объекта.

Решение

Первая таблица задает вероятности $p(B_1), p(B_2), p(B_3), p(B_4)$, а вторая - вероятности $p_{B_1}(A), p_{B_2}(A), p_{B_3}(A), p_{B_4}(A)$ (событие B_i состоит в попадании в цель i снаряда, событие A состоит в разрушении мишени). По формуле (1) находим

$$p(A) = 0,4 \cdot 0,5 + 0,26 \cdot 0,7 + 0,22 \cdot 0,8 + 0,03 \cdot 0,99 = 0,5877.$$

Пример 11. Статистика запросов кредитов в банке такова: 10% - государственные органы, 20% - другие банки, остальные - физические лица. Вероятности того, что взятый кредит не будет возвращён, составляют 0,01, 0,05 и 0,2, соответственно. Определить, какая доля кредитов в среднем не возвращается.

Решение

Пусть событие A состоит в том, что взятый кредит не возвращается, гипотеза B_1 - в том, что запрос на этот кредит поступил от государственного органа, гипотеза B_2 - в том, что запрос на кредит поступил от другого банка, гипотеза B_3 - в том, что запрос на кредит поступил от физического лица. По условию вероятности гипотез составляют

$$p(B_1) = 0.1; \quad p(B_2) = 0.2; \quad p(B_3) = 1 - p(B_1) - p(B_2) = 0.7.$$

Апостериорные вероятности, в свою очередь, по условию равны

$$p(B_1) = 0,01; \quad p(B_2) = 0,005; \quad p(B_3) = 0,2.$$

По формуле полной вероятности

$$\begin{aligned} p(A) &= p(B_1) \cdot p_{B_1}(A) + p(B_2) \cdot p_{B_2}(A) + p(B_3) \cdot p_{B_3}(A) = \\ &= 0.01 \cdot 0.1 + 0.05 \cdot 0.02 + 0.2 \cdot 0.7 = 0.151. \end{aligned}$$

Упражнения

1. Партия транзисторов, среди которых 10 % дефектных, поступила на проверку. Схема проверки такова, что с вероятностью 0,95 обнаруживается дефект (если он есть) и существует ненулевая вероятность 0,03 того, что исправный транзистор будет признан дефектным. Какова вероятность того, что случайно выбранный из партии транзистор будет признан дефектным?

2. Прибор, установленный на борту самолета, может работать в двух режимах: в условиях нормального крейсерского полета и в условиях перегрузки при взлете и посадке. Крейсерский режим полета осуществляется в 80% всего времени полета, условия перегрузки - в 20 %. Вероятность выхода прибора из строя за время полета в нормальном режиме равна 0,1, в условиях перегрузки - 0,4. Вычислить надежность прибора за время полета.

3. На шахматную доску ставят наудачу двух слонов, белого и черного. Какова вероятность того, что слоны побьют друг друга?

4. В продажу поступают телевизоры трех заводов. Продукция первого завода содержит 20 % телевизоров со скрытым дефектом, второго - 10 % и третьего - 5 %. Какова вероятность приобрести исправный телевизор, если в

~~50 %~~ поступило 30 % телевизоров с первого завода, 20 % - со второго и 50 % - с третьего?

5. В ящике лежат 20 теннисных мячей, в том числе 15 новых и 5 ~~игр~~. Для игры наудачу выбираются два мяча и после игры возвращаются ~~обратно~~. Затем для второй игры также наудачу извлекаются еще два мяча. Какова вероятность того, что вторая игра будет проводиться новыми мяча-
~~ми~~?

6. Три стрелка, вероятности попадания которых при одном выстреле в мишень в неизменных условиях постоянны и, соответственно, равны $p_1 = 0,8$, $p_2 = 0,7$, $p_3 = 0,6$, делают по одному выстрелу в одну и ту же мишень. Вычислить вероятность события $A = \{\text{в мишени окажется ровно две пробоины}\}$, приняв в качестве гипотез элементарные исходы данного опыта.

7. Из десяти студентов, пришедших сдавать экзамен по теории вероятностей и взявших билеты, Иванов и Петров знают 20 билетов из 30, Сидоров плохо занимался весь семестр и успел повторить только 15 билетов, остальные студенты знают все 30 билетов. По прошествии отведенного времени на подготовку экзаменатор наудачу вызывает отвечать одного из студентов. Какова вероятность того, что вызванный сдал экзамен, если знание билета гарантирует сдачу экзамена с вероятностью 0,85, а при незнании билета можно сдать экзамен лишь с вероятностью 0,1?

8. В первой урне находится 6 белых и 4 черных шара, во второй - 3 белых и 2 черных. Из первой урны наудачу извлекают сразу 3 шара, и шары того цвета, которые окажутся в большинстве, опускают во вторую урну и тщательно перемешивают. После этого из второй урны наудачу извлекают один шар. Какова вероятность того, что этот шар белый?

4.2. Теперь приступаем к обсуждению формул Байеса.

Теорема 5 (теорема Байеса). Пусть события $B_1, B_2, \dots, B_n$ попарно несовместны и пусть событие A может наступить только вместе с одним из событий $B_1, B_2, \dots, B_n$. Известны вероятности $p(B_1), p(B_2), \dots, p(B_n)$ событий

$B_1, B_2, \dots, B_n$ и условные вероятности $p_{B_1}(A), p_{B_2}(A), \dots, p_{B_n}(A)$ события A при условиях $B_1, B_2, \dots, B_n$. Также известно, что событие A наступило. Тогда вероятности событий $B_1, B_2, \dots, B_n$ при условии, что событие A наступило, находятся по формулам

$$p_A(B_i) = \frac{p(B_i) \cdot p_{B_i}(A)}{p(B_1) \cdot p_{B_1}(A) + p(B_2) \cdot p_{B_2}(A) + \dots + p(B_n) \cdot p_{B_n}(A)}, \quad i = 1, 2, \dots, n. \quad (8)$$

Комментарий к теореме 5

1. Знаменатель в правой части формулы (8) совпадает с правой частью формулы (7) и равен $p(A)$.

2. Формула (8) дает вероятности гипотезы B_i , при которой наступило событие A .

Доказательство. Согласно теореме умножения вероятностей, имеем

$$p(AB_i) = p_A(B_i) \cdot p(A) = p(B_i) \cdot p_{B_i}(A).$$

Отсюда

$$p_A(B_i) = \frac{p(B_i) \cdot p_{B_i}(A)}{p(A)}. \quad (9)$$

Подставляя в знаменатель правой части равенства (9) вместо $p(A)$ правую часть формулы (7), получаем соотношение (8).

Формулы (8) называются **формулами Байеса (или формулами гипотез)**.

Пример 12. Поломка прибора (событие A) может быть вызвана одной из трех причин B_1, B_2, B_3 , вероятности которых $p(B_1) = 0,7$, $p(B_2) = 0,2$, $p(B_3) = 0,1$. При наличии этих причин поломка прибора происходит с вероятностями $p_{B_1}(A) = 0,1$, $p_{B_2}(A) = 0,2$, $p_{B_3}(A) = 0,2$. Известно, что прибор вышел из строя. Найти вероятности $p_A(B_1)$, $p_A(B_2)$, $p_A(B_3)$.

Решение. Используя формулы (8), получим

$$P_1(B_1) = \frac{0,7 \cdot 0,1}{0,7 \cdot 0,1 + 0,2 \cdot 0,2 + 0,1 \cdot 0,2} = \frac{0,07}{0,13} = \frac{7}{13};$$

$$P_1(B_2) = \frac{0,2 \cdot 0,2}{0,7 \cdot 0,1 + 0,2 \cdot 0,2 + 0,1 \cdot 0,2} = \frac{0,04}{0,13} = \frac{4}{13};$$

$$P_1(B_3) = \frac{0,1 \cdot 0,2}{0,7 \cdot 0,1 + 0,2 \cdot 0,2 + 0,1 \cdot 0,2} = \frac{0,02}{0,13} = \frac{2}{13}.$$

Из результатов вычислений видно, что апостериорные вероятности отличаются от априорных.

Пример 13. В условиях примера 11 начальнику кредитного отдела известно, что получено факсимильное сообщение о неисполнении обязательств по возврату кредита, в котором очень плохо пропечаталось имя клиента. Найти вероятность того, что кредит не возвращает какой-либо клиент.

Решение

Пусть событие A состоит в том, что взятый кредит не возвращается, гипотеза B_1 - в том, что запрос на этот кредит поступил от государственного органа, гипотеза B_2 - в том, что запрос на кредит поступил от другого банка, гипотеза B_3 - в том, что запрос на кредит поступил от физического лица. По условию вероятности гипотез составляют

$$p(B_1) = 0,1; \quad p(B_2) = 0,2; \quad p(B_3) = 1 - p(B_1) - p(B_2) = 0,7.$$

Апостериорные вероятности, в свою очередь, по условию равны

$$p(B_1) = 0,01; \quad p(B_2) = 0,05; \quad p(B_3) = 0,2.$$

По формуле Байеса

$$P_A(B_2) = \frac{p(B_2) \cdot p_{B_2}(A)}{p(A)} = \frac{0,2}{0,151} \approx 0,66,$$

где вероятность $p(A) = 0,151$ рассчитана по формуле полной вероятности в примере 10.

Пример 14. Страховая компания занимается страхованием жизни. 10% застрахованных в этой компании являются курильщиками. Если застрахо-

ванный не курит, вероятность его смерти на протяжении года равна 0,01. Если же он курильщик, то эта вероятность равна 0,05.

Какова доля курильщиков среди тех застрахованных, которые умерли в течение года?

Решение

Введем события:

$B_1 = \{\text{застрахованный - курильщик}\};$

$B_2 = \{\text{застрахованный - не курильщик}\};$

$A = \{\text{застрахованный умер в течение года}\}.$

Условие задачи означает, что

$$p(B_1) = 0,1; \quad p_{B_1}(A) = 0,05; \quad p_{B_2}(A) = 0,01.$$

Кроме того, поскольку события B_1 и B_2 образуют полную группу попарно несовместимых событий, $p(B_2) = 1 - p(B_1) = 1 - 0,1 = 0,9$.

Интересующая нас вероятность – это $p_A(B_1)$. Используя формулу Байеса, мы имеем:

$$p_A(B_1) = \frac{p(B_1) \cdot p_{B_1}(A)}{p(B_1) \cdot p_{B_1}(A) + p(B_2) \cdot p_{B_2}(A)} = \frac{0,1 \cdot 0,05}{0,1 \cdot 0,05 + 0,9 \cdot 0,01} = \frac{5}{14} \approx 0,36.$$

Упражнения

1. Страховая компания продает договора страхования жизни трех категорий: стандартные, привилегированные и ультрапривилегированные. 50% всех застрахованных являются стандартными, 40% - привилегированными и 10% - ультрапривилегированными. Вероятность смерти в течение года для стандартного застрахованного равна 0,010, для привилегированного - 0,005, а для ультрапривилегированного ² - 0,001.

Чему равна вероятность того, что умерший застрахованный является ультрапривилегированным ?

² Привилегированность договора/застрахованного означает меньший риск для страховой компании (по результатам андеррайтинга)

2. Партия транзисторов, среди которых 10 % дефектных, поступила на ~~проверку~~. Схема проверки такова, что с вероятностью 0,95 обнаруживается ~~дефект~~ (если он есть) и существует ненулевая вероятность 0,03 того, что исправный транзистор будет признан дефектным. Случайно выбранный из ~~партии~~ транзистор был признан дефектным. Какова вероятность того, что на ~~самом~~ деле транзистор исправен?

3. Изучается три вида дефектов запоминающих устройств, выполнен-
~~ных на~~ интегральных схемах: дефекты схем обрамления (гипотеза B_1); де-
фекты, вызванные паразитными связями между ячейками (гипотеза B_2); и
дефекты адресных шин (гипотеза B_3). Известно, что

$$p(B_1) = 0,1; \quad p(B_2) = 0,6; \quad p(B_3) = 0,3.$$

Диагностика запоминающих устройств производится с помощью набора тес-
тов $T_1, T_2, \dots, T_n$, каждый из которых проверяет определенное состояние ячеек
памяти. Наблюдаемый результат - состояние выбранной ячейки по отно-
шению к каждому тесту. Пусть диагностика произведена и наблюдался не-
который результат (произошло событие A). Известно до опыта, что

$$p_{B_1}(A) = 0,4, \quad p_{B_2}(A) = 0,2, \quad p_{B_3}(A) = 0,3.$$

Установить, какая из гипотез имеет наибольшую апостериорную вероят-
ность (т.е. какой из дефектов наиболее вероятен).

4. В специализированную больницу поступают в среднем 50% боль-
ных с заболеванием K , 20% с заболеванием M , 30% с заболеванием L . Ве-
роятность полного излечения болезни K равна 0,7, для болезней L и M эти
вероятности равны, соответственно, 0,8 и 0,9. Больной, поступивший в боль-
ницу, был выписан здоровым. Найти вероятность того, что этот больной
страдал заболеванием K .

5. Продукция, изготавливаемая в цехе, проверяется двумя контроле-
рами. Вероятность того, что деталь попадет на проверку к первому контро-
леру равна 0,6; ко второму - 0,4. Вероятность принять качественную деталь
за бракованную равна для первого контролера 0,06, для второго контролера

эта вероятность 0,02. Среди забракованных деталей оказалась качественная. Найти вероятность того, что она проверялась первым контролером.

6. В первой урне 1 белый и 2 черных шара, во второй – 100 белых и 100 черных шаров. Из второй урны переложили в первую один шар, а затем из первой урны вынули наугад один шар. Какова вероятность того, что вынутый шар ранее находился во второй урне, если известно, что он белый.

7. Среди поступающих на сборку деталей с первого станка 0,1% бракованных, со второго 0,2%, с третьего 0,25%, с четвертого 0,5%. Производительности их относятся, соответственно, как 4:3:2:1. Взятая наудачу деталь оказалась стандартной. Найти вероятность того, что она изготовлена на третьем станке.

8. В коробке находятся две неотличимые по внешнему виду и по весу игральные кости: одна правильная, с одинаковыми вероятностями выпадения всех шести цифр при случайном подбрасывании; другая неправильная, с неравномерным распределением массы по объему. При случайном подбрасывании неправильной игровой кости шестерка появилась с вероятностью $1/3$, единица с вероятностью $1/9$, остальные цифры выпадают с одинаковой вероятностью. Наудачу извлеченная из коробки игральная кость была подброшена, и в результате выпало 6 очков. Найти вероятность того, что была подброшена правильная игральная кость.

9. В группе из 25 человек, пришедших сдавать экзамен по теории вероятности, имеется 10 отличников, 7 подготовленных хорошо, 5 удовлетворительно и 3 человека плохо подготовлены. Отличники знают все 25 вопросов программы, хорошо подготовленные – 20, подготовленных удовлетворительно – 15, и плохо подготовленные знают лишь 10 вопросов. Вызванный наудачу студент ответил на два заданных вопроса. Найти апостериорные вероятности гипотез:

$$H_1 = \{\text{студент подготовлен отлично или хорошо}\};$$

$$H_2 = \{\text{студенты подготовлены удовлетворительно}\};$$

$\Xi = \{\text{студент подготовлен плохо}\}.$

Вопросы для самопроверки

1. Сформулируйте аксиомы 1-3, которые составляют основу всей теории вероятностей.
2. Все теоремы теории вероятностей, включая самые сложные, выводятся из аксиомы 1-3 формально-логическим путем. Укажите несколько примеров такого вывода.
3. Дайте определения независимости (зависимости) одного события от другого.
4. Что называется условной вероятностью и как она обозначается?
5. Приведите теорему умножения вероятностей.
6. Можно ли утверждать: - «Если случайное событие B не зависит от A , то и A не зависит от B ».
7. Как поступают в практических вопросах для установления независимости одного события от другого?
8. Когда несколько событий называют попарно независимыми? Приведите примеры.
9. Когда несколько событий называют независимыми в совокупности? Приведите примеры.
10. Если несколько событий независимы попарно, то следует ли отсюда их независимость в совокупности?
11. Привести пример, показывающий, что из попарной независимости событий A, B, C не следует их независимость в совокупности.
12. Приведите обобщенную теорему умножения вероятностей.
13. Дайте определение полной группы событий. Чему равна сумма вероятностей событий, образующих полную группу?
14. Приведите другое определение противоположных событий с помощью понятия полной группы событий.
15. Сформулируйте теорему о полной вероятности. Прокомментируйте эту теорему.
16. Сформулируйте теорему Байеса. Прокомментируйте эту теорему.
17. Почему теоремы полной вероятности и Байеса (или гипотез) считают следствием теорем сложения и умножения вероятностей?

§3. Последовательность независимых испытаний. Предельные теоремы в схеме Бернулли

Ключевые слова: последовательность независимых испытаний, схема Бернулли, формула Бернулли, наивероятнейшее число, локальная теорема Муавра-Лапласа, интегральная теорема Муавра-Лапласа, правило «трёх сигм», практически достоверные события, теорема Пуассона.

При практическом применении теории вероятностей часто приходится встречаться с задачами, в которых одно и то же испытание или аналогичные испытания повторяются неоднократно. В результате каждого испытания может появиться или не появиться некоторое событие A , причем нас интересует не результат каждого отдельного испытания, а общее число появлений события A в результате испытаний. Например, если производится серия выстрелов по одной и той же цели, нас, как правило, интересует не результат каждого выстрела, а общее число попаданий. Такие задачи рассматриваются на этой и следующей лекциях. Оказывается, при определенных условиях, они решаются весьма просто.

1. Последовательность независимых испытаний (схема Бернулли)

Пусть производится серия из n испытаний, в каждом из которых событие A может наступить, а может и не наступить. Пусть при этом выполнено следующее условие: вероятность p наступления события A в каждом испытании постоянна, т.е. не зависит ни от номера испытания, ни от результатов предыдущих испытаний.

Это условие означает, что последовательность испытаний независима (вероятность p не зависит от результатов предыдущих испытаний).

Определение. Последовательность испытаний, удовлетворяющих указанному условию, называется *последовательностью независимых испытаний (или схемой Бернулли)*. Схема Бернулли полностью определяется двумя числами - натуральным числом n , означающим количество испытаний, и

~~Случай~~ p ($0 < p < 1$), означающим вероятность наступления события A в ~~испытании~~ испытании (безразлично, в каком по счету).

Примеры. Следующие серии опытов представляют собой конкретные ~~испытания~~ схемы Бернулли:

1. Монету подбрасывают n , раз; вероятность появления герба в одном ~~испытании~~ испытании $p = 1/2$.

2. Производят n выстрелов по мишени. Предполагается, что вероятность попадания в мишень при каждом выстреле постоянна и равна p .

Отметим, однако, что если в процессе стрельбы стрелок пристрелялся и стал лучше поражать мишень, то такая последовательность испытаний не является схемой Бернулли.

3. Из кучи зерна отбирают n зерен для проверки их на всхожесть. Вероятность того, что каждое зерно при проверке дает положительный результат, постоянна (так будет, например, в том случае, когда куча зерна большая, а зерна отбирают наугад после перемешивания).

В связи со схемой Бернулли рассматривают такие задачи:

1. Найти вероятность $P_n(k)$ того, что в серии из n испытаний событие A наступит ровно k раз. Решение этой задачи дает формула Бернулли.

2. Найти вероятность $P_n(k_1, k_2)$ того, что в серии из n испытаний количество k наступлений события A будет находиться в пределах $k_1 \leq k \leq k_2$.

3. Решить задачу 1 для больших чисел n и k (формула Бернулли, дающая решение задачи 1, неудобна для вычислений при больших n и k). Задача 3 решается с помощью локальной теоремы Муавра-Лапласа.

4. Решить задачу 2 для больших чисел n, k_1, k_2 (формула Бернулли мало пригодна для вычислений $P_n(k_1, k_2)$ при больших n, k_1, k_2). Задача решается с помощью интегральной теоремы Муавра-Лапласа.

2. Формула Бернулли

Теорема 1. Вероятность $P_n(k)$ того, что в последовательности из n испытаний в схеме Бернулли событие A наступит ровно k раз, выражается формулой

$$P_n(k) = C_n^k p^k q^{n-k},$$

где $C_n^k = \frac{n!}{(n-k)!k!}$ число сочетаний из n элементов по k ; p - вероятность наступления события A в одном испытании; $q = 1 - p$ - вероятность ненаступления события A в одном испытании.

Доказательство. Рассмотрим последовательность из k плюсов и $n-k$ минусов, расположенных в произвольном, но фиксированном порядке. Каждая такая последовательность задает событие при « n -кратном испытании по схеме Бернулли: знак « $+$ » или « $-$ » на k -м месте последовательности означает, соответственно, наступление или ненаступление события A при k -м испытании. Вероятность такого события (расположение k плюсов и $n-k$ минусов в произвольном, но фиксированном порядке) в силу теоремы умножения вероятностей равна $p^k q^{n-k}$ и не зависит от порядка плюсов и минусов в рассматриваемой последовательности. При этом последовательности с различным расположением k плюсов и $n-k$ минусов определяют различные попарно несовместные события. Количество последовательностей из k плюсов и $n-k$ минусов равно числу сочетаний из n элементов по k . Действительно, последовательность будет полностью определена, если из множества номеров $\{1, 2, 3, \dots, n\}$ выбрано k штук и плюсы последовательности поставлены на места с номерами из выбранного множества.

Отсюда по теореме сложения вероятностей получаем

$$P_n(k) = C_n^k p^k q^{n-k}.$$

где, как известно, число сочетаний из n элементов по k выражается формулой

$$C_n^k = \frac{n!}{(n-k)!k!}.$$

Примеры

1. Найти вероятность того, что при 10-кратном бросании монеты выпадет ровно 3 герба.

Решение

Здесь $n = 10$, $k = 3$, $p = \frac{1}{2}$. Согласно формуле Бернулли, получим

$$P_{10}(3) = C_{10}^3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} \left(\frac{1}{2}\right)^{10} = \frac{120}{1024}.$$

2. Пусть вероятность поражения мишени при одном выстреле равна $1/3$. Найти вероятность того, что из 6 выстрелов три поразят мишень.

Решение

Используя формулу Бернулли при $n = 6$, $k = 3$, $p = \frac{1}{3}$, $q = \frac{2}{3}$, находим

$$P_6(3) = C_6^3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} \cdot \frac{8}{729} = \frac{160}{729}.$$

3. Пусть вероятность того, что взятое наудачу из кучи зерно окажется всхожим, равна 0.9. Какова вероятность того, что из 7 отобранных зерен ровно 5 окажутся всхожими?

Решение

Имеем

$$P_7(5) = C_7^5 0,9^5 \cdot 0,1^2 = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 21 \cdot 0,0059049 = 0,124.$$

4. В схеме Бернулли, связанной с бросанием монеты, вычислить вероятности $P_{10}(k)$, где $k = 0, 1, 2, \dots, 10$ (т.е. вероятности того, что в 10 испытаниях герб выпадет ровно k раз).

Решение. Используя формулу Бернулли при $p=q=\frac{1}{2}$, $k=0,1,2,\dots,10$,

получим

$$P_{10}(0) = \frac{1}{1024}, \quad P_{10}(1) = \frac{10}{1024}, \quad P_{10}(2) = \frac{45}{1024}, \quad P_{10}(3) = \frac{120}{1024}, \quad P_{10}(4) = \frac{210}{1024},$$

$$P_{10}(5) = \frac{252}{1024}, \quad P_{10}(6) = \frac{210}{1024}, \quad P_{10}(7) = \frac{120}{1024}, \quad P_{10}(8) = \frac{45}{1024}, \quad P_{10}(9) = \frac{10}{1024},$$

$$P_{10}(10) = \frac{1}{1024}.$$

Результаты вычислений иллюстрирует рис.1. Как видно из рисунка, наибольшей из вероятностей $P_{10}(k)$ является $P_{10}(5) \approx 0,25$. Сравнительно велики и значения $P_{10}(4)$ и $P_{10}(6)$ ($\approx 0,21$); в то же время «крайние» значения k дают $P_{10}(0) = P_{10}(10) \approx 0,001$.

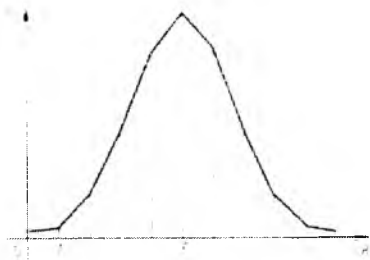


Рис. 1

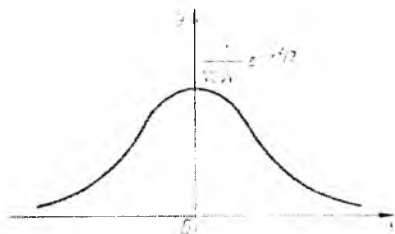


Рис. 2

Обратим внимание на характерный вид изображенной на рисунке ломаной, имеющей пик в точке $k=5$. В дальнейшем нам часто придется иметь

дело с кривой $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (рис. 2). Она называется гауссовой кривой (или

кривой нормального распределения) и играет исключительно важную роль в теории вероятностей.

Тот факт, что ломаная на рис. 1 и кривая на рис. 2 имеют значительное сходство, не случаен. Причины этого явления раскрываются локальной теоремой Муавра-Лапласа.

Для вычисления вероятностей $P_n(k_1, k_2)$ того, что в схеме Бернулли из n испытаний количество m наступлений события A будет находиться в пределах $k_1 \leq m < k_2$, можно использовать формулу

$$P_n(k_1, k_2) = P_n(k_1) + P_n(k_1 + 1) + \dots + P_n(k_2 - 1). \quad (1)$$

Событие, о котором идет речь, является суммой попарно несовместных событий B_i ($i = k_1, k_1 + 1, \dots, k_2 - 1$), состоящих в том, что в n испытаниях событие A наступит ровно i раз; затем, используя теорему сложения вероятностей, получаем формулу (1).

В частности, вероятность того, что в n испытаниях событие наступит: а) менее k раз; б) более k раз; в) не менее k раз; г) не более k раз, находят, соответственно, по формулам:

$$\begin{aligned} &P_n(0) + P_n(1) + \dots + P_n(k-1); \\ &P_n(k+1) + P_n(k+2) + \dots + P_n(n); \\ &P_n(k) + P_n(k+1) + \dots + P_n(n); \\ &P_n(0) + P_n(1) + \dots + P_n(k). \end{aligned}$$

3. Наивероятнейшее число появлений события в независимых испытаниях

Число k_0 (наступления события в независимых испытаниях, в каждом из которых вероятность появления события равна p) называют наивероятнейшим, если вероятность того, что событие наступит в этих испытаниях k_0 раз, превышает (или, но в крайней мере, не меньше) вероятности остальных возможных исходов испытаний.

Наивероятнейшее число k_0 определяют из двойного неравенства

$$np - q \leq k_0 \leq np + p,$$

причем:

а) если число $np - q$ - дробное, то существует одно наивероятнейшее число k_0 ;

б) если число $np - q$ – целое, то существует два наимвероятнейших числа, а именно: k_0 и $k_0 + 1$;

в) если число np – целое, то наимвероятнейшее число $k_0 = np$.

Пример. Испытывается каждый из 15 элементов некоторого устройства. Вероятность того, что элемент выдержит испытание, равна 0.9. Найти наимвероятнейшее число элементов, которые выдержат испытание.

Решение. По условию, $n = 15$, $p = 0.9$, $q = 0.1$. Найдем наимвероятнейшее число k_0 из двойного неравенства

$$np - q \leq k_0 \leq np + p.$$

Подставив данные задачи, получим

$$15 \cdot 0.9 - 0.1 \leq k_0 < 15 \cdot 0.9 + 0.9, \text{ или } 13.5 < k_0 < 14.4.$$

Так как k_0 – целое число и поскольку между числами 13,4 и 14,4 заключено одно целое число, а именно 14, то искомое наимвероятнейшее число $k_0 = 14$.

Пример. Найти наимвероятнейшее число появления герба в 10 испытаниях (см. пример 4 из п.2).

Решение

$np - q = 10 \cdot 0.5 - 0.5 = 4.5$ – дробное число; существует одно наимвероятнейшее число k_0 . Имеем $4.5 \leq k_0 \leq 5.5$. Следовательно, $k_0 = 5$. Нетрудно заметить, что расчеты, проведенные в п.2, это подтверждают, т.е., наибольшей из вероятностей $P_{10}(k)$ является $P_{10}(5) \approx 0.25$.

Пример. Известно, что из числа зрителей определённой телепрограммы 70% смотрят и рекламные блоки. Группы, состоящие из трёх наугад выбранных телезрителей, опрашивают относительно содержания рекламного блока. Рассчитать вероятности числа лиц в группе, которые смотрят рекламные блоки.

Решение

Вероятность того, что наугад выбранный зритель данной телепрограммы смотрит и рекламные блоки, согласно статистическому определению ве-

вероятности, равна $p = 0,7$. Интерпретируя опрос трёх телезрителей как три испытания Бернулли и считая наступил событие A ситуацией, когда телезритель смотрит рекламные блоки, найдём искомые вероятности по формуле Бернулли

$$P_3(k) = C_3^k 0,7^k \cdot 0,3^{3-k} \quad (k = 0, 1, 2, 3),$$

в которой $n = 3$, $p = 3$. Отсюда имеем:

$$P_3(0) = C_3^0 0,7^0 \cdot 0,3^3 = 0,027; \quad P_3(1) = C_3^1 0,7^1 \cdot 0,3^2 = 0,189;$$

$$P_3(2) = C_3^2 0,7^2 \cdot 0,3^1 = 0,441; \quad P_3(3) = C_3^3 0,7^3 \cdot 0,3^0 = 0,343.$$

Пример. В условиях предыдущего примера найти наивероятнейшее число лиц в группе, которые смотрят рекламные блоки.

Решение

Наивероятнейшее число k_0 лиц в группе, которые смотрят рекламные блоки, подчиняется неравенствам

$$np - q \leq k_0 \leq np + p,$$

в которых $n = 3$, $p = 3$, т.е.

$$3 \cdot 0,7 - 0,3 \leq k_0 \leq 3 \cdot 0,7 + 0,7 \text{ или } 1,8 \leq k_0 \leq 2,8,$$

откуда $k_0 = 2$. Это подтверждается и решением предыдущего примера.

Упражнения

1. Стоимость проезда в автобусе равна 3, месячный проездной билет на автобус стоит 120, а штраф за безбилетный проезд составляет 10. Петя 24 раза в месяц ездит на автобусе в институт и обратно. Он не покупает проездного билета, никогда не платит за проезд и считает, что вероятность быть пойманным и заплатить штраф равна 0,05. Сравнить стоимость проездного билета с наиболее вероятной величиной штрафа за 48 поездок.

2. В брокерской конторе для стимулирования прибыльности торговли применяется следующая система премирования сотрудников. Если сотрудник не достигал установленного дневного уровня прибыли на протяжении более трёх дней за две недели (10 рабочих дней), он теряет свою премию. Ве-

роятность того, что сотрудник выполнит требуемую норму прибыли, составляет 0,85. Найти число премий, потерянных 100 сотрудниками этой брокерской конторы за год (50 рабочих недель).

3. Среди 12 проверяемых ревизором договоров семь оформлены неправильно. Найти вероятность того, что среди пяти договоров, произвольно отобранных ревизором для проверки, окажутся неправильно оформленными: а) ровно три договора; б) не менее трёх договоров.

4. Что вероятнее: выиграть в бильярд у равносильного противника три партии из четырёх или пять партий из восьми?

5. Что вероятнее: выиграть в бильярд у равносильного противника не менее трёх партий из четырёх или не менее пяти партий из восьми?

6. В течение месяца данная акция может подорожать на 1% с вероятностью 0,7 и подешеветь на 1% с вероятностью 0,3. Предполагая ежемесячные изменения цены независимыми, рассчитать вероятности того, что за три месяца цена акции возрастет: а) в $(1,01)^3$ раза; б) в $0,99 \cdot (1,01)^2$ раза.

7. Из 1000 опрошенных 700 человек поддерживают некоторую правительственную программу. Найти минимальную численность группы, в которой с вероятностью, не меньшей 0,9, хотя бы один респондент не поддерживает эту программу.

8. Среди билетов лотереи половина выигрышных. Найти минимальное число билетов, чтобы с вероятностью, не меньшей 0,99, быть уверенным в выигрыше хотя бы по одному билету.

9. В городе работают 1000 коммерческих банков, из которых 330 допускают нарушения налогового законодательства. Определить число банков, которые должна отобрать для проверки налоговая инспекция, чтобы с вероятностью, не меньшей 0,99, среди них оказался хотя бы один нарушитель законодательства.

10. В условиях задачи 9 налоговая инспекция проводит проверку 12 банков, выбирая их случайным образом. Выбранные банки проверяются независимо друг от друга. Допущенные в проверяемом банке нарушения могут

быть выявлены инспекцией с вероятностью 0,8. Найти вероятность того, что в ходе этой проверки будет выявлен хотя бы один нарушитель налогового законодательства.

11. Банк имеет пять отделений. Ежедневно с вероятностью 0,3 каждое отделение, независимо от других, может заказать на следующий день крупную сумму денег. В конце рабочего дня один из вице-президентов банка знакомится с поступившими заявками. Найти вероятности следующих событий: а) поступили ровно две заявки; б) поступила хотя бы одна заявка; в) среди поступивших двух заявок есть заявка от первого отделения.

12. Игральную кость бросают пять раз. Найти вероятность того, что дважды появится число, кратное трём.

Пример. (Задача о разделе ставки¹). Петя и Маша часто играют в бильярд друг с другом, причём Петя выигрывает в два раза чаще, чем Маша.

Исходя из этого, они оценили свои вероятности победить как $\frac{2}{3}$ для Пети и

$\frac{1}{3}$ для Маши и начали турнир на следующих условиях: каждый выигрыш приносит одно очко. Петя для победы должен набрать двенадцать очков, а Маша - шесть. После того, как Петя набрал восемь очков, а Маша - четыре, игру пришлось прекратить, и победу решили присудить тому, у кого вероятность окончательного выигрыша больше. Определить, кому присудили победу в этом турнире.

Решение

Очевидно, максимальное количество партий, которое осталось сыграть Пете и Маше, равно пяти (либо Петя выиграет три раза, а Маша - два раза, либо Маша выиграет один раз, а Петя - четыре раза). Поэтому событие, заключающееся в выигрыше Пети (а значит, проигрыше Маши), состоит в том,

¹ Такая задача возникает при определении доли инвестора, который хочет «выйти» из незавершенного проекта.

что Маша из пяти партий не выиграет ни одной или выиграет всего одну. Поэтому вероятность выигрыша Пети равна вероятности того, что в пяти испытаниях Бернулли, в каждом из которых успех интерпретируется как выигрыш Машей очередной партии (т.е. вероятность успеха в каждом испытании составляет $p = \frac{1}{3}$), наступит 0 или 1 успех:

$$\begin{aligned} P\{\text{выигрыш Пети}\} &= P\{0 \text{ или } 1 \text{ выигрыш Маши из } 5 \text{ партий}\} = \\ &= P_5(0) + P_5(1) = C_5^0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^5 + C_5^1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^4 = \frac{112}{243}. \end{aligned}$$

При этом

$$\begin{aligned} P\{\text{выигрыш Маши}\} &= P\{\overline{\{\text{выигрыш Пети}\}}\} = \\ &= 1 - P\{\text{выигрыш Пети}\} = 1 - \frac{112}{243} = \frac{131}{243}. \end{aligned}$$

Поэтому в данном случае победу должны были присудить Маше.

Пример (Задача Банаха). Известный математик Стефан Банах всегда носил с собой две коробки спичек, в каждой из которых первоначально было n спичек. Каждый раз, когда он хотел зажечь спичку, Банах доставал наугад одну из коробок. Найти вероятность того, что когда он в первый раз вынимал пустую коробку, в другой коробке оказывалось ровно r спичек, где $r = \overline{0, n}$.

Решение

Спички брались всего $(2n - r)$ раз, причём n раз из коробки, оказавшейся пустой. Это соответствует n успехам в $(2n - r)$ независимых испытаниях, поэтому вероятность

$$p(A) = \frac{C_{2n-r}^n}{2^{2n-r}}.$$

Упражнение

Доказать формулу $np - q \leq k_0 \leq np + p$ для наивероятнейшего числа наступления события в независимых испытаниях.

4. Полиномиальная схема. Более сложная схема с n независимых испытаний получается, когда при каждом испытании возможно появление одного из r попарно несовместных $A_1, A_2, \dots, A_r$ исходов ($A_1 + A_2 + \dots + A_r = \Omega$). Если $p_1, p_2, \dots, p_r$ – вероятности этих исходов, тогда $p_1 + p_2 + \dots + p_r = 1$. Полагая $P_n(m_1, m_2, \dots, m_r)$ – вероятность того, что A_1 произошло ровно m_1 раз, A_2 произошло ровно m_2 раз, и т.д., A_r произошло ровно m_r раз (нетрудно убедиться, что $m_1 + m_2 + \dots + m_r = n$) имеем

$$P_n(m_1, m_2, \dots, m_r) = \frac{n!}{m_1! m_2! \dots m_r!} p_1^{m_1} \cdot p_2^{m_2} \cdot \dots \cdot p_r^{m_r}. \quad (2)$$

Формула (3) носит название *полиномиальной*; описанная схема независимых испытаний с r исходами также называется полиномиальной. При $r = 2$ эта схема превращается в биномиальную схему Бернулли.

Пример. В урне находятся 8 белых, 5 красных и 2 черных шара. Проводится 5 извлечений с возвращением по одному шару. Рассматриваются события:

$A = \{\text{появился следующий состав шаров: 3 белых и по одному остальных цветов}\};$

$B = \{\text{появилось ровно 3 белых шара}\};$

$C = \{\text{появилось 3 белых шара и по одному остальных цветов; причем белые шары появились подряд}\}.$

Определить их вероятности.

Решение

Событие A соответствует полиномиальной схеме при $n = 5, r = 3$,

$p_1 = 8/15, p_2 = 5/15, p_3 = 2/15$, поэтому

$$P(A) = P_5(3, 1, 1) = \frac{5!}{3!1!1!} \left(\frac{8}{15}\right)^3 \left(\frac{5}{15}\right) \left(\frac{2}{15}\right) = \frac{8}{9} \cdot \left(\frac{8}{15}\right)^3 \approx 0,1348.$$

Событие B соответствует полиномиальной схеме при $n = 5, r = 2$,

$p_1 = 8/15, q_1 = 7/15$, поэтому

$$p(B) = P_5(3) = C_5^3 \left(\frac{8}{15}\right)^3 \left(\frac{7}{15}\right)^2 = 10 \cdot \left(\frac{8}{15}\right)^3 \cdot \left(\frac{7}{15}\right)^2 \approx 0,3304.$$

Событие C соответствует комбинированной схеме, в которой в каких-либо трех последовательных испытаниях белый шар выпал трижды, а в остальных двух испытаниях по одному разу выпали черный и красный шары, поэтому

$$p(C) = 3 \cdot P_3(3) \cdot P_2(0,1,1) = 3 \cdot \left(\frac{8}{15}\right)^3 \cdot \frac{2!}{0!1!1!} \cdot \left(\frac{8}{15}\right)^0 \cdot \frac{5}{15} \cdot \frac{2}{15} \approx 0,0404.$$

Упражнения

1. Каждый из десяти аспирантов группы случайным образом и независимо от остальных выбирает один из четырех дней наступающей недели (понедельник, вторник, среду или четверг) для работы в библиотеке в отделе текущей периодики. Найти вероятности следующих событий:

A = (в понедельник в библиотеку явится один аспирант, во вторник - два, в среду - три, в четверг - четыре аспиранта);

B = {все десять аспирантов соберутся в библиотеке в четверг};

C = {пятеро из аспирантов появятся в библиотеке в первые два дня недели и пятеро - в следующие два дня};

D = {в понедельник в библиотеке появятся ровно три аспиранта}.

2. Два равносильных шахматиста играют матч из 12 партий. В каждой партии возможно три исхода: ω_1 = {выиграл первый игрок (проиграл второй)}; ω_2 = {ничья}; ω_3 = {выиграл второй (проиграл первый)}. Пусть

$$p(\omega_1) = p(\omega_3) = 0,2; \quad p(\omega_2) = 1 - p(\omega_1) - p(\omega_3) = 0,6.$$

Найти вероятности следующих событий:

A = {первый игрок выиграл 3 партии, проиграл 3 партии и остальные свел вничью};

B = {один из игроков выиграл 4 партии и проиграл 3 партии};

C = {сыграно 6 результативных партий}.

5. Прямые теоремы в схеме Бернулли

Формула Бернулли $P_n(k) = C_n^k p^k q^{n-k}$, выражающая $P_n(k)$ через n и k , становится неудобной при больших n : в этом случае она вызывает вычисление C_n^k .

5.1. Существует удобный в практическом отношении способ вычисления вероятностей $P_n(k)$ - приближенный, но достаточно точный при больших n . Его описание дано в следующей теореме.

Теорема 2 (локальная теорема Муавра-Лапласа). При больших значениях n в схеме Бернулли справедливо приближенное равенство

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x), \quad (3)$$

$$x = \frac{k - np}{\sqrt{npq}}, \text{ а } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Комментарий к теореме 2

1. Локальная теорема Муавра-Лапласа является глубоким математическим фактом, ее доказательство связано с использованием нетривиальных и сложных построений.

2. Функция $\varphi(x)$, упоминаемая в теореме, табулирована: таблицы значений этой функции приведены в каждом учебнике по теории вероятностей. Эта функция четная; ее график называется нормальной или гауссовой кривой и изображен на рис. 3.

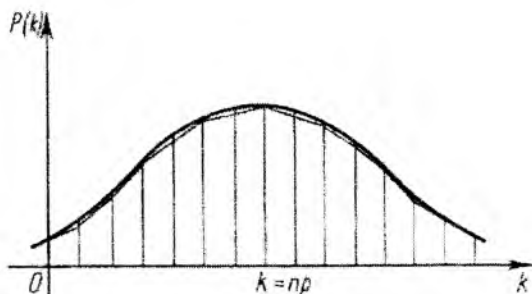


Рис. 3

3. Заметим, что $P_n(k)$ стремится к нулю при $n \rightarrow \infty$. Наибольшая из вероятностей $P_n(k)$ достигается при $k \approx np$ (k - ближайшее к np целое число). В этом случае

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}}.$$

Пример. Вычислить вероятность того, что при 100-кратном бросании монеты герб выпадет: а) ровно 50 раз; б) ровно 60 раз.

Решение

а) Здесь $n=100$, $k=50$, $p=0,5$, $q=0,5$. Используя формулу (1), получим

$$P_{100}(50) = \frac{1}{\sqrt{100 \cdot 0,5 \cdot 0,5}} \varphi(x) = \varphi(x) \cdot \frac{1}{5}, \text{ где } x = \frac{50 - 0,5 \cdot 100}{\sqrt{100 \cdot 0,5 \cdot 0,5}} = 0.$$

Следовательно, $P_{100}(50) = \frac{1}{5} \cdot \varphi(0) = \frac{1}{5} 0,3989 = 0,079$ (значение $\varphi(0)$ найдено по таблице).

$$\text{б) Аналогично находим } P_{100}(60) = \frac{1}{5} \varphi(x), \text{ где } x = \frac{60 - 0,5 \cdot 100}{\sqrt{100 \cdot 0,5 \cdot 0,5}} = \frac{10}{5} = 2.$$

$$\text{Таким образом, } P_{100}(60) = \frac{1}{5} \varphi(2) = \frac{1}{5} 0,0540 = 0,0108.$$

Пример. Строительная фирма для привлечения инвестиций в строительство нового дома собирается воспользоваться банковским кредитом. Вероятность того, что какой-либо банк в ответ на поступление бизнес-плана примет положительное решение о кредитовании фирмы, равна 0,3. Строительная фирма обратилась в 100 банков. Найти вероятности того, что решения о предоставлении кредитов этой фирме примут: а) один банк; б) 15 банков; в) 30 банков; г) 50 банков.

Решение

Данную ситуацию можно рассматривать как серию из $n=100$ испытаний Бернулли, в которых появления события в каждом испытании считается

банк решения о кредитовании. Вероятность появления события в единичном испытании равна по условию $p = 0,3$. Поскольку число испытаний велико можно воспользоваться локальной теоремой Муавра – Лапласа.

$$P_{100}(1) \approx \frac{1}{\sqrt{100 \cdot 0,3 \cdot 0,7}} \cdot \varphi\left(\frac{1-100 \cdot 0,3}{\sqrt{100 \cdot 0,3 \cdot 0,7}}\right) = \frac{1}{\sqrt{21}} \cdot \varphi\left(\frac{1-30}{\sqrt{21}}\right) = \\ = 0,22 \cdot \varphi(-6,33) \approx 0,22 \cdot 0 = 0;$$

$$P_{100}(15) \approx \frac{1}{\sqrt{21}} \varphi\left(\frac{15-30}{\sqrt{21}}\right) = 0,22 \varphi(-3,27) \approx 0,22 \cdot 0,0020 = 0,00044;$$

$$P_{100}(30) \approx \frac{1}{\sqrt{21}} \cdot \varphi\left(\frac{30-30}{\sqrt{21}}\right) = 0,22 \cdot \varphi(0) \approx 0,22 \cdot 0,03989 = 0,088;$$

$$P_{100}(50) \approx \frac{1}{\sqrt{21}} \cdot \varphi\left(\frac{50-30}{\sqrt{21}}\right) = 0,22 \cdot \varphi(4,36) \approx 0,22 \cdot 0 = 0.$$

Упражнение

Вероятность появления успеха в каждом из независимых испытаний равна 0,25. Найти вероятность того, что в серии из 300 испытаний успех наступит ровно 75 раз.

Из формулы (3) вытекает, что график функции $P_n(k)$ приближенно совпадает с графиком функции $f = \frac{\varphi(x)}{\sqrt{npq}}$, где $x = \frac{k - np}{\sqrt{npq}}$, k - целое число.

Это означает, что график функции $P_n(k)$ приближенно совпадает с гауссов-

ской кривой $y = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, сдвинутой вправо на np и сжатой по вер-

тикали в $\sqrt{npq}$ раз. При этом график $P_n(k)$ обладает характерной чертой - наличием пика в точке $k \approx np$ (рис. 3). В учебниках по теории вероятностей можно встретить более строгую формулировку локальной теоремы Муавра - Лапласа.

5.2. Вычисление вероятностей $P_n(k_1, k_2)$ в схеме Бернулли по формуле

$P_n(k) = C_n^k p^k = q^{n-k}$ при больших n является еще более затруднительным,

чем использование формулы Бернулли для вычисления $P_n(k)$. Заметим, что в практическом отношении вероятности $P_n(k_1, k_2)$ имеют большее значение, чем $P_n(k)$. Действительно, при больших n часто бывает не столь существенным знать то обстоятельство, что событие A произойдет ровно k раз, но важно знать, что количество наступлений этого события будет находиться в заданных пределах. Так, при проверке семян на всхожесть не столь важно знать, что из выбранных 1000 семян ровно 907 окажутся всхожими, но важно знать, что всхожесть семян находится в пределах от 900 до 950.

Как отмечалось выше, вероятности $P_n(k)$ при больших n малы. Вероятности $P_n(k_1, k_2)$ могут быть сколь угодно близки к единице.

Удобный приближенный способ вычисления вероятностей $P_n(k_1, k_2)$ в схеме Бернулли дает следующая теорема.

Теорема 3 (интегральная теорема Муавра - Лапласа). При больших значениях n в схеме Бернулли имеет место приближенное равенство

$$P_n(k_1, k_2) \approx \Phi(k_2') - \Phi(k_1'), \quad (4)$$

$$\text{где } k_1' = \frac{k_1 - np}{\sqrt{npq}}, \quad k_2' = \frac{k_2 - np}{\sqrt{npq}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx.$$

Комментарий к теореме 3.

1. Функция $\Phi(x)$ называется функцией Лапласа; она табулирована. Таблицы функции $\Phi(x)$ даны в каждом учебнике по теории вероятностей. Эта функция нечетная.

2. Отметим, что

$$\Phi(x) = 0, \quad \Phi(1) = 0,3413, \quad \Phi(2) = 0,4772, \quad \Phi(3) = 0,4986, \quad \Phi(\infty) = 0,5.$$

Таким образом, если в формуле (4) положить $k_2' = 3$, $k_1' = -3$, то получим $P_n(k_1, k_2) = 0,9973$.

Существует более строгая формулировка интегральной теоремы Муавра - Лапласа.

Пример. Вычислить вероятность того, что при 100-кратном бросании

монеты количество гербов будет находиться в следующих пределах:

а) [45;55]; б) [40;60]; в) [35;65].

Решение

Здесь $p = 0,5$, $q = 0,5$, $n = 100$. $\sqrt{npq} = \sqrt{100 \cdot 0,5 \cdot 0,5} = 5$.

а) $k_1 = \frac{45 - 50}{5} = -1$, $k_2 = \frac{55 - 50}{5} = 1$; $P_{100}(45,55) \approx \Phi(1) - \Phi(-1) = 2\Phi(1) = 0,6826$.

б) $k_1 = \frac{40 - 50}{5} = -2$, $k_2 = \frac{60 - 50}{5} = 2$; $P_{100}(40,60) \approx \Phi(2) - \Phi(-2) = 2\Phi(2) = 0,9545$.

в) $k_1 = \frac{35 - 50}{5} = -3$; $k_2 = \frac{65 - 50}{5} = 3$; $P_{100}(35,65) \approx \Phi(3) - \Phi(-3) = 2\Phi(3) = 0,9973$.

Из результатов вычислений видно, что вероятности рассматриваемых событий достаточно велики, в особенности последняя вероятность, равная 0,9973.

События, имеющие большую вероятность, называются практически достоверными.

В этом случае считается, что в результате опыта событие обязательно наступит. Насколько должна быть велика вероятность, чтобы событие считалось практически достоверным? Это зависит от характера задачи: во всякой ли задаче замена случайного события практически достоверным? Это зависит от характера задачи: во всякой задаче замена случайного события практически достоверным содержит «элемент риска». Ясно, что в различных условиях допустимый риск различен. Все же часто останавливаются на вероятности 0,9973. Мы также примем за определение практически достоверного события такое случайное событие, вероятность которого не меньше, чем $2\Phi(3) = 0,9973$.

5.3. Рассмотрим схему Бернулли с большим количеством n испытаний; обозначим через σ число $\sqrt{npq}$. Из интегральной теоремы Муавра - Лапласа вытекает, что

$$P_n(np - 3\sigma, np + 3\sigma) = 0,9973. \quad (5)$$

Действительно, при $k_1 = np - 3\sigma$, $k_2 = np + 3\sigma$ имеем $k_1' = -3$, $k_2' = 3$ и

$$P_n(k_1, k_2) \approx \Phi(3) - \Phi(-3) = 2\Phi(3) = 0,9973.$$

Формула (5) позволяет для каждой схемы Бернулли указать интервал (k_1, k_2) такой, что количество наступлений события A принадлежит этому интервалу с вероятностью 0,9973; иными словами, событие $k_1 \leq m < k_2$ *практически достоверно*. Формула (3) называется правилом «трех сигм», а интервал (k_1, k_2) , где $k_1 = np - 3\sqrt{npq}$, $k_2 = np + 3\sqrt{npq}$ - трехсигмовым интервалом.

Заметим, что трехсигмовый интервал оказывается удивительно узким. Если любому здравомыслящему человеку, не знакомому с теорией вероятностей, предложить угадать интервал, в который с практической достоверностью попадет количество наступлений событий при последовательных испытаниях, то, как правило, в ответе будет дан гораздо более широкий интервал.

Пример. Некоторая система состоит из 10000 (независимых) элементов. Вероятность выхода из строя одного элемента равна 0,5. Пусть n - количество вышедших из строя элементов системы. Найти трехсигмовый интервал.

Решение

Имеем $n = 10000$, $p = 0,5$, $q = 0,5$. Тогда $\sigma = \sqrt{10000 \cdot 0,5 \cdot 0,5} = 50$, $k_1 = np - 3\sigma = 5000 - 150$, $k_2 = np + 3\sigma = 5000 + 150$. Итак, с вероятностью 0,9973 можно утверждать, что количество вышедших из строя элементов находится в пределах 5000 ± 150 (событие *практически достоверное*).

В частности, если взять запас в 5000 элементов для замены вышедших из строя, то в 50% случаев этого запаса не хватит. Если же увеличить этот запас всего на 3%, т.е. взять 5150 элементов, то его хватит наверняка (т. е. с вероятностью большей, чем 0,9973). Оценка трехсигмового интервала этого примера «на глаз», «по здравому смыслу» приводит, как правило, к большому преувеличению истинного значения.

С помощью интегральной теоремы Муавра - Лапласа можно пояснить, почему и в каком смысле вероятность p события A в одном испытании совпадает (приближенно) с частотой $\frac{m}{n}$ наступления события A в n испытаниях. Действительно, с вероятностью 0,9973 выполняется неравенство

$$np - 3\sqrt{npq} \leq m < np + 3\sqrt{npq},$$

откуда после деления всех его частей на n получим

$$p - 3\sqrt{\frac{pq}{n}} \leq \frac{m}{n} < p + 3\sqrt{\frac{pq}{n}}.$$

Так как $3\sqrt{\frac{pq}{n}} \rightarrow \infty$ при $n \rightarrow \infty$, то частота $\frac{m}{n}$ с практической достоверностью при больших n так угодно мало отличается от p .

Следствие. Вообще говоря, используя интегральную теорему Муавра-Лапласа легко можно получить вероятность отклонения относительной частоты от постоянной вероятности в n независимых испытаниях в более общем случае, т.е. формулу

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Пример. Вероятность того, что деталь не стандартна, $p = 0,1$. Найти вероятность того, что среди случайно отобранных 400 деталей относительная частота появления нестандартных деталей отклонится от вероятности $p = 0,1$ по абсолютной величине не более, чем на 0,03.

Решение. По условию $n = 400$; $p = 0,1$; $q = 0,9$; $\varepsilon = 0,03$. Требуется найти вероятность $P\left(\left|\frac{m}{400} - 0,1\right| \leq 0,03\right)$.

Пользуясь формулой

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right),$$

имеем:

$$P\left(\left|\frac{m}{400} - 0,1\right| \leq 0,03\right) \approx 2\Phi\left(0,03 \sqrt{\frac{400}{0,1 \cdot 0,9}}\right) = 2\Phi(2).$$

По таблице значений функции Лапласа находим $2\Phi(2) = 0,9544$.

Итак, искомая вероятность приближенно равна 0,9544.

Смысл полученного результата таков: если взять достаточно большое число проб по 400 деталей в каждой, то примерно в 95,44% этих проб отклонение относительной частоты от постоянной вероятности $p = 0,1$ по абсолютной величине не превысит 0,03.

Пример. Вероятность того, что деталь не стандартна $p = 0,1$. Найти, сколько деталей надо отобрать, чтобы с вероятностью равной 0,9544 можно было утверждать, что относительная частота появления нестандартных деталей (среди отобранных) отклонится от постоянной вероятности p по абсолютной величине не более, чем на 0,03.

Решение

По условию

$$p = 0,1; \quad q = 0,9; \quad \varepsilon = 0,03; \quad P\left(\left|\frac{m}{n} - 0,1\right| \leq 0,03\right) = 0,9544.$$

Требуется найти n .

Воспользуемся формулой

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

В силу условия,

$$2\Phi\left(0,03\sqrt{\frac{n}{0,1 \cdot 0,9}}\right) = 2\Phi(0,1\sqrt{n}) = 0,9544.$$

Следовательно, $\Phi(0,1\sqrt{n}) = 0,4772 = 0,4772$. По таблице значений функции Лапласа находим

$$\Phi(2) = 0,4772.$$

Для отыскания числа n получаем уравнение

$$0,1\sqrt{n} = 2.$$

Отсюда искомое число деталей $n = 400$.

Смысл полученного результата таков: если взять достаточно большое число проб по 400 деталей, то в 95,44% этих проб относительная частота появления нестандартных деталей будет отличаться от постоянной вероятности $p = 0,1$ по абсолютной величине не более, чем на 0,03, т.е. относительная частота будет заключена в границах от 0,07 ($0,1 - 0,03 = 0,07$) до 0,13 ($0,1 + 0,03 = 0,13$).

Другими словами, число нестандартных деталей в 95,44% проб будет заключено от 28 (7% от 400) до 52 (13% от 400).

Если взять лишь одну пробу из 400 деталей, то с большой уверенностью можно ожидать, что в этой пробе будет нестандартных деталей не менее 28 и не более 52. Возможно, хотя и маловероятно, что нестандартных деталей окажется меньше 28, либо больше 52.

Более строгая формулировка утверждения о близости частоты и вероятности дана в теореме Бернулли (один из вариантов закона больших чисел), которую рассмотрим в одном из следующих параграфов.

5.4. Представляет интерес схема Бернулли с малой вероятностью p появления события A в одном испытании и с большим количеством n испытаний. Пусть при большом n малая вероятность p такова, что $np = \lambda$,

где λ - некоторое число. Вероятность $P_n(k)$ в такой схеме Бернулли описывается следующей теоремой.

Теорема 4 (теорема Пуассона). Пусть $n \rightarrow \infty$, $\lambda > 0$ постоянно и $p = \frac{\lambda}{n}$.

Тогда в схеме Бернулли из n независимых испытаний, в каждом из которых вероятность наступления события A равна p , имеет место приближенное равенство

$$P_n(k) = P(k) \approx e^{-\lambda} \frac{\lambda^k}{k!}. \quad (6)$$

Комментарий к теореме 3. Обратим внимание на следующее обстоятельство: вероятность наступления события A ровно k раз не зависит от n , что выглядит неправдоподобно. Это можно объяснить так. Пусть n велико: увеличивая n в μ раз и уменьшая p во столько же раз (так что np не изменяется), мы в самом деле имеем $P_n(k, p) \approx P_{\mu n}\left(k, \frac{p}{\mu}\right)$. Таким образом, независимость вероятности рассматриваемого события от n объясняется тем, что она вычислена в разных схемах Бернулли.

Теорему примем без доказательства.

Пример. Учебник издан тиражом 100000 экземпляров. Вероятность того, что учебник сброшюрован неправильно, равна 0,0001. Найти вероятность того, что тираж содержит ровно 5 бракованных книг.

Решение. По условию, $n=100000$, $p=0,0001$, $k=5$. События, состоящие в том, что книги сброшюрованы неправильно, независимы, число n велико, а вероятность p мала, поэтому воспользуемся формулой (4). Найдём $\lambda=np=100000 \cdot 0,0001=10$. Тогда

$$P_{100000}(5) = e^{-10} \frac{10^5}{5!} = 10^{-5} \cdot \frac{0,000045}{120} = 0,0375.$$

Пример. На лекции по теории вероятностей присутствует 200 человек. Вероятность того, что день рождения случайно выбранного студента приходится на определённый день года, составляет $1/36$. Найти вероятность того,

что один человек из присутствующих родился 1 января, и два человека родились 8 марта.

Решение

Пусть событие A состоит в том, что случайно выбранный студент родился 1 января, событие N - в том, что k человек из 200 родились 1 января.

Тогда по условию $p = P(A) = \frac{1}{365}$. Предположим, что опрос $n = 200$ студентов относительно даты их рождения удовлетворяет условиям, которые накладываются на испытания Бернулли, в каждом из которых событие A может наступить, а может и не наступить. Тогда, поскольку $n = 200$ велико, а произведение $np = \frac{200}{365} = 0,548$, для подсчёта вероятности события N можно воспользоваться формулой Пуассона

$$P_{200}(k) \approx e^{-0,548} \frac{(0,548)^k}{k!}$$

и при $k = 1$ получаем

$$P_{200}(k) \approx e^{-0,548} (0,548) = 0,317.$$

Пусть событие M - в том, что m человек из 200 родились 8 марта. Тогда в соответствии с формулой умножения вероятностей, $p(NM) = p(N)p_N(M)$, где $p_N(M) = P_{n-k}(m)$ - вероятность того, что из $(n-k)$ студентов m родились 8 марта. Так как число $(n-k) = 200 - 1 = 199$ велико, а $(n-k)p = \frac{198}{365} = 0,542$, для расчёта вероятности события M можно вновь воспользоваться формулой Пуассона

$$P_{n-k}(m) \approx e^{-0,542} \frac{(0,542)^m}{m!}.$$

При $n = 200$, $k = 1$, $m = 2$ получаем

$$p_N(M) = P_{199}(2) = e^{-0,542} \frac{(0,542)^2}{2} = 0,086,$$

поэтому искомая вероятность $p(NM) = 0,317 \cdot 0,086 = 0,027$.

Упражнения

1. В условиях предыдущего примера найти вероятность того, что число родившихся 1 января и 8 марта не больше двух.
2. Владельцы кредитных карт ценят их и теряют весьма редко - вероятность потерять кредитную карту в течение недели для случайно выбранного вкладчика составляет 0,001. Банк выдал кредитные карты 2000 клиентам. Найти: а) вероятность того, что за предстоящую неделю будет утеряна ровно одна кредитная карта; б) вероятность того, что за предстоящую неделю будет утеряна хотя бы одна кредитная карта; в) наиболее вероятное число кредитных карт, теряемых за месяц.

Вопросы для самопроверки

1. Что называется схемой Бернулли?
2. Приведите примеры последовательности испытаний, которые не образуют схему Бернулли.
3. Какие задачи рассматриваются в связи со схемой Бернулли?
4. Сформулируйте теорему Бернулли.
5. На чем основывается доказательство теоремы Бернулли?
4. Что называется наивероятнейшим числом?
5. В чем заключаются затруднения, возникающие при вычислении вероятностей в схеме Бернулли при больших n ?
6. Сформулируйте локальную теорему Муавра-Лапласа. Приведите свойства функции $\varphi(x)$, которая упоминается в этой теореме.
7. Исходя из локальной теоремы Муавра-Лапласа прокомментируйте поведения вероятностей $P_n(k)$, при $n \rightarrow \infty$ и $k \approx np$.
8. Напишите вид гауссовской функции.
9. В чем заключается сходство функции $P_n(k)$ и гауссовской функции?
10. Сформулируйте интегральную теорему Муавра-Лапласа.
11. Как называется функция $\Phi(x)$, упоминаемая в интегральной теореме Муавра-Лапласа? Назовите ее свойства.
12. Что представляет собой «трехсигмовый интервал»?
13. Сформулируйте теорему Пуассона. Прокомментируйте теорему.

§4. Случайные величины и законы их распределения

Ключевые слова: случайная величина, дискретная случайная величина, непрерывная случайная величина, ряд распределения, функция распределения, функция плотности распределения, биномиальное и пуассоновское распределения, нормальное, равномерное и показательное распределения.

Случайные события могут быть представлены через случайные величины. Понятия «случайная величина» расширяет область применения методов теории вероятностей в решении практических задач. Поэтому понятие «случайной величины» является одним из важнейших понятий теории вероятностей.

1. Определение случайной величины. Задание дискретной случайной величины

Определение 1. Случайной величиной называется функция $X = X(\omega)$, определенная на некотором множестве элементарных событий Ω .

Если Ω конечно, или счетно, то на функцию $X(\omega)$ не накладывается никаких ограничений, т.е. это может быть любая функция. Если Ω континуум, то $X(\omega)$ должна быть такой, что для любого события вида

$$A = \{\omega : X(\omega) \leq x_0\},$$

где x_0 — любое число и $x_0 \in R^1$, могла быть определена его вероятность

$$p(A) = p\{X < x_0\}.$$

Следует отметить, что при изучении случайных величин и их свойств почти никогда не вспоминают об области их определения, т.е. о Ω .

Все встречающиеся в природе процессы и объекты, так или иначе, характеризуются численными значениями своих параметров. Причем по цело-

му ряду причин эти значения должны рассматриваться нами как случайные. Поэтому понятие случайные величины имеет очень большую практическую значимость. Примеров случайных величин можно привести бесконечное количество – это и температура воздуха на завтра, и курс доллара на банковских торгах, и количество покупателей в магазине, и масса, даже стандартно упакованного, пакета некоторого продукта, и т.д. Фактически теория случайных величин составляет основную часть содержания теории вероятностей, и всю математическую статистику.

Определение 2. Конкретные значения, которые принимала случайная величина в ходе соответствующих экспериментов или наблюдений, называются *реализациями* данной случайной величины.

Например, X – курс доллара, установленный Центральным банком Узбекистана, это случайная величина. А, $x_1 = 8092.13$ курс на 12 сентября 2017 года, $x_2 = 8092.13$ курс на 17 апреля 2018 года - это реализации данной случайной величины.

Чаще всего случайная величина обозначается заглавными латинскими буквами, например, X, Y, Z аргумент при этом обычно опускают. А их реализации – соответствующими строчными буквами – x, y, z , и т.д.

Определение 3. Говорят, что случайная величина X задана, или задан ее закон распределения, если для любого множества $B \subset R^1$ определена вероятность

$$P\{x \in B\},$$

т.е. вероятность того, что при очередном эксперименте реализация x окажется лежащей в этом множестве.

Определение 4. Случайная величина X называется *дискретной*, если множество ее возможных значений конечно, или счетно.

Если случайная величина X имеет конечное количество возможных значений, то задать ее можно простым перечислением этих значений x_i , $i = \overline{1, n}$, и их соответствующих вероятностей

$$p\{X = x_i\} = p_i.$$

Очевидно, что последнее означает, что « p_i есть вероятность события $X = x_i$ ». Ясно, что события

$$X = x_1, X = x_2, \dots, X = x_n \quad (*)$$

попарно несовместны, а их сумма есть событие достоверное (при каждом осуществлении опыта величина X принимает одно и только одно из своих значений), т.е. наступает одно и только одно из событий (*), т.е.

$$\sum_{i=1}^n p\{X = x_i\} = \sum_{i=1}^n p_i = 1$$

(*) на единицу как-то распределена между значениями случайной величины, отсюда и термин «*распределение*»).

Такое перечисление обычно записывают в виде таблицы (табл. 1), которую называют *рядом распределения* данной случайной величины.

Таблица 1

x_1	x_1	x_2	x_2	...	x_n
p_1	p_1	p_2	p_3	...	p_n

Итак, простейшей формой задания закона распределения дискретной случайной величины X является ряд распределений данной случайной величины.

Пример 1. Пусть X – случайная величина – числа очков при подбрасывании игральной кости. Это дискретная случайная величина, а ее ряд распределения имеет вид, указанный в таблице 2.

Таблица 2

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Сложив вероятности, получим

$$\sum_{i=1}^6 p_i = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1.$$

Пример 2. Для случайной величины X - числа бросаний монеты из примера 2.4 §3 при $n=10$, $p=1/2$ таблица распределения такова:

x_i	0	1	2	3	4	5	6	7	8	9	10
p_i	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$

Сложение вероятностей дает

$$\sum_{i=0}^{11} p_i = \frac{1}{1024} + \frac{10}{1024} + \frac{45}{1024} + \frac{120}{1024} + \frac{210}{1024} + \frac{252}{1024} + \frac{210}{1024} + \frac{120}{1024} + \frac{45}{1024} + \frac{10}{1024} + \frac{1}{1024} = 1.$$

Если случайная величина имеет счетное количество возможных значений, то ее задают с помощью формульного описания этих значений и их вероятностей. При этом, формульные выражения зависят от i , т.е. номера значения

$$p\{X = x_i(i)\} = p_i(i), \quad i = \overline{1, \infty}.$$

При этом должно выполняться

$$\sum_{i=1}^{\infty} p_i = 1.$$

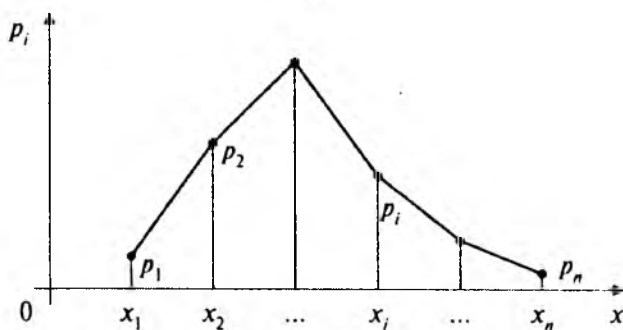


Рис. 1

Ряд распределения может быть изображен графически, если по оси абсцисс откладывать значения случайной величины, а по оси ординат - соответствующие им вероятности. Соединение полученных точек образует ломаную,

индивидуальную многоугольником или полигоном распределения вероятностей (рис. 1)

Пример 3. В лотерее разыгрываются: автомобиль стоимостью 5000 ден. ед., 4 телевизора стоимостью 250 ден. ед., 5 видеомэагнитофонов стоимостью 200 ден. ед. Всего продается 1000 билетов по 7 ден. ед. Составить закон распределения чистого выигрыша, полученного участником лотереи, купившим один билет.

Решение

Возможные значения случайной величины X - чистого выигрыша на один билет - равны $0 - 7 = -7$ ден. ед. (если билет не выиграл), $200 - 7 = 193$, $250 - 7 = 243$, $5000 - 7 = 4993$ ден. ед. (если на билет выпал выигрыш, соответственно, видеомэагнитофона, телевизора или автомобиля). Учитывая, что из 1000 билетов число невыигравших составляет 990, а указанных выигрышей соответственно 5, 4 и 1, и используя классическое определение вероятности, получим:

$$p\{X = -7\} = \frac{990}{1000} = 0.990; \quad p\{X = 193\} = \frac{5}{1000} = 0.005;$$

$$p\{X = 243\} = \frac{4}{1000} = 0.004; \quad p\{X = 4993\} = \frac{1}{1000} = 0.001,$$

т. е. ряд распределения

x_i	-7	193	243	4993
p_i	0,990	0,005	0,004	0,001

$$\text{и } \sum_{i=1}^4 p_i = 0,990 + 0,005 + 0,004 + 0,001 = 1.$$

Пример 4. Вероятности того, что студент сдаст семестровый экзамен в сессию по дисциплинам А и Б, равны, соответственно, 0,7 и 0,9. Составить закон распределения числа семестровых экзаменов, которые сдаст студент.

Решение

Возможные значения случайной величины X - числа сданных экзаменов - 0, 1, 2.

Пусть A_i - событие, состоящее в том, что студент сдаст i -й экзамен ($i = 1, 2$). Тогда вероятности того, что студент сдаст в сессию 0, 1, 2 экзамена, будут, соответственно, равны (считаем события A_1 и A_2 независимыми):

$$p\{X=0\} = p(\overline{A_1} \overline{A_2}) = p(\overline{A_1}) p(\overline{A_2}) = (1-0.7)(1-0.9) = 0.3 \cdot 0.1 = 0.03;$$

$$p\{X=1\} = p(A_1 \overline{A_2} + \overline{A_1} A_2) = p(A_1) p(\overline{A_2}) + p(\overline{A_1}) p(A_2) = \\ = 0.7 \cdot 0.1 + 0.3 \cdot 0.9 = 0.34;$$

$$p\{X=2\} = p(A_1 A_2) = p(A_1) p(A_2) = 0.7 \cdot 0.9 = 0.63.$$

Итак, ряд распределения

x_i	1	2	3
p_i	0.03	0.34	0.63

и $\sum_{i=1}^3 p_i = 0.03 + 0.34 + 0.63 = 1.$

На рис. 2 полученный ряд распределения представлен графически в виде многоугольника (полигона) распределения вероятностей.

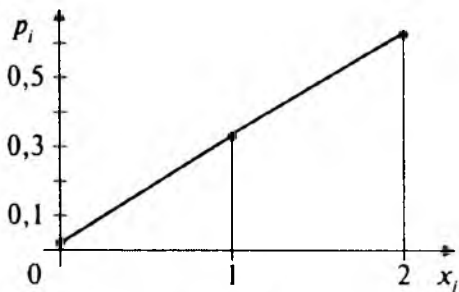


Рис. 2

Упражнения

1. Клиенты банка, не связанные друг с другом, не возвращают кредиты в срок с вероятностью 0.1. Составить закон распределения числа возвращенных в срок кредитов из 5 выданных.

2. Найти закон распределения числа пакетов трех акций, по которым владельцем будет получен доход, если вероятность получения дохода по каждому из них равна, соответственно, 0.5; 0.6; 0.7.

3. Контрольная работа состоит из трех вопросов. На каждый вопрос приведено 4 ответа, один из которых правильный. Составить закон распределения числа правильных ответов при простом угадывании.

4. В среднем по 10% договоров страховая компания выплачивает страховые суммы в связи с наступлением страхового случая. Составить закон распределения числа таких договоров среди наудачу выбранных четырех.

5. В билете три задачи. Вероятность правильного решения первой задачи равна 0,9, второй - 0,8, третьей - 0,7. Составить закон распределения числа правильно решенных задач в билете.

6. Вероятность попадания в цель при одном выстреле равна 0,8 и уменьшается с каждым выстрелом на 0,1. Составить закон распределения числа попаданий в цель, если сделано три выстрела.

7. Произведено два выстрела в мишень. Вероятность попадания в мишень первым стрелком равна 0,8, вторым - 0,7. Составить закон распределения числа попаданий в мишень (Каждый стрелок делает по одному выстрелу).

8. Из пяти гвоздик две белые. Составить закон распределения и найти функцию распределения случайной величины, выражающей число белых гвоздик среди двух одновременно взятых.

9. Из 10 телевизоров на выставке 4 оказались фирмы «Сони». Наудачу для осмотра выбрано 3. Составить закон распределения числа телевизоров фирмы «Сони» среди 3 отобранных.

10. Среди 15 собранных агрегатов 6 нуждаются в дополнительной смазке. Составить закон распределения числа агрегатов, нуждающихся в дополнительной смазке, среди пяти наудачу отобранных из общего числа.

Пример 5. Пусть X - случайная величина числа бросаний монеты до первого выпадения герба. Ясно, что эта случайная величина имеет счетное количество возможных значений, и

$$p\{X = i\} = \frac{1}{2^i}, \quad i = \overline{1, \infty}.$$

Ряд распределения:

x_i	1	2	3	...	i	...
p_i	$1/2$	$1/2^2$	$1/2^3$	...	$1/2^i$	...

Ряд $\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i$ представляет собой бесконечно убывающую геометрическую прогрессию, первый член которой $b_1 = \frac{1}{2}$, а знаменатель $q = \frac{1}{2}$. Этот ряд сходится, его сумма $S = \frac{b_1}{1-q} = 1$. Таким образом,

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1.$$

Следует отметить, что реальных примеров счетных случайных величин не очень много.

Пример 6. Изделия испытывают на прочность при работе в перегрузочных режимах. Вероятность для каждого изделия пройти испытание равна $5/6$ и не зависит от исходов испытаний других изделий. Испытания заканчиваются сразу же после того, как появится первое изделие, не выдержавшее проверку на прочность. Найти ряд распределения случайной величины X - числа производимых испытаний.

Решение

Вероятность для каждого изделия пройти испытание равна $p = 5/6$; вероятность того, что изделие не пройдет испытание, равна $q = 1 - p = 1/6$. Испытания заканчиваются на k -м изделии ($k = 1, 2, 3, \dots$), если первые $k-1$ изделий пройдут испытания, а k -е изделие не выдержит испытания.

Случайная величина X - число проводимых испытаний, ее возможные значения: $1, 2, 3, \dots, k, \dots$ (теоретически число испытаний может быть бесконечно большим). Вероятности возможных значений величины X найдем по теореме умножения вероятностей независимых событий:

$$P\{X = k\} = \left(\frac{5}{6}\right)^{k-1} \left(\frac{1}{6}\right), \quad (k = 1, 2, 3, \dots).$$

Рассматриваемая случайная величина X имеет следующий ряд распределения:

x_i	1	2	3	...	k	...
p_i	$\frac{1}{6}$	$\frac{5}{6^2}$	$\frac{5^2}{6^3}$	...	$\frac{5^{k-1}}{6^k}$	...

Особенность рассматриваемой случайной величины X состоит в том, что теоретически последовательность ее возможных значений является неограниченной. Число производимых испытаний может быть бесконечно большим, однако вероятность того, что это произойдет, стремится к нулю:

$$\lim_{k \rightarrow \infty} p\{X = k\} = \frac{5^{k-1}}{6^k} = 0.$$

Убедимся в том, что полученная последовательность вероятностей характеризует закон распределения X , т. е. что $\sum_{i=1}^{\infty} p_i = 1$.

Найдем

$$\sum_{i=1}^{\infty} p_i = \frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} + \dots + \frac{5^{k-1}}{6^k} + \dots = \frac{1}{6} \left(1 + \frac{5}{6} + \frac{5^2}{6^2} + \dots + \frac{5^{k-1}}{6^{k-1}} + \dots \right) = \frac{1}{6} \sum_{k=1}^{\infty} \left(\frac{5}{6} \right)^{k-1}.$$

Ряд $\sum_{k=1}^{\infty} \left(\frac{5}{6} \right)^{k-1}$ представляет собой бесконечно убывающую геометрическую прогрессию, первый член которой $b_1 = 1$, а знаменатель $q = \frac{5}{6}$. Этот ряд сходится, его сумма $S = \frac{b_1}{1-q} = 6$. Таким образом,

$$\sum_{i=1}^{\infty} p_i = \frac{1}{6} S = \frac{1}{6} \cdot 6 = 1.$$

Закон распределения полностью характеризует случайную величину, указывая возможные значения и вероятности, с которыми эти значения появляются в результате испытаний. Для первого из рассмотренных законов распределений (пример 1) все значения равновероятны (пример 1), а для второго (пример 2) значения резко различаются по своим вероятностям: значение 10 имеет вероятность, в 252 раза меньшую, чем значение 5. Это, в частности,

означает, что случайная величина принимает значение 5 в 252 раза чаще, чем 10 и т.д.

Упражнения

1. В рекламных целях торговая фирма вкладывает в каждую десятую единицу товара денежный приз размером 1 тыс. д.е. Составить закон распределения случайной величины-размера выигрыша при пяти сделанных покупках.

2. Вероятность поражения вирусным заболеванием куста земляники равна 0,2. Составить закон распределения числа кустов земляники, зараженных вирусом, из четырех посаженных кустов.

3. Стрелок ведет стрельбу по цели с вероятностью попадания при каждом выстреле 0,2. За каждое попадание он получает 5 очков, а в случае промаха очков ему не начисляют. Составить закон распределения числа очков, полученных стрелком за 3 выстрела.

4. Возможные значения случайной величины таковы: $x_1 = 2$, $x_2 = 5$, $x_3 = 8$. Известны вероятности первых двух возможных значений: $p_1 = 0,4$, $p_2 = 0,15$. Найти вероятность x_3 .

5. Игральная кость брошена 3 раза. Написать закон распределения числа появлений шестерки.

6. Составить закон распределения вероятностей числа появлений события A в трех независимых испытаниях, если вероятность появления события в каждом испытании равна 0,6.

7. Пряжильщица обслуживает 1000 веретен. Вероятность обрыва нити на одном веретене в течение одной минуты равна 0,004. Найти вероятность того, что в течение одной минуты обрыв произойдет в пяти веретенах.

8. Найти среднее число опечаток на странице рукописи, если вероятность того, что страница рукописи содержит хотя бы одну опечатку, равна 0,95. Предполагается, что число опечаток распределено по закону Пуассона.

Указание: задача сводится к отысканию параметра λ из уравнения

9. Коммутатор учреждения обслуживает 100 абонентов. Вероятность того, что в течение одной минуты абонент позвонит на коммутатор, равна 0,01. Какое из двух событий вероятнее: в течение одной минуты позвонит 3 абонента; позвонит 4 абонента?

10. Рукопись объемом в 1000 страниц машинописного текста содержит 1000 опечаток. Найти вероятность того, что наудачу взятая страница содержит: а) хотя бы одну опечатку; б) ровно 2 опечатки; в) не менее двух опечаток. Предполагается, что число опечаток распределено по закону Пуассона.

Ниже ознакомимся с двумя важными примерами дискретных случайных величин. Соответствующие им законы носят названия: *биномиальное распределение*, *пуассоновское распределение*.

Определение 6. Распределение случайной величины X , равной количеству наступлений события A в схеме Бернулли из n испытаний, называется *биномиальным распределением*.

В этом распределении значению $k \in \{0, 1, 2, \dots, n\}$ случайной величины X соответствует вероятность $P_n(k) = C_n^k p^k q^{n-k}$, где p – вероятность наступления события A в одном испытании, $q = 1 - p$.

Комментарий к определению 6

Биномиальное распределение дискретно (т.е. является распределением дискретной случайной величины X).

Биномиальное распределение широко используется в теории и практике статистического контроля продукции, при описании функционирования систем массового обслуживания, в теории стрельбы и в других областях.

Примером биномиального распределения служит последняя таблица из примера 2. Здесь $n = 10$, $p = 1/2$.

Определение 7. Распределение случайной величины X , принимающей значения $k \in \{0, 1, 2, \dots\}$ с вероятностями $\frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda > 0$ - некоторый параметр, называется *пуассоновским распределением* или *распределением Пуассона*.

Комментарий к определению 7

Пуассоновское распределение дискретно, т. е. является распределением дискретной случайной величины.

По пуассоновскому распределены, например, число рождения четверяшек, число сбоев на автоматической линии, число отказов сложной системы в «нормальном режиме», число «требуемых на обслуживание», поступивших в единицу времени в системах массового обслуживания и др.

2. Непрерывная случайная величина. Функция распределения

Определение 5. Случайная величина называется *непрерывной*, если область ее возможных значений имеет более чем счетную мощность, т.е. включает, хотя бы один непрерывный интервал числовой оси.

На практике, как правило, встречаются непрерывные случайные величины. Непрерывными считают даже денежные суммы, человеческие ресурсы, и т.д. Тем более таковыми являются различные физические величины, или относительные экономические показатели, и т.д.

Ясно, что задать непрерывную случайную величину **простым перечислением ее значений и их вероятностей** невозможно.

Отметим, что если при рассмотрении дискретных случайных величин мы могли ограничиться событиями, представляющимися в виде суммы конечного или счетного множества элементарных событий $X = x_i$, то при переходе к непрерывным случайным величинам нам следует, прежде всего, расширить класс событий. В необходимости такого расширения можно убедиться на примере. Пусть с испытательной целью определяется полное время работы электрической лампы; для этого выпущенную заводом лампу эксплуа-

тируют без перерыва до выхода ее из строя. Результатом такого испытания является величина X – срок службы лампы. Очевидно, эта величина является случайной – предсказать заранее ее значение невозможно. Элементарным событием в данном примере будет любое событие вида $X = a$, где a – неотрицательное число. Однако в отличие от дискретного случая каждое отдельное элементарное событие не представляет теперь большого интереса. Действительно, возможных значений для X существует несчетное множество, между тем в любой серии испытаний мы имеем дело всегда с конечным числом ламп. Поэтому ясно, что данное фиксированное значение a в серии испытаний, как правило, не будет встречаться вообще или же будет наблюдаться чрезвычайно редко. Другими словами, вероятность события $X = a$, будет равна нулю.

В то же время события, выражаемые при помощи неравенств скажем, $X > 1000$ (лампа перегорела, не прослужив 1000 часов), представляются значительно более важными. Вероятности таких событий дают существенную информацию о распределении значений величины X и тем самым – о качестве ламп. Разумеется, вслед за событиями такого рода мы должны привлечь к рассмотрению и их комбинации, получаемые при помощи конечного или счетного числа операций сложения, умножения и перехода к противоположному событию. В первую очередь, с помощью такого рода событий можно ввести понятия распределения непрерывной случайной величины.

Оказывается, что любое множество числовой оси можно представить в виде не более чем счетного объединения непересекающихся интервалов вида

$$[a; b], (a; b], [a; b), (a; b),$$

или их дополнений. Такие интервалы называют *простыми множествами*. Поэтому достаточно определить вероятность попадания реализации некоторой случайной величины в любой такой интервал, и тогда, на основании теоремы сложения вероятностей, эта случайная величина будет задана. Рассмотрим, как это можно сделать.

Определение 8. *Функцией распределения* случайной величины X называется такая функция $F(x)$, что для любого $x \in R^1$ выполняется

$$F(x) = p\{X \leq x\}.$$

Покажем, что с помощью функции распределения можно определить вероятность попадания реализации случайной величины в любое простое множество:

1) для $(a; b]$ это почти очевидно, так как

$$p\{X \leq b\} = p\{X \leq a\} + p\{a < X \leq b\} \text{ или } p\{a < X \leq b\} = p\{X \leq b\} - p\{X \leq a\},$$

тогда

$$p\{a < X \leq b\} = F(b) - F(a);$$

2) для $[a; b]$. Пусть x_i — некоторая числовая последовательность, стремящаяся к a слева, т.е. это последовательность чисел меньших a , и таких что

$$\lim_{i \rightarrow \infty} x_i = a,$$

но тогда и для последовательности интервалов справедливо

$$(x_i, b] \rightarrow [a; b],$$

а тем самым

$$p\{a < X \leq b\} = F(b) - \lim_{i \rightarrow \infty} F(x_i)$$

если только соответствующий предел слева функции $F(x)$ существует. Но для большинства реальных случайных величин это выполняется;

3) аналогично и для $[a; b)$. $(a; b)$.

Таким образом, функция распределения полностью задает любую случайную величину, в том числе и непрерывную.

Для дискретной случайной величины, имеющей ряд распределения

x_i	x_1	x_2	x_3	...	x_n
p_i	p_1	p_2	p_3	...	p_n

• функция распределения имеет кусочно-постоянный, ступенчатый вид.

Пример 7. Дискретная случайная величина X задана рядом распределения

x_i	-2	0	3	7
p_i	0,3	0,1	0,5	0,1

Найти функцию распределения $F(x)$ и построить график этой функции.

Решение

1. На промежутке $-\infty < x \leq -2$ случайная величина X не принимает ни одного значения, меньшего числа -2 , поэтому

$$F(x) = F(-2) = p\{X < -2\} = 0.$$

2. На промежутке $-2 < x \leq 0$ величина X принимает одно значение $X = 0$, функция распределения $F(x) = F(0) = p\{X < 0\}$. Вероятность того, что X меньше 0, равна 0,3. На рассматриваемом промежутке

$$F(x) = F(0) = p\{X < 0\} = 0,3.$$

3. На промежутке $0 < x \leq 3$ случайная величина X принимает одно значение $X = 3$, функция распределения $F(x) = F(3) = p\{X < 3\}$. При $X < 3$ случайная величина X может принять или значение -2 с вероятностью 0,3, или же значение 0 с вероятностью 0,1, поэтому, одно из этих значений, безразлично какое, X может принять (по теореме сложения вероятностей несовместных событий) с вероятностью $0,3 + 0,1 = 0,4$. Таким образом,

$$F(x) = F(3) = p\{X < 3\} = p\{X = -2\} + p\{X = 0\} = 0,3 + 0,1 = 0,4.$$

4. На промежутке $3 < x \leq 7$ случайная величина X принимает одно значение $X = 7$, функция распределения $F(x) = F(7) = p\{X < 7\}$. При $X < 7$ величина X может принять или значение -2 с вероятностью 0,3, или значение 0 с вероятностью 0,1, или значение 3 с вероятностью 0,5, поэтому одно из этих значений, безразлично какое, X может принять (по теореме сложения

вероятностей несовместных событий) с вероятностью $0.3 + 0.1 + 0.5 = 0.9$.

Таким образом, на промежутке

$$\begin{aligned} F(x) = F(3) &= p\{X < 7\} = p\{X = -2\} + p\{X = 3\} + p\{X = 0\} = \\ &= 0.3 + 0.1 + 0.5 = 0.9. \end{aligned}$$

5. На промежутке $7 < x < \infty$ величина X может принять любое из всех своих возможных значений, поэтому функция распределения

$$F(x) = p\{X < \infty\} = 1.$$

Итак, рассматриваемая случайная величина X на всей числовой оси характеризуется следующей функцией распределения:

$$F(x) = \begin{cases} 0 & \text{при } x \leq -2, \\ 0.3 & \text{при } -2 < x \leq 0, \\ 0.4 & \text{при } 0 < x \leq 3, \\ 0.9 & \text{при } 3 < x \leq 7, \\ 1 & \text{при } x > 7. \end{cases}$$

График этой функции приведен на рис. 3.

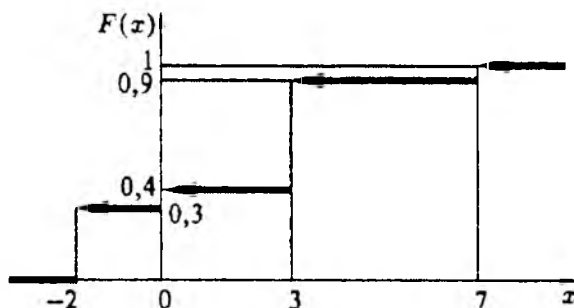


Рис. 3

Заметим, что при подходе слева к точкам разрыва функция сохраняет свое значение (про такую функцию говорят, что она непрерывна слева).

Этот пример позволяет прийти к утверждению, что функция распределения любой дискретной случайной величины есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным

значениям случайной величины и равны вероятностям этих значений. Сумма скачков функции $F(x)$ равна 1.

Пример 8. Дискретная случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 2, \\ 0.5 & \text{при } 2 < x \leq 4, \\ 0.7 & \text{при } 4 < x \leq 8, \\ 1 & \text{при } x > 8. \end{cases}$$

Найти вероятность того, что X примет значение не меньшее 4 и меньшее 8.

Решение

Искомую вероятность $p\{4 \leq X < 8\}$ найдем по формуле

$$p\{4 \leq X < 8\} = F(8) - F(4) = 0.7 - 0.5 = 0.2.$$

Упражнения

1. Дискретная случайная величина X задана рядом распределения:

x_i	1	2	3	4
p_i	0,4	0,3	0,1	0,2

Найти функцию распределения $F(x)$ и построить график этой функции.

2. Дискретная случайная величина X задана рядом распределения:

x_i	-1	3	6	7	8
p_i	0,1	0,4	0,3	0,1	0,1

Чему равна вероятность $p\{3 < X \leq 7\}$.

3. Дискретная случайная величина X задана рядом распределения:

x_i	1	3	5	7	9
p_i	0,25	a	b	c	0,15

и вероятность $p\{1 \leq X \leq 5\} = 0,6$. Тогда чему равны значения a , b и c .

4. Для дискретной случайной величины X ,

x_i	2	3	4	5
p_i	p_1	p_2	p_3	p_4

функция распределения имеет вид:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 2, \\ 0,25 & \text{при } 2 < x \leq 3, \\ 0,40 & \text{при } 3 < x \leq 4, \\ 0,75 & \text{при } 4 < x \leq 5, \\ 1 & \text{при } x > 5. \end{cases}$$

Найти значения вероятностей p_1 , p_2 , p_3 и p_4 .

Для так называемой *равномерно распределенной* на отрезке $[a; b] \subset R^1$ случайной величины. функция распределения задается следующим кусочно-аналитическим выражением

$$F(x) = \begin{cases} 0 & \text{при } x < a, \\ \frac{x-a}{b-a} & \text{при } a \leq x \leq b, \\ 1 & \text{при } x > b. \end{cases}$$

Определение 9. Случайная величина называется *абсолютно непрерывной*, если ее функция распределения непрерывна на всей числовой оси.

Множеством возможных значений абсолютно непрерывной случайной величины, обычно является один единственный непрерывный интервал числовой оси, быть может, с бесконечными границами. Большинство встречающихся на практике случайные величины являются абсолютно непрерывными, поэтому в дальнейшем, говоря «непрерывная случайная величина», мы будем подразумевать абсолютно непрерывную случайную величину.

Достаточно очевидны следующие важные свойства функции распределения.

Свойства функции распределения:

- 1) $0 \leq F(x) \leq 1$, для любого $x \in R^1$;
- 2) $F(x)$ – неубывающая функция;

1) $F(x) \rightarrow 0$, при $x \rightarrow -\infty$; $F(x) \rightarrow 1$, при $x \rightarrow +\infty$.

3. Функция плотности распределения случайной величины

Следующее понятие является чрезвычайно важным.

Определение 10. Функция $f(x)$ называется *функцией плотности распределения* случайной величины X (или просто *плотностью*), если для любого $(a, b] \subset R^1$ выполняется

$$p\{x \in (a, b]\} = \int_a^b f(x) dx.$$

Ясно, что плотность тоже полностью задает случайную величину. Из сказанного выше следует

$$\int_a^b f(x) dx = p\{x \in (a, b]\} = F(b) - F(a),$$

т.е. функция распределения является первообразной для плотности, т.е.

$$F'(x) = f(x), \quad F(x) = \int_{-\infty}^x f(\xi) d\xi.$$

Таким образом, зная плотность всегда можно найти функцию распределения, и наоборот. Однако на практике для задания случайной величины почти всегда используется плотность. Это связано со следующим обстоятельством.

Вспомним геометрический смысл определенного интеграла (рис. 4), а именно, что он численно равен площади криволинейной трапеции, ограниченной соответствующим участком графика функции на интервале интегрирования и самим этим интервалом.

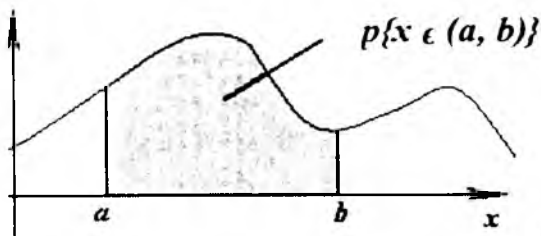


Рис. 4

Таким образом, на тех участках числовой оси, где больше значения функции плотности, там и больше площадь под ее графиком, а тем самым и вероятность появления реализаций данной случайной величины. Поэтому график плотности весьма хорошо иллюстрирует закон распределения данной случайной величины. Этого нельзя сказать о графике функции распределения.

Пример 9. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 2, \\ (x-2)^2 & \text{при } 2 < x \leq 3, \\ 1 & \text{при } x > 3. \end{cases}$$

Найти:

а) функцию плотности распределения $f(x)$;

б) графики функций $F(x)$ и $f(x)$;

в) по известной функции $F(x)$ и по найденной функции $f(x)$ вероятность того, что в результате испытания X примет значение, не меньшее 2,1 и меньшее 2,5.

Дать геометрическую интерпретацию величины найденной вероятности $P\{2,1 \leq X < 2,5\}$.

Решение

а) Плотность распределения вероятностей $f(x)$ равна производной от функции распределения $F(x)$, поэтому имеем $f(x) = ((x-2)^2)' = 2(x-2)$ при $2,1 \leq x < 2,5$ и $f(x) = 0$ при $x \leq 2$ и $x > 3$. Таким образом, функция плотности распределения характеризуется выражением:

$$f(x) = \begin{cases} 0 & \text{при } x \leq 2, \\ 2(x-2) & \text{при } 2 < x \leq 3, \\ 0 & \text{при } x > 3. \end{cases}$$

б) Графики функции $F(x)$ и $f(x)$ представлены ниже (рис. 5 и 6):

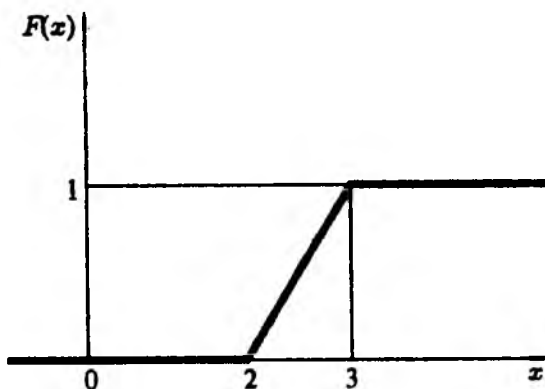


Рис. 5

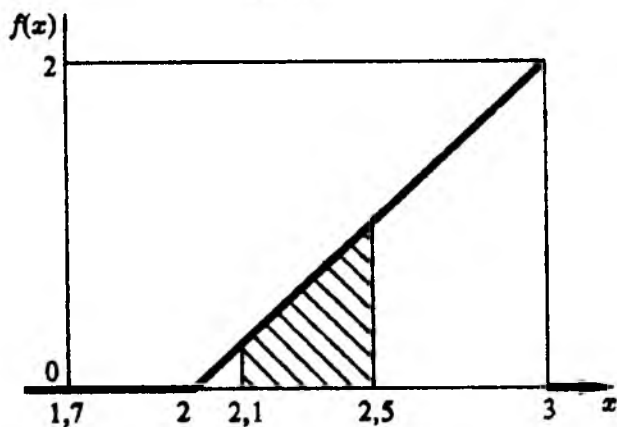


Рис. 6

в) Вероятность попадания случайной величины в интервал $[2,1; 2,5]$ определим как приращение функции распределения в этом интервале:

$$p\{2,1 < X < 2,5\} = F(2,5) - F(2,1) = (0,5)^2 - (0,1)^2 = 0,24.$$

Эту же вероятность найдем по известной функции плотности распределения:

$$p\{2,1 < X < 2,5\} = \int_{2,1}^{2,5} f(x) dx = 2 \int_{2,1}^{2,5} (x-2)^2 dx = (x-2)^2 \Big|_{2,1}^{2,5} = 0,24.$$

Полученную вероятность $p\{2,1 \leq X < 2,5\}$ можно интерпретировать геометрически как площадь заштрихованной криволинейной трапеции, приведенной на рис. 6.

Пример 10. Случайная величина X задана функцией плотности распределения

$$f(x) = \begin{cases} 0 & \text{при } x \leq \frac{\pi}{4}, \\ 2 \sin 2x & \text{при } \frac{\pi}{4} < x \leq \frac{\pi}{2}, \\ 0 & \text{при } x > \frac{\pi}{2}. \end{cases}$$

Найти функцию распределения $F(x)$.

Решение

Если $x \leq \frac{\pi}{4}$, то $f(x) = 0$, следовательно,

$$F(x) = \int_{-\infty}^x f(\xi) d\xi = \int_{-\infty}^x 0 \cdot d\xi = 0.$$

Если $\frac{\pi}{4} < x \leq \frac{\pi}{2}$, то

$$F(x) = \int_{-\infty}^x 0 \cdot d\xi = \int_{\frac{\pi}{4}}^x 2 \sin 2\xi d\xi = \cos 2\xi \Big|_{\frac{\pi}{4}}^x = -\cos 2x.$$

Если $x > \frac{\pi}{2}$, то

$$F(x) = \int_{-\infty}^x 0 \cdot d\xi = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} 2 \sin 2x dx + \int_{\frac{\pi}{2}}^x 0 \cdot d\xi = 0 - \cos 2x \Big|_{\frac{\pi}{4}}^{\frac{\pi}{2}} + 0 = 1.$$

Таким образом, функция распределения характеризуется выражением:

$$F(x) = \begin{cases} 0 & \text{при } x \leq \frac{\pi}{4}, \\ -\cos 2x & \text{при } \frac{\pi}{4} < x \leq \frac{\pi}{2}, \\ 1 & \text{при } x > \frac{\pi}{2}. \end{cases}$$

Ниже приведем функции плотности трех важных непрерывных случайных величин.

Упражнения

1. График функции плотности распределения $f(x)$ случайной величины X имеет вид, изображенный на рис. 7.

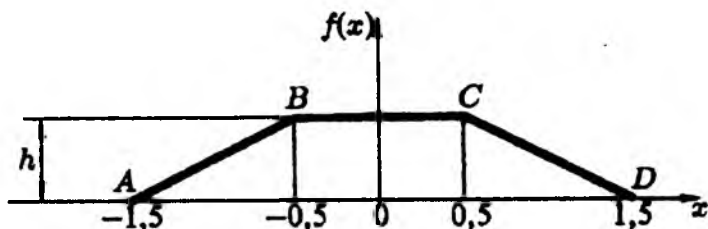


Рис. 7

Найти аналитическое выражение для $f(x)$ на всей числовой оси.

2. Случайная величина X подчинена закону Симпсона (закону равнобедренного треугольника) на отрезке $x \in [-c; c]$ (рис. 8). Найти: а) функцию плотности распределения вероятностей этой случайной величины; б) вероятность попадания величины X в интервал $(c/2; c)$.

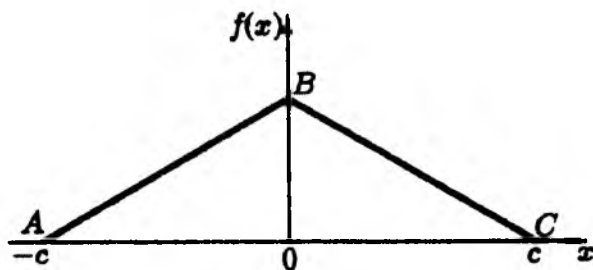


Рис. 8

Равномерно распределенная на отрезке $[a; b] \subset R^1$ случайная величина. Выражения для функции плотности распределения

$$f(x) = \begin{cases} 0 & \text{при } x < a, \\ \frac{1}{b-a} & \text{при } a \leq x \leq b, \\ 0 & \text{при } x > b. \end{cases}$$

Ее график имеет вид рис. 9.

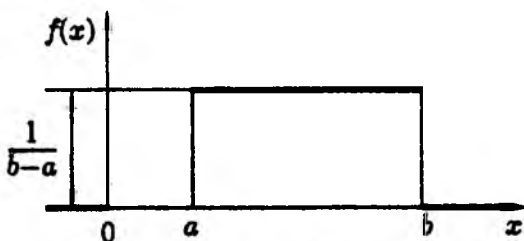


Рис. 9

А график ее функции распределения

$$F(x) = \begin{cases} 0 & \text{при } x < a, \\ \frac{x-a}{b-a} & \text{при } a \leq x \leq b, \\ 1 & \text{при } x > b \end{cases}$$

имеет вид рис.10.

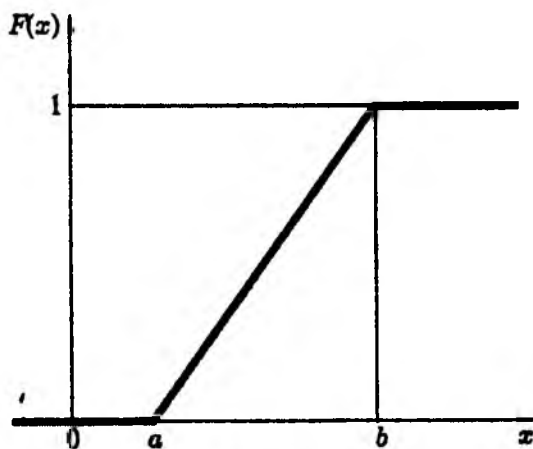


Рис. 10

Равномерный закон распределения используется при анализе ошибок округления при проведении числовых расчетов (например, ошибка округления числа до целого распределена равномерно на отрезке $[-0,5; +0,5]$), в ряде задач массового обслуживания, при статистическом моделировании наблюдений, подчиненных заданному распределению. Так, случайная величина X , распределенная по равномерному закону на отрезке $[0;1]$, называемая «случайным числом от 0 до 1», служит исходным материалом для получения случайных величин с любым законом распределения.

Пример 11. Поезда метрополитена идут регулярно с интервалом 2 мин. Пассажир выходит на платформу в случайный момент времени. Какова вероятность того, что ждать пассажиру придется не больше полминуты.

Решение

Случайная величина X – время ожидания поезда на временном (в минутах) отрезке $[0;2]$ имеет равномерный закон распределения

$$\varphi(x) = \begin{cases} 0 & \text{при } x < 0, \\ \frac{1}{2} & \text{при } 0 \leq x \leq 2, \\ 0 & \text{при } x > 2. \end{cases}$$

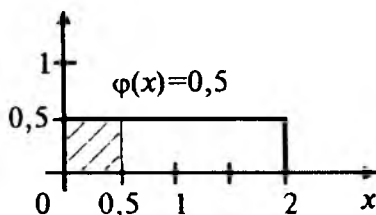


Рис. 11

Поэтому вероятность того, что пассажиру придется ждать не более полминуты, равна $1/4$ от равной единице площади прямоугольника (рис. 11), т.е.

$$P\{X \leq 0,5\} = \int_0^{0,5} \frac{1}{2} dx = \frac{1}{2} x \Big|_0^{0,5} = \frac{1}{4}.$$

Нормальное распределение, называемое еще *гауссовским* распределением, играет важнейшую роль в теории вероятностей, математической статистике и на практике. Главная особенность, выделяющая его среди других законов, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях (см. §6).

Говорят, что случайная величина X имеет *нормальный закон распределения*, если ее плотность задается выражением

$$\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где $\sigma > 0$, a - заданные числа, параметры этого распределения. Графики функции плотности (рис. 12а) и функции распределения нормального распределения (рис. 12б) с параметрами σ^2 и a , т.е. $N(a; \sigma^2)$ приведены ниже.

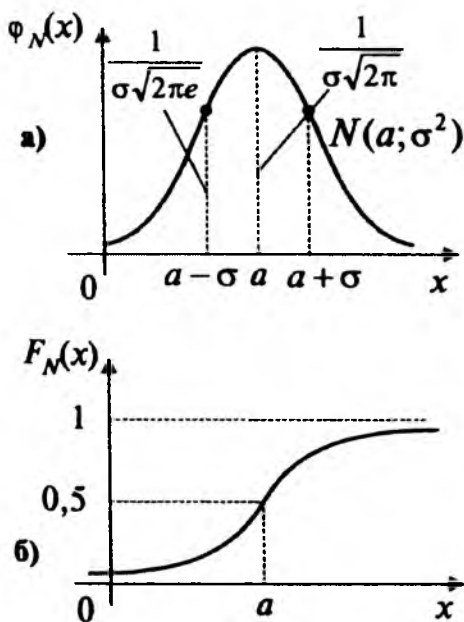


Рис. 12

Говорят, что случайная величина X имеет *показательный (экспоненциальный) закон распределения*, если ее плотность задается выражением

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0. \end{cases}$$

где λ - некоторый параметр.

Функция распределения $F(x)$ показательной случайной величины X имеет вид

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{при } x > 0, \\ 0 & \text{при } x < 0. \end{cases}$$

Графики функции плотности (рис. 13а) и функции распределения нормального распределения (рис. 13б) приведены ниже.

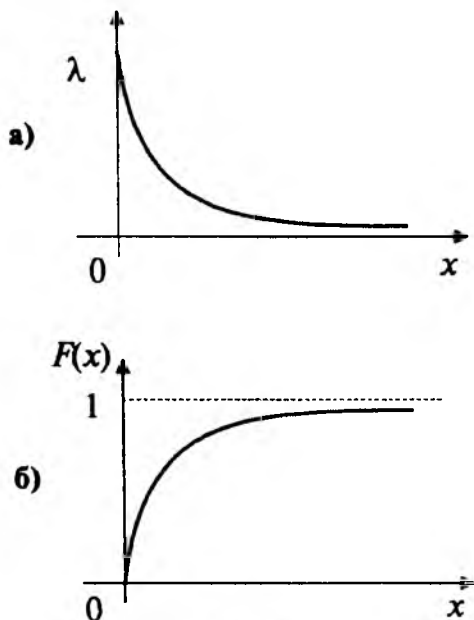


Рис. 13

Показательный закон распределения играет большую роль в теории массового обслуживания и теории надежности. Так, например, интервал времени T между двумя соседними событиями в простейшем потоке имеет показательное распределение с параметром λ - интенсивностью потока.

Свойства функции плотности распределения:

1) $f(x) \geq 0$ во всех точках $x \in R^1$, где существует $F'(x)$;

$$2) \int_{-\infty}^{+\infty} f(x) dx = 1;$$

3) $f(x) \rightarrow 0$ при $x \rightarrow \pm\infty$.

Пример 12. Случайная величина X распределена по показательному закону, заданному плотностью распределения вероятностей $f(x) = 0,5e^{-0,5x}$ при $x \geq 0$; при $x < 0$ функция $f(x) = 0$. Найти вероятность того, что в результате испытания X примет значение из интервала (2; 4).

Решение

Искомая вероятность того, что X примет значение из интервала (2; 4), равна

$$P\{2 < X < 4\} = \int_2^4 0,5e^{-0,5x} dx = -e^{-0,5x} \Big|_2^4 = e^{-0,5 \cdot 2} - e^{-0,5 \cdot 4} = 0,3678 - 0,1353 = 0,2325.$$

Пример 13. Случайная величина X , все возможные значения которой принадлежат интервалу $(0; \pi/3)$, задана в этом интервале функцией плотности распределения $f(x) = C \sin 3x$. Найти коэффициент C .

Решение

Если функция $f(x)$ представляет собой функцию распределения непрерывной случайной величины X , заданной в интервале $(a; b)$, то выполняется условие: $\int_a^b f(x) dx = 1$. Неизвестный коэффициент C в выражении функции $f(x) = C \sin 3x$ определим так, чтобы это условие выполнялось для данной случайной величины X . Для этого найдем

$$\int_0^{\frac{\pi}{3}} f(x) dx = \int_0^{\frac{\pi}{3}} C \cdot \sin 3x dx = -\frac{C}{3} (\cos \pi - \cos 0) = -\frac{C}{3} (-1 - 1) = \frac{2C}{3}$$

и приравняем этот результат к единице: $\frac{2C}{3} = 1$. Из последнего равенства получим $C = \frac{3}{2}$.

Упражнение

Дана функция $y = \frac{A}{x^3}$. Найти значение постоянного множителя A , при котором эта функция могла бы характеризовать плотность распределения вероятностей случайной величины X при условии, что все возможные значения величины X находятся на луче $(2; +\infty)$.

Вопросы для самопроверки

1. Что называется случайной величиной?
2. Приведите примеры случайных величин.
3. Какие бывают случайные величины?
4. Что представляют возможные значения: а) дискретной случайной величины? б) непрерывной случайной величины?
5. Что называется законом распределения случайной величины?
6. Что называется рядом распределения дискретной случайной величины?
7. Что называется функцией распределения непрерывной случайной величины? Приведите свойства функции распределения.
8. Что называется функцией плотности распределения непрерывной случайной величины? Приведите свойства функции плотности распределения.
9. Что представляет собой биномиальное распределение?
10. Что представляет собой пуассоновское распределение?
11. Что называется равномерным законом распределения?
12. Что называется нормальным законом распределения?
13. Что называется показательным (экспоненциальным) законом распределения?

§5. Числовые характеристики случайных величин

Ключевые слова: математическое ожидание, дисперсия, среднеквадратическое отклонение, ковариация, коэффициент корреляции, моменты случайных величин, начальный момент k -го порядка, относительный момент k -го порядка, центральный момент k -го порядка, коэффициент асимметрии, эксцесс (коэффициент эксцесса), начальный момент порядка k, l , относительный момент порядка k, l , - центральный момент порядка k, l .

Из предыдущего параграфа известно, что исчерпывающей характеристикой случайных величин является их закон распределения: ряд распределения для дискретных случайных величин; функция распределения или функция плотности распределения для непрерывных случайных величин. Но далеко не в каждой задаче нужно знать весь закон распределения. В ряде случаев можно обойтись одним или несколькими числами, отражающими наиболее важные особенности закона распределения: например, числом, имеющим смысл «среднего значения» случайной величины, или же числом, характеризующим средний размер отклонения случайной величины от своего среднего значения, и т.д. Такого рода числа называют *числовыми характеристиками случайной величины* (или соответствующего закона распределения). Их роль в теории вероятностей чрезвычайно велика; многие задачи удается решить до конца, оставляя в стороне законы распределения и оперируя только числовыми характеристиками.

Определение 1. Любая числовая величина, так или иначе характеризующая закон распределения некоторой случайной величины, называется *параметром* этого распределения.

Из бесконечного количества всевозможных параметров, важнейшее значение имеют два: *математическое ожидание* и *дисперсия* случайной величины.

1. Математическое ожидание и дисперсия случайной величины

Определение 2. Математическим ожиданием дискретной случайной величины X , имеющей ряд распределения

Таблица 1

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

называется число

$$M(X) = \sum_{i=1}^n x_i p_i.$$

Математическое ожидание дискретной случайной величины X имеет следующий вероятностный смысл.

Пусть проведено N испытаний случайной величины X , в результате чего получены значения $x_1, x_2, \dots, x_N$. Среднее арифметическое этих чисел

$$\frac{x_1 + x_2 + \dots + x_N}{N}$$

при больших n близко к $M(X)$. Строгую формулировку этого утверждения приведем в §6 (теорема Чебышева).

Здесь мы приведем нестрогое рассуждение, поясняющее причины такого явления. Если величина X дискретна и ее ряд распределения имеет вид табл.1, то в результате N испытаний мы получим $p_1 N$ раз значение x_1 ; $p_2 N$ раз - значение x_2 ; ...; $p_n N$ раз - значение x_n . Действительно, если вероятность события A равна p , то в N испытаниях событие A наступит примерно Np раз (согласно статистическому определению вероятности (см. §1)). После N испытаний сумма значений будет приближенно равна

$$x_1 p_1 N + x_2 p_2 N + \dots + x_n p_n N.$$

Среднее арифметическое полученных в результате испытаний значений равно

$$\frac{x_1 p_1 N + x_2 p_2 N + \dots + x_n p_n N}{N} = \sum_{i=1}^n x_i p_i,$$

что совпадает с $M(X)$.

В связи с этим математическое ожидание называют также средним значением случайной величины.

Пример 1. Пусть X — количество очков при бросании игральной кости. Ряд распределения этой величины имеет вид

x_i	1	2	3	4	5	6
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Найдем математическое ожидание

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Пример 2. По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,515. Составить закон распределения случайной величины X — числа мальчиков в семье из 4 детей. Найти математическое ожидание этой случайной величины.

Решение

Число мальчиков в семье из $n = 4$ представляет случайную величину X со множеством значений $X = m = 0, 1, 2, 3, 4$ вероятности, которые определяются по формуле Бернулли:

$$p\{X = m\} = C_n^m p^m q^{n-m}, \text{ где } q = 1 - p.$$

В нашем случае $n = 4$, $p = 0,515$, $q = 1 - p = 1 - 0,515 = 0,485$.

Вычислим

$$p\{X = 0\} = C_4^0 0,515^0 \cdot 0,485^4 = 0,055;$$

$$p\{X = 1\} = C_4^1 0,515^1 \cdot 0,485^3 = 0,235;$$

$$p\{X = 2\} = C_4^2 0,515^2 \cdot 0,485^2 = 0,375;$$

$$p\{X = 3\} = C_4^3 0,515^3 \cdot 0,485^1 = 0,265;$$

$$p\{X = 4\} = C_4^4 0,515^4 \cdot 0,485^0 = 0,070.$$

Ряд распределения имеет вид

x_i	0	1	2	3	4
p_i	0.055	0.235	0.375	0.265	0.070

Убеждаемся, что $\sum_{i=1}^5 p_i = 0,055 + 0,235 + 0,375 + 0,265 + 0,070 = 1$.

Найдем математическое ожидание

$$M(X) = 0 \cdot 0,055 + 1 \cdot 0,235 + 2 \cdot 0,375 + 3 \cdot 0,265 + 4 \cdot 0,070 = 2,06.$$

Математическое ожидание дискретной случайной величины X , имеющей счетное множество значений, равно

$$M(X) = \sum_{i=1}^{\infty} x_i p_i,$$

причем математическое ожидание существует, если ряд в правой части равенства сходится абсолютно.

Так как этот ряд может и расходиться, то соответствующая случайная величина может и не иметь математического ожидания. Например, случайная величина X с рядом распределения

x_i	2	2^2	2^3	...	2^i	...
p_i	$1/2$	$1/2^2$	$1/2^3$	...	$1/2^i$	...

не имеет математического ожидания, ибо сумма ряда равна ∞ . На практике, как правило, множество возможных значений случайной величины распространяется лишь на ограниченный участок оси абсцисс и, значит, математическое ожидание существует.

Пример 3. После ответа студента на вопросы экзаменационного билета экзаменатор задает студенту дополнительные вопросы. Преподаватель прекращает задавать дополнительные вопросы, как только студент обнаруживает незнание заданного вопроса. Вероятность того, что студент ответит на любой заданный дополнительный вопрос равна 0,9. Требуется:

а) составить ряд распределения случайной дискретной величины X - числа дополнительных вопросов, которые задаст преподаватель студенту;

б) найти математическое ожидание этой случайной величины.

Решение

а) Дискретная случайная величина X - число заданных дополнительных вопросов - имеет следующие возможные значения: $1, 2, 3, \dots, k, \dots$. Найдем вероятности этих возможных значений.

Величина X примет возможное значение 1 (экзаменатор задаст только один вопрос), если студент не ответит на первый вопрос. Вероятность этого возможного значения равна $1 - 0,9 = 0,1$. Таким образом,

$$p\{X = 1\} = 0,1.$$

Величина X примет возможное значение 2 (экзаменатор задаст только два вопроса), если студент ответит на первый вопрос (вероятность этого события равна 0,9) и не ответит на второй (вероятность этого события равна 0,1). Таким образом,

$$p\{X = 2\} = 0,9 \cdot 0,1 = 0,09.$$

Аналогично найдем

$$p\{X = 3\} = 0,9^2 \cdot 0,1 = 0,081; \quad p\{X = k\} = 0,9^{k-1} \cdot 0,1; \dots$$

Напишем искомый закон распределения:

x_k	1	2	3	...	k	...
p_k	0,1	0,09	0,081	...	$0,9^{k-1} \cdot 0,1$	...

Проверяем, что

$$\begin{aligned} \sum_{i=1}^{\infty} p_i &= 0,1 + 0,1 \cdot 0,9 + 0,1 \cdot 0,9^2 + \dots + 0,1 \cdot 0,9^{k-1} + \dots = \\ &= 0,1(1 + 0,9 + 0,9^2 + \dots) = 0,1 \cdot \frac{1}{1 - 0,9} = \frac{0,1}{0,1} = 1. \end{aligned}$$

Здесь использовали формулу суммы сходящегося ($|q| < 1$) геометрического ряда: $S = \frac{u}{1 - q}$ при $u = 1$, $q = 0,9$.

Вычислим математическое ожидание

$$M(X) = \sum_{i=1}^{\infty} x_i p_i = 1 \cdot 0,1 + 2 \cdot 0,09 + 3 \cdot 0,081 + \dots + k \cdot 0,1 \cdot 0,9^{k-1} + \dots = \\ = 0,1(1 + 2 \cdot 0,9 + 3 \cdot 0,9^2 + \dots + k \cdot 0,9^{k-1} + \dots).$$

Для вычисления суммы полученного ряда воспользуемся формулой:

$$S(x) = 1 + 2x + 3x^2 + \dots + kx^{k-1} + \dots = (x + x^2 + \dots + x^k + \dots)' = \left(\frac{x}{1-x^2} \right)' = \frac{1}{(1-x)^2},$$

(т.е. сумма данного ряда является производной сходящегося геометрического ряда со знаменателем $|q| = |x| < 1$). При $x = 0,9$ имеем: $S(0,9) = \frac{1}{(1-0,9)^2} = 100$.

Тогда $M(X) = 0,1 \cdot 100 = 10$.

Упражнения

1. Торговый агент имеет 5 телефонных номеров потенциальных покупателей и звонит им до тех пор, пока не получит заказ на покупку товара. Вероятность того, что потенциальный покупатель сделает заказ, равна 0,4. Составить ряд распределения числа телефонных разговоров, которые предстоит провести агенту. Найти математическое ожидание этой случайной величины.

2. Охотник, имеющий 4 патрона, стреляет по дичи до первого попадания или до израсходования всех патронов. Вероятность попадания при первом выстреле равна 0,6, при каждом последующем - уменьшается на 0,1. Необходимо: а) составить ряд распределения числа патронов, израсходованных охотником; б) найти математическое ожидание этой случайной величины.

3. Из поступивших в ремонт 10 часов 7 нуждаются в общей чистке механизма. Часы не рассортированы по виду ремонта. Мастер, желая найти часы, нуждающиеся в чистке, рассматривает их поочередно и, найдя такие часы, прекращает дальнейший просмотр. Составить ряд распределения числа просмотренных часов. Найти математическое ожидание этой случайной величины.

4. Имеются 4 ключа, из которых только один подходит к замку. Составить ряд распределения числа попыток открывания замка, если испробованный ключ в последующих попытках не участвует. Найти математическое ожидание этой случайной величины.

5. Радист вызывает корреспондента, причем каждый последующий вызов производится лишь в том случае, если предыдущий вызов не принят. Вероятность того, что корреспондент примет вызов, равна 0,4. Составить ряд распределения числа вызовов, если: а) число вызовов не более 5; б) число вызовов не ограничено. Найти математическое ожидание этой случайной величины.

Определение 3. Пусть X - непрерывная случайная величина и $f(x)$ - ее функция плотности распределения. Математическим ожиданием непрерывной случайной величины X называется число

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

(если этот интеграл сходится).

Математическое ожидание непрерывной случайной величины имеет такой же вероятностный смысл, что и математическое ожидание дискретной случайной величины.

Пример 4. Функция $f(x)$ задана в виде:

$$f(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ \frac{A}{x^4} & \text{при } x > 1. \end{cases}$$

Найти: а) значение постоянной A , при которой $f(x)$ будет функцией плотности некоторой случайной величины X ; б) математическое ожидание случайной величины X .

Решение

а) Для того чтобы $f(x)$ была функцией плотности случайной величины

λ , она должна быть неотрицательна, т.е. $f(x) \geq 0$ или $\frac{\lambda}{x^4} \geq 0$, откуда $\lambda \geq 0$

для нее $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Следовательно,

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x) dx &= \int_{-\infty}^1 0 \cdot dx + \int_1^{+\infty} \frac{\lambda}{x^4} dx = 0 + \lim_{b \rightarrow +\infty} \int_1^b \frac{\lambda}{x^4} dx = \frac{\lambda}{3} \lim_{b \rightarrow +\infty} \left(-\frac{1}{x^3} \Big|_1^b \right) = \\ &= \frac{\lambda}{3} \lim_{b \rightarrow +\infty} \left(-\frac{1}{x^3} \Big|_1^b \right) = \frac{\lambda}{3} \lim_{b \rightarrow +\infty} \left(1 - \frac{1}{b^3} \right) = \frac{\lambda}{3} = 1.\end{aligned}$$

откуда $\lambda = 3$.

б) По формуле $M(X) = \int_{-\infty}^{+\infty} xf(x) dx$ вычислим математическое ожидание

$$\begin{aligned}M(X) &= \int_{-\infty}^{+\infty} xf(x) dx = \int_{-\infty}^1 x \cdot 0 \cdot dx + \int_1^{+\infty} x \frac{3}{x^4} dx = 0 + 3 \lim_{b \rightarrow +\infty} \int_1^b \frac{dx}{x^3} = \\ &= 3 \lim_{b \rightarrow +\infty} \left(-\frac{1}{2x^2} \Big|_1^b \right) = \frac{3}{2} \lim_{b \rightarrow +\infty} \left(1 - \frac{1}{b^2} \right) = \frac{3}{2}.\end{aligned}$$

Пример 5. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{x^2}{4} & \text{при } 0 < x \leq 2, \\ 1 & \text{при } x > 2. \end{cases}$$

Найти математическое ожидание случайной величины X .

Решение

Сначала найдем функцию плотности

$$f(x) = F'(x) = \begin{cases} 0 & \text{при } x \leq 0, x > 2, \\ \frac{x}{2} & \text{при } 0 < x \leq 2. \end{cases}$$

Теперь по формуле $M(X) = \int_{-\infty}^{+\infty} xf(x) dx$ вычислим математическое ожи-

дание

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^0 x \cdot 0 \cdot dx + \int_0^2 x \frac{x}{2} dx + \int_2^{+\infty} x \cdot 0 \cdot dx = 0 + \frac{x^3}{6} \Big|_0^2 + 0 = \frac{2^3}{6} = \frac{4}{3}.$$

Упражнения

1. Дана функция

$$f(x) = \begin{cases} 0 & \text{при } x < 0, \\ Cxe^{-x} & \text{при } x \geq 0. \end{cases}$$

При каком значении параметра C эта функция является функцией плотности распределения некоторой непрерывной случайной величины X .

Найти математическое ожидание и дисперсию случайной величины X .

2. Случайная величина X распределена по закону Коши:

$$f(x) = \frac{A}{1+x^2}.$$

Найти: а) коэффициент A ; б) математическое ожидание случайной величины X .

3. Случайная величина X распределена по закону Лапласа:

$$f(x) = Ae^{-\lambda|x|}.$$

Найти: а) коэффициент A ; б) математическое ожидание случайной величины X .

Свойства математического ожидания

1) $M(\lambda) = \lambda$, λ – любая константа (постоянное число формально можно рассматривать как случайную величину, принимающую единственное значение λ – с вероятностью единица);

2) $M(\lambda X) = \lambda M(X)$ для любой случайной величины X и произвольного числа λ .

3) $M(X+Y) = M(X) + M(Y)$ для произвольных случайных величин X, Y ;

4) $M(XY) = M(X)M(Y)$ для произвольных независимых случайных величин X и Y (см. п. 2.)

Доказательство этих свойств опускаем.

Отметим еще раз, что математическое ожидание является (постоянным, не зависящим от опыта) числом, характеризующим определенное свойство случайной величины, а именно - устойчивость среднего арифметического полученного в результате испытаний значений.

Эта характеристика является важной, но далеко не полной.

Следующее понятие также сопоставляет случайной величине некоторое число, характеризующее определенное свойство этой величины.

Определение 4. *Дисперсией* случайной величины X называется число

$$D(X) = M(X - M(X))^2,$$

т.е. математическое ожидание случайного квадрата отклонения значений случайной величины от ее математического ожидания.

Таким образом, дисперсия характеризует величину рассеяния или разброса возможных значений случайной величины вокруг ее среднего значения. Ясно, что это тоже очень важный параметр.

Заметим, что в качестве характеристики рассеяния нельзя брать математическое ожидание отклонения случайной величины от ее математического ожидания $M(X - a)$ (где $a = M(X)$), ибо согласно свойствам математического ожидания эта величина равна нулю для любой случайной величины.

Выбор дисперсии, определяемой по формуле $D(X) = M(X - M(X))^2$ в качестве характеристики рассеяния значений случайной величины X оправдывается также тем, что математическое ожидание квадрата отклонения случайной величины X от постоянной величины C минимально именно тогда, когда эта постоянная C равна математическому ожиданию $M(X) = a$.

т.е.

$$\min_C M(X - C)^2 = M(X - a)^2 = D(X).$$

Используя свойства математического ожидания, легко доказать следующую, более удобную для практических расчетов, формулу дисперсии

$$D(X) = M(X^2 - 2X \cdot M(X) + [M(X)]^2) = M(X^2) - 2[M(X)]^2 + [M(X)]^2 =$$

$$= M(X^2) - [M(X)]^2.$$

т.е.

$$M(X^2) = \sum_i x_i^2 p_i - \text{ для дискретной случайной величины;}$$

$$M(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \text{ для непрерывной случайной величины.}$$

Таким образом, справедливы следующие формулы, упрощающие вычисление дисперсии:

$$D(X) = \sum_i x_i^2 p_i - \left[\sum_i x_i p_i \right]^2 - \text{ для дискретной случайной величины;}$$

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left[\int_{-\infty}^{+\infty} x f(x) dx \right]^2 - \text{ для непрерывной случайной ве-}$$

личины.

Дисперсия, по сути, является квадратическим показателем. Иногда более удобно использовать аналогичный линейный параметр, называемый *среднеквадратическим отклонением* случайной величины

$$\sigma(X) = \sqrt{D(X)}.$$

Пример 6. Пусть ряд распределения величины X имеет вид

x_i	-1	5
p_i	0,3	0,7

Найти дисперсию и среднеквадратическое отклонение этой случайной величины.

Решение

Найдем

$$M(X) = (-1) \cdot 0,3 + 5 \cdot 0,7 = 3,2;$$

$$M(X^2) = (-1)^2 \cdot 0,3 + 5^2 \cdot 0,7 = 17,8.$$

Тогда

$$D(X) = M(X^2) - [M(X)]^2 = 17,8 - (3,2)^2 = 7,56;$$

$$\sigma = \sqrt{D(X)} = \sqrt{7,56} \approx 2,7495.$$

Пример 7. Найти дисперсию и среднеквадратическое отклонение случайной величины X из примера 4.

Решение

Вначале найдем

$$M(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 \left(\frac{3}{x^4} \right) dx = 3.$$

Вычисление интеграла производим аналогично вычислению математического ожидания в примере 4. Теперь учитывая, что $M(X) = \frac{3}{2}$ (см. пример 4) имеем:

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left[\int_{-\infty}^{+\infty} x f(x) dx \right]^2 = 3 - \left(\frac{3}{2} \right)^2 = \frac{3}{4};$$

$$\sigma = \sqrt{D(X)} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}.$$

Упражнения

1. Дискретная случайная величина X задана рядом распределения

x_i	-1	0	3
p_i	0,4	0,4	0,2

Найти дисперсию и среднеквадратическое отклонение этой случайной величины.

2. Дисперсия дискретной случайной величины X , заданной рядом распределения, равна 0,06.

x_i	1	x_2
p_i	0.4	0.6

Найти значение $x_2 > 0$.

3. Непрерывная случайная величина X задана функцией плотности

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{3x^2}{64} & \text{при } 0 < x \leq 4, \\ 0 & \text{при } x > 4. \end{cases}$$

Найти дисперсию и среднеквадратическое отклонение этой случайной величины.

4. Непрерывная случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{x}{7} & \text{при } 0 < x \leq 7, \\ 1 & \text{при } x > 7. \end{cases}$$

Найти дисперсию и среднеквадратическое отклонение этой случайной величины.

5. Непрерывная случайная величина X задана функцией распределения вероятностей:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{x^2}{36} & \text{при } 0 < x \leq 6, \\ 1 & \text{при } x > 6. \end{cases}$$

Найти дисперсию и среднеквадратическое отклонение этой случайной величины.

Приведем свойства дисперсии и среднеквадратического отклонения.

Свойства дисперсии

1) $D(X) \geq 0$, где X – любая случайная величина;

2) $D(\lambda) = 0$, где λ – любая константа (постоянное число формально можно рассматривать как случайную величину, принимающую единственное значение λ – с вероятностью единица);

3) $D(\lambda X) = \lambda^2 D(X)$ для любой случайной величины X и произвольного числа λ .

4) $D(X+Y) = D(X) + D(Y)$ для независимых случайных величин X и Y (см. п. 2.).

Доказательство этих свойств опускаем.

Свойства среднеквадратического отклонения

1) $\sigma(\lambda) = 0$, т.е. среднеквадратическое отклонение постоянной равно нулю.

2) $\sigma(\lambda X) = |\lambda| \sigma(X)$ для любой случайной величины X и произвольного числа λ .

3) $\sigma(X+Y) = \sqrt{[\sigma(X)]^2 + [\sigma(Y)]^2}$ для независимых случайных величин X и Y (см. п. 2.).

Эти свойства непосредственно вытекают из свойств дисперсии и определения среднеквадратического отклонения.

2. Независимость случайных величин. Математические операции над случайными величинами

Определение 5. Две случайные величины называются *независимыми*, если закон распределения одной из них не меняется от того, какие возможные значения приняла другая величина.

Так, если дискретная случайная величина X может принимать значения x_i , $i = 1, 2, \dots, n$, а случайная величина Y - значения y_j , $j = 1, 2, \dots, n$, то независимость дискретных случайных величин X и Y означает независимость событий $\{X = x_i\}$ и $\{Y = y_j\}$ при любых $i = 1, 2, \dots, n$ и $j = 1, 2, \dots, n$. В противном случае случайные величины называются *зависимыми*.

Например, если имеются билеты двух различных денежных лотерей, то случайные величины X и Y , выражающие, соответственно, выигрыш по каждому билету (в денежных единицах), будут независимыми, так как при любом выигрыше по билету одной лотереи (например, при $X = x_i$) закон распределения выигрыша по другому билету (Y) не изменится. Если же случай-

ные величины X и Y выражают выигрыш по билетам одной денежной лотереи, то в этом случае X и Y являются зависимыми, ибо любой выигрыш по одному билету $\{X = x_i\}$ приводит к изменению вероятностей выигрыша по другому билету (Y), т.е. к изменению закона распределения Y .

Определим *математические операции* над дискретными случайными величинами.

Пусть даны две случайные величины - X и Y :

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

y_j	y_1	y_2	...	y_n
p_j	p_1	p_2	...	p_m

Произведением kX случайной величины X на постоянную величину k называется случайная величина, которая принимает значения kx_i с теми же вероятностями p_i .

m -й степенью случайной величины X , т.е. X^m , называется случайная величина, которая принимает значения x_i^m с теми же вероятностями p_i .

Пример 8. Дана случайная величина X :

x_i	-2	1	2
p_i	0,5	0,3	0,2

Найти закон распределения случайных величин: а) $Y = 3X$; б) $Z = X^2$.

Решение

а) Значения случайной величины Y будут: $3(-2) = -6$; $3 \cdot 1 = 3$; $3 \cdot 2 = 6$ с теми же вероятностями 0,5; 0,3; 0,2, т.е.

y_i	-6	3	6
p_i	0,5	0,3	0,2

б) Значения случайной величины Z будут: $(-2)^2 = 4$, $1^2 = 1$, $2^2 = 4$ с теми же вероятностями 0,5; 0,3; 0,2. Так как значение $Z = 4$ может быть получено возведением в квадрат значений (-2) с вероятностью 0,5 и $(+2)$ с вероятностью 0,2, то по теореме сложения $p\{Z = 4\} = 0,5 + 0,2 = 0,7$. Итак, закон распределения случайной величины Z :

z_i	1	4
p_i	0,3	0,7

Суммой (разностью или произведением) случайных величин X и Y называется случайная величина, которая принимает все возможные значения вида $x_i + y_j$ ($x_i - y_j$ или $x_i \cdot y_j$), где $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$, с вероятностями p_{ij} того, что случайная величина X примет значение x_i , а Y - значение y_j :

$$p_{ij} = p\left[\{X = x_i\} \{Y = y_j\}\right].$$

Если случайные величины X и Y независимы, т.е. независимы любые события $\{X = x_i\}$, $\{Y = y_j\}$, то по теореме умножения вероятностей для независимых событий

$$p_{ij} = p\{X = x_i\} \cdot p\{Y = y_j\} = p_i \cdot p_j.$$

Замечание. Приведенные выше определения операций над дискретными случайными величинами нуждаются в уточнении, так как в ряде случаев одни и те же значения x_i^m , $x_i \pm y_j$, $x_i y_j$ могут получаться разными способами при различных значениях x_i , y_j , вообще говоря, с различными вероятностями p_i, p_j (см. примеры 8б и 9).

Пример 9. Даны законы (ряды) распределения двух независимых случайных величин X и Y :

x_i	0	2	4
p_i	0.5	0.2	0.3

y_i	-2	0	2
p_i	0,1	0.6	0.2

Найти закон распределения случайных величин: а) $Z = X - Y$; б) $U = XY$.

Решение

а) Для удобства нахождения всех значений разности $Z = X - Y$ и их вероятностей составим вспомогательную таблицу, в каждой клетке которой поместим в левом углу значения разности $Z = X - Y$, а в правом углу - вероятности этих значений, полученные в результате перемножения вероятностей соответствующих значений случайных величин X и Y .

	y_j	-2	0	2
	p_j	0,1	0,6	0,3
x_i	p_i			
0	0.5	2 0,05	0 0,30	-2 0,15
2	0,2	4 0.02	2 0.12	0 0.06
4	0.3	6 0,03	4 0,18	2 0.09

Например, если $X = 4$ (последняя строка таблицы), а $Y = -2$ (третий столбец таблицы), то случайная величина $Z = X - Y$ принимает значение

$$Z = 4 - (-2) = 6$$

с вероятностью

$$p\{Z = 6\} = p\{X = 4\} p\{Y = -2\} = 0,3 \cdot 0,1 = 0,03.$$

Эти числа $Z = 6$ и $p = 0.03$ находятся в клетке на пересечении последней строки и третьего столбца.

Так как среди 9 значений Z имеются повторяющиеся, то соответствующие вероятности их складываем по теореме сложения вероятностей. Например, значение $Z = X - Y = 2$ может быть получено, когда $X = 0$, $Y = -2$ (с вероятностью 0,05); $X = 2$, $Y = 0$ (с вероятностью 0,12); $X = 4$, $Y = 2$ (с вероятностью 0,09), поэтому

$$p\{Z = 2\} = 0.05 + 0.12 + 0.09 = 0.26 \text{ и т.д.}$$

В результате получим ряд распределения для Z :

z_k	-2	0	2	4	6
p_k	0.15	0.36	0.26	0.20	0.03

Убеждаемся в том, что условие $\sum_{i=1}^5 p_i = 1$ выполнено.

б) Ряд распределения для $U = XY$ находится аналогично п. а).

u_k	-8	-4	0	4	8
p_k	0.03	0.02	0.80	0.06	0.09

Замечание. Выше ввели понятие независимости случайных величин X и Y , основанное на независимости связанных с ними событий $\{X = x_i\}$ и $\{Y = y_j\}$ при любых i и j . Ниже можно дать общее определение независимых непрерывных случайных величин, основанное на независимости событий $\{X < x\}$ и $\{Y < y\}$. Напомним, что необходимость введения события такого рода мы обсудили в §4.

Определение 6. Непрерывные величины X и Y независимы, если независимы события $\{X < x\}$ и $\{Y < y\}$, где x и y две любые действительные числа.

Иначе говоря, величины X и Y независимы, если при любых x и y справедливо равенство

$$p\{X < x, Y < y\} = p\{X < x\} \cdot p\{Y < y\}, \quad (1)$$

или в эквивалентной записи

$$F(x, y) = F_1(x)F_2(y), \quad (2)$$

где $F(x, y)$, $F_1(x)$ и $F_2(y)$, обозначают, соответственно, функции распределения для системы (X, Y) , величины X и величины Y .

Из соотношения (2), очевидно, следует

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = [F(x_2) - F(x_1)] \cdot [F(y_2) - F(y_1)],$$

т. е.

$$p\{x_1 \leq X < x_2, y_1 \leq Y < y_2\} = p\{x_1 \leq X < x_2\} p\{y_1 \leq Y < y_2\}. \quad (3)$$

Это значит, что независимыми являются не только события $\{X < x\}$ и $\{Y < y\}$, но и любые два события вида $\{x_1 \leq X < x_2\}$, $\{y_1 \leq Y < y_2\}$.

Формула (3) вскрывает смысл определения независимости величин X и Y : вероятность того, что величина Y примет значение из того или иного промежутка, не зависит от того, в каком промежутке окажется величина X . На более простом языке это означает:

Величины X и Y *принимают свои значения независимо друг от друга*.

В случае, когда система (X, Y) - дискретного типа, условие независимости величин X и Y можно представить в более обозримом виде:

$$p\{X = \alpha, Y = \beta\} = p\{X = \alpha\} p\{Y = \beta\}, \quad (4)$$

где α - любое возможное значение случайной величины X , а β - любое возможное значение величины Y .

Доказательство эквивалентности (4) и (3) (для системы дискретного типа) предоставляем провести в качестве упражнения.

Иллюстрацией независимости случайных величин может служить следующий пример.

Пример 10. Пусть в опыте с двукратным бросанием игральной кости величина X есть число очков, выпадающих при первом бросании, а Y - число очков при втором бросании. В этом случае равенство (4) выполняется для любых двух чисел α, β из множества $\{1, 2, 3, 4, 5, 6\}$, и, следовательно, величины X и Y независимы.

Вернемся снова к общей теории.

Допустим, что система (X, Y) имеет некоторую функцию плотности $f(x, y)$. Постараемся для этого случая привести условие независимости X и Y к более простому виду.

Функцию плотности для X обозначим $f_1(x)$, а функцию плотности для Y - $f_2(y)$. Тогда условие независимости X и Y - условие (3) - перепишется в виде:

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dx dy = \int_{x_1}^{x_2} f_1(x) dx \int_{y_1}^{y_2} f_2(y) dy$$

или

$$\iint_G f(x, y) dx dy = \iint_G f_1(x) f_2(y) dx dy. \quad (5)$$

где G обозначает прямоугольник, определенный неравенствами

$$x_1 \leq X < x_2, \quad y_1 \leq Y < y_2.$$

Таким образом, для любого прямоугольника G со сторонами, параллельными координатным осям, выполняется соотношение (5). Но в таком случае подынтегральные выражения должны совпадать:

$$f(x, y) = f_1(x) f_2(y). \quad (6)$$

Соотношение (6) можно рассматривать как условие независимости величин X и Y , для случая, когда (X, Y) имеет некоторую функцию плотности.

Комментарий к свойствам $M(X)$, $D(X)$, $\sigma(X)$

1. Свойство 3 математического ожидания является на первый взгляд удивительным. Это свойство справедливо для произвольной пары случайных величин X , Y (безразлично, зависимых или независимых). Как известно, в случае зависимых величин совместное распределение пары не определяется распределением слагаемых. Формула $M(X+Y) = M(X) + M(Y)$ определяет $M(X+Y)$ без использования закона распределения $X+Y$ (он в принципе неизвестен). Оказывается, что для вычисления $X+Y$ достаточно иметь (неполную) информацию о распределении $X+Y$, которую дают распределения X и Y .

2. Свойство 4 математического ожидания справедливо лишь для независимых случайных величин. Естественно измерять «степень зависимости» между X и Y разностью $M(XY) = M(X)M(Y)$ (которая равна нулю в том случае, когда величины независимы). На этой идее основаны понятия *ковариации и коэффициента корреляции* (см. ниже (см. п. 3)).

3. Свойства $M(X)$, $D(X)$, $\sigma(X)$ позволяют легко вычислить среднее значение, дисперсию и среднеквадратическое отклонение *биномиальной случайной величины* (см. §4).

Пусть X - количество наступлений события A в серии из n независимых испытаний схемы Бернулли. Очевидно, что

$$X = X_1 + X_2 + \dots + X_n,$$

где X_i - количество наступлений события A в i -м испытании ($i = 1, 2, \dots, n$).

Так в каждом испытании событие A наступает с вероятностью p и не наступает с вероятностью q , то ряд распределения случайной величины X имеет вид

0	1
q	p

(одно наступление события A - с вероятностью p и ноль наступления события $\bar{A}$ - с вероятностью q). Очевидно, что

$$M(X_i) = 0 \cdot q + 1 \cdot p = p. \quad D(X_i) = (0^2 \cdot q + 1^2 \cdot p) - p^2 = p - p^2 = p(1 - p) = pq.$$

Далее, используя свойство 3) математического ожидания, имеем

$$M(X) = M(X_1 + \dots + X_n) = M(X_1) + \dots + M(X_n) = \underbrace{p + p + \dots + p}_{n \text{ слагаемых}} = np. \quad (7)$$

Наконец, используя свойство 4) дисперсии, находим

$$D(X) = D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n) = \underbrace{pq + pq + \dots + pq}_{n \text{ слагаемых}} = npq. \quad (8)$$

откуда

$$\sigma(X) = \sqrt{npq}. \quad (9)$$

4. Из свойства математического ожидания и дисперсии в качестве следствия вытекает важный теоретико-вероятностный факт, лежащий в основе законов больших чисел (см. §6, теорема Чебышева).

Пусть $X_1, X_2, \dots, X_n$ - независимые случайные величины с одинаковыми математическими ожиданиями $M(X_i) = a$ и среднеквадратическим отклонением $\sigma(X_i) = \sigma$ т.е. $D(X_i) = \sigma^2$. Тогда для случайной величины

$$X = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

справедливы равенства

$$M(X) = a, \quad D(X) = D\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

$$\sigma(X) = \frac{\sigma}{\sqrt{n}}.$$

Таким образом, дисперсия случайной величины X уменьшается с ростом n , рассеяние значений X относительно $M(X)$ уменьшается, сама величина X теряет случайный характер.

3. Ковариация и коэффициент корреляции

Как уже было отмечено (см. п. 2), с помощью числа

$$k(X, Y) = M(XY) - M(X)M(Y) \quad (10)$$

можно измерять степень зависимости случайных величин X и Y . Свойство

4) математического ожидания означает, что $k(X, Y) = 0$ для независимых случайных величин. Естественно считать, что чем больше $k(X, Y)$ по абсолютной величине, тем больше степень зависимости. Так как $k(X, Y)$ имеет размерность XY , то при изменении единицы масштаба его значение будет подвержено изменению. Чтобы избежать этого, введем коэффициент

$$r(X, Y) = \frac{k(X, Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{M(XY) - M(X)M(Y)}{\sigma(X) \cdot \sigma(Y)}, \quad (11)$$

где $\sigma(X)$, $\sigma(Y)$ - среднеквадратические отклонения случайных величин X и Y .

Коэффициент $r(X, Y)$ является безразмерным: он не зависит от единиц измерения величин X и Y .

Определение 7. Пусть X , Y - случайные величины. XY - их произведение, $M(X)$, $M(Y)$, $M(XY)$ - математические ожидания этих величин, $\sigma(X)$, $\sigma(Y)$ - среднеквадратические отклонения величин X и Y . Коэффициент $k(X, Y)$, определенный формулой (9), называется *коэффициентом ковариации*, а коэффициент $r(X, Y)$, определенный формулой (10), - *коэффициентом корреляции*.

Для теории вероятностей и ее приложений большее значение имеет коэффициент корреляции (основная причина этого - его безразмерность).

Свойства коэффициента корреляции

- 1) $r(X, Y) = 0$ для независимых случайных величин X и Y .
- 2) $-1 \leq r(X, Y) \leq 1$ для любых двух случайных величин X и Y .

3) Если $|r(X, Y)| = 1$, то случайные величины X и Y связаны соотношением

$$Y = aX + b, \quad (12)$$

где a и b - некоторые постоянные.

Обратно, если X и Y связаны условием (11), то $|r(X, Y)| = 1$ ($r(X, Y) = -1$ при $a < 0$ и $r(X, Y) = 1$ при $a > 0$).

Доказательство свойств коэффициента опускаем.

Комментарий к свойствам коэффициента корреляции

Свойства 1) - 3) означают, что коэффициент корреляции измеряет степень зависимости случайных величин X , Y в следующем смысле.

Для независимых величин X и Y коэффициент корреляции $r(X, Y)$ равен нулю. Крайние возможные значения $r(X, Y)$, равные 1 и - 1, соответствуют функциональной зависимости между X и Y , имеющей вид $Y = aX + b$. Функциональная зависимость между X и Y - самый тесный вид зависимости.

В общем случае независимость величин X и Y означает, что условное распределение величины Y при заданном значении $X = X_0$ совпадает с безусловным распределением Y ; если же Y является функцией от X , то при $X = X_0$ она принимает вполне определенное значение, так что при условии $X = X_0$ величина Y даже не является случайной.

Зависимостям, близким к зависимости вида $Y = aX + b$ соответствуют значения $r(X, Y)$, близкие к 1 или - 1 (при $a > 0$ или $a < 0$, соответственно). Если величины X и Y слабо зависимы, то значения $r(X, Y)$ близки к нулю.

Следует иметь в виду, что существуют зависимые величины X и Y , коэффициент корреляции которых равен нулю; их называют *некоррелированными*.

Если величины X и Y связаны нелинейной функциональной зависимостью, то $r(X, Y)$ может отличаться от 1 и -1.

Итак, коэффициент корреляции измеряет степень линейной зависимости между случайными величинами X и Y .

Для дискретных случайных величин коэффициенты $k(X, Y)$ и $r(X, Y)$ можно вычислить

$$k(X, Y) = \sum x_i y_j p_{ij} - \left(\sum x_i p_i \right) \left(\sum y_j q_j \right). \quad (13)$$

$$r(X, Y) = \frac{\sum x_i y_j p_{ij} - \left(\sum x_i p_i \right) \left(\sum y_j q_j \right)}{\sqrt{\sum x_i^2 p_i - \left(\sum x_i p_i \right)^2} \sqrt{\sum y_j^2 q_j - \left(\sum y_j q_j \right)^2}}. \quad (14)$$

Здесь x_i и y_j - значения случайных величин X и Y , p_i и q_j - соответствующие им вероятности, p_{ij} - вероятность совместного появления событий $\{X = x_i\}$ и $\{Y = y_j\}$.

Суммы, входящие в правые части равенств (13) и (14), можно выразить через математическое ожидание и среднеквадратическое отклонение:

$$\sum x_i y_j p_{ij} = M(X, Y), \quad \sum x_i p_i = M(X), \quad \sum y_j q_j = M(Y),$$

$$\sqrt{\sum x_i^2 p_i - \left(\sum x_i p_i \right)^2} = \sqrt{D(X)} = \sigma(X),$$

$$\sqrt{\sum y_j^2 q_j - \left(\sum y_j q_j \right)^2} = \sqrt{D(Y)} = \sigma(Y).$$

Эти формулы вытекают из определения математического ожидания и дисперсии, и величины XY .

Полезно также использовать равенства

$$k(X, Y) = \sum (x_i - M(X))(y_j - M(Y)) p_{ij}, \quad (15)$$

$$r(X, Y) = \frac{\sum (x_i - M(X))(y_i - M(Y))p_{ij}}{\sigma(X) \cdot \sigma(Y)}. \quad (16)$$

Согласно формуле (15), коэффициент ковариации равен математическому ожиданию произведения случайных величин $X - M(X)$ и $Y - M(Y)$.

Вывод формул (15), (16) осуществить самостоятельно.

Справедливы формулы, дающие выражение коэффициентов ковариации и корреляции через функции распределения непрерывных случайных величин X , Y и через функцию их совместного распределения. Эти формулы аналогичны (15), (16).

Пример 11. Определить ковариацию и коэффициент корреляции случайных величин X и Y имеющие следующие ряды распределения:

x_i	1	2
p_i	0,8	0,2

y_i	-1	0	1	2
p_i	0,2	0,3	0,3	0,2

Решение

Найдем математические ожидания и средние квадратические отклонения этих случайных величин:

$$M(X) = 1 \cdot 0,8 + 2 \cdot 0,2 = 1,2; \quad M(X^2) = 1^2 \cdot 0,8 + 2^2 \cdot 0,2 = 1,6;$$

$$D(X) = 1,6 - (1,2)^2 = 0,16; \quad \sigma(X) = \sqrt{D(X)} = \sqrt{0,16} = 0,4;$$

$$M(Y) = (-1) \cdot 0,2 + 0 \cdot 0,3 + 1 \cdot 0,3 + 2 \cdot 0,2 = 0,5;$$

$$M(Y^2) = (-1)^2 \cdot 0,2 + 0^2 \cdot 0,3 + 1^2 \cdot 0,3 + 2^2 \cdot 0,2 = 1,3;$$

$$D(Y) = 1,3 - (0,5)^2 = 1,05; \quad \sigma(Y) = \sqrt{D(Y)} = \sqrt{1,05} = 1,025.$$

Для нахождения математического ожидания $M(XY)$ произведения случайных величин X и Y нужно составить закон распределения произведения

двух дискретных случайных величин (как это сделано выше при решении примера 9), а затем по нему найти $M(XY)$.

Закон распределения XY имеет вид:

$(x, y)_k$	-2	-1	0	1	2	4
p_k	0,1	0,1	0,3	0,3	0,15	0,05

Тогда имеем:

$$M(XY) = (-2) \cdot 0,1 + (-1) \cdot 0,1 + 0 \cdot 0,3 + 1 \cdot 0,3 + 2 \cdot 0,15 + 4 \cdot 0,05 = 0,5.$$

Вычислим ковариацию по формуле (10):

$$k(X, Y) = M(XY) - M(X)M(Y) = 0,5 - 1,2 \cdot 0,5 = -0,1.$$

Вычислим коэффициент корреляции по формуле (11):

$$r(X, Y) = \frac{k(X, Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{-0,1}{0,4 \cdot 1,025} = -0,244,$$

т.е. между случайными величинами X и Y существует отрицательная линейная зависимость; следовательно, при увеличении (уменьшении) одной из случайных величин другая имеет некоторую тенденцию уменьшаться (увеличиваться).

4. Моменты случайных величин

Пусть X - случайная величина (для определенности - дискретная); x_i - ее значения; p_i - соответствующие им вероятности ($i = 1, 2, \dots, n$); $M(X)$ - математическое ожидание X ; a - произвольное число.

Рассмотрим следующие суммы:

$$m_k = \sum_{i=1}^n x_i^k p_i, \quad m'_k = \sum_{i=1}^n (x_i - a)^k p_i, \quad m''_k = \sum_{i=1}^n (x_i - M(X))^k p_i. \quad (17)$$

m_k называют *начальным моментом* k -го порядка, m'_k - *относительным моментом* k -го порядка (относительно числа a), m''_k - *центральный моментом*. Заметим, что важнейшие характеристики случайной величины X

математическое ожидание и дисперсия - представляют собой моменты этой величины, а именно, $M(X) = m_1$, $D(X) = m_2''$.

Существуют и другие характеристики случайной величины, описывающие ее различные свойства: эти характеристики, так же как $M(X)$ и $D(X)$, сравнительно просто выражаются через начальные и центральные моменты.

Выше отмечено, что математическое ожидание $M(X)$, или первый начальный момент, характеризует среднее значение или положение распределения случайной величины X на числовой оси; дисперсия $D(X)$ или второй центральный момент - степень рассеяния распределения X относительно $M(X)$. Моменты высших порядков служат для более подробного описания распределения.

Третий центральный момент служит для характеристики асимметрии (скошенности) распределения. Он имеет размерность куба случайной величины. Чтобы получить безразмерную величину, ее делят на σ^3 , где σ - среднее квадратическое отклонение случайной величины X . Полученная величина A называется *коэффициентом асимметрии* случайной величины:

$$A = \frac{m_3'''}{\sigma^3}.$$

Если распределение симметрично относительно математического ожидания, то коэффициент асимметрии $A = 0$.

На рис. 1 показаны две кривые распределения: I и II. Кривая I имеет положительную (правостороннюю) асимметрию ($A > 0$), а кривая II - отрицательную (левостороннюю) ($A < 0$).

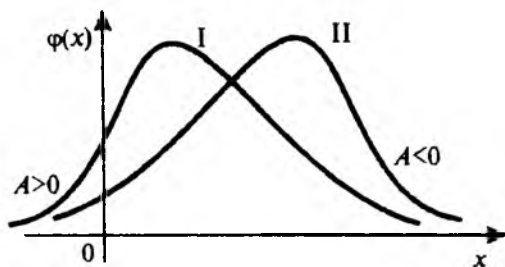


Рис. 1

Четвертый центральный момент служит для характеристики крутости (островершинности или плосковершинности) распределения.

Эксцессом (или коэффициентом эксцесса) случайной величины называется число

$$E = \frac{m_4^*}{\sigma^4} - 3.$$

Число 3 вычитается из отношения $\frac{m_4^*}{\sigma^4}$ потому, что для наиболее часто встречающегося нормального распределения (о нем идет речь в §6) отношение $\frac{m_4^*}{\sigma^4} = 3$. Кривые, более островершинные, чем нормальная, обладают положительным эксцессом, более плосковершинные - отрицательным эксцессом (рис.2).

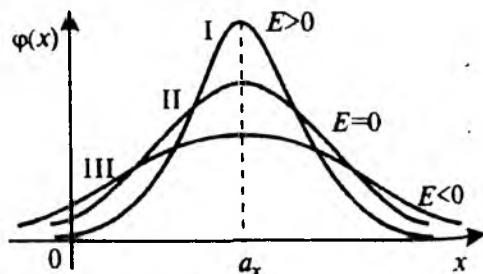


Рис. 2

Вычисление центральных моментов несколько облегчается, если сначала найти относительные (относительно некоторого подходящего числа) моменты, а затем преобразовать их в центральные по формулам:

$$\begin{aligned}m_2'' &= m_2' - (m_1')^2, \\m_3'' &= m_3' - 3m_2'm_1' + 2(m_1')^3, \\m_4'' &= m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m_1')^4.\end{aligned}\tag{18}$$

Вывод формул (18) предоставляем читателю самостоятельно.

Пример 12. Найти коэффициент асимметрии и эксцесс случайной величины, распределенной по так называемому закону Лапласа с функцией плотности $\varphi(x) = \frac{1}{2}e^{-|x|}$.

Решение

Так как распределение случайной величины X симметрично относительно оси ординат, то все нечетные как начальные, так и центральные моменты равны 0, т.е. $m_1 = 0$, $m_3 = 0$, $m_5'' = 0$ и силу $A = \frac{m_3''}{\sigma^3}$ коэффициент асимметрии $A = 0$.

Для нахождения эксцесса необходимо вычислить четные начальные моменты m_2 и m_4 :

$$m_2 = \int_{-\infty}^{+\infty} x^2 \varphi(x) dx = \int_{-\infty}^{+\infty} x^2 \left(\frac{1}{2} e^{-|x|} \right) dx = 2 \cdot \frac{1}{2} \int_0^{+\infty} x^2 e^{-x} dx = 2.$$

Следовательно,

$$D(X) = m_2'' = m_2 - m_1^2 = 2 - 0^2 = 2 \text{ и } \sigma(X) = \sqrt{D(X)} = \sqrt{2}.$$

$$m_4 = \int_{-\infty}^{+\infty} x^4 \varphi(x) dx = \int_{-\infty}^{+\infty} x^4 \left(\frac{1}{2} e^{-|x|} \right) dx = 2 \cdot \frac{1}{2} \int_0^{+\infty} x^4 e^{-x} dx = 24 \text{ и } m_4'' = 24.$$

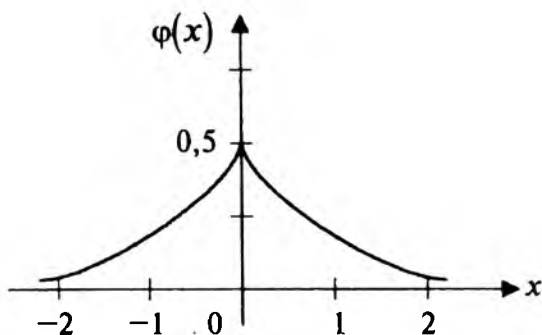


Рис. 3

Теперь вычислим эксцесс

$$E = \frac{m_4^*}{\sigma^4} - 3 = \frac{24}{(\sqrt{2})^4} - 3 = 3.$$

Эксцесс распределения положителен, что говорит об островершинности кривой распределения $\varphi(x)$ (рис. 3).

Для пары дискретных случайных величин X, Y рассмотрим следующие суммы:

$$m_{kl} = \sum x_i^k y_j^l p_{ij}, \quad m'_{kl} = \sum (x_i - a)^k (y_j - b)^l p_{ij}, \quad (19)$$

$$m''_{kl} = \sum (x_i - M(X))^k (y_j - M(Y))^l p_{ij}, \quad (k, l = 0, 1, 2, \dots).$$

Они имеют следующие названия: m_{kl} - *начальный момент* порядка k, l ; m'_{kl} - *относительный момент* порядка k, l (относительно чисел a и b); m''_{kl} - *центральный момент* порядка k, l .

При $k = 0$ или $l = 0$ моменты пары превращаются в моменты случайных величин.

Коэффициент ковариации, как видно из сравнения формул (19) и (15), совпадает с m''_{11} математическое ожидание произведения XY равно m_{11} .

Моменты пары (X, Y) можно истолковать как математические ожидания некоторых функций от случайных величин X, Y , а именно:

$$m_{kl} = M(X^k Y^l), \quad m'_{kl} = M((X - a)^k (Y - b)^l)$$

$$m''_{kl} = \sum (X - M(X))^k (Y - M(Y))^l p_{ij}, \quad (k, l = 0, 1, 2, \dots).$$

Эти формулы определяют моменты не только для дискретных, но и для непрерывных величин.

Можно написать формулы, дающие явное выражение моментов через функции совместного распределения пары случайных величин. Написания этих формул предоставляем читателю.

Многие важные характеристики пары случайных величин X и Y достаточно просто выражаются через моменты этой пары величины.

Вопросы для самопроверки

1. Разъясните необходимость введения числовых характеристик для случайных величин.
2. Дайте определение понятия математического ожидания.
3. Какой вероятностный смысл имеет математическое ожидание?
4. Дайте определение дисперсии.
5. Какой вероятностный смысл имеет дисперсия?
6. Что называется среднеквадратичным отклонением?
7. Приведите свойства математического ожидания и прокомментируйте их.
8. Приведите свойства математической дисперсии и прокомментируйте их.
9. Приведите свойства среднеквадратического отклонения и прокомментируйте их.
10. Что представляет собой коэффициент корреляции?
11. Приведите свойства коэффициента корреляции и прокомментируйте их.
12. Что называют: начальным моментом k -го порядка, относительным моментом k -го порядка, центральным моментом?
13. Что называют: начальным моментом порядка k, l ; относительным моментом порядка k, l ; центральным моментом порядка k, l ?

§6. Наиболее известные дискретные и непрерывные законы распределения. Закон больших чисел и центральная предельная теорема

Ключевые слова: биномиальное распределение, пуассоновское распределение, геометрическое распределение, гипергеометрическое распределение, равномерный закон распределения, показательный закон распределения, нормальный закон распределения, логарифмически-нормальный закон распределения, логарифмически-нормальное распределение. χ^2 – распределение, распределение Стюдента, распределение Фишера-Снедекора, закон больших чисел, неравенство Чебышева, теорема Бернулли, теорема Чебышева, центральная предельная теорема Ляпунова, характеристическая функция, преобразования Фурье.

1. Наиболее известные дискретные и непрерывные законы распределения

В этом пункте этого параграфа описаны основные законы распределения дискретных и непрерывных случайных величин, используемых для построения теоретико-вероятностных моделей реальных социально-экономических явлений.

1.1. Наиболее известные дискретные законы распределения

Напомним, что в §4 мы ознакомились с двумя важными примерами дискретных случайных величин: *биномиальное распределение, пуассоновское распределение.*

Ниже приведем определения, некоторые числовые характеристики этих и других известных дискретных законов распределения.

Биномиальный закон распределения

Определение 1. Распределение случайной величины X , равной количеству наступлений события A в схеме Бернулли из n испытаний, называется *биномиальным распределением*.

В этом распределении значению $k \in \{0, 1, 2, \dots, n\}$ случайной величины X соответствует вероятность $P_n(k) = C_n^k p^k q^{n-k}$, где p – вероятность наступления события A в одном испытании, $q = 1 - p$.

Ряд распределения биномиального закона имеет вид:

x_i	0	1	2	...	k	...	n
p_i	q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$	...	$C_n^k p^k q^{n-k}$	...	p^n

Очевидно, что определение биномиального закона корректно, так как основное свойство ряда распределения $\sum_{i=0}^n p_i = 1$ выполнено, ибо $\sum_{i=0}^n p_i$ есть не что иное, как сумма всех членов разложения бинома Ньютона:

$$q^n + C_n^1 p q^{n-1} + C_n^2 p^2 q^{n-2} + \dots + C_n^k p^k q^{n-k} + p^n = (p + q)^n = 1$$

(отсюда и название закона - биномиальный).

Теорема. Математическое ожидание случайной величины X , распределенной по биномиальному закону,

$$M(X) = np.$$

а ее дисперсия

$$D(X) = npq.$$

Доказательство см. в §4.

Биномиальное распределение широко используется в теории и практике статистического контроля продукции, при описании функционирования систем массового обслуживания, в теории стрельбы и в других областях.

Пример 1. По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,515. Найти математическое ожидание и дисперсию случайной величины X - числа мальчиков в семье из 4 детей.

Решение

Закон распределения случайной величины X - числа мальчиков в семье из 4 детей - биномиальный с параметрами $n = 4$, $p = 0,515$. Тогда

$$M(X) = np = 4 \cdot 0,515 = 2,06 \text{ и } D(X) = 4 \cdot 0,515 \cdot 0,485 = 0,999.$$

Мы получили бы такие же результаты, если математическое ожидание и дисперсию вычислили по ряду (закону) распределения этой случайной величины, которую мы составили при решении примера 2 из §5.

Упражнение

В магазин поступила обувь с двух фабрик в соотношении 2:3. Куплено 4 пары обуви. Найти закон распределения числа купленных пар обуви, изготовленной первой фабрикой. Найти математическое ожидание и среднее квадратическое отклонение этой случайной величины.

Закон распределения Пуассона

Определение 2. Распределение случайной величины X , принимающей значения $k \in \{0, 1, 2, \dots\}$ с вероятностями $\frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda > 0$ - некоторый параметр, называется *пуассоновским распределением* или *распределением Пуассона*.

Ряд распределения закона Пуассона имеет вид:

x_i	0	1	2	...	k	...
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$	...	$\frac{\lambda^k}{k!} e^{-\lambda}$	...

Очевидно, что определение закона Пуассона корректно, так как основное свойство ряда распределения $\sum_{i=0}^{\infty} p_i = 1$ выполнено, ибо сумма ряда

$$\begin{aligned} \sum_{i=0}^{\infty} p_i &= e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} + \dots + \frac{\lambda^k}{k!} e^{-\lambda} + \dots = \\ &= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^k}{k!} + \dots \right) = e^{-\lambda} \cdot e^{\lambda} = 1. \end{aligned}$$

Здесь учтено, что в скобках записано разложение в ряд функции e^x при $x = \lambda$.

Теорема. Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, совпадают и равны параметру λ этого закона, т.е.

$$M(X) = \lambda, \quad D(X) = \lambda.$$

Найдем математическое ожидание случайной величины X

$$M(X) = \sum_{i=1}^{\infty} x_i p_i = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \\ = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^k}{k!} + \dots \right) = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

Для того чтобы найти дисперсии случайной величины X сначала вычислим

$$M(X^2) = \sum_{i=1}^{\infty} x_i^2 p_i = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \\ = e^{-\lambda} \sum_{k=1}^{\infty} \frac{((\lambda-1)+1)^k}{(k-1)!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \\ = \lambda^2 e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \dots \right) + \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \dots \right) = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.$$

Теперь

$$D(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

При достаточно больших n (вообще при $n \rightarrow \infty$) и малых значениях p ($p \rightarrow 0$) при условии, что произведение np - постоянная величина ($np \rightarrow \lambda = const$), закон распределения Пуассона является хорошим приближением биномиального закона, так как в этом случае функция вероятностей Пуассона хорошо аппроксимирует функцию вероятностей, определяемую по формуле Бернулли. Иначе, при $p \rightarrow 0$, $n \rightarrow \infty$, $np \rightarrow \lambda = const$ закон распределения Пуассона является **предельным случаем** биномиального закона. Так как при этом вероятность p события A в каждом испытании мала, то закон распределения Пуассона называют часто **законом редких явлений**.

Наряду с «предельным» случаем биномиального распределения закон Пуассона может возникнуть и в ряде других ситуаций.

Можно показать, что для простейшего потока событий число событий, попадающих на произвольный отрезок времени, есть случайная величина, имеющая пуассоновское распределение.

По пуассоновскому распределены, например, число рождения четверняшек, число сбоев на автоматической линии, число отказов сложной систе-

мл в «нормальном режиме», число «требований на обслуживание», поступивших в единицу времени в системах массового обслуживания и др.

Отметим еще, что если случайная величина представляет собой сумму двух независимых случайных величин, распределенных каждая по закону Пуассона, то она также распределена по этому закону.

Упражнение

Доказать, что сумма двух независимых случайных величин, распределенных по закону Пуассона с параметрами λ_1 и λ_2 , также распределена по закону Пуассона с параметром $\lambda = \lambda_1 + \lambda_2$.

Геометрическое распределение

Определение 3. Распределение случайной величины X , принимающей значения $k \in \{1, 2, \dots\}$ с вероятностями pq^{k-1} , где $0 < p < 1$, называется *геометрическим распределением*.

Ряд геометрического распределения имеет вид:

x_i	1	2	3	...	k	...
p_i	p	pq	pq^2	...	pq^{k-1}	...

Нетрудно видеть, что вероятности p_i образуют геометрическую прогрессию с первым членом p и знаменателем q (отсюда название «геометрическое распределение»).

Определение геометрического распределения корректно, так как сумма ряда

$$\sum_{i=1}^{\infty} p_i = p + pq + \dots + pq^{k-1} + \dots = p(1 + q + \dots + q^{k-1} + \dots) = p \cdot \frac{1}{1-q} = \frac{p}{p} = 1.$$

Случайная величина $X = k$ имеющая геометрическое распределение, представляет собой число k испытаний, проведенных по схеме Бернулли, с вероятностью p наступления события в каждом испытании до первого положительного исхода.

Теорема. Математическое ожидание случайной величины X , имеющей геометрическое распределение с параметром p ,

$$M(X) = \frac{1}{p},$$

и ее дисперсия

$$D(X) = \frac{q}{p^2}.$$

Доказательство теоремы, связанное с суммированием членов бесконечного ряда предоставляем провести читателю самостоятельно.

Пример 2. Радиотехник вызывает корреспондента, причем каждый последующий вызов производится лишь в том случае, если предыдущий вызов не принят. Вероятность того, что корреспондент примет вызов, равна 0,4. Составить ряд распределения числа вызовов. Найти математическое ожидание и дисперсию этой случайной величины.

Решение

Так, например, число вызовов радиотехником корреспондента до тех пор, пока вызов не будет принят, есть случайная величина, имеющая геометрическое распределение с параметром $p = 0,4$. Поэтому ряд распределения имеет вид:

x_i	1	2	3	4	...	k	...
p_i	0,4	0,24	0,144	0,0864	...	$0,4 \cdot (0,6)^{k-1}$	...

Теперь непосредственно можно вычислять математическое ожидание и дисперсию этой случайной величины. Математическое ожидание и дисперсию можно найти, пользуясь утверждением теоремы:

$$M(X) = \frac{1}{p} = \frac{1}{0,4} = 2,5 \text{ и } D(X) = \frac{0,6}{(0,4)^2} = 3,75.$$

Упражнение

Проводится проверка большой партии деталей до обнаружения бракованной (без ограничения числа проверенных деталей). Составить закон распределения числа проверенных деталей. Найти его математическое ожидание

и дисперсию, если известно, что вероятность брака для каждой детали равна 0.1.

Гипергеометрическое распределение

Определение 4. Распределение случайной величины X , принимающей значения $k \in \{0, 1, 2, \dots, \min\{n, K\}\}$ с вероятностями $\frac{C_K^N C_{N-K}^{n-k}}{C_N^n}$, где $k \leq N$, $n \leq N$; n, K, N – натуральные числа, называется *гипергеометрическим распределением*.

Гипергеометрическое распределение имеет случайная величина $X = k$ число объектов, обладающих заданным свойством, среди n объектов, случайно извлеченных (без возврата) из совокупности N объектов, K из которых обладают этим свойством.

Теорема. Математическое ожидание случайной величины X , имеющей гипергеометрическое распределение с параметрами n, K, N , есть

$$M(X) = n \frac{K}{N},$$

а ее дисперсия

$$D(X) = n \frac{K}{N-1} \left(1 - \frac{K}{N}\right) \left(1 - \frac{n}{N}\right).$$

Доказательство опускаем.

Пример 3. Среди 10 изготовленных приборов 3 неточных. Найти математическое ожидание и дисперсию случайной величины X – число неточных приборов среди взятых наудачу четырех.

Решение

Нетрудно заметить, что случайная величина X – число неточных приборов среди взятых наудачу четырех имеет гипергеометрическое распределение с параметрами $n = 4, K = 3, N = 10$.

Математическое ожидание и дисперсию можно найти, пользуясь утверждением теоремы:

$$M(X) = n \frac{K}{N} = 4 \cdot \frac{3}{10} = 1,2 \text{ и}$$

$$D(X) = n \frac{K}{N-1} \left(1 - \frac{K}{N}\right) \left(1 - \frac{n}{N}\right) = 4 \cdot \frac{3}{9} \left(1 - \frac{3}{10}\right) \left(1 - \frac{4}{10}\right) = 0,56.$$

Упражнение

В лотерее «Спортлото 6 из 45» денежные призы получают участники, угадавшие 3, 4, 5 и 6 видов спорта из отобранных случайно 6 видов из 45 (размер приза увеличивается с увеличением числа угаданных видов спорта). Найти закон распределения случайной величины X - числа угаданных видов спорта среди случайно отобранных шести. Какова вероятность получения денежного приза? Найти математическое ожидание и дисперсию случайной величины X .

Случайную величину $X = k$, распределенную по биномиальному закону, можно интерпретировать как число k объектов, обладающих данным свойством, из общего числа n объектов, случайно извлеченных из некоторой воображаемой бесконечной совокупности, доля p объектов которой обладает этим свойством. Поэтому гипергеометрическое распределение можно рассматривать как модификацию биномиального распределения для случая конечной совокупности, состоящей из N объектов, K из которых обладают этим свойством.

Можно показать, что при $N \rightarrow \infty$ функция вероятностей гипергеометрического распределения стремится к соответствующей функции биномиального закона.

Гипергеометрическое распределение широко используется в практике статистического приемочного контроля качества промышленной продукции, в задачах, связанных с организацией выборочных обследований, и других областях.

1.2. Наиболее известные непрерывные законы распределения

Напомним, что в §4 мы ознакомились с тремя важными примерами непрерывных случайных величин: *равномерный, показательный и нормальный законы распределения*.

Ниже приведем некоторые числовые характеристики этих распределений, а также определения и числовые характеристики других известных непрерывных законов распределения.

Равномерный закон распределения

Определения, функцию распределения, функцию плотности равномерного закона мы рассмотрели в §4.

Рассмотрим теорему, устанавливающую теоретико-вероятностный смысл параметров равномерного закона.

Теорема. Математическое ожидание случайной величины X , имеющей равномерный закон распределения, есть

$$M(X) = \frac{a+b}{2},$$

а ее дисперсия

$$D(X) = \frac{(b-a)^2}{12}.$$

Доказательство. Выражения для функции плотности равномерного закона (см. §4) имеет вид:

$$f(x) = \begin{cases} 0 & \text{при } x < a, \\ \frac{1}{b-a} & \text{при } a \leq x \leq b, \\ 0 & \text{при } x > b. \end{cases}$$

Вычислим математическое ожидание случайной величины X :

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{xdx}{b-a} = \frac{1}{b-a} \left(\frac{x^2}{2} \Big|_a^b \right) = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$\begin{aligned} D(X) &= \int_{-\infty}^{+\infty} (x - M(X))^2 f(x) dx = \int_a^b \left(x - \frac{a+b}{2} \right)^2 \frac{dx}{b-a} = \\ &= \frac{1}{3(b-a)} \left(x - \frac{a+b}{2} \right)^3 \Big|_a^b = \frac{1}{3(b-a)} \left[\frac{(b-a)^3}{8} - \frac{(a-b)^3}{8} \right] = \frac{(b-a)^2}{12}. \end{aligned}$$

Показательный (экспоненциальный) закон распределения

Определения, функцию распределения, функцию плотности показательного (экспоненциального) закона мы также рассмотрели в §4.

Рассмотрим теорему, устанавливающую теоретико-вероятностный смысл параметров показательного (экспоненциального) закона.

Теорема. Математическое ожидание случайной величины X , имеющей показательный закон распределения, есть

$$M(X) = \frac{1}{\lambda},$$

а ее дисперсия

$$D(X) = \frac{1}{\lambda^2}.$$

Доказательство. Выражения для функции плотности показательного (экспоненциального) закона (см. §4) имеет вид:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где λ - некоторый параметр.

Найдем математическое ожидание случайной величины X , используя при вычислении метод интегрирования по частям:

$$\begin{aligned} M(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \lim_{b \rightarrow +\infty} \int_0^b x \lambda e^{-\lambda x} dx = \lim_{b \rightarrow +\infty} \left(- \int_0^b x d e^{-\lambda x} \right) = \\ &= \lim_{b \rightarrow +\infty} \left(-x e^{-\lambda x} \Big|_0^b + \int_0^b e^{-\lambda x} dx \right) = \lim_{b \rightarrow +\infty} \left(-b e^{-\lambda b} - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^b \right) = 0 - \frac{1}{\lambda} \lim_{b \rightarrow +\infty} (e^{-\lambda b} - 1) = \frac{1}{\lambda}. \end{aligned}$$

Аналогично вычислим, что $M(X^2) = \frac{2}{\lambda^2}$. Тогда имеем

$$D(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}.$$

Нормальный закон распределения

Определения, функцию распределения, функцию плотности нормального закона уже мы рассмотрели в §4. Там мы отметили, что нормальный за-

кон распределения наиболее часто встречается на практике. Главная особенность, выделяющая его среди других законов, состоит в том, что он является **предельным законом**, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях (см. ниже в п.2).

Напомним определения нормального закона распределения.

Определение 5. Непрерывная случайная величина X имеет *нормальный закон распределения (закон Гаусса)* с параметрами a и σ^2 , если ее функция плотности имеет вид:

$$\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (1)$$

Комментарий к определению 5

Термин «нормальный» не совсем удачный. Многие признаки подчиняются нормальному закону, например, рост человека, дальность полета снаряда и т.п. Но если какой-либо признак подчиняется другому, отличному от нормального, закону распределения, то это вовсе не говорит о «ненормальности» явления, связанного с этим признаком.

Кривую нормального закона распределения называют *нормальной или гауссовой кривой*. На рис. 12 а, б §4 приведены нормальная кривая с параметрами σ^2 и a , т.е. $N(a; \sigma^2)$ и график функции распределения случайной величины X , имеющей нормальный закон. Обратим внимание на то, что нормальная кривая симметрична относительно прямой $x = a$, имеет максимум в точке $x = a$, равный $\frac{1}{\sigma\sqrt{2\pi}}$, т.е. $\varphi_{\max}(a) = \frac{1}{\sigma\sqrt{2\pi}} \approx \frac{0,3989}{\sigma}$ и две точки перегиба $x = a \pm \sigma$ ординатой $\varphi_{\text{пер}}(a) = \frac{1}{\sigma\sqrt{2\pi}e} \approx \frac{0,2420}{\sigma}$. Можно заметить, что в выражении плотности нормального закона параметры обозначены буквами a и σ^2 , которыми мы обозначаем математическое ожидание $M(X)$ и дисперсию $D(X)$. Такое совпадение неслучайно.

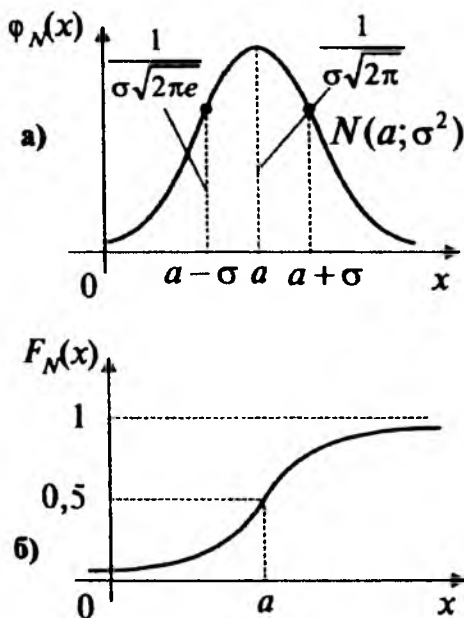


Рис. 1

Рассмотрим теорему, устанавливающую теоретико-вероятностный смысл параметров нормального закона.

Теорема. Математическое ожидание случайной величины X , распределенной по нормальному закону, равно параметру a этого закона, т.е.

$$M(X) = a.$$

а ее дисперсия - параметру σ^2 , т.е.

$$D(X) = \sigma^2.$$

Доказательство. Математическое ожидание случайной величины X :

$$M(X) = \int_{-\infty}^{+\infty} x \varphi_N(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Произведем замену переменной, положив $t = \frac{x-a}{\sigma\sqrt{2}}$. Тогда $x = a + \sigma\sqrt{2}t$

и $dx = \sigma\sqrt{2}dt$, пределы интегрирования не меняются и, следовательно,

$$M(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} (a + \sigma\sqrt{2}t) e^{-t^2} \sigma\sqrt{2} dt = \\ = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2} dt + \frac{a}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = 0 + \frac{a}{\sqrt{\pi}} \cdot \sqrt{\pi} = a.$$

Первый интеграл равен нулю как интеграл от нечетной функции по симметричному относительно начала координат промежутку, а второй интеграл

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi} \text{ - интеграл Эйлера - Пуассона.}$$

Дисперсия случайной величины X :

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 \varphi_X(x) dx = \int_{-\infty}^{+\infty} (x-a)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Сделаем ту же замену переменной $x = a + \sigma\sqrt{2}t$ как при вычислении предыдущего интеграла. Тогда

$$D(X) = \int_{-\infty}^{+\infty} \sigma^2 2t \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2} \sigma\sqrt{2} dt = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t dt e^{-t^2}.$$

Применяя метод интегрирования по частям, получим

$$D(X) = -\frac{\sigma^2}{\sqrt{\pi}} t e^{-t^2} \Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = 0 + \frac{\sigma^2}{\sqrt{\pi}} \cdot \sqrt{\pi} = \sigma^2.$$

Выясним, как будет меняться нормальная кривая при изменении параметров a и σ^2 (или σ). Если $\sigma = const$, и меняется параметр a ($a_1 < a_2 < a_3$), т.е. центр симметрии распределения, то нормальная кривая будет смещаться вдоль оси абсцисс, не меняя формы (рис. 2).

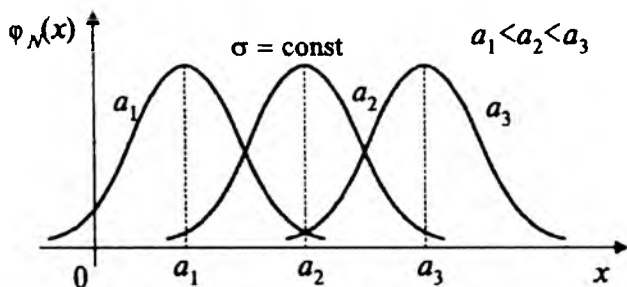


Рис. 2

Если $a = \text{const}$ и меняется параметр σ^2 (или σ), то меняется ордината максимума кривой $\varphi_{\max}(a) = \frac{1}{\sigma\sqrt{2\pi}}$. При увеличении σ ордината максимума кривой уменьшается, но так как площадь под любой кривой распределения должна оставаться равной единице, то кривая становится более плоской, растягиваясь вдоль оси абсцисс; при уменьшении σ , напротив, нормальная кривая вытягивается вверх, одновременно сжимаясь с боков. На рис. 3 показаны нормальные кривые с параметрами σ_1 , σ_2 и σ_3 , где $\sigma_1 < \sigma_2 < \sigma_3$. Таким образом, параметр a (он же математическое ожидание) характеризует *положение*, а параметр σ^2 (он же дисперсия) - *форму* нормальной кривой.

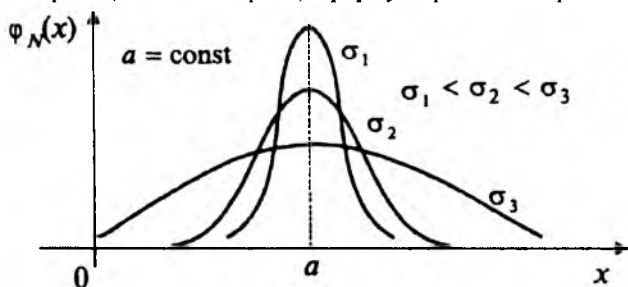


Рис. 3

Нормальный закон распределения случайной величины с параметрами $a = 0$, $\sigma^2 = 1$, т.е. $N(0;1)$, называется *стандартным или нормированным*, а соответствующая нормальная кривая - *стандартной или нормированной*.

Сложность непосредственного нахождения функции распределения случайной величины, распределенной по нормальному закону, по формуле

$F(x) = \int_{-\infty}^x f(\xi) d\xi$ и вероятности ее попадания на некоторый промежуток по

формуле $p\{a < X \leq b\} = \int_a^b f(x) dx$ связана с тем, что интеграл от функции (1)

является «неберущимся» в элементарных функциях. Поэтому их выражают через функцию

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad (2)$$

- функцию (интеграл вероятностей) Лапласа, для которой составлены таблицы. Напомним, что функция Лапласа уже встречалась нам при рассмотрении интегральной теоремы Муавра - Лапласа. Там же были рассмотрены и ее свойства. Геометрически функция Лапласа представляет собой площадь под стандартной нормальной кривой на отрезке $[-x; x]$ (рис. 4).

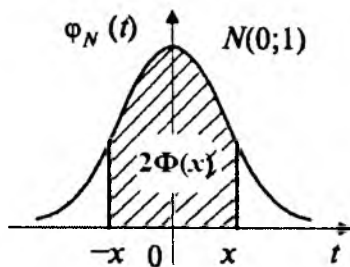


Рис. 4

Теорема. Функция распределения случайной величины X , распределенной по нормальному закону, выражается через функцию Лапласа $\Phi(x)$ по формуле:

$$F_N(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right). \quad (3)$$

Доказательство рекомендуется провести самостоятельно.

Геометрически функция распределения представляет собой площадь под нормальной кривой на интервале $(-\infty, x)$ (рис. 5). Как видим, она состоит из двух частей: первой, на интервале $(-\infty, a)$, равной $1/2$, т.е. половине всей площади под нормальной кривой, и второй, на интервале (a, x) равной $\Phi\left(\frac{x-a}{\sigma}\right)$.

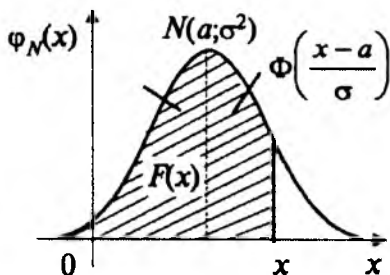


Рис. 5

Рассмотрим свойства случайной величины, распределенной по нормальному закону.

1. Вероятность попадания случайной величины X , распределенной по нормальному закону, в интервал $[x_1; x_2]$ равна

$$p\{x_1 \leq X \leq x_2\} = \Phi(t_2) - \Phi(t_1), \quad (4)$$

где

$$t_1 = \frac{x_1 - a}{\sigma}, \quad t_2 = \frac{x_2 - a}{\sigma}. \quad (5)$$

Доказательство. Учитывая, что вероятность $p\{x_1 \leq X \leq x_2\}$ есть приращение функции распределения на отрезке $[x_1; x_2]$, и формулу (3), получим

$$\begin{aligned} p\{x_1 \leq X \leq x_2\} &= F(x_2) - F(x_1) = \\ &= \left(\frac{1}{2} + \Phi\left(\frac{x_2 - a}{\sigma}\right) \right) - \left(\frac{1}{2} + \Phi\left(\frac{x_1 - a}{\sigma}\right) \right) = \Phi(t_2) - \Phi(t_1), \end{aligned}$$

где t_1 и t_2 определяются по формуле (5) (рис. 6)

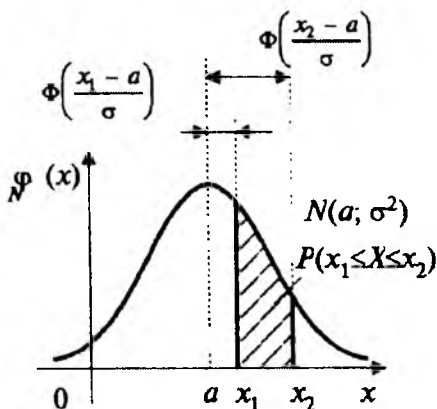


Рис. 6

2. Вероятность того, что отклонение случайной величины X , распределенной по нормальному закону, от математического ожидания a не превысит величину $\Delta > 0$ (по абсолютной величине), равна

$$p\{|X - a| \leq \Delta\} = 2\Phi(t), \quad (6)$$

где

$$t = \frac{\Delta}{\sigma}. \quad (7)$$

Доказательство. $p\{|X - a| \leq \Delta\} = p\{a - \Delta \leq X \leq a + \Delta\}$. Учитывая (4) и (5), а также свойство нечетности функции Лапласа, получим

$$\begin{aligned} p\{|X - a| \leq \Delta\} &= \left(\Phi\left(\frac{(a + \Delta) - a}{\sigma}\right) - \Phi\left(\frac{(a - \Delta) - a}{\sigma}\right) \right) = \\ &= \left(\Phi\left(\frac{\Delta}{\sigma}\right) - \Phi\left(-\frac{\Delta}{\sigma}\right) \right) = \left(\Phi\left(\frac{\Delta}{\sigma}\right) + \Phi\left(\frac{\Delta}{\sigma}\right) \right) = 2\Phi(t), \end{aligned}$$

где $t = \frac{\Delta}{\sigma}$ рис. 7.

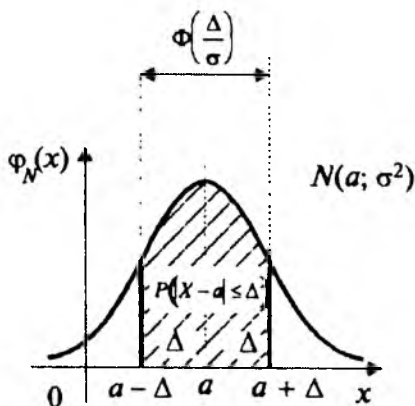


Рис. 7

Замечание. Рассмотренная в §3 приближенная интегральная формула Муавра – Лапласа (формула (4) из §3) следует из свойства (4) нормально распределенной случайной величины при $x_1 = k'_1$, $x_2 = k'_2$, $a = np$ и $\sigma = \sqrt{npq}$. так как биномиальный закон распределения случайной величины $X = m$ с параметрами n и p , для которого получена эта формула, при $n \rightarrow \infty$ стремится к нормальному закону (см. п. 2).

Аналогично и следствие $P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right)$, и другие следствия интегральной формулы Муавра-Лапласа (см. §3) для числа $X = m$ появления события в n независимых испытаниях и его частоты $\frac{m}{n}$ вытекают из свойств (4) и (6) нормального закона.

Вычислим по формуле (6) вероятности $p\{|X - a| \leq \Delta\}$ при различных значениях Δ (используем соответствующую таблицу из приложений). Получим при

$$\Delta = \sigma: p\{|X - a| \leq \sigma\} = 2\Phi(1) = 0,6826;$$

$$\Delta = \sigma: p\{|X - a| \leq 2\sigma\} = 2\Phi(2) = 0,9544;$$

$$\Delta = 3\sigma: p\{|X - a| \leq 3\sigma\} = 2\Phi(3) = 0,9973$$

(рис. 8).

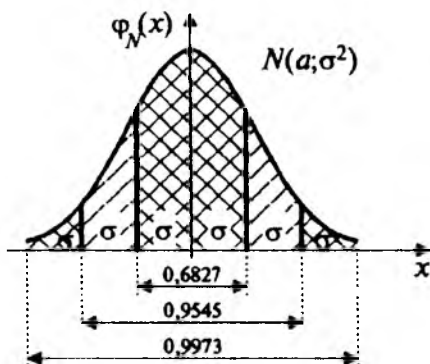


Рис. 8

Отсюда вытекает «правило трех сигм»:

Если случайная величина X имеет нормальный закон распределения с параметрами σ^2 и a , т.е. $N(a; \sigma^2)$, то практически достоверно, что ее значения заключены в интервале $(a - 3\sigma; a + 3\sigma)$.

Нарушение «правила трех сигм», т.е. отклонение нормально распределенной случайной величины X больше, чем на 3σ (по абсолютной величине), является событием практически невозможным, так как его вероятность весьма мала:

$$p\{|X - a| > 3\sigma\} = 1 - p\{|X - a| \leq 3\sigma\} = 1 - 0.9973 = 0.0027.$$

Найдем *коэффициент асимметрии* и *эксцесс* случайной величины X , распределенной по нормальному закону.

Очевидно, в силу симметрии нормальной кривой относительно вертикальной прямой $x = a$, проходящей через центр распределения $a = M(X)$ *коэффициент асимметрии нормального распределения* $A = 0$.

Эксцесс нормально распределенной случайной величины X найдем по формуле $E = \frac{m_4''}{\sigma^4} - 3$ (см. §5), т.е.

$$E = \frac{m_4''}{\sigma^4} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0,$$

где

$$m_4'' = \int_{-\infty}^{+\infty} (x - a)^4 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = 3\sigma^4.$$

Вычисление интеграла опускаем.

Таким образом, эксцесс нормального распределения равен нулю, и крутость других распределений определяется по отношению к нормальному (об этом мы уже упоминали в п.4 § 5).

Пример 4. Полагая, что рост мужчин определенной возрастной группы есть нормально распределенная случайная величина X с параметрами $a = 173$ и $\sigma^2 = 36$.

Найти: а) выражение функции плотности распределения случайной величины X ; б) доли костюмов 4-го роста (176 -182 см) и 3-го роста (170 -176 см), которые нужно предусмотреть в общем объеме производства для данной возрастной группы.

Сформулировать «правило трех сигм» для случайной величины X .

Решение

а) По формулам (1) и (3) запишем

$$\varphi_X(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(x-173)^2}{2 \cdot 36}};$$

$$F_X(x) = \frac{1}{2} + \frac{1}{6\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-173)^2}{2 \cdot 36}} = \frac{1}{2} + \Phi\left(\frac{x-173}{6}\right).$$

б) Доля костюмов 4-го роста (176 - 182 см) в общем объеме производства определится по формуле (4) как вероятность

$$p\{176 \leq X \leq 182\} = \Phi(t_2) - \Phi(t_1) = \Phi(1.50) - \Phi(0.50) = 0.4232 - 0.1915 = 0.2417,$$

$$\text{где } t_1 = \frac{176-173}{6} = 0.50, \quad t_2 = \frac{182-173}{6} = 1.50 \quad (\text{рис. 8}).$$

Долю костюмов 3-го роста (170-176 см) можно было определить аналогично по формуле (4), но проще это сделать по формуле (6), если учесть, что данный интервал симметричен относительно математического ожидания $a = M(X) = 173$, т.е. неравенство $176 \leq X \leq 182$ равносильно неравенству $|X - 173| \leq 3$:

$$p\{176 \leq X \leq 182\} = p\{|X - 173| \leq 3\} = 2\Phi\left(\frac{3}{6}\right) = 2\Phi(0.50) = 2 \cdot 0.1915 = 0.3830$$

(рис. 9).

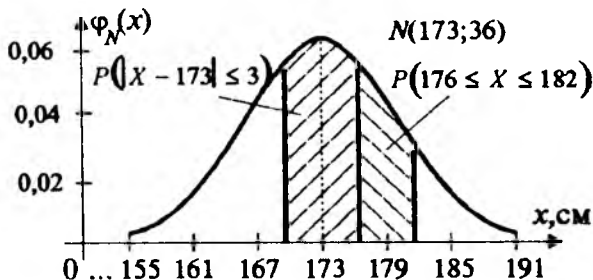


Рис. 9

«Правило трех сигм» для случайной величины X можно сформулировать следующим образом:

Практически достоверно, что рост мужчин данной возрастной группы заключен в границах от $a - 3\sigma = 173 - 3 \cdot 6 = 155$ до $a + 3\sigma = 173 + 3 \cdot 6 = 191$, т.е. $155 \leq X \leq 191$ (см).

Упражнения

1. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед. 1. Найти вероятность того, что цена акции: а) не выше 15,3 ден. ед.; б) не ниже 15,4 ден. ед.; в) от 14,9 до 15,3 ден. ед. 2. С помощью правила трех сигм найти границы, в которых будет находиться текущая цена акции.

2. Цена некой ценной бумаги нормально распределена. В течение последнего года 20% рабочих дней она была ниже 88 ден. ед., а 75% - выше 90 ден. ед. Найти:

а) математическое ожидание и среднее квадратическое отклонение цены ценной бумаги; б) вероятность того, что в день покупки цена будет заключена в пределах от 83 до 96 ден. ед.; в) с надежностью 0,95 определить максимальное отклонение цены ценной бумаги от среднего (прогнозного) значения (по абсолютной величине).

В силу особенностей нормального закона распределения, отмеченных в начале п. 1 и ниже, он занимает центральное место в теории и практике вероятностно-статистических методов. Большое теоретическое значение нормального закона состоит в том, что с его помощью получен ряд важных распределений: *логарифмически-нормальное распределение, χ^2 – распределение, распределение Стюдента, распределение Фишера-Снедекора и др.*

Из них ниже приведем определения и числовые характеристики логарифмически-нормального распределения, а другие законы из перечисленных рассматривать не будем. С ними можно ознакомиться в более объемной и специальной учебной литературе по теории вероятностей и математической статистике. Отметим, что эти законы составляют необходимый аппарат для

построения в дальнейшем статистических критериев и оценок, применяемых в математической статистике.

Логарифмически-нормальное распределение

Определение 6. Непрерывная случайная величина X имеет логарифмически-нормальное (сокращенно логнормальное распределение), если ее логарифм подчинен нормальному закону.

Так как при $x > 0$ неравенства $X < x$ и $\ln X < \ln x$ равносильны, то функция распределения логнормального распределения совпадает с функцией нормального распределения для случайной величины $\ln X$, т.е.

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(t-\ln a)^2}{2\sigma^2}} dt. \quad (8)$$

Дифференцируя (8) по x , получим выражение для функции плотности логнормального распределения (рис. 10)

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\ln a)^2}{2\sigma^2}}. \quad (9)$$

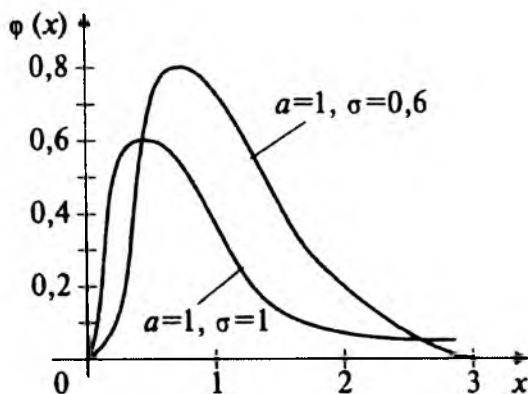


Рис. 10

Можно доказать, что числовые характеристики случайной величины X , распределенной по логнормальному закону (9), имеют вид: математическое ожидание

$$M(X) = ae^{\frac{\sigma^2}{2}},$$

а дисперсия

$$D(X) = a^2 e^{\sigma^2} (e^{\sigma^2} - 1).$$

Логнормальное распределение используется для описания распределения доходов, банковских вкладов, цен активов, месячной заработной платы, посевных площадей под разные культуры, долговечности изделий в режиме износа и старения и др.

Пример 5. Проведенное исследование показало, что вклады населения в данном банке могут быть описаны случайной величиной X , распределенной по логнормальному закону (9) с параметрами $a = 530$, $\sigma^2 = 0.64$.

Найти: а) средний размер вклада; б) долю вкладчиков, размер вклада которых составляет не менее 1000 ден. ед.

Решение

а) Найдем средний размер вклада, т.е.

$$M(X) = ae^{\frac{\sigma^2}{2}} = 530e^{\frac{0.64}{2}} = 730 \text{ (ден.ед.)}.$$

б) Доля вкладчиков, размер вклада которых составляет не менее 1000 ден. ед., есть

$$p\{X \geq 1000\} = 1 - p\{X < 1000\} = 1 - F(1000).$$

При определении $F(1000)$ воспользуемся тем, что функция логнормального распределения случайной величины X совпадает с функцией нормального распределения случайной величины $\ln X$, т.е. с учетом (3) имеем:

$$F(x) = \frac{1}{2} + \Phi\left(\frac{\ln x - \ln a}{\sigma}\right);$$

$$F(1000) = \frac{1}{2} + \Phi\left(\frac{\ln 1000 - \ln 530}{\sqrt{0.64}}\right) = \frac{1}{2} + \Phi(0.79) = 0.5 + 0.2852 = 0.7852.$$

Теперь

$$p\{X \geq 1000\} = 1 - F(1000) = 1 - 0.7852 = 0.2148.$$

Упражнение

Месячный доход семей можно рассматривать как случайную величину, распределенную по логнормальному закону. Полагая, что математическое ожидание этой случайной величины равно 1000, а среднеквадратическое отклонение 800, найти долю семей, имеющих доход: а) не менее 1000; б) менее 500.

2. Закон больших чисел и центральная предельная теорема

2.1. Закон больших чисел

Из повседневного опыта известно, что массовые случайные явления обладают свойствами устойчивости средних. Это означает, что при независимых испытаниях случайной величины X среднее арифметическое

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

полученных значений при больших n стабилизируется. Случайные колебания значений каждого испытания взаимно компенсируются и случайная величина $\frac{x_1 + x_2 + \dots + x_n}{n}$, где x_i есть i -е испытание ($i=1, 2, \dots, n$) величины

X , при больших n теряет свой случайный характер. Теоремы, описывающие такие ситуации, называются *законами больших чисел*.

Мы строго сформулируем и докажем два варианта закона больших чисел - теоремы Бернулли и Чебышева.

В основе доказательства этих теорем лежит неравенство Чебышева, составляющего содержание следующей леммы.

Лемма (неравенство Чебышева). Пусть X - произвольная случайная величина; $M(X)$ и $D(X)$ - соответственно, ее математическое ожидание и дисперсия, $\varepsilon > 0$ - произвольное число. Тогда справедливо неравенство

$$P\{|X - M(X)| < \varepsilon\} > 1 - \frac{D(X)}{\varepsilon^2}, \quad (10)$$

где $p\{|X - M(X)| < \varepsilon\}$ означает вероятность того, что отклонение случайной величины X от своего математического ожидания меньше, чем ε .

Комментарий к лемме

Неравенство (10) и теорема об устойчивости среднего арифметического (см. ниже) доказаны П. Л. Чебышевым.

Доказательство. Пусть X - дискретная случайная величина, распределение, которой задано таблицей

x_1	x_2	...	x_n
p_1	p_2	...	p_n

Имеем

$$D(X) = \sum_{i=1}^n (x_i - MX)^2 p_i = \sum^I (x_i - M(X))^2 p_i + \sum^{II} (x_i - M(X))^2 p_i,$$

где $\sum (x_i - M(X))^2 p_i$ означает сумму всех слагаемых вида $(x_i - M(X))^2 p_i$ таких, что $(x_i - M(X))^2 < \varepsilon^2$, а $\sum^{II} (x_i - M(X))^2 p_i$ - сумму всех слагаемых вида $(x_i - M(X))^2 p_i$ таких, что $(x_i - M(X))^2 \geq \varepsilon^2$. При этих условиях $\sum^{II} (x_i - M(X))^2 p_i \geq \sum^I (x_i - M(X))^2 p_i$, $\sum^I (x_i - M(X))^2 p_i \geq 0$, откуда

$$D(X) \geq \varepsilon^2 \sum^{II} p_i,$$

где под знаком $\sum^{II}$ собраны вероятности всех тех значений x_i , для которых $(x_i - M(X))^2 \geq \varepsilon^2$. Поэтому $\sum^{II} p_i = p(|x_i - M(X)| \geq \varepsilon)$ и следовательно,

$$DX \geq \varepsilon^2 p(|X - MX| \geq \varepsilon) = \varepsilon^2 [1 - p(|X - MX| < \varepsilon)].$$

Из последнего соотношения получим

$$p\{|X - M(X)| < \varepsilon\} > 1 - \frac{D(X)}{\varepsilon^2}.$$

Аналогично доказывается неравенство Чебышева для непрерывной случайной величины X с функцией плотности распределения $f(x)$.

Замечание. Неравенство Чебышева $P(|X - M(X)| < \varepsilon) > 1 - \frac{D(X)}{\varepsilon^2}$ можно записать в другом виде

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}.$$

Теорема (теорема Бернулли). Пусть k - количество наступлений события A в серии из n испытаний схемы Бернулли, p - вероятность наступления события A в одном испытании. Тогда

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} = 1. \quad (11)$$

Комментарий к теореме Бернулли

Теорема Бернулли утверждает, что вероятность малого (меньшего, чем ε) отклонения вероятности p от частоты $\frac{k}{n}$ велика (при большом n). Иными словами, почти всегда будет наблюдаться малое отклонение частоты наступления события A в n испытаниях от вероятности наступления A в одном испытании.

В частности, теорема объясняет, почему при многократном бросании монеты количество гербов составляет примерно половину от числа бросаний.

Теорема Бернулли была исторически первым, строго доказанным математическим фактом из числа тех утверждений, которые носят название закона больших чисел. Доказательство дано швейцарским математиком Я. Бернулли.

Доказательство теоремы Бернулли. Как известно, для случайной величины $X = k$, имеющее биномиальное распределение

$$M(X) = np, \quad D(X) = npq.$$

Следовательно,

$$M\left(\frac{k}{n}\right) = p, \quad D\left(\frac{k}{n}\right) = \frac{1}{n^2} \cdot npq = \frac{pq}{n}.$$

Здесь использованы свойства математического ожидания и дисперсии: $M(\lambda X) = \lambda MX$, $D(\lambda X) = \lambda^2 DX$.

Запишем неравенство Чебышева для $X = \frac{k}{n}$:

$$p\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} > 1 - \frac{pq}{n\varepsilon^2}. \quad (12)$$

Правая часть неравенства (12) при $n \rightarrow \infty$ стремится к единице, поэтому и

$$\lim_{n \rightarrow \infty} p\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} = 1.$$

Теорема 2 (теорема Чебышева). Пусть X_i ($i = 1, 2, \dots, n$) - попарно независимые случайные величины, имеющие одинаковые распределения: $M(X_i) = a$, $D(X_i) = \sigma^2$. Тогда имеет место соотношение

$$\lim_{n \rightarrow \infty} p\left\{\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right\} = 1. \quad (13)$$

Комментарий к теореме Чебышева

Можно считать, что дана одна случайная величина X , которая (независимо) испытывается n раз; случайное значение i -го испытания определяет случайную величину X_i . Теорема Чебышева утверждает, что малое (меньшее, чем ε) отклонение среднего арифметического $\frac{X_1 + X_2 + \dots + X_n}{n}$ от математического ожидания a весьма вероятно. Иными словами, почти всегда будет наблюдаться малое отклонение (при больших n).

Доказательство теоремы Чебышева. Имеем

$$M\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \frac{1}{n}(M(X_1) + M(X_2) + \dots + M(X_n)) = \frac{1}{n}na,$$

$$D\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \frac{1}{n^2}(D(X_1) + D(X_2) + \dots + D(X_n)) = \frac{1}{n^2}\sigma^2 n = \frac{\sigma^2}{n}.$$

Здесь использованы формулы $M(\lambda X) = \lambda MX$, $D(\lambda X) = \lambda^2 DX$.

Неравенство Чебышева для $X = \frac{X_1 + X_2 + \dots + X_n}{n}$ дает

$$p \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{\sigma^2}{n\varepsilon^2} \right| < \varepsilon \right\} > 1 - \frac{\sigma^2}{n\varepsilon^2}. \quad (14)$$

Правая часть неравенства (14) при $n \rightarrow \infty$ стремится к единице; отсюда и следует утверждение (13).

Теорему Бернулли можно рассматривать как частный случай теоремы Чебышева, если считать, что $X_i = k_{(i)}$ ($i = 1, 2, \dots, n$), где $k_{(i)}$ - количество наступлений события A в i -м испытании схемы Бернулли. Тогда частота наступления события A есть

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} (k_{(1)} + k_{(2)} + \dots + k_{(n)}) = \frac{k}{n}; \quad a = p; \quad \sigma^2 = np.$$

Пример 6. Для определения средней продолжительности горения электроламп в партии из 200 одинаковых ящиков было взято на выборку по одной лампе из каждого ящика. Оценить вероятность того, что средняя продолжительность горения отобранных 200 электроламп отличается от средней продолжительности горения ламп во всей партии не более чем на 5 ч (по абсолютной величине), если известно, что среднее квадратическое отклонение продолжительности горения ламп в каждом ящике меньше 7 ч.

Решение

Пусть X_i - продолжительность горения электролампы, взятой из i -го ящика (ч). По условию дисперсия $D(X_i) < 7^2 = 49$. Очевидно, что средняя продолжительность горения отобранных ламп равна

$$\frac{X_1 + X_2 + \dots + X_{200}}{200},$$

а средняя продолжительность горения ламп во всей партии

$$\frac{M(X_1) + M(X_2) + \dots + M(X_{200})}{200}.$$

Тогда вероятность искомого события:

$$p \left\{ \left| \frac{X_1 + X_2 + \dots + X_{200}}{200} - \frac{M(X_1) + M(X_2) + \dots + M(X_{200})}{200} \right| < 5 \right\} > 1 - \frac{49}{200 \cdot 5^2} \approx 0.9902,$$

т.е. не менее, чем 0,9902.

Пример 7. Сколько надо провести измерений данной величины, чтобы с вероятностью не менее 0,95 гарантировать отклонение средней арифметической этих измерений от истинного значения величины не более, чем на 1 (по абсолютной величине), если среднеквадратическое отклонение каждого из измерений не превосходит 5?

Решение

Пусть X_i – результат i -го измерения ($i = 1, 2, \dots, n$) – истинное значение величины, т.е. $M(X_i) = a$ при любом i . Необходимо найти n , при котором

$$P\left\{\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < 1\right\} > 0,95.$$

Данное неравенство выполняется, если

$$1 - \frac{D(X)}{n\varepsilon^2} = 1 - \frac{5^2}{n \cdot 1^2} > 0,95.$$

Откуда $\frac{25}{n} < 0,05$ и $n > \frac{25}{0,05} = 500$, т.е. потребуется не менее 501 измерений.

Упражнения

1. Выход цыплят в инкубаторе составляет в среднем 70% числа заложенных яиц. Сколько нужно заложить яиц, чтобы с вероятностью, не меньшей 0,95, ожидать, что отклонение числа вылупившихся цыплят от математического ожидания их не превышало 50 (по абсолютной величине)? Решить задачу с помощью: а) неравенства Чебышева; б) интегральной теоремы Муавра - Лапласа.

2. Опыт работы страховой компании показывает, что страховой случай приходится примерно на каждый пятый договор. Оценить с помощью неравенства Чебышева необходимое количество договоров, которые следует заключить, чтобы с вероятностью 0,9 можно было утверждать, что доля страховых случаев отклонится от 0,1 не более чем на 0,01 (по абсолютной величине). Уточнить ответ с помощью следствия из интегральной теоремы Муавра - Лапласа.

2.2. Центральная предельная теорема

Кроме законов больших чисел, описывающих устойчивость средних значений и изложенных в п. 21, в теории вероятностей имеет место еще одно замечательное явление. Как и законы больших чисел, это явление заключается в том, что при большом количестве случайных слагаемых, каждое из которых вносит лишь небольшой вклад в общую сумму, распределение каждого из слагаемых не влияет на суммарный результат. Точнее, при указанных условиях вид распределения суммы не зависит от распределения слагаемых.

Более строгое утверждение сформулировано в следующей теореме.

Теорема (центральная предельная теорема Ляпунова). Пусть

$$X_1, X_2, \dots, X_n$$

- одинаково распределенные независимые случайные величины с математическим ожиданием $M(X_i) = a$ и дисперсией $D(X_i) = \sigma^2$. Тогда при большом n распределение суммы $Y = X_1 + X_2 + \dots + X_n$ близко к нормальному распределению.

Комментарии к теореме Ляпунова

1. Когда говорят, что последовательность распределений $Z_1, Z_2, \dots, Z_n$ стремится к некоторому распределению Z , имеют в виду, что функции плотности $f_i(x)$ распределений Z_i стремятся к функции плотности $f(x)$ распределения Z .

2. Так как $M(Y) = \sum M(X_i) = na$ и $D(Y) = \sum D(X_i) = n\sigma^2$, то величины $a = \frac{M(X)}{n}$, $\sigma^2 = \frac{D(X)}{n}$ малы при больших n . Величины X_i вносят «равномерно малый вклад», о чем шла речь выше.

3. Утверждение о нормальном законе распределения суммы

$$X_1 + X_2 + \dots + X_n$$

справедливо при менее ограничительных условиях, чем те, которые фигурируют в условии теоремы. В частности, справедлив более сильный вариант

теоремы Ляпунова, устанавливающий, что сумма $X_1 + X_2 + \dots + X_n$ имеет нормальное распределение при весьма общих предположениях относительно величин $X_1, X_2, \dots, X_n$.

Доказательство центральной предельной теоремы использует аппарат характеристических функций и в общих чертах следует такой схеме. Каждой случайной величине соответствует характеристическая функция, сумме случайных величин соответствует произведение характеристических функций. Это произведение при неограниченном увеличении числа n слагаемых стремится к некоторой функции, которая оказывается характеристической функцией нормального распределения. Отсюда и следует утверждение центральной предельной теоремы (Впрочем, важным обстоятельством, пропущенным в вышеуказанных рассуждениях, является тот факт, что если последовательность характеристических функций $f_i(x)$ сходится к функции $f(x)$, то последовательность распределений, соответствующих функциям $f_i(x)$, сходится к распределению, соответствующему функции $f(x)$; в действительности именно доказательство этого факта является самым трудным местом в доказательстве предельной теоремы).

Понятие характеристической функции распределения, играющее столь важную роль в доказательстве центральной теоремы, является сложным понятием математики. Оно связано с важным общематематическим понятием преобразования Фурье. Это выходит за рамки математического образования студентов-экономистов. Поэтому выше вместо строгого доказательства теоремы мы ограничивались приведением схемы доказательства в общих чертах.

Вопросы для самопроверки

1. Перечислите наиболее известные дискретные законы распределения.
2. Приведите определения, некоторые числовые характеристики наиболее известных дискретных законов распределения.

3. Перечислите наиболее известные непрерывные законы распределения.

4. Приведите определения, некоторые числовые характеристики наиболее известных непрерывных законов распределения.

5. Сформулируйте лемму (неравенства) Чебышева.

6. Приведите обе формы записи неравенства Чебышева.

7. Сформулируйте теорему Чебышева и прокомментируйте ее.

8. Сформулируйте теорему Бернулли.

9. Сформулируйте центральную предельную теорему Ляпунова и прокомментируйте ее.

10. На чем основывается доказательство центральной предельной теоремы Ляпунова.

§7. Статистические оценки параметров и требования, предъявляемые к ним

Ключевые слова: генеральная совокупность, случайная выборка, повторная и бесповторная выборки, статистическая оценка параметра, несмещенная оценка, смещенная оценка, состоятельная оценка, эффективная оценка, выборочная средняя, выборочная дисперсия, доверительный интервал, стандартные статистические распределения, гистограмма, размах выборки, полигон частот, оценки метода моментов, выборочные начальные моменты, выборочные центральные моменты, оценки метода максимального правдоподобия, функция правдоподобия, логарифмическая функция правдоподобия, функция правдоподобия Фишера.

1. Основные понятия выборочного метода

Математическая статистика является одним из наиболее широко используемых на практике разделов прикладной математики.

Определение 1. Генеральной совокупностью называется вероятностное пространство и определенная на этом пространстве случайная величина X .

Проще говоря, под генеральной совокупностью подразумевают всю совокупность изучаемых объектов, или все множество возможных значений изучаемой случайной величины.

Определение 2. Случайной выборкой или просто выборкой объема n значений случайной величины X называется набор чисел

$$x_1, x_2, \dots, x_n,$$

являющихся реализациями случайной величины X , полученными в ходе n соответствующих независимых наблюдений или экспериментов.

Иногда выборку также называют статистическими данными, или статистикой реализаций случайной величины X , или просто статистикой.

Различают повторные и бесповторные выборки. Говорят о собственно-случайных, механических, серийных и типических выборках.

Основной задачей математической статистики является оценка и анализ параметров распределения изучаемой случайной величины, или самого вида этого распределения (непараметрическое оценивание) по данным выборки ее значений. Часто отмечают, что основная задача математической статистики является, в некотором смысле, обратной к задачам теории вероятностей.

Основными целями оценивания являются:

- 1) прогнозирование поведения изучаемой случайной величины в будущем;
- 2) проверка соответствия значений полученных оценок некоторым регламентированным характеристикам.

И то и другое часто может служить обоснованием выбора наиболее оптимального варианта управленческих решений.

Определение 3. Истинное значение того или иного параметра распределения изучаемой случайной величины X называют его теоретическим, или генеральным значением.

Определение 4. Статистической оценкой некоторого параметра Θ распределения случайной величины X называется функция, определенная на множестве выборок значений этой случайной величины

$$\Theta^* = \Theta^*(x_1, x_2, \dots, x_n).$$

значения которой, в некотором статистическом смысле, близки к теоретическому значению θ .

Определение 5. Конкретное значение статистической оценки на данной конкретной выборке называют выборочным значением этой оценки, или точечным значением, или выборочной оценкой.

Важно четко понимать разницу между теоретическим значением параметра и его выборочными оценками. Теоретическое значение того или иного

параметра распределения случайной величины в общем случае определяется только его плотностью распределения, т.е. бесконечной информацией. Данные же выборки всегда конечны. Поэтому никогда нет возможности найти точное теоретическое значение параметра.

Очень важно также, что выборочное значение параметра само является случайной величиной, поскольку рассчитывается по данным случайной выборки.

Иногда приходится говорить не о конкретной выборке, а вообще, абстрактно. Как, например, при выводе различных статистических формул и доказательств теорем. Тогда приходится использовать следующее понятие.

Определение 6. Теоретической выборкой объема n значений случайной величины X будем называть совокупность независимых случайных величин

$$X_1, X_2, \dots, X_n,$$

каждая из которых имеет то же распределение, что и X , т.е. $X = X_i$, $i = \overline{1, n}$.

2. Требования, предъявляемые к статистическим оценкам или свойства статистических оценок

Для оценок, которые предполагается использовать на практике, очевидно, очень желательны следующие свойства.

Определение 7. Оценка Θ^* некоторого параметра Θ называется несмещенной, если

$$M(\Theta^*) = \Theta.$$

Несмещенность оценки означает, что она не дает какой-либо регулярной (постоянной) ошибки. Иногда на практике используют и смещенные оценки.

Определение 8. Оценка Θ_n^* называется состоятельной, если для любого $\varepsilon > 0$ справедливо

$$\lim_{n \rightarrow \infty} P\{|\Theta_n^* - \Theta| \leq \varepsilon\} = 1.$$

Таким образом, состоятельность оценки означает, что чем больше объем выборки, тем относительно точнее выборочная оценка. Примеры несостоятельных оценок достаточно редки.

Справедлива теорема.

Теорема 1. Если Θ_n^* - несмещенная оценка параметра Θ и

$$\lim_{n \rightarrow \infty} D(\Theta_n^*) = 0,$$

то Θ_n^* - состоятельна.

Доказательство этой теоремы легко вытекает из неравенства Чебышева.

Определение 9. Оценка называется эффективной, если она имеет минимальную дисперсию среди всех других оценок данного параметра.

3. Оценка математического ожидания случайной величины

Из теории вероятностей известно, что среди всех параметров распределения случайной величины важнейшую роль играют математическое ожидание и дисперсия. Напомним, что нормальное распределение полностью задается этими двумя параметрами. Поэтому и в математической статистике их оценки занимают центральное место.

Поэтому, для решения основной задачи математической статистики зачастую бывает достаточно оценить именно эти два параметра.

Пусть X - некоторая случайная величина. Как всегда, будем обозначать

$$M(X) = a, \quad D(X) = \sigma^2,$$

где a, σ - неизвестны. Пусть имеется выборка значений этой случайной величины

$$x_1, x_2, \dots, x_n.$$

Вспоминая определение и смысл математического ожидания в качестве его оценки естественно предложить следующую

$$a^* = \bar{X} = \frac{\sum_{i=1}^n x_i}{n}.$$

Величину $\bar{X}$ называют выборочным средним, и действительно используют в качестве оценки математического ожидания. Изучим ее свойства как статистической оценки.

Оказывается

$$M(\bar{X}) = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right).$$

Переходя к теоретической выборке и используя свойства математического ожидания, имеем

$$M(\bar{X}) = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n M(X_i) = \frac{\sum_{i=1}^n a}{n} = a.$$

Таким образом, выборочное среднее - несмещенная оценка дисперсии.

Рассмотрим состоятельность. Используя свойства дисперсии, имеем

$$D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{\sum_{i=1}^n D(X_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0, \text{ при } n \rightarrow \infty.$$

Тогда по теореме 1, $\bar{X}$ - состоятельная оценка математического ожидания.

Рассмотрим эффективность. Все возможные линейные оценки математического ожидания имеют вид

$$\Theta^*(C_1, C_2, \dots, C_n) = \sum_{i=1}^n C_i x_i,$$

где C_i , - любые весовые коэффициента, т.е. $\sum_{i=1}^n C_i = 1$.

Отметим, что все такие оценки и несмещены, и состоятельны.

Найдем, при каких значениях коэффициентов C_i дисперсия соответствующей взвешенной суммы будет наименьшей, т.е. имеем задачу

$$D(\Theta^*(C_1, C_2, \dots, C_n)) \rightarrow \min.$$

при ограничении

$$\sum_{i=1}^n C_i = 1.$$

Это задача на условный экстремум, которую мы решим методом множителей Лагранжа. Составляем функцию Лагранжа:

$$L = D(\Theta^*(C_1, C_2, \dots, C_n)) + \lambda \left(\sum_{i=1}^n C_i - 1 \right),$$

где

$$D(\Theta^*(C_1, C_2, \dots, C_n)) = D\left(\sum_{i=1}^n C_i x_i\right) = \sum_{i=1}^n C_i^2 D(X_i) = \sigma^2 \sum_{i=1}^n C_i^2.$$

Тогда

$$L = \sigma^2 \sum_{i=1}^n C_i^2 + \lambda \left(\sum_{i=1}^n C_i - 1 \right).$$

Система необходимых условий минимума функции Лагранжа

$$\begin{cases} \frac{\partial L}{\partial C_i} = 2C_i \sigma^2 + \lambda = 0, i = \overline{1, n}, \\ \frac{\partial L}{\partial \lambda} = \sum_{i=1}^n C_i - 1 = 0. \end{cases} \Rightarrow$$

$$\Rightarrow \sum_{i=1}^n \frac{\lambda}{2\sigma^2} = \sum_{i=1}^n C_i - 1 = -1, \lambda = -\frac{2\sigma^2}{n}, \Rightarrow C_i = \frac{1}{n}.$$

Таким образом, минимум дисперсии достигается при тех самых коэффициентах, которые входят в X , т.е. X - эффективная оценка.

4. Оценки дисперсии случайной величины

Возможны два случая:

1) a - известно. Тогда в качестве оценки дисперсии естественно предложить следующую функцию

$$S_a^2 = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2.$$

Изучаем несмещенность

$$M(S_a^2) = M\left(\frac{1}{n} \sum_{i=1}^n (x_i - a)^2\right) = \left(\frac{1}{n} \sum_{i=1}^n M(X_i - a)^2\right) = \frac{1}{n} \sum_{i=1}^n D(X_i) = D(X) = \sigma^2,$$

т.е. рассматриваемая оценка является несмещенной.

Рассмотрим состоятельность этой оценки

$$\begin{aligned} \lim_{n \rightarrow \infty} D(S_a^2) &= \lim_{n \rightarrow \infty} D\left(\frac{1}{n} \sum_{i=1}^n (x_i - a)^2\right) = \lim_{n \rightarrow \infty} \frac{D\left(\sum_{i=1}^n X_i^2 - 2aX_i - a^2\right)}{n^2} = \\ &= \lim_{n \rightarrow \infty} \frac{nD(X^2 - 2aX)}{n^2} = 0, \end{aligned}$$

если только

$$D(X^2 - 2aX) < \infty,$$

т.е. дисперсия случайной величины $(X^2 - 2aX)$ конечна, что является весьма не обременительным предположением.

Эффективность доказывается по схеме, аналогичной той, что была применена для математического ожидания.

2) a - неизвестно. При этом, естественно предложить оценку

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2.$$

Изучаем ее несмещенность

$$\begin{aligned} M(\tilde{S}^2) &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = M\left(\frac{1}{n} \sum_{i=1}^n (X_i - a - (\bar{X} - a))^2\right) = \\ &= M\left(\frac{1}{n} \sum_{i=1}^n \left((X_i - a)^2 - 2(\bar{X} - a)(X_i - a) + (\bar{X} - a)^2\right)\right) = \\ &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \frac{2}{n} (\bar{X} - a) \sum_{i=1}^n (X_i - a) + (\bar{X} - a)^2\right) = \\ &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - 2(\bar{X} - a)(\bar{X} - a) + (\bar{X} - a)^2\right) = \\ &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - (\bar{X} - a)^2\right) = M\left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \left(\frac{1}{n} \sum_{i=1}^n (X_i - a)\right)^2\right) = \end{aligned}$$

$$= M \left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \frac{1}{n^2} \sum_{i=1}^n (X_i - a) \sum_{j=1}^n (X_j - a) \right) =$$

$$= M \left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - a)(X_j - a) \right) = \dots,$$

здесь при $i \neq j$,

$$M((X_i - a)(X_j - a)) = 0,$$

так как X_i - независимые случайные величины

$$\frac{1}{n} \sum_{i=1}^n M(X_i - a)^2 - \frac{1}{n^2} M \left(\sum_{i=1}^n (X_i - a) \right)^2 = \frac{1}{n} \cdot n \cdot \sigma^2 - \frac{1}{n^2} \cdot n \cdot \sigma^2 = \sigma^2 - \frac{\sigma^2}{n} \neq \sigma^2.$$

Таким образом, рассматриваемая оценка является смещенной. Однако увидим, что это легко исправить, перейдя к оценке

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1},$$

которую так и называют несмещенной или исправленной оценкой дисперсии.

Аналогично рассмотренному выше, не сложно доказать состоятельность как $\tilde{S}^2$, так и S^2 .

Очевидно, дисперсия оценки S^2 больше, чем у $\tilde{S}^2$, и поэтому S^2 не является эффективной. Тем не менее, свойство несмещенности считается практически более значимым, и для оценки дисперсии обычно используется именно S^2 . Однако и $\tilde{S}^2$ находит практическое применение, в частности, величина $\tilde{S}$ называется стандартной ошибкой выборки.

5. Оценка доли признака

Для весьма широкого спектра задач важным является следующий параметр.

Определение 10. Генеральной долей признака называется величина

$$p = \frac{M}{N},$$

где N - общее количество объектов, составляющих некоторую генеральную совокупность. M - количество объектов из этой совокупности, обладающих некоторым соответствующим признаком. Например, N - количество всех работников на предприятии, M - число женщин из них.

Во многих маркетинговых, социологических и прочих исследованиях требуется оценить генеральную долю признака. Например, долю потенциальных покупателей, предпочитающих некоторый определенный товар. При этом невозможно протестировать всю совокупность N , а следует сделать выводы только на основании некоторой выборки объема n .

Определение 11. Выборочной долей признака называется величина

$$w = \frac{m}{n},$$

где n - общее количество протестированных объектов из соответствующей генеральной совокупности, m - количество объектов из n , обладающих интересующим нас признаком.

Следует ли считать, что w можно использовать, как оценку p ?

Во-первых, вспомним, что выборки бывают повторные и бесповторные. При оценке доли признака, тип выборки имеет большое значение. В случае повторной выборки, тестируемый объект возвращается в генеральную совокупность, и может оказаться выбранным еще, и еще раз. Например, на бензозаправке решили выяснить, какова доля иномарок среди всех автомобилей. Ясно, что при этом за исследуемый день возможно неоднократное посещение одной и той же иномаркой данной бензозаправки. В случае бесповторной выборки тестируемый объект помечается и в дальнейшем не учитывается, т.е. исключается из генеральной совокупности. Чаще имеют место именно такие ситуации.

Не сложно понять, что при бесконечном количестве объектов в генеральной совокупности разницы между повторной и бесповторной выборками нет.

Рассмотрим сначала более простой случай повторной выборки.

Тогда каждый раз вероятность выбора объекта, обладающего признаком, по классическому определению вероятности, составляет

$$p = \frac{M}{N},$$

а вероятность, что из n отборов объектов с признаком попадется ровно m , выражается формулой Бернулли, т.е. в этом случае $\frac{m}{n}$ - случайная величина имеющая, так называемое, дробно биномиальное распределение, для которого, как известно,

$$M\left(\frac{m}{n}\right) = p,$$

т.е. w - несмещенная оценка генеральной доли. Дисперсия дробно биномиальной случайной величины

$$D\left(\frac{m}{n}\right) = \frac{pq}{n} \rightarrow 0,$$

при $n \rightarrow \infty$, т.е. w - состоятельная оценка.

В случае бесповторной выборки эксперимент соответствует «урновой» схеме, и m имеет, так называемое, гипергеометрическое распределение. Тогда

$$M\left(\frac{m}{n}\right) = \frac{1}{n} \sum_{i=0}^n i \frac{C_{N-n}^{n-i}}{C_N^n},$$

и можно доказать, что

$$M\left(\frac{m}{n}\right) = \frac{M}{N} = p,$$

т.е. опять w - несмещенная оценка. При этом

$$D\left(\frac{m}{n}\right) = \frac{pq}{n} \cdot \frac{N-n}{N-1} \rightarrow 0,$$

при $n \rightarrow \infty$. А, точнее говоря, предельно возможное значение для n это N и при $n = N$ эта дисперсия равна нулю, т.е. опять w - состоятельная оценка.

6. Стандартные статистические распределения и их критические границы

Во многих формулах математической статистики используются, так называемые, стандартные статистические распределения. К ним, прежде всего,

относят: стандартное гауссовское распределение, а также распределения Пирсона, Стьюдента и Фишера.

Определение 12. Говорят, что случайная величина X имеет *стандартное, гауссовское распределение*, если

$$X \sim N(0,1).$$

Определение 13. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ - независимые стандартные гауссовские случайные величины, тогда распределение случайной величины

$$z = \sum_{i=1}^n \varepsilon_i^2$$

называют *распределением Пирсона* с n степенями свободы и пишут

$$z \sim \chi^2(n).$$

График плотности χ^2 - распределения зависит от числа степеней свободы, однако схематически он имеет вид, представленный на рис. 1.

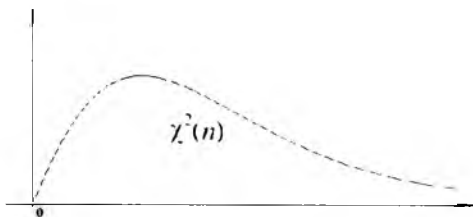


Рис. 1

Определение 14. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ - независимые стандартные гауссовские случайные величины, тогда распределение случайной величины

$$z = \frac{\varepsilon_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}}$$

называют *распределением Стьюдента* с n степенями свободы, и пишут

$$z \sim t(n).$$

График плотности t - распределения зависит от числа степеней свободы, однако в целом он подобен стандартному гауссовскому.

Замечание. Распределение Стьюдента с n степенями свободы можно было бы определить и так

$$t = \frac{\varepsilon_0}{\sqrt{\frac{1}{n} \chi^2}}$$

Определение 14. Пусть, $z_1 \sim \chi^2(n)$, $z_2 \sim \chi^2(m)$ - независимые случайные величины, тогда распределение случайной величины

$$z = \frac{\frac{1}{n} z_1}{\frac{1}{m} z_2}$$

называют *распределением Фишера* с n степенями свободы числителя и m знаменателя, и пишут

$$z = F\left(\frac{n}{m}\right).$$

График плотности F - распределения зависит от числа степеней свободы, однако в целом он подобен распределению Пирсона.

Определение 15. Будем называть распределение симметричным, если график его плотности симметричен относительно оси ординат.

Так, например, стандартное гауссовское и распределение Стьюдента - симметричные распределения, а Пирсона и Фишера - нет.

Огромную важность имеют следующие понятия.

Определение 16. Односторонней критической границей уровня значимости α некоторого распределения $P(x)$ случайной величины X (рис. 2), называется такое число z_α , что

$$P\{x \leq z_\alpha\} = 1 - \alpha.$$

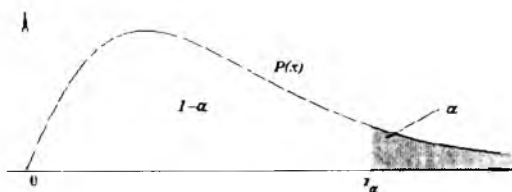


Рис. 2

Определение 17. Двухсторонней критической границей уровня значимости α некоторого симметричного распределения $P(x)$ случайной величины X (рис. 3) называется такое число $z_{\alpha/2}$, что

$$P\{-z_{\alpha/2} \leq x \leq z_{\alpha/2}\} = 1 - \alpha.$$

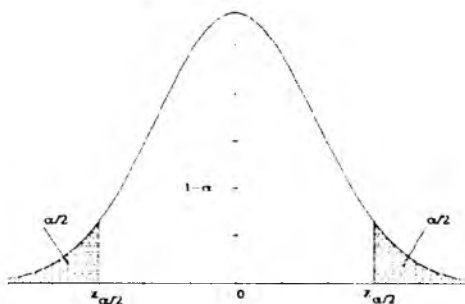


Рис. 3

Отметим, что односторонняя критическая граница уровня значимости $\alpha/2$ некоторого симметричного распределения одновременно является его двухсторонней границей уровня значимости α .

Критические границы различных уровней значимости, стандартных статистических распределений приводятся в виде справочных таблиц, в соответствующей литературе (например, в учебниках и задачниках по математической статистике).

7. Понятие доверительного интервала

Как уже говорилось, конкретное значение оценки некоторого параметра, рассчитанное по числовым данным конкретной выборки, называют точечной оценкой данного параметра.

Поскольку у нас нет информации о точности этих оценок, то на практике конкретные управленческие решения, выработанные на основе сведений о значениях точечных оценок, не являются достаточно обоснованными. То есть, пока мы не установили меры близости между теоретическим значением

интересующего нас параметра и его точечной оценкой, невозможно говорить о допустимости практического использования последней.

Поэтому большую важность в теории статистического оценивания имеет следующее понятие.

Определение 18. Доверительным интервалом уровня значимости α некоторого параметра Θ называется любой интервал $[a, b] \subset R^1$ такой, что

$$P\{\Theta \in [a, b]\} = 1 - \alpha.$$

Величину $\gamma = 1 - \alpha$ при этом называют доверительной вероятностью или уровнем доверия.

Иначе говоря, доверительный интервал - это интервал, в котором с заданной (желаемой) вероятностью содержится теоретическое значение оцениваемого параметра. Обычно строят доверительные интервалы при $\alpha = 0,1$ или чаще всего при $\alpha = 0,05$, или, в особо ответственных случаях, при $\alpha = 0,01$.

Доверительные интервалы часто называют интервальными оценками данного параметра.

Ясно, что доверительный интервал уже несет информацию о мере точности наших знаний об истинном значении параметра. Именно эта информация и может действительно служить основанием для тех или иных конкретных практических выводов.

8. Доверительный интервал для математического ожидания нормальной случайной величины

Рассмотрим задачу построения интервальной оценки для математического ожидания нормально распределенной случайной величины $X \sim N(a, \sigma^2)$.

Во-первых, отметим, что для выборочного среднего

$$M(\bar{X}) = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{nM(X)}{n} = a,$$

$$D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{\sigma^2}{n},$$

т.е.

$$\bar{X} \sim N\left(a, \frac{\sigma^2}{n}\right).$$

Тогда, из свойств математического ожидания и дисперсии, получаем

$$\frac{\sqrt{n}(\bar{X} - a)}{\sigma} \sim N(0,1)$$

- стандартная гауссовская случайная величина.

И если $u_{\alpha/2}$ - двухсторонняя критическая граница уровня значимости α стандартного гауссовского распределения, то с вероятностью $\gamma = 1 - \alpha$ выполняется неравенство

$$-u_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - a)}{\sigma} \leq u_{\alpha/2}.$$

Отсюда с той же вероятностью выполняется неравенство

$$\bar{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \leq a \leq \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha/2},$$

которое и определяет искомый доверительный интервал для математического ожидания нормально распределенной случайной величины при известной дисперсии σ^2 .

Рассмотрим гораздо более важный для практики случай, когда дисперсия неизвестна.

Справедлива теорема.

Теорема 2. Пусть случайная величина $X \sim N(a, \sigma^2)$ и задана выборка ее значений $x_1, x_2, \dots, x_n$, тогда

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi^2(n-1),$$

где S^2 - несмещенная оценка дисперсии случайной величины X по этой выборке.

Поскольку

$$(n-1) \frac{S^2}{\sigma^2} = \frac{\sum (x_i - \bar{X})^2}{\sigma^2},$$

то для случайной величины q

$$q = \frac{\sqrt{n}(\bar{X} - a)}{S} = \frac{\sqrt{n}(\bar{X} - a)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}} = \frac{\frac{\sqrt{n}(\bar{X} - a)}{\sigma}}{\sqrt{\frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sigma^2}}} = \frac{N(0,1)}{\sqrt{\frac{1}{n-1} \chi^2(n-1)}},$$

т.е.

$$q \sim t(n-1).$$

Отсюда получаем, что с вероятностью $\gamma = 1 - \alpha$ выполняется неравенство

$$-t(n-1) \leq \frac{n(\bar{X} - a)}{S} \leq t(n-1),$$

где $t_{\alpha/2}(n-1)$ - табличное значение двухсторонней критической границы уровня значимости α распределения Стьюдента с $(n-1)$ степенями свободы.

Или, с той же вероятностью, неравенство

$$\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \leq a \leq \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}},$$

которое и определяет искомый доверительный интервал для математического ожидания нормально распределенной случайной величины при неизвестной дисперсии.

Отметим, что тот же интервал можно выразить через формулу

$$\bar{X} - t_{\alpha/2}(n-1) \frac{\bar{S}}{\sqrt{n-1}} \leq a \leq \bar{X} + t_{\alpha/2}(n-1) \frac{\bar{S}}{\sqrt{n-1}}.$$

Все полученные доверительные интервалы строились в предположении, что X имеет нормальное распределение. В этом случае $\bar{X}$ имеет в точности нормальное распределение, что и является, в действительности, единственным необходимым условием корректности всех выкладок. Однако и при любом распределении X при $n \rightarrow \infty$, $\bar{X}$ будет асимптотически нор-

мальной случайной величиной в соответствии с центральной предельной теоремой. Эта близость, разумеется, зависит от объема выборки n , но можно сказать, что уже при $n > 5 \div 10$ ее можно считать достаточной.

Пример

Производитель шин заинтересован в получении оценки средней износоустойчивости шин одной особой модели. Над 10 случайно выбранными шинами произвели специальные испытания, оказалось, что средняя длина пробега $\bar{X} = 22500$ миль, при стандартном отклонении $\bar{S} = 3000$ миль. Найдем 95 и 99 % - ные доверительные интервалы для $M(X) = a$, X - длина пробега.

Решение

По таблицам находим $t_{0,025}(10-1) = 2,26$; $t_{0,005}(9) = 3,25$; тогда с вероятностью:

- $\gamma = 0,95$ износоустойчивости будет лежать в пределах

$$22500 \pm 2,26 \cdot \frac{3000}{\sqrt{9}} = 22500 \pm 2260 \text{ миль, отклонение от } \bar{X} \text{ при этом составляет } \pm 10 \% ;$$

- $\gamma = 0,99$ в пределах $22\,500 \pm 3250$, отклонение $\pm 14\%$.

Практически считается, что чем меньше размах доверительного интервала, тем оценка достовернее. В разных случаях считается по-разному, но обычно исходят из того, что при размахе до 10 % от значения $\bar{X}$ полученная информация достаточно точно отражает реальную ситуацию. При размахе более 20 – 30 % достоверность прогноза мала. Очевидно, уменьшить размах интервала можно увеличив объем выборки.

9. Доверительный интервал для дисперсии нормального распределения

Пусть опять случайная величина $X \sim N(a, \sigma^2)$ и задана выборка ее значений $x_1, x_2, \dots, x_n$.

Рассмотрим вопрос построения доверительного интервала для ее дисперсии. Опять разберем два случая:

1) a - известно, тогда

$$n \frac{S_a^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - a}{\sigma} \right)^2 \sim \chi^2(n),$$

поскольку

$$\frac{X_i - a}{\sigma} \sim N(0,1).$$

Тогда с вероятностью $\gamma = 1 - \alpha$ выполняется неравенство

$$\chi_{1-\alpha/2}^2(n) \leq \frac{nS_a^2}{\sigma^2} \leq \chi_{\alpha/2}^2(n),$$

где $\chi_{\alpha/2}^2(n)$, $\chi_{1-\alpha/2}^2(n)$ соответствующие односторонние критические границы распределения Пирсона. Отсюда и получаем требуемый доверительный интервал уровня и значимости α , для истинного значения σ^2

$$\frac{nS_a^2}{\chi_{\alpha/2}^2(n)} \leq \sigma^2 \leq \frac{nS_a^2}{\chi_{1-\alpha/2}^2(n)}.$$

2) a - неизвестно, тогда по теореме 2

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi^2(n-1).$$

И аналогично, доверительный интервал с уровнем доверия $\gamma = 1 - \alpha$ имеет вид

$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}.$$

10. Доверительный интервал для генеральной доли признака

При достаточно большом количестве наблюдений (считается, что при $n > 20$) на основании центральной предельной теоремы, можно считать, что выборочная доля

$$w = \frac{m}{n},$$

имеет распределение, достаточно близкое к нормальному, параметры которого были указаны в п. 5.

Тогда, например, для повторной выборки, с вероятностью $1-\alpha$, выполняются неравенства

$$-u_{\alpha/2} \leq \frac{w-p}{\sqrt{\frac{pq}{n}}} \leq u_{\alpha/2} \Rightarrow w - u_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq p \leq w + u_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}.$$

Последнее из них дает границы искомого доверительного интервала в неявном виде. Для их явного выражения решаем соответствующее квадратное уравнение

$$(w-p)^2 \leq (u_{\alpha/2})^2 \frac{p(1-p)}{n}.$$

Откуда получаем, что эти границы задаются выражениями

$$p_{1,2} = \frac{1}{1 + \frac{u_{\alpha/2}^2}{n}} \left(w + \frac{u_{\alpha/2}^2}{2n} \pm u_{\alpha/2} \sqrt{\frac{w(1-w)}{n} + \left(\frac{u_{\alpha/2}}{2n} \right)^2} \right).$$

Для бесповторной выборки аналогично получаем следующие границы интервальной оценки

$$p_{1,2} = \frac{1}{1 + \frac{u_{\alpha/2}^2}{n} \theta} \left(w + \frac{u_{\alpha/2}^2}{2n} \theta \pm u_{\alpha/2} \sqrt{\frac{w(1-w)}{n} \theta + \left(\frac{u_{\alpha/2}}{2n} \theta \right)^2} \right), \quad \text{где } \theta = \frac{N-n}{N-1}.$$

Если объем данных выборки не достаточно велик ($n < 20$), то для построения границ p_1 и p_2 искомого интервала строят специальные уравнения на основе конкретных выражений для вероятностей значений биномиального или гипергеометрического распределений.

11. Определение необходимого объема выборки

Формулы доверительных интервалов рассмотренных параметров позволяют ответить на следующий весьма важный вопрос: выборку какого объ-

смысла мы должны иметь, чтобы оценить интересующий нас параметр с заданной точностью?

1. Для оценки математического ожидания. Отметим, что размах интервала

$$\lim_{n \rightarrow \infty} \left(t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right) \rightarrow 0,$$

поскольку

$$\lim_{n \rightarrow \infty} t_{\alpha}(n) \rightarrow u_{\alpha}, \quad \lim_{n \rightarrow \infty} S \rightarrow \sigma.$$

Итак, с ростом n доверительный интервал сжимается. Для определения достаточного объема выборки:

- а) сначала находят S по некоторой пробной выборке;
- б) для объема пробной выборки (или большего) принимают соответствующую критическую границу t -распределения;
- в) далее корректируют объем.

Пример. Пусть было отобрано $n = 25$ пакетов некоторого стандартно расфасованного продукта, средний вес которых оказался $\bar{X} = 1020$ г при стандартном отклонении $\tilde{S} = 12$ г. Каким должен быть объем выборки, чтобы установить 99 %- ный интервал с размахом не более ± 5 г.

Решение

Принимаем, например, $t_{0,005}(n-1) = t_{0,005}(40) = 2,797$.

Получаем

$$2,797 \frac{12}{\sqrt{n}} \leq 5 \Rightarrow \sqrt{n} \geq 6,71 \Rightarrow n \geq 45,06 \Rightarrow n = 46.$$

Далее можно скорректировать $t_{0,005}(45) = 2,41$ и т.д.

При известной генеральной дисперсии задача решается еще проще.

2. Для оценки дисперсии. Как известно

$$\lim_{n \rightarrow \infty} \chi_{\alpha}^2(n) \rightarrow n \left(1 - \frac{2}{9n} + u_{\alpha} \sqrt{\frac{2}{9n}} \right),$$

т.е. и для дисперсии размах интервала с ростом n уменьшается. С помощью несложных численных процедур можно подобрать необходимое n . Опять следует использовать данные пробной выборки.

3. Для оценки генеральной доли. Задача решается непосредственным выбором достаточно большого n , если только этот объем выборки может быть практически реализован.

12. Оценка функции распределения

Определение 19. Выборка значений случайной величины, упорядоченная по возрастанию, называется *вариационным рядом*.

В качестве оценки $F_n^*(x)$ функции распределения $F(x)$ случайной величины X естественно предложить следующую кусочно-постоянную функцию

$$F_n^*(x) = \begin{cases} 0 & \text{при } x < x_1, \\ \frac{k-1}{n} & \text{при } x_{k-1} < x < x_k, \\ 1 & \text{при } x > x_n, \end{cases}$$

где x_k – элементы вариационного ряда.

Функция $F_n^*(x)$ называется *эмпирической функцией распределения*. Заметим, она соответствует всем необходимым свойствам функции распределения.

Необходимо определить точность этой оценки. В качестве меры отклонения $F_n^*(x)$ от $F(x)$ обычно рассматривают величину

$$F_n^*(x) D_n = \max_{1 \leq k \leq n} d_k,$$

где

$$d_k = \max_{1 \leq k \leq n} \left\{ \frac{k}{n} - F(x_k), F(x_k) - \frac{k-1}{n} \right\}$$

т.е. D_n – это наибольшее отклонение $F_n^*(x)$ от $F(x)$ в точках вариационного ряда.

Справедлива следующая теорема.

Теорема 3. Если $F_n^*(x)$ непрерывна, то при $z > 0$ и $n \rightarrow \infty$

$$p\{\sqrt{n}D_n < z\} \rightarrow K(z).$$

В этой теореме $K(z)$ - так называемое *распределение Колмогорова*, которое имеет свойство

$$\lim_{z \rightarrow \infty} K(z) = 1.$$

Из теоремы 3 получаем, что с вероятностью $p = 1 - \alpha$ выполняется неравенство

$$|F_n^*(x) - F(x)| \leq \frac{z_\alpha}{\sqrt{n}}, \quad \forall x \in R^1,$$

где z_α - соответствующая критическая граница распределения Колмогорова. Последнее неравенство, очевидно, может выполнять роль доверительного интервала для $F(x)$. Например, при $\alpha = 0.05$

$$F_n^*(x) - \frac{1,358}{\sqrt{n}} \leq F(x) \leq F_n^*(x) + \frac{1,358}{\sqrt{n}}.$$

Существуют и другие меры уклонения $F_n^*(x)$ от $F(x)$, например, мера Мисса.

13. Оценка функции плотности распределения

Плотность распределение также может быть оценена, причем несколькими способами:

1. *Гистограмма*. Находят $x_{\min}$ и $x_{\max}$ - минимальный и максимальный элементы выборки. Интервал $[x_{\min}, x_{\max}]$, длина которого называется *размахом выборки*, делят на k подинтервалов одинаковой длины

$$\Delta = \frac{x_{\max} - x_{\min}}{k}.$$

Рекомендуется, согласно известной формуле Стерджеса, принимать

$$k = 1 + 3,322 \cdot \ln n,$$

так как здесь имеется ввиду ближайшее целое число.

Далее, находят величины v_i - количество элементов выборки, попадающих в i -й подинтервал

$$[x_{\min} + \Delta i, x_{\min} + \Delta(i+1)), \quad i = \overline{1, k-1}.$$

Затем над каждым таким интервалом строят прямоугольник высотой

$$h_i = \frac{v_i}{n},$$

где n - объем выборки (рис. 4).

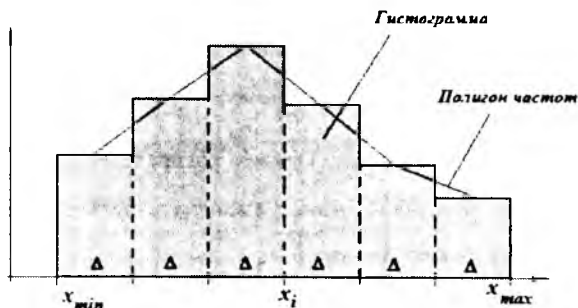


Рис. 4

2. **Полигон частот** – это ломаная, соединяющая середины соседних ступеней гистограммы. Если случайная величина X непрерывна, то полигон частот лучше оценивает плотность распределения.

Известна теорема, дающая оценку точности аппроксимации плотности посредством полигона частот.

Теорема 4 (Смирнова). Для любого $(a, b) \in R$

$$\lim_{n \rightarrow \infty} P \left\{ \max_{a \leq x \leq b} \frac{|P^*(x) - P(x)|}{\sqrt{P(x)}} < \frac{\lambda + \frac{z}{\lambda}}{\sqrt{n\Delta}} \right\} = e^{-2e^{-z}},$$

где λ – решение уравнения $\Phi(\lambda) = \frac{1}{2} - \frac{1}{2\sqrt{n}}$, $P(x)$ - теоретическая плотность распределения.

14. Метод моментов

Рассмотренные выше точечные оценки параметров строились эвристически и лишь затем доказывалось, что они действительно являются оценками интересующих нас параметров. Оказывается, имеются регулярные методы построения статистических оценок. Одним из них является метод моментов.

Определение 20. Пусть случайная величина X имеет плотность распределения $P(x)$, тогда величины

$$m_k = \int_{-\infty}^{\infty} x^k P(x) dx, \quad d_k = m_k^* = \int_{-\infty}^{\infty} (x - M(X))^k P(x) dx$$

называются, соответственно, *теоретическими начальными и центральными моментами* k -го порядка.

Отметим, что математическое ожидание является начальным моментом первого порядка, а дисперсия – *центральным* второго порядка.

Определения 21. Пусть имеется выборка $x_1, x_2, \dots, x_n$ значений случайной величины X . Тогда величины

$$m_k^* = \frac{\sum_{i=1}^n x_i^k}{n}, \quad d_k^* = \frac{\sum_{i=1}^n (x_i - \bar{X})^k}{n}$$

называются, соответственно, *выборочными начальными и центральными моментами* k -го порядка.

Пусть $P(x, \gamma)$ – плотность распределения случайной величины X , и γ – неизвестный параметр этой плотности. Достаточно часто теоретические моменты могут быть выражены явно через неизвестный параметр γ . Идея метода моментов состоит в приравнивании теоретических и выборочных моментов. Решая соответствующие уравнения и получают искомую оценку неизвестного параметра. Если неизвестных параметров несколько, то приравнивая несколько моментов получают целую систему уравнений, которую также достаточно часто удается разрешить относительно оцениваемых пара-

метров. Полученные таким образом статистические оценки называют *оценками метода моментов*.

Примеры

1. Пусть X равномерно распределена на отрезке $[c, d] \in R$. Ясно, что c, d являются параметрами данного распределения. Построим для них оценки, используя метод моментов.

Ранее мы видели, что для равномерного распределения

$$M(X) = \frac{c+d}{2}, \quad D(X) = \frac{(d-c)^2}{12}.$$

Приравнявая $M(X)$ и $D(X)$ с соответствующими выборочными моментами получаем систему уравнений

$$\begin{cases} \frac{c+d}{2} = \bar{X}, \\ \frac{(d-c)^2}{12} = \tilde{S}^2. \end{cases}$$

решая которую находим

$$\begin{cases} c = \bar{X} - \sqrt{3}\tilde{S}, \\ d = \bar{X} + \sqrt{3}\tilde{S}. \end{cases}$$

Полученные равенства и являются оценками метода моментов искоемых параметров.

2. Аналогично для нормального распределения получаем систему

$$\begin{cases} a = \bar{X}, \\ \sigma^2 = \tilde{S}^2, \end{cases}$$

которая в целом соответствует полученным ранее оценкам. Однако видим, что оценка дисперсии получилась смещенной. Это происходит часто и при использовании других методов построения оценок и не считается серьезным недостатком, так как смещенные оценки обычно удается «исправить».

Достоинством метода моментов является его относительно простая вычислительная реализация. Важнейшим недостатком – неоднозначность получения оценок. Приравнявая различные моменты мы, вообще говоря, можем получить различные оценки для одних и тех же параметров. Кроме того,

оценки метода моментов часто менее эффективны, чем некоторых других методов, например, следующего.

15. Метод максимального правдоподобия

Пусть $P(x, \gamma)$ – плотность распределения случайной величины X , где γ – неизвестный, подлежащий оценке параметр. И пусть имеется выборка $x_1, x_2, \dots, x_n$ значений этой случайной величины. Идея метода максимального правдоподобия состоит в том, что в качестве оценки γ следует принять такое значение, при котором вероятность появления именно этой имеющейся выборки максимальна.

А эта вероятность, условно говоря, пропорциональна функции

$$L(\gamma) = P(x_1, \gamma)P(x_2, \gamma) \cdots P(x_n, \gamma),$$

которую называют *функцией правдоподобия Фишера*. Таким образом, мы приходим к задаче на нахождение экстремума функции $L(\gamma) \rightarrow \max$.

Необходимое условие экстремума дает уравнение

$$\frac{\partial L(\gamma)}{\partial \gamma} = 0,$$

которое, будучи разрешенным относительно γ , и определяет оценку метода максимального правдоподобия. Если неизвестных параметров несколько $\gamma_1, \dots, \gamma_k$, то, аналогично, получаем систему необходимых условий экстремума

$$\begin{cases} \frac{\partial L(\gamma_1, \dots, \gamma_k)}{\partial \gamma_1} = 0, \\ \dots\dots\dots \\ \frac{\partial L(\gamma_1, \dots, \gamma_k)}{\partial \gamma_k} = 0, \end{cases}$$

из которой выражаются необходимые оценки.

Часто оказывается удобнее использовать не $L(\gamma)$, а

$$\tilde{L}(\gamma) = \ln(L(\gamma)) = \ln(P(x_1, \gamma)P(x_2, \gamma) \cdots P(x_n, \gamma)) = \sum_{i=1}^n \ln(P(x_i, \gamma)),$$

которая называется *логарифмической функцией правдоподобия*. В силу свойств логарифма, и $L(\gamma)$, и $\tilde{L}(\gamma)$ достигают максимума при одном и том же значении.

Примеры

1. Найдем оценки максимального правдоподобия для параметров нормального распределения. Функция правдоподобия имеет вид

$$L(a, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma^2}},$$

а логарифмическая функция правдоподобия

$$\bar{L}(a, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2.$$

Ясно, что удобнее максимизировать последнюю. Система необходимых условий экстремума

$$\begin{cases} \frac{\partial \bar{L}(a, \sigma^2)}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a) = 0, \\ \frac{\partial \bar{L}(a, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - a)^2 = 0. \end{cases}$$

Выражая из этой системы a и σ^2 получаем искомые оценки

$$\begin{cases} a^* = \frac{1}{n} \sum_{i=1}^n x_i, \\ (\sigma^*)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2 = 0. \end{cases}$$

т.е. получили хорошо знакомый результат. В данном случае, оценки максимального правдоподобия совпадают с оценками метода моментов, что на самом деле случается редко.

2. Найдем оценки максимального правдоподобия границ интервала распределения равномерно распределенной случайной величины.

Подлежащая максимизации функция правдоподобия в данном случае имеет вид

$$L(c, d) = \frac{1}{(d - c)^n} \rightarrow \max$$

и, очевидно, что должны выполняться ограничения

$$c \leq x_{\min}, \quad x_{\max} \leq d,$$

где $x_{\min}$, $x_{\max}$ – соответственно, минимальный и максимальный элементы выборки. Решением задачи, а значит искомыми оценками являются значения

$$c^* = x_{\min}, \quad d^* = x_{\max}.$$

Как видим, они не совпадают с полученными ранее оценками метода моментов.

Вопросы для самопроверки

1. В чем заключается основная задача математической статистики?
2. Что такое выборочное среднее?
3. Какие вы знаете оценки дисперсии?
4. Что такое генеральная доля?
5. В чем недостаток точечных оценок?
6. Какие стандартные статистические распределения вы знаете?
7. Постройте оценку максимального правдоподобия для параметра экспоненциального распределения.
8. Постройте доверительный интервал для математического ожидания экспоненциального распределения.
9. Что такое распределение Колмогорова, и для чего оно может быть использовано?
10. Зачем нужны методы построения статистических оценок?

§8. Статистические и корреляционные зависимости

Ключевые слова: функциональная зависимость, статистическая (вероятностная или стохастическая) зависимость, корреляционная зависимость, условные средние, уравнения регрессии, функция регрессии, линия регрессии, выборочный коэффициент корреляции, выборочное уравнение прямой линии регрессии, метод наименьших квадратов, выборочный коэффициент корреляции.

1. Функциональная, статистическая и корреляционная зависимости

Во многих задачах требуется установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других величин.

Рассмотрим сначала зависимость (связь) Y от одной случайной (или неслучайной) величины X .

В некоторых случаях эта связь является настолько тесной что, зная, какое значение приняла величина X , можно однозначно предсказать значение Y ; это означает, что связь между величинами X и Y - функциональная. Возможен, однако, и другой крайний случай, когда зависимость между X и Y отсутствует вовсе, т.е. величины X и Y независимы. Точное определение независимости случайных величин было дано ранее.

В общем случае связь между величинами X и Y находит свое выражение в том, что при фиксированном значении x величины X , величина Y остается случайной, но с законом распределения, зависящим от X . Иначе говоря, каждому значению $X = x$ отвечает свой закон, распределения величины Y . Рассмотренные выше крайние случаи – функциональная зависимость и полная независимость – вполне укладываются в эту общую схему; функциональная зависимость $Y = f(X)$ означает, что при фиксированном значении $X = x$ величина Y принимает единственное значение $f(x)$ (с вероятностью 1), а полная независимость означает, что при любом значении x величины X закон распределения величины Y - один и тот же (он не зависит от выбранного нами значения величины X).

Связь между двумя случайными величинами, проявляющаяся тем, что изменение одной из них влечет за собой изменение закона распределения другой, называется статистической (или вероятностной или стохастической).

Вероятностная связь между двумя случайными величинами X и Y появляется обычно тогда, когда имеются общие случайные факторы, влияющие как на X , так и на Y (наряду с другими факторами, неодинаковыми для X и Y). Например, если X представляет собой некоторую функцию от случайных величин U и V :

$$X = f(U, V),$$

и Y есть функция от той же самой величины и другой случайной величины W :

$$X = f(U, V).$$

то величины X и Y будут связаны между собой вероятностной связью.

Определение. Статистической называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой; в этом случае статистическую зависимость называют корреляционной.

Приведем пример случайной величины Y , которая не связана с величиной X функционально, а связана корреляционно. Пусть Y - урожай зерна, X - количество удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т.е. Y не является функцией от X . Это объясняется влиянием случайных факторов (осадки, температура воздуха и др.). Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т.е. Y связан с X корреляционной зависимостью.

2. Условные средние. Корреляционная зависимость

Уточним определение корреляционной зависимости, для чего введем понятие условной средней.

Предположим, что изучается связь между случайной величиной Y и случайной величиной X . Пусть каждому значению X соответствует несколько значений Y . Например, пусть при $x_1 = 2$ величина Y приняла значения: $y_1 = 5$, $y_2 = 6$, $y_3 = 10$. Найдем среднее арифметическое этих чисел:

$$\bar{y}_2 = \frac{5 + 6 + 10}{3} = 7.$$

Число $\bar{y}_2$ называют условным средним; черточка над буквой y служит обозначением среднего арифметического, а число 2 указывает, что рассматриваются те значения Y , которые соответствуют $x_1 = 2$.

Применительно к примеру предыдущего пункта эти данные можно истолковать так: на каждый из трех одинаковых участков земли внесли по 2 единицы удобрений и сняли, соответственно, 5; 6 и 10 единиц зерна: средний урожай составил 7 соответствующих единиц.

Определение. Условным средним $\bar{y}_x$ называют среднее арифметическое значений Y , соответствующих значению $X = x$.

Если каждому значению x соответствует одно значение условной средней, то, очевидно, условная средняя есть функция от x ; в этом случае говорят, что случайная величина Y зависит от X корреляционно.

Определение. Корреляционной зависимостью Y от X называют функциональную зависимость условной средней $\bar{y}_x$ от x :

$$\bar{y}_x = f(x). \quad (1)$$

Определение. Уравнение (1) называют уравнением регрессии Y на X ; функцию $f(x)$ называют регрессией Y на X , а ее график - линией регрессии Y на X .

Аналогично определяется условная средняя $\bar{x}_y$, и корреляционная зависимость X от Y .

Условным средним $\bar{x}_y$ значений X , соответствующих $Y = y$.

Определение. Корреляционной зависимостью X от Y называют функциональную зависимость условной средней $\bar{x}_y$ от y :

$$\bar{x}_y = \varphi(y). \quad (2)$$

Определение. Уравнение (2) называют уравнением регрессии X и Y функцию $\varphi(y)$ называют регрессией X на Y , а ее график - линией регрессии X на Y .

3. Две основные задачи теории корреляции

Первая задача теории корреляции - установить форму корреляционной связи, т.е. вид функции регрессии (линейная, квадратичная показательная и т. д.). Наиболее часто функции регрессии оказываются линейными. Ес-

ли обе функции регрессии $f(x)$ и $\varphi(x)$ линейны, то корреляцию называют линейной; в противном случае - нелинейной. Очевидно, при линейной корреляции обе линии регрессии являются прямыми линиями.

Вторая задача теории корреляции - оценить тесноту (силу) корреляционной связи. Теснота корреляционной зависимости Y от X оценивается по величине рассеяния значений Y вокруг условного среднего $\bar{y}_x$. Большое рассеяние свидетельствует о слабой зависимости Y от X либо об отсутствии зависимости. Малое рассеяние указывает наличие достаточно сильной зависимости; возможно даже, что Y и X связаны функционально, но под воздействием второстепенных случайных факторов эта связь оказалась размытой, в результате чего при одном и том же значении x величина Y принимает различные значения.

Аналогично (по величине рассеяния значений X вокруг условного среднего $\bar{x}_y$) оценивается теснота корреляционной связи X от Y .

4. Отыскание параметров выборочного уравнения прямой линии регрессии по несгруппированным данным

Допустим, что количественные признаки X и Y связаны линейной корреляционной зависимостью. В этом случае обе линии регрессии будут прямыми.

Предположим, что для отыскания уравнений этих прямых проведено n независимых испытаний, в результате которых получены n пар чисел:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Поскольку наблюдаемые пары чисел можно рассматривать как случайную выборку из генеральной совокупности всех возможных значений случайной величины (X, Y) , то величины и уравнения, найденные по этим данным, называют выборочными.

Для определенности будем искать выборочное уравнение прямой линии регрессии Y на X .

Рассмотрим простейший случай: различные значения x признака X и соответствующие им значения y признака Y наблюдались по одному разу. Очевидно, что группировать данные нет необходимости. Также нет надобности использовать понятие условной средней, поэтому искомое уравнение

$$\bar{y}_x = kx + b$$

можно записать так:

$$Y = kx + b.$$

Угловым коэффициентом прямой линии регрессии Y на X принято называть выборочным коэффициентом регрессии Y на X и обозначать через ρ_{yx} .

Итак, будем искать выборочное уравнение прямой линии регрессии Y на X вида:

$$Y = \rho_{yx}x + b. \quad (1)$$

Поставим своей задачей подобрать параметры ρ_{yx} и b так, чтобы точки $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, построенные по данным наблюдений на плоскости $ХОУ$, как можно ближе лежали вблизи прямой (1).

Уточним смысл этого требования. Назовем отклонением разность

$$Y_i - y_i \quad (i = 1, 2, \dots, n),$$

где Y_i - вычисленная по уравнению (1) ордината, соответствующая наблюдаемому значению x_i ; y_i - наблюдаемая ордината, соответствующая x_i .

Подберем параметры ρ_{yx} и b так, чтобы сумма квадратов отклонений была минимальной (в этом состоит сущность метода наименьших квадратов).

Так как каждое отклонение зависит от отыскиваемых параметров, то и сумма квадратов отклонений есть функция F этих параметров (времененно вместо ρ_{yx} будем писать ρ):

$$F(\rho, b) = \sum_{i=1}^n (Y_i - y_i)^2, \quad \text{или} \quad F(\rho, b) = \sum_{i=1}^n (\rho x_i + b - y_i)^2.$$

Для отыскания минимума приравняем нулю соответствующие частные производные:

$$\frac{\partial F}{\partial \rho} = 2 \sum_{i=1}^n (\rho x_i + b_i - y_i) x_i ;$$

$$\frac{\partial F}{\partial b} = 2 \sum_{i=1}^n (\rho x_i + b_i - y_i)$$

(для простоты записи вместо $\sum_{i=1}^n$ будем писать $\sum$).

Выполнив элементарные преобразования, получим систему двух линейных уравнений относительно ρ и b

$$\begin{cases} (\sum x^2)\rho + (\sum x)b = \sum xy, \\ (\sum x)\rho + nb = \sum y. \end{cases} \quad (2)$$

Решив эту систему, найдём искомые параметры:

$$\begin{aligned} \rho_{xy} &= \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2}, \\ b &= \frac{\sum x^2 \cdot \sum y - \sum x \cdot \sum xy}{n \sum x^2 - (\sum x)^2}. \end{aligned} \quad (3)$$

Аналогично можно найти выборочное уравнение прямой линии регрессии X на Y :

$$\bar{x}_y = \rho_{yx} y + c,$$

где ρ_{yx} – выборочный коэффициент регрессии X на Y .

Пример. Найти выборочное уравнение прямой линии регрессии Y на X по данным $n=5$ наблюдений:

x	1	1,5	3	4,5	5
y	1,25	1,4	1,5	1,75	2,25

Решение

Составим расчетную табл. 1.

Таблица 1

x_i	y_i	x_i^2	$x_i y_i$
1,00	1,25	1,00	1,250
1,50	1,40	2,25	2,100
3,00	1,50	9,00	4,500
4,50	1,75	20,25	4,875
5,00	2,25	25,00	11,250
$\sum x_i = 15$	$\sum y_i = 8,15$	$\sum x_i^2 = 57,50$	$\sum x_i y_i = 26,975$

Найдем искомые параметры, для чего подставим вычисленные по таблице суммы в соотношения (3):

$$\rho_{yx} = \frac{5 \cdot 26,975 - 15 \cdot 8,15}{5 \cdot 57,5 - 15^2} = 0,202;$$

$$b = \frac{57,5 \cdot 8,15 - 15 \cdot 26,975}{5 \cdot 57,5 - 15^2} = 1,024.$$

Напишем искомое уравнение регрессии:

$$y = 0,202x + 1,024.$$

Для того чтобы получить представление, насколько хорошо вычисленные по этому уравнению значения Y_i согласуются с наблюдаемыми значениями y_i , найдем отклонения $Y_i - y_i$. Результаты вычислений сведены в табл. 2.

Таблица 2

x_i	Y_i	y_i	$Y_i - y_i$
1,00	1,226	1,25	-0,024
1,50	1,327	1,40	-0,073
3,00	1,630	1,50	0,130
4,50	1,993	1,75	0,083
5,00	2,034	2,25	-0,216

Как видно из таблицы, не все отклонения достаточно малы. Это объясняется малым числом наблюдений.

5. Корреляционная таблица

При большом числе наблюдений одно и то же значение x может встретиться n_x раз, одно и то же значение y может встретиться n_y раз, одна и та же пара чисел (x, y) может наблюдаться n_{xy} раз. Поэтому данные наблюдений группируют, т.е. подсчитывают частоты n_x , n_y , n_{xy} . Все сгруппированные данные записывают в виде таблицы, которую называют корреляционной.

Поясним устройство корреляционной таблицы на примере (табл. 3).

Таблица 3

X	10	20	30	40	n_y
Y					
0,4	5	-	7	14	26
0,6	-	2	6	4	12
0,8	3	19	-	-	22
n_x	8	21	13	18	$n = 60$

В первой строке таблицы указаны наблюдаемые значения (10; 20; 30; 40) признака X , а в первом столбце - наблюдаемые значения (0,4; 0,6; 0,8) признака Y . На пересечении строк и столбцов вписаны частоты n_{xy} наблюдаемых пар значений признаков. Например, частота 5 указывает, что пара чисел (10; 0,4) наблюдалась 5 раз. Все частоты помещены в прямоугольнике, клетки которого выделены. Черточка означает, что соответственная пара чисел, например, (20; 0,4) не наблюдалась.

В последнем столбце записаны суммы частот строк. Например, сумма частот первой строки прямоугольника, клетки которого выделены, равна $n_y = 5 + 7 + 14 = 26$; это число указывает, что значение признака Y , равное 0,4 (в сочетании с различными значениями признака X) наблюдалось 26 раз.

В последней строке записаны суммы частот столбцов. Например, число 8 указывает, что значение признака X , равное 10 (в сочетании с различными значениями признака Y) наблюдалось 8 раз.

В клетке, расположенной в нижнем правом углу таблицы, помещена сумма всех частот (общее число всех наблюдений n). Очевидно

$$\sum n_x = \sum n_y = n.$$

В нашем примере $\sum n_x = 8 + 21 + 13 + 18 = 60$ и $\sum n_y = 26 + 12 + 22 = 60$.

6. Отыскание параметров выборочного уравнения прямой линии регрессии по сгруппированным данным. Выборочный коэффициент корреляции

В п. 4 для определения параметров уравнения прямой линии регрессии Y на X была получена система уравнений (2):

$$\begin{cases} (\sum x^2)\rho_{yx} + (\sum x)b = \sum xy, \\ (\sum x)\rho_{yx} + nb = \sum y. \end{cases} \quad (4)$$

Предполагалось, что значения X и соответствующие им значения Y наблюдались по одному разу. Теперь же допустим, что получено большое число данных (практически для удовлетворительной оценки искомых параметров должно быть хотя бы 50 наблюдений), среди них есть повторяющиеся, и они сгруппированы в виде корреляционной таблицы. Запишем систему (4) так, чтобы она отражала данные корреляционной таблицы. Воспользуемся тождествами:

$$\sum x = n\bar{x} \text{ следствие } \bar{x} = \frac{\sum x}{n};$$

$$\sum y = n\bar{y} \text{ следствие } \bar{y} = \frac{\sum y}{n};$$

$$\sum x^2 = n\overline{x^2} \text{ следствие } \overline{x^2} = \frac{\sum x^2}{n}.$$

$\sum n_x = \sum n_y$ (учтено, что пара чисел (x, y) наблюдалась n_{xy} раз).

Подставив правые части тождеств в систему (4) и сократив обе части второго уравнения на n , получим:

$$\begin{cases} (\overline{nx^2})\rho_{yx} + (n\bar{x})b = \sum n_{xy}xy, \\ (\bar{x})\rho_{yx} + b = \bar{y}. \end{cases} \quad (5)$$

Решив эту систему, найдем параметры ρ_{yx} и b и, следовательно, искомое уравнение:

$$\bar{y}_x = \rho_{yx}x + b. \quad (*)$$

Однако более целесообразно, введя новую величину - коэффициент корреляции, написать уравнение регрессии в ином виде. Сделаем это.

Найдем b из второго уравнения (5):

$$b = \bar{y} - \rho_{yx}x.$$

Подставив правую часть этого равенства в уравнение (*), получим:

$$\bar{y}_x - \bar{y} = \rho_{yx} (x - \bar{x}). \quad (6)$$

Найдем из системы (4) коэффициент регрессии, учитывая, что

$$s^2 - (\bar{x})^2 = \sigma_x^2:$$

$$\rho_{yx} = \frac{\sum n_{xy} - n\bar{x}\bar{y}}{n[s^2 - (\bar{x})^2]} = \frac{\sum n_{xy} - n\bar{x}\bar{y}}{n\sigma_x^2}.$$

Умножим обе части равенства на дробь $\frac{\sigma_x}{\sigma_y}$:

$$\rho_{yx} \cdot \frac{\sigma_x}{\sigma_y} = \frac{\sum n_{xy} - n\bar{x}\bar{y}}{n\sigma_x\sigma_y}.$$

Обозначим правую часть равенства через r_* и назовем ее выборочным коэффициентом корреляции:

$$\rho_{yx} \cdot \frac{\sigma_x}{\sigma_y} = r_*, \quad \text{или} \quad \rho_{yx} = r_* \cdot \frac{\sigma_y}{\sigma_x}.$$

Подставив правую часть этого равенства в (6), окончательно получим выборочное уравнение прямой линии регрессии Y на X вида

$$\bar{y}_x - \bar{y} = r_* \frac{\sigma_y}{\sigma_x} (x - \bar{x}).$$

Замечание 1. Аналогично находят выборочное уравнение прямой линии регрессии X на Y вида

$$\bar{x}_y - \bar{x} = r_* \frac{\sigma_x}{\sigma_y} (y - \bar{y}),$$

где

$$r_* \frac{\sigma_y}{\sigma_x} = \rho_{xy}.$$

Замечание 2. Выборочный коэффициент корреляции имеет важное самостоятельное значение. Как следует из предыдущего, выборочный коэффициент корреляции определяется равенством

$$r_* = \frac{\sum n_{xy} - n\bar{x}\bar{y}}{n\sigma_x\sigma_y},$$

где x, y - варианты (наблюдавшиеся значения) признаков X и Y ;

n_{xy} - частота наблюдавшейся пары вариант (x, y) ;

n - объем выборки (сумма всех частот);

$\bar{x}, \bar{y}$ - выборочные средние;

σ_x, σ_y - выборочные среднеквадратические отклонения.

Ниже приведем свойства выборочного коэффициента корреляции из которых следует, что он служит для оценки тесноты линейной корреляционной зависимости.

1°. Абсолютная величина выборочного коэффициента корреляции не превосходит единицы.

2°. Если выборочный коэффициент корреляции равен нулю и выборочные линии регрессии прямые. Тогда X и Y не связаны линейной корреляционной зависимостью.

Замечание. Если выборочный коэффициент корреляции равен нулю, то признаки X и Y могут быть связаны нелинейной корреляционной или даже функциональной зависимостью.

3°. Если $|r_s| = 1$, то наблюдаемые значения признаков связаны линейной функциональной зависимостью.

4°. С возрастанием абсолютной величины выборочного коэффициента корреляции линейная корреляционная зависимость становится более тесной и при $|r_s| = 1$ переходит в функциональную зависимость.

Из приведенных свойств вытекает смысл $|r_s|$: выборочный коэффициент корреляции характеризует *тесноту линейной связи* между количественными признаками в выборке: чем ближе $|r_s|$ к 1, тем связь сильнее; чем ближе $|r_s|$ к 0, тем связь слабее.

Вопросы для самопроверки

1. Дайте определение функциональной зависимости.
2. Дайте определение статистической зависимости.

3. Что называется условным средним?
4. Дайте определение корреляционной зависимости.
5. Дайте определения уравнения регрессии.
6. В чем состоит задача теории корреляции?
7. Что представляет собой метод наименьших квадратов (МНК)?
8. Что такое наблюдаемая ордината в МНК?
9. Напишите параметры выборочного уравнения прямой линии регрессии в случае, когда данные негруппированы?
10. Поясните устройство корреляционной таблицы.
11. Напишите параметры выборочного уравнения прямой линии регрессии в случае, когда данные сгруппированы?
12. Как определяется выборочный коэффициент корреляции?
13. Приведите свойства выборочного коэффициента корреляции?
14. Что характеризует коэффициент корреляции?

§9. Критерий согласия Пирсона и Колмогорова-Смирнова

Ключевые слова: статистическая гипотеза, простая гипотеза, сложная гипотеза, основная или нулевая гипотеза, альтернативная или конкурирующая гипотеза, критическая граница, статистическая гипотеза, критерий согласия Пирсона, критерий Колмогорова-Смирнова.

1. Общая схема проверки статистических гипотез

Определение 1. Статистической гипотезой называется любое высказывание о конкретных значениях параметров распределения некоторой случайной величины, или о виде этого распределения.

В первом случае гипотеза называется параметрической, во втором - непараметрической.

Нередко конкретные управленческие решения могут или должны быть приняты на основе анализа статистической информации о рассматриваемом процессе или явлении. Причем эти решения напрямую определяются тем, каковы конкретные параметры этого процесса.

Примеры

1. Станок производит фасовку некоторого продукта в стандартные упаковки весом по 1 кг. Средний вес 50-ти, случайным образом отобранных упаковок оказался равным $\bar{X} = 0,98$ кг. при $S = 0,05$ кг. Можно ли считать, что имеющееся отклонение является результатом случайности и что станок настроен правильно? Или следует останавливать производство для переналадки станка?

2. В сельскохозяйственном районе опробуются два новых сорта пшеницы. Средняя урожайность первого составила 22 ц/га, второго – 22,5 ц/га. Следует ли считать, что второй сорт действительно урожайнее первого? Или имеющуюся разницу можно объяснить всегда присутствующими случайными причинами.

3. Имеется статистика страховых случаев по поводу угонов автомобилей различных марок. Следует ли считать, что частота случаев зависит от марки автомобилей? Очевидно, что если да, то страховая компания должна учитывать это в условиях страхования.

Оказывается, в этих и во многих других подобных случаях, ответ на поставленный вопрос можно свести к проверке соответствующей статистической гипотезы. Именно имеющиеся в математической статистике методики проверки таких гипотез делают ее одним из самых практически значимых разделов прикладной математики.

Общую схему проверки статистических гипотез кратко сформулировать непросто. Но можно выделить в ней следующие моменты:

1. Во-первых, формулируют, так называемую, основную или нулевую гипотезу. Ее обычно обозначают H_0 . Различают простые и сложные гипотезы.

Определение 2. Гипотеза называется простой, если она состоит в равенстве одного или нескольких параметров заданным числам. Если множество допустимых, для справедливости гипотезы, значений параметров состоит более, чем из одного элемента, то ее называют сложной. Например,

$H_0 : a = 5$ - простая гипотеза,

$H_0 : a = 5, \sigma^2 = 10$ - простая гипотеза,

$H_0 : 4 \leq a \leq 5$ - сложная гипотеза.

2. Одновременно рассматривается некоторая гипотеза называемая альтернативной или конкурирующей. ее обычно обозначают H_1 . Различают односторонние и двухсторонние конкурирующие гипотезы. Например.

$H_1 : a > 5$ - односторонняя гипотеза,

$H_1 : a < 5$ - односторонняя,

$H_1 : a \neq 5$ - двусторонняя сложная гипотеза.

Выбор вида альтернативной гипотезы определяется смыслом задачи.

3. Всегда имеется некоторая величина Θ , которая рассчитывается из данных выборки, и поэтому ее называют выборочной статистикой или критерием проверки гипотезы. Причем из теории бывает известно, какое распределение $P(\Theta)$ будет иметь данная величина, если верна нулевая гипотеза.

4. Из данных конкретной выборки находится расчетное значение $\Theta_{\text{расч}}$, и если оно плохо соответствует теоретическому распределению $P(\Theta)$, то отсюда делается вывод, что в действительности случайная величина Θ имеет другое распределение, а значит, и нулевая гипотеза H_0 не верна. Она отклоняется, и принимается альтернативная гипотеза H_1 .

5. При этом всегда имеется вероятность сделать неправильный вывод. Ошибкой I-го рода называют отклонение, на самом деле истинной, нулевой гипотезы. Вероятность этой ошибки называют уровнем значимости проверки гипотезы, и обычно обозначают α . Ошибкой II-го рода называют принятие на самом деле ложной, нулевой гипотезы. Вероятность этой ошибки обычно обозначают β , а величину $1 - \beta$ называют мощностью критерия. В общем случае α и β не связаны каким-либо однозначным соотношением, хотя для конкретных критериев это возможно. Считается, что хороший критерий должен обладать свойством $\alpha < \beta$.

Определение 3. Область Q при попадании, в которую выборочной статистики $\Theta_{рас}$ отвергается основная гипотеза, называется критической областью.

Говорят об уровне значимости критической области. В соответствии с видом альтернативной гипотезы различают односторонние и двухсторонние критические области.

Иногда говорят, что критерием проверки статистической гипотезы называется правило построения критической области. Отметим еще, что иногда критерием называют теоретическое распределение $P(\Theta)$.

2. Критерий согласия

Выборки, полученные из нормально распределенной генеральной совокупности, обеспечивают наивысшую достоверность статистических выводов только в тех случаях, когда они получены из нормально распределенной генеральной совокупности. При отклонениях от нормального распределения точность оптимальных критериев существенно падает, поэтому, чтобы уверенно применять оптимальные критерии необходимо проверить предположение о нормальном распределении генеральной совокупности. Для этого используются критерии согласия. Здесь нулевая гипотеза H_0 представляет собой утверждение о том, что распределение генеральной совокупности, из которой получена выборка, не отличается от нормального. Существует несколько разновидностей критериев согласия. Рассмотрим те из них, которые получили наибольшее распространение на практике.

2.1. Предварительная проверка соответствия нормальному распределению

Критерии согласия требуют достаточно большой вычислительной работы, поэтому целесообразно перед тем, как их использовать, проверить с помощью более простых методов соответствие имеющихся экспериментальных данных нормальному распределению. Эти методы, естественно, облада-

ны меньшей мощностью и позволяют установить только значительные расхождения с нормальным распределением, но если такие расхождения будут установлены, то необходимость в применении более точных, но более сложных критериев, как правило, отпадает.

Для предварительной проверки эмпирического распределения на нормальность можно использовать основные свойства нормального распределения, изложенные ранее. При этом эмпирическое распределение представляется в виде вариационного ряда или гистограммы (см. §7). Если в качестве параметров α и σ нормального распределения принять их выборочные оценки $\bar{X}$ и S , то для проверки можно использовать следующие свойства нормального распределения:

1) практически все отклонения от среднего значения (99,7 %) должны быть меньше $\pm 3S$;

2) примерно 2/3 всех отклонений (68,3 %) должны быть меньше $\pm S$;

3) половина всех отклонений от среднего значения должна быть меньше $\pm 0,675S$;

4) можно использовать такое свойство нормального распределения, что его коэффициенты асимметрии и эксцесса равны нулю.

Для проверки по этому свойству необходимо вычислить выборочные оценки этих параметров по формулам:

$$A = \frac{\sum_{i=1}^k n_i (x_i - \bar{X})^3}{nS^3}, \quad E = \frac{\sum_{i=1}^k n_i (x_i - \bar{X})^4}{nS^4},$$

где n_i – частоты интервалов группировки; k – число интервалов группировки; S – выборочное стандартное отклонение.

Значения коэффициентов A и E сравниваются с критическими значениями на уровне значимости α , и если критические значения превышены, то делается вывод о том, что распределение генеральной совокупности, из которой получена выборка, не согласуется с нормальным. В противном случае

модель нормального распределения может быть принята. Таблица критических значений A и E содержится в литературе по статистике. Здесь не будем подробно останавливаться на этих приближенных критериях. Отметим лишь еще раз, что они могут использоваться только совместно с более точными критериями, рассмотренными ниже.

2.2. Критерий согласия χ^2 – Пирсона

Критерий согласия χ^2 – Пирсона разработан лучше других критериев и чаще других используется. Он основан на сравнении эмпирических частот интервалов группировки с теоретическими (ожидаемыми) частотами, рассчитываемыми по формулам нормального распределения.

Условия применения: объем выборки $n \geq 40$, выборочные данные сгруппированы в интервальный вариационный ряд с числом интервалов не менее 7, ожидаемые (теоретические) частоты интервалов не должны быть меньше 5.

Гипотеза $H_0: f(x) = f'(x)$ – плотность распределения $f(x)$ генеральной совокупности, из которой взята выборка, соответствует теоретической модели $f'(x)$ нормального распределения.

Альтернатива $H_1: f(x) \neq f'(x)$.

Уровень значимости: α .

Порядок применения:

1. Формулируется гипотеза, выбирается уровень значимости α .
2. Получается выборка объема $n \geq 40$ независимых наблюдений и представляется эмпирическое распределение в виде интервального вариационного ряда.
3. Рассчитываются выборочные характеристики $\bar{X}$ и S . Их используют в качестве генеральных параметров a и σ нормального распределения, с которым предстоит сравнить эмпирическое распределение.

4. Вычисляются значения теоретических частот n'_i попадания в i -интервал группировки. Для этого необходимо вероятность попадания в этот интервал, определенную по формулам (4), (5) §6, умножить на объем выборки n

$$n'_i = n \left[\Phi_0 \left(\frac{x_{*i} - \bar{x}}{S} \right) - \Phi_0 \left(\frac{x_{ni} - \bar{x}}{S} \right) \right], \quad (1)$$

где $\Phi_0(u)$ – функции Лапласа (см. табл. Приложения); x_{*i} и x_{ni} верхняя и нижняя границы интервала группировки.

Если окажется, что вычисленные ожидаемые частоты n'_i некоторых интервалов группировки меньше 5, то соседние интервалы объединяются так, чтобы сумма их ожидаемых частот была больше или равна 5. Соответственно складываются и эмпирические частоты объединяемых интервалов.

5. Значение χ^2 - критерия рассчитывается по формуле:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}, \quad (2)$$

где n_i – эмпирические частоты; n'_i ожидаемые (теоретические) частоты; k – число интервалов группировки после объединения.

6. Из соответствующей таблицы Приложения находится критическое значение χ^2_α критерия согласия χ^2 – Пирсона для уровня значимости α и числа степеней свободы $\nu = k - 3$.

7. **Вывод:** $\chi^2 \geq \chi^2_\alpha$ если, то эмпирическое распределение не соответствует нормальному распределению на уровне значимости α , в противном случае нет оснований отрицать это соответствие.

Пример 1. Воспользуемся данными табл. 1, где представлены результаты в беге на 100 м группы школьников $n = 50$ для проверки соответствия эмпирического распределения нормальному распределению.

Исходные данные помещены в графы 2, 3 табл. 2 (графа 2 – границы интервалов группировки, графа 3 – эмпирические частоты интервалов). В табл. 1 верхние границы уменьшены на 0,1 для удобства подсчета частот. В табл. 2 верхние границы оставлены без изменений.

Таблица 1

Табличное представление данных о результате в беге на 100 м

Но- мер ин- тер- вала i	Границы интер- валов $x_{ni} - x_{ei}$	Среднее значение интерва- лов, x_i	Распреде- ление данных	Час- тоты n_i	Накоп- лен- ные час- тоты n_{xi}	Относи- тельные частоты f_i	Отноше- ния n_{xi} к объему выборки F_i
1	12,4-13,1	12,8	/	1	1	0,02	0,02
2	13,2-13,9	13,6	//	2	3	0,04	0,06
3	14,0-14,7	14,4	////////	9	12	0,18	0,24
4	14,8-15,5	15,2	////////////////	15	27	0,30	0,54
5	15,6-16,3	16,0	///////////////// /	17	44	0,34	0,88
6	16,4-17,1	16,8	////	5	49	0,10	0,98
7	17,2-17,9	17,6	/	1	50	0,02	1,00
	Сумма			50		1,00	

1. Формулируем гипотезу $H_0: f(x) = f'(x)$, выбираем уровень значимости $\alpha = 0,05$.

2. Получаем выборку объема $n = 50$, строим интервальный вариационный ряд с числом интервалов $k = 7$ (см. табл. 1).

3. Выборочные характеристики по этим данным рассчитаны в примере 3.6:

$$\bar{X} = 15,4, \quad S = 0,9.$$

4. Вычисляем значения теоретических частот по формуле (1) с использованием соответствующей таблицы Приложения. Предварительно нормируем границы интервалов группировки:

$$u_{ni} = \frac{x_{ni} - \bar{X}}{S}, \quad u_{ni} = \frac{x_{ni} - \bar{X}}{S}.$$

Нормированные границы занесены в графу 4, а вычисленные теоретические частоты – в графу 5 табл. 2.

Поскольку для интервалов с номерами 1, 2, 7 теоретические частоты оказались меньше 5, объединяем интервалы 1 и 2 с 3-м, а интервал 7 с 6-м интервалами. Суммируем эмпирические и ожидаемые частоты интервалов, которые мы объединили. После объединения получилось $k = 4$ интервала.

Таблица 2

Расчет критерия χ^2 – Пирсона

№ п/п	Границы интерва- лов $x_{ni}; x_{oi}$	Частоты интерва- лов n_i	Нормирован- ные границы интервалов u_{ni}, u_{oi}	Теоретиче- ские частоты n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$ n'_i
1	2	3	4	5	6	7
1	12,4; 13,2	12 1 2 9	-3,33; -2,44	0,345 2,603 9,603	-0,551	0,024
2	13,2; 14,0		-2,44; -1,56			
3	14,0; 14,8		-1,56; -0,67			
4	14,8; 15,6	15	-0,67; 0,22	16,780	-1,780	0,189
5	15,6; 16,4	17	0,22; 1,11	13,973	3,027	0,658
6	16,4; 17,2	6 5 1	1,11; 2,0	5,538 1,04	-0,578	0,051
7	17,2; 18,0		2,0; 2,89			
Сумма		50		49,882		0,922

5. Значение критерия χ^2 , определяемое по формуле (2), равно:

$$\chi^2 = 0,922.$$

Промежуточные расчеты отражены в графах 6 и 7 табл. 2.

6. Из соответствующей таблицы Приложения находим для уровня значимости $\alpha = 0,05$ и числа степеней свободы $\nu = 4 - 3 = 1$:

$$\chi^2_{0,05} = 2,84.$$

7. **Вывод:** поскольку $\chi^2 < \chi^2_{0,05}$ считаем, что эмпирическое распределение соответствует нормальному на уровне значимости 0,05.

2.3. Критерий λ Колмогорова - Смирнова

Другим критерием, часто используемым для проверки гипотезы о нормальности распределения, является критерий λ Колмогорова - Смирнова. Здесь гипотеза H_0 формулируется по отношению к функциям распределения $F(x)$ и $F'(x)$. $F(x)$ функция распределения генеральной совокупности, из которой получена выборка, а $F'(x)$ - функция непрерывного теоретического распределения (нормального распределения).

Условия применения: объем выборки $n \geq 35$, эмпирическое распределение представлено в виде интервального вариационного ряда.

Гипотеза $H_0 : F(x) = F'(x)$.

Альтернатива $H_1 : F(x) \neq F'(x)$.

Уровень значимости: α .

Порядок применения:

1. Формулируется гипотеза H_0 , назначается уровень значимости α .
2. Получают выборку объема $n \geq 35$ независимых наблюдений, она группируется в интервальный вариационный ряд.
3. Вычисляются выборочные характеристики $\bar{X}$ и S .
4. Рассчитываются значения эмпирических накопленных частот n_{xi} и теоретических накопленных частот n'_{xi} по формуле:

$$n'_{xi} = n \left[\frac{1}{2} + \Phi_0 \left(\frac{x_i - \bar{X}}{S} \right) \right], \quad (3)$$

где n – объем выборки; $\Phi_0(u)$ – функция Лапласа; x_i – срединные значения интервалов группировки.

5. Вычисляются значения критерия λ :

$$\lambda = \frac{\max_i |n_{xi} - n'_{xi}|}{\sqrt{n}}, \quad (4)$$

где $\max |n_{x_i} - n'_{x_i}|$ – максимальное значение модуля (абсолютной величины) разности между эмпирическими n_{x_i} и n'_{x_i} накопленными частотами.

6. Определяется критическое значение критерия λ_α Колмогорова – Смирнова при уровне значимости α . Для стандартных уравнений значимости критические значения равны:

$$\lambda_{0,05} = 0,895; \quad \lambda_{0,01} = 1,035.$$

Они соответствуют рассматриваемому варианту применения критерия Колмогорова – Смирнова, когда для вычисления теоретических накопленных частот используются выборочные характеристики $\bar{x}$ и S в качестве параметров μ и σ нормального распределения.

7. **Вывод:** если $\lambda \geq \lambda_\alpha$, то эмпирическое распределение не соответствует нормальному на уровне значимости α , в противном случае принимается гипотеза о согласии распределения генеральной совокупности с нормальным распределением.

Пример 2

Воспользуемся данными предыдущего примера 1 для проверки их соответствия нормальному распределению по критерию Колмогорова – Смирнова.

В табл. 3 в столбцах 2, 3 приведены срединные значения интервалов группировки и эмпирические накопленные частоты, взятые из табл. 1.

Таблица 3

Расчет критерия λ Колмогорова – Смирнова

№ п/п	Срединные значения интервалов x_i	Накопленные частоты n_{x_i}	Нормированные срединные зна- чения u_i	Теоретические накопленные частоты u'_{x_i}	$\frac{ n_{x_i} - n'_{x_i} }{\sqrt{n}}$
1	2	3	4	5	6
1	12,8	1	-2,89	0,098	0,127
2	13,6	3	-2,0	1,138	0,263
3	14,4	12	-1,11	6,675	0,753
4	15,2	27	-0,22	20,648	0,898
5	16,0	44	0,67	37,428	0,929
6	16,8	49	1,56	47,03	0,279
7	50	50	2,44	49,633	0,052

1. Формулируем гипотезу $H_0: F(x) = F'(x)$ и выбираем уровень значимости $\alpha = 0.05$.

2. Имеем выборку объема $n = 50$, сгруппированную в интервальный вариационный ряд с семью интервалами.

3. Выборочные характеристики рассчитаны в предыдущем примере:

$$\bar{X} = 15.4; \quad S = 0.9.$$

4. Эмпирические накопленные частоты приведены в графе 3, а теоретические, рассчитанные по формуле (3) – в графе 5.

5. Значение критерия λ составляет

$$\lambda = \frac{\max |n_x - n'_x|}{\sqrt{n}} = 0.929.$$

6. Критическое значение для $\alpha = 0.05$ равно $\lambda_{0.05} = 0.895$.

7. **Вывод:** поскольку $\lambda > \lambda_{0.05}$, мы вынуждены отклонить гипотезу о том, что выборка взята из нормально распределенной генеральной совокупности.

Оба рассмотренных критерия χ^2 – Пирсона и λ Колмогорова–Смирнова применимы в одних и тех же условиях ($n \geq 40$). Сравнение мощностей этих критериев для общего случая затруднительно, но из опыта известно, что критерий λ Колмогорова – Смирнова является более мощным (чаще обнаруживает отклонения от нормальности), если среднее и дисперсия теоретического нормального распределения оцениваются по выборке. Рассмотренные выше примеры 1 и 2 подтверждают это: для одних и тех же данных на одинаковом уровне значимости критерий λ Колмогорова – Смирнова обнаружил несоответствие нормальному распределению, а χ^2 – критерий позволяет принять гипотезу о нормальности.

Упражнение

Проверить гипотезу о нормальном распределении выборки из 55 наблюдений.

11,8	20,3	15,4	17,2	19,2	23,3	18,1	21,9	15,3	16,8	13,2
17,8	20,4	16,5	19,7	20,5	14,3	20,1	16,8	14,7	20,8	19,5
11,8	15,3	19,3	17,8	16,2	15,7	22,8	21,9	12,5	10,1	21,1
18,6	18,3	14,7	14,5	18,1	18,4	13,9	19,1	18,5	20,2	23,8
19,1	16,7	20,4	19,5	17,2	19,6	17,8	21,3	17,5	19,4	17,8

Принять $\alpha = 0,1$.

Вопросы для самопроверки

1. Какие вы можете привести примеры статистических гипотез? Какие из них простые гипотезы?
2. Что такое статистический критерий и его теоретическое распределение?
3. Что такое мощность критерия?
4. Что такое критическая область?
5. Сформулируйте кратко общую схему проверки статистических гипотез.
6. Какие вопросы решает критерий согласия?
7. Какие критерий согласия приводятся выше?
8. На чем основан критерий согласия χ^2 – Пирсона?
9. Сформулируйте порядок применения критерий согласия χ^2 – Пирсона.
10. На чем основан критерий λ Колмогорова - Смирнова. По отношению к чему формулируется гипотеза H_0 ?
11. Сформулируйте порядок применения критерия λ Колмогорова - Смирнова.
12. Что можно сказать по поводу сравнения мощностей этих критериев?

ПРИЛОЖЕНИЯ

Вероятностные таблицы

Таблица значений плотности стандартного нормального распределения

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3988	0,3987	0,3986	0,3984	0,3982	0,3979	0,3976	0,3973
0,1	0,3969	0,3965	0,3960	0,3955	0,3950	0,3944	0,3938	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3866	0,3856	0,3846	0,3836	0,3825
0,3	0,3813	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3711	0,3697
0,4	0,3682	0,3667	0,3652	0,3637	0,3621	0,3605	0,3588	0,3572	0,3555	0,3538
0,5	0,3520	0,3502	0,3484	0,3466	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2896	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1827	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0789	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0619	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0296	0,0290
2,3	0,0283	0,0277	0,0271	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0162	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0016	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0094	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0059	0,0058	0,0056	0,0054	0,0053	0,0051	0,0050	0,0049	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0035	0,0034	0,0033
3,1	0,0032	0,0031	0,0031	0,0029	0,0029	0,0028	0,0027	0,0026	0,0025	0,0024
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Таблица значений функции Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999

Критические точки распределения Стьюдента

α k	0,1	0,05	0,02	0,01	0,002	0,001
1	6,3137	12,706	31,82	63,656	318,31	636,62
2	2,9199	4,3026	6,9645	9,9248	22,327	31,599
3	2,3533	3,1824	4,5407	5,8409	10,215	12,924
4	2,1318	2,7764	3,7469	4,6040	7,1732	8,6103
5	2,0150	2,5705	3,3649	4,0321	5,8934	6,8688
6	1,9431	2,4469	3,1426	3,7074	5,2076	5,9588
7	1,8945	2,3646	2,9979	3,4994	4,7853	5,4079
8	1,8595	2,3060	2,8964	3,3553	4,5008	5,0413
9	1,8331	2,2621	2,8214	3,2498	4,2968	4,7809
10	1,8124	2,2281	2,7637	3,1692	4,1437	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370
12	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
14	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405
15	1,7530	2,1314	2,6025	2,9467	3,7328	4,0728
16	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,5372	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
32	1,6930	2,0369	2,4487	2,7385	3,3653	3,6218
34	1,6909	2,0322	2,4411	2,7284	3,3479	3,6007
36	1,6883	2,0281	2,4345	2,7195	3,3326	3,5821
38	1,6860	2,0244	2,4286	2,7116	3,3190	3,5657
40	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
42	1,6820	2,0181	2,4185	2,6981	3,2960	3,5377
44	1,6802	2,0154	2,4141	2,6923	3,2861	3,5258
46	1,6787	2,0129	2,4102	2,6870	3,2771	3,5150
48	1,6772	2,0106	2,4066	2,6822	3,2689	3,5051
50	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
55	1,6730	2,0040	2,3961	2,6682	3,2561	3,4764
60	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
65	1,6686	1,9971	2,3851	2,6536	3,2204	3,4466
70	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019

Критические точки распределения χ^2

α k	0,01	0,025	0,05	0,95	0,975	0,99
1	6.6	5,0	3,8	0.0039	0.00098	0.00016
2	9,2	7,4	6,0	0.103	0.051	0.020
3	11,3	9,4	7,8	0.352	0.216	0.115
4	13,3	11,1	9,5	0.711	0.484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

ГЛАВА 4. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

§ 10. Построение математических моделей некоторых простейших экономических задач

Ключевые слова: математическая модель, ограничения типа равенств, ограничения типа неравенств, целевая функция, линейное программирование, нелинейное программирование, дискретное программирование, целочисленное программирование, динамическое программирование, стохастическое программирование, параметрическое программирование, дробно-линейное программирование, блочное программирование, сетевое (потокосовое) программирование, многоиндексное программирование, булево программирование, комбинаторное программирование, квадратичное программирование, биквадратичное программирование, сепарабельное программирование, выпуклое программирование, скалярная оптимизация, векторная оптимизация, задачи многокритериального подхода.

1. Введение

Каждое разумное действие является в определённом смысле оптимальным, ибо оно выбирается после сравнения с другими вариантами на основе интуиции и опыта человека. Интерес к задачам наилучшего выбора был всегда высоким.

Многие экономические задачи и процессы обладают многовариантными решениями. Следует отметить, что иногда число допустимых вариантов или практически необозримо, или бесконечно. Конечно, нахождение оптимальных решений (выбор наилучших вариантов) не основывается на «здравом смысле», интуиции и опыте специалиста. Нахождение оптимальных решений - выбор наилучших вариантов (максимума, минимума или вообще оптимального в том или ином смысле) необходимо осуществлять с помощью точных математических методов с применением компьютерной технологии. Эти точные математические методы составляют содержание предмета «Ма-

математическое программирование»¹. Более подходящим по смыслу названием было бы «методы оптимизации», но не слишком удачно переведенный с английского² термин прочно вошел в обиход.

Математическое программирование – эта отрасль математики, которая является теоретической основой решения задач о нахождении оптимальных решений.

Нахождение оптимальных решений экономических задач можно разбить на следующие этапы:

1. Построение математической модели³.
2. Нахождение оптимального метода одним из математических методов.
3. Практическое внедрение.

Построение математической модели состоит из такого функционального описания изучаемого экономического процесса математическими соотношениями, которое отражало бы его сущность. Иными словами в математической модели должны учитываться существенные особенности задачи, а также ограничивающие условия, которые могут повлиять на результат.

Можно считать, что процесс построения математической модели любой задачи начинается с ответов на следующие вопросы [5]:

1. Для определения каких величин должна быть построена модель, т.е. как идентифицировать переменные данной задачи?
2. Какие ограничения должны быть наложены на переменные, чтобы выполнялись условия, характерные для моделируемой системы?
3. В чем состоит цель задачи, для достижения которой из всех допустимых переменных нужно выбрать те, которые будут соответствовать оптимальному (наилучшему) решению задачи?

¹ Этот термин не следует смешивать с термином «программирование», обозначающим составление программы, осуществляющей определенный вычислительный процесс на компьютере.

² Выражение «mathematic programming» лучше было бы перевести как «математическое планирование».

³ Математическая модель – описание какого-либо класса явлений внешнего мира (в том числе, экономических явлений), выраженное с помощью математической символики.

Следует отметить, что в экономической теории на первоначальном этапе ее развития редко использовались математические формулировки. Тем не менее, многие классические доктрины экономики в словесной, завуалированной форме по существу содержали определенные математические утверждения; это иногда оставалось скрытым и от самих приверженцев таких доктрин. По мере развития экономической науки внутренне присущие ей математические черты постепенно проявлялись все сильнее. На более поздней стадии ее истории некоторые экономисты стали предлагать даже полную математизацию экономической теории. По-видимому, самым выдающимся представителем зарождавшегося математического направления в экономической теории был французский экономист Леон Вальрас. Именно он в конце XIX столетия заложил основы теории общего экономического равновесия [6].

Эта теория основана на том, что между соответствующими величинами в экономике существует не односторонняя причинно-следственная связь, а многосторонняя взаимозависимость, которую математически можно представить некоторой системой соотношений между этими величинами. Исследования Вальраса остаются блестящим примером математического подхода к экономическим явлениям. Большой его заслугой является ныне общепризнанная идея о возможности функционального описания экономических явлений системами уравнений, неравенств или и тех и других вместе.

В общем случае, задачу математического программирования можно поставить следующим образом:

Пусть в R^n заданы функции $f(X)$, $g_1(X)$, $g_2(X)$, ..., $g_m(X)$ и $b_i \in R^1$.

Требуется найти точку, $X^0 \in R^n$ удовлетворяющую условиям:

$$g_1(X) = b_1, g_2(X) = b_2, \dots, g_m(X) = b_m,$$

такую что

$$f(X_0) = \min_{\substack{g_i(X) = b_i \\ i=1, m}} f(X).$$

Кратко задачу математического программирования в общем случае можно записать так:

$$f(X) \rightarrow \min, \quad (1)$$

$$g_i(X) = b_i, \quad i = 1, 2, \dots, m. \quad (2)$$

При этом система уравнений (2), определяющая допустимое множество (или множество планов) задачи, называется системой ограничений задачи математического программирования, а функция $f(X)$ называется целевой функцией, а также критерием качества.

Нетрудно понять, что задача математического программирования не всегда имеет решение.

Простое достаточное условие существования решения основано на теореме Вейерштрасса и состоит в следующем: если допустимое множество замкнуто, непустое и ограничено (непустой компакт), а функция $f(X)$ на нем полунепрерывна снизу, то задача математического программирования имеет решение.

Замечание. Символ $f(X) \rightarrow \min$, в записи задачи математического программирования используется вместо слов «минимизировать функцию $f(X)$ ». Далее указываются ограничения, определяющие допустимое множество.

2. Составные части «Математического программирования»

В зависимости от особенностей целевой функции $f(X)$ и функций, задающих ограничения $g_i(X)$, задачи математического программирования делятся на ряд типов.

Если целевая функция $f(X)$ и функции $g_i(X)$ ($i = 1, 2, \dots, m$), входящие в систему ограничений, линейны (первой степени) относительно входя-

ших в задачу неизвестных x_j , то такой раздел математического программирования называется **линейным программированием**.

Методы и модели линейного программирования широко применяются при оптимизации процессов во всех отраслях народного хозяйства:

- при разработке производственной программы предприятия, распределение ее по исполнителям;
- при размещении заказов между исполнителями и по временным интервалам;
- при определении наилучшего ассортимента выпускаемой продукции, в задачах перспективного, текущего и оперативного планирования и управления;
- при планировании грузопотоков, определении плана товарооборота и его распределении: в задачах развития и размещения производительных сил, баз и складов систем обращения материальных ресурсов и т. д.

Особенно широкое применение методы и модели линейного программирования получили при решении задач экономии ресурсов (выбор ресурсосберегающих технологий, составление смесей, раскрой материалов), производственно-транспортных и других задач.

Начало линейному программированию было положено в 1939 г. математиком-экономистом Л. В. Канторовичем в работе «Математические методы организации и планирования производства». Появление этой работы открыло новый этап в применении математики в экономике. Спустя десять лет американский математик Дж. Данциг разработал эффективный метод решения данного класса задач - симплекс-метод. Термин **«линейное программирование»** впервые появился в 1951 г. в работах Дж. Данцига и Т. Купманса.

Линейное программирование и межотраслевой баланс характеризуют линейные взаимосвязи элементов народного хозяйства.

Однако при более глубоком исследовании в ряде задач появляются и связи нелинейного характера, когда с изменением одного элемента другие

и изменяются непропорционально первому. Например, даже простейшая транспортная задача принимает нелинейный вид, если стоимость перевозки единицы груза зависит от их общего количества. Поэтому вслед за разработкой моделей линейного программирования начались интенсивные исследования нелинейных моделей.

Если в задаче математического программирования целевая функция $f(X)$ и (или) хотя бы одна из функций системы ограничений $g_i(X)$ нелинейна, то такой раздел называется **нелинейным программированием**.

Методы и модели нелинейного программирования могут применяться при решении перечисленных выше задач, когда хотя бы одна из функций $f(X)$, $g_i(X)$ нелинейна. Кроме того, методы нелинейного программирования получили широкое применение при расчете экономически выгодных партий запуска деталей в производство, при определении экономически выгодной партии поставки, поставочного комплекта, размеров запасов, распределении ограниченных ресурсов, размещении производительных сил, в тарном хозяйстве, при решении многих производственно-экономических задач и т.д.

Если на все или некоторые переменные x_j , наложено условие дискретности, например, целочисленности $x_j \in N_0$, $N_0 = \{0, 1, 2, \dots\}$, то такие задачи рассматриваются в разделе математического программирования, называемом **дискретным программированием**, в частности, **целочисленным программированием**.

Методами **целочисленного программирования** решается широкий круг задач оптимизации с неделимостями, комбинаторного типа, с логическими условиями, с разрывной целевой функцией и т.д. В частности, задачи выбора (о назначениях), о контейнерных перевозках (о рюкзаке), о маршрутизации (коммивояжера), теории расписаний, комплектных поставок и комплектования, размещения производственно-складской структуры и т. п.

Если параметры целевой функции и (или) системы ограничений изменяются во времени или целевая функция имеет аддитивный

$$f(X) = \sum_{j=1}^n f_j(x_j),$$

либо мультипликативный вид

$$f(X) = \prod_{j=1}^n f_j(x_j),$$

или сам процесс выработки решения имеет многошаговый характер, то такие задачи решаются методами **динамического программирования**. Методами динамического программирования могут решаться задачи перспективного и текущего планирования, управления производством, поставками и запасами в условиях изменяющегося спроса, распределения ограниченных ресурсов, в частности, размещения капитальных вложений, замены оборудования, обновления и восстановления элементов сложных человеко-машинных организационных систем и т.д.

В перечисленных выше разделах математического программирования предполагается, что вся информация о протекании процессов заранее известна и достоверна. Такие методы оптимизации называются **детерминированными** или методами обоснования решений в условиях определенности.

Если параметры, входящие в функцию цели, или ограничения задачи являются случайными величинами или если приходится принимать решения в условиях риска, неполной информации, то говорят о проблеме стохастической оптимизации, а соответствующий раздел называется **стохастическим программированием**. К нему, в первую очередь, следует отнести методы и модели выработки решений в условиях конфликтных ситуаций (математическая теория игр), в условиях неполной информации (экспертные оценки), в условиях риска (статистические решения) и др. Позднее появились другие типы задач, учитывающих специфику целевой функции и системы ограничений, в связи с чем возникли **параметрическое, дробно-линейное, блочное, сетевое (потокосное), многоиндексное, булевское, комбинаторное** и другие

типы программирования. В случае нелинейностей специфика задач породила квадратичное, биквадратичное, сепарабельное, выпуклое и другие типы программирования. Появились численные методы отыскания оптимальных решений: градиентные, штрафных и барьерных функций, возможных направлений, линейной.

Следует отметить, что задачи математического программирования с одной целевой функцией решаются методами *скалярной оптимизации*. Однако реальные ситуации настолько сложны, что нередко приходится одновременно учитывать несколько целевых функций, которые должны принимать экстремальные значения. Например, дать продукцию больше, высокого качества и с минимальными затратами. Задачи, где находят решение по нескольким целевым функциям, относятся к *векторной оптимизации* – это так называемые задачи многокритериального подхода.

3. Линейное программирование

Линейным программированием называют раздел математического программирования, в котором при формулировке задач используются только линейные функции.

Экономические приложения наиболее характерны для линейного программирования. Ниже рассмотрим математические модели некоторых экономических задач.

Построение математических моделей некоторых экономических задач

Рассмотрим процесс построения математических моделей задач линейного программирования на примерах.

Пример 1. Задача использования сырья. Для изготовления двух видов продукции P_1 , P_2 используют три вида сырья: S_1 , S_2 , S_3 . Запасы сырья, количество единиц сырья, затрачиваемых на изготовление единицы продукции, а также величина прибыли, получаемая от реализации единицы продук-

ции, приведены в табл. 1. Необходимо составить такой план выпуска продукции, чтобы при её реализации получить максимальную прибыль.

Обозначим через x_1 количество единиц продукции P_1 , а через x_2 количество единиц продукции P_2 . Тогда, учитывая количество единиц сырья, затрачиваемых на изготовление единицы продукции, а также запасы сырья, получим систему ограничений

$$\begin{cases} 2x_1 + 5x_2 \leq 20, \\ 8x_1 + 5x_2 \leq 40, \\ 5x_1 + 6x_2 \leq 30, \end{cases}$$

которая показывает, что количество сырья, расходуемое на изготовление продукции, не может превысить имеющихся запасов. Если продукция P_1 не выпускается, то $x_1 = 0$; в противном случае $x_1 > 0$. То же самое получаем и для продукции P_2 . Таким образом, на неизвестные x_1 и x_2 должно быть наложено ограничение неотрицательности: $x_1 \geq 0$, $x_2 \geq 0$.

Таблица 1

Вид сырья	Запас сырья	Количество единиц сырья, идущих на изготовление единицы продукции	
		P_1	P_2
S_1	20	2	5
S_2	40	8	5
S_3	30	5	6
Прибыль от единицы продукции, (ден. ед.)		50	40

Конечную цель решаемой задачи – получение максимальной прибыли при реализации продукции – выразим как функцию двух переменных x_1 и x_2 . Реализация x_1 единиц продукции вида P_1 и x_2 единиц продукции вида P_2 даёт, соответственно, $50x_1$ и $40x_2$ (ден. ед.) прибыли, суммарная прибыль

$$f = 50x_1 + 40x_2 \text{ (ден. ед.)}.$$

Условиями не оговорена неделимость единицы продукции, поэтому x_1 и x_2 (план выпуска продукции) могут быть и дробными числами, следовательно, задача имеет бесконечное множество вариантов планов (значение x_1 и x_2 , которые удовлетворяют системе ограничений). Необходимо найти такие неотрицательные значения x_1 и x_2 , при которых функция f достигает максимума, то есть найти максимальное значение линейной функции

$$f = 50x_1 + 40x_2$$

при ограничениях

$$\begin{cases} 2x_1 + 5x_2 \leq 20, \\ 8x_1 + 5x_2 \leq 40, \\ 5x_1 + 6x_2 \leq 30, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Построенная линейная функция называется функцией цели и совместно с системой ограничений образует математическую модель рассматриваемой экономической задачи.

Задачу использования сырья можно легко обобщить, если при выпуске n видов продукции используются m видов сырья. Обозначим через S_i ($i=1,2,\dots,m$) виды сырья; b_i – запасы сырья i -го вида; P_j ($j=1,2,\dots,n$) – виды продукции; a_{ij} – количество единиц i -го сырья, идущего на изготовление единицы j -й продукции; c_j – величину прибыли, получаемой при реализации единицы продукции. Условия задачи запишем в табл. 2.

Таблица 2

Вид сырья	Запас сырья	Количество единиц i -го сырья, идущих на изготовление единицы j -й продукции			
		P_1	P_2	...	P_n
S_1	b_1	a_{11}	a_{12}	...	a_{1n}
S_2	b_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...	...
S_m	b_m	a_{m1}	a_{m2}	...	a_{mn}
Прибыль от единицы продукции, (ден. ед.)		P_1	P_2	...	P_n

Таблица 3

Сырье	Расход сырья на 1 ед. продукции P_1	Расход сырья на 1 ед. продукции P_2	Запас сырья, ед.
<i>A</i>	2	3	9
<i>B</i>	3	2	13

Опыт работы показал, что суточный спрос на продукцию P_1 никогда не превышает спроса на продукцию P_2 более чем на 1 ед. Кроме того, известно, что спрос на продукцию P_2 никогда не превышает 2 ед. в сутки.

Оптовые цены единицы продукции равны: 3 ден.ед. – для P_1 и 4 ден.ед. для P_2 .

Какое количество продукции каждого вида должно производить предприятие, чтобы доход от реализации продукции был максимальным?

Решение

Ответы на вышеперечисленные вопросы 1, 2, 3 п.1 могут быть сформулированы для данной задачи так: предприятию требуется определить объемы производства каждого вида продукции в тоннах, максимизирующие доход в ден. ед. от реализации продукции, с учетом ограничений на спрос и расход исходных продуктов.

Для построения математической модели остается только идентифицировать переменные и представить цель и ограничения в виде математических функций этих переменных.

Предположим, что предприятие изготовит x_1 единиц продукции P_1 и x_2 единиц продукции P_2 . Поскольку производство продукции P_1 и P_2 ограничено имеющимися в распоряжении предприятия сырьем каждого вида и спросом на данную продукцию, а также учитывая, что количество изготавливаемых изделий не может быть отрицательным, должны выполняться следующие неравенства:

$$\begin{cases} 2x_1 + 3x_2 \leq 9, \\ 3x_1 + 2x_2 \leq 13, \\ x_1 - x_2 \leq 1, \\ x_2 \leq 2. \end{cases} \quad (3)$$

$$x_1 \geq 0, x_2 \geq 0. \quad (4)$$

Доход от реализации x_1 единиц продукции P_1 и x_2 единиц продукции P_2 составит $f = 3x_1 + 4x_2$. Нас интересует максимальный доход, т.е.

$$f = 3x_1 + 4x_2 \rightarrow \max. \quad (5)$$

Итак, (5), (4), (3) является математической моделью задачи об ассортименте продукции.

Таким образом, мы приходим к следующей математической задаче: среди всех неотрицательных решений данной системы линейных неравенств требуется найти такое, при котором функция f принимает максимальное значение.

Пример 3. Использование мощностей оборудования. Предприятие имеет m моделей машин различных мощностей. Задан план по времени и номенклатуре: T - время работы каждой машины; продукции j -го вида должно быть выпущено не менее N_j - единиц.

Необходимо составить такой план работы оборудования, чтобы обеспечить минимальные затраты на производство, если известны производительность каждой i -й машины по выпуску j -го вида продукции b_{ij} и стоимость единицы времени, затрачиваемого i -й машиной на выпуск j -го вида продукции c_{ij} .

Другими словами, задача для предприятия состоит в следующем: требуется определить время работы i -й машины по выпуску j -го вида продукции x_{ij} , обеспечивающее минимальные затраты на производство при соблюдении ограничений по общему времени работы машин T и заданному количеству продукции N_j .

Задача решается на минимум затрат на производство:

$$f = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min.$$

По условию задачи машины работают заданное время T , поэтому данное ограничение можно представить в следующем виде:

$$\sum_{j=1}^n x_{ij} = T, \quad i = \overline{1, m}.$$

Ограничение по заданному количеству продукции выглядит следующим образом:

$$\sum_{i=1}^m b_{ij} x_{ij} \geq N_j, \quad j = \overline{1, n}.$$

Необходимо также учесть неотрицательность переменных $x_{ij} \geq 0$.

Задача поставлена так, чтобы израсходовать все отведенное время работы машины, т.е. обеспечить полную загрузку машины. При этом количество выпускаемой продукции каждого вида должно быть, по крайней мере, не менее N_j . Однако в некоторых случаях не допускается превышение плана по номенклатуре, тогда ограничения математической модели изменяются следующим образом:

$$f = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min,$$

$$\sum_{j=1}^n x_{ij} \leq T, \quad i = \overline{1, m};$$

$$\sum_{i=1}^m b_{ij} x_{ij} = N_j, \quad j = \overline{1, n};$$

$$x_{ij} \geq 0, \quad i = \overline{1, m}; \quad j = \overline{1, n}.$$

Пример 4. Минимизация дисбаланса на линии сборки. Промышленная фирма производит изделие, представляющее собой сборку из m различных узлов. Эти узлы изготавливаются на n заводах.

Из-за различий в составе технологического оборудования производительность заводов по выпуску j -го узла неодинакова и равна b_{ij} . Ка-

ждый i – й завод располагает максимальным суммарным ресурсом времени в течение недели для производства t узлов, равного величине T_i .

Задача состоит в максимизации выпуска изделий, что по существу эквивалентно минимизации дисбаланса, возникающего вследствие некомплектности поставки по одному или по нескольким видам узлов.

В данной задаче требуется определить еженедельные затраты времени (в часах) на производство j – го узла на i – м заводе, не превышающие в сумме временные ресурсы i – го завода и обеспечивающие максимальный выпуск изделий.

Пусть x_{ij} – недельный фонд времени (в часах), выделяемый на заводе i для производства узла j . Тогда объемы производства узла j будут следующими:

$$\sum_{i=1}^n b_{ij} x_{ij}, \quad j = 1, 2, \dots, n.$$

Так как в конечной сборке каждый из комплектующих узлов представлен в одном экземпляре, количество конечных изделий должно быть равно количеству комплектующих узлов, объем производства которых минимален:

$$\min \left(\sum_{i=1}^n b_{ij} x_{ij}, \quad j = 1, 2, \dots, m \right).$$

Условие рассматриваемой задачи устанавливает ограничение на фонд времени, которым располагает завод i .

Таким образом, математическая модель может быть представлена в следующем виде.

$$\begin{aligned} f = \min \left(\sum_{i=1}^n b_{ij} x_{ij}, \quad j = 1, 2, \dots, m \right) &\rightarrow \max \\ \sum_{j=1}^m x_{ij} &\leq T_i, \quad i = 1, 2, \dots, n; \\ x_{ij} &\geq 0, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m. \end{aligned}$$

Эта модель не является линейной, но ее можно привести к линейной форме с помощью простого преобразования. Пусть y – количество изделий:

$$y = \min \left(\sum_{i=1}^n b_{ij} x_{ij} = N_j, \quad j = 1, 2, \dots, m \right).$$

Этому выражению с математической точки зрения эквивалентна следующая формулировка: максимизировать $f = y$ при ограничениях

$$\begin{aligned} \sum_{i=1}^n b_{ij} x_{ij} - y &\geq 0, \quad j = 1, 2, \dots, m; \\ \sum_{j=1}^m x_{ij} &\leq T_i, \quad i = 1, 2, \dots, n; \\ x_{ij} &\geq 0, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m; \quad y \geq 0. \end{aligned}$$

Пример 5. Задача составления кормовой смеси или задача о диете.

Пусть крупная фирма имеет возможность покупать m различных видов сырья и приготавливать различные виды смесей (продуктов). Каждый вид сырья содержит разное количество питательных компонентов (ингредиентов).

Лабораторией фирмы установлено, что продукция должна удовлетворять по крайней мере некоторым минимальным требованиям с точки зрения питательности (полезности). Перед руководством фирмы стоит задача определить количество каждого i -го сырья, образующего смесь минимальной стоимости при соблюдении требований к общему расходу смеси и ее питательности.

Решение

Введем условные обозначения:

x_i - количество i -го сырья в смеси;

m - количество видов сырья;

n - количество ингредиентов в сырье;

a_{ij} - количество ингредиента j , содержащегося в единице i -го вида сырья;

b_j - минимальное количество ингредиента j , содержащегося в единице смеси;

c_i - стоимость единицы сырья i ;

q - минимальный общий вес смеси, используемый фирмой.

Задача может быть представлена в виде

$$f = \sum_{i=1}^m c_i x_i \rightarrow \min$$

при следующих ограничениях:

- на общий расход смеси:

$$\sum_{i=1}^m x_i \geq q;$$

- на питательность смеси:

$$\sum a_{ij} x_i \geq b_j, \quad \sum_{i=1}^m x_i, \quad j = 1, 2, \dots, n;$$

- на неотрицательность переменных:

$$x_i \geq 0, \quad i = 1, 2, \dots, m.$$

Пример 6. Задача составления жидких смесей. Еще один класс моделей, аналогичных рассмотренным выше, возникает при решении экономической проблемы, связанной с изготовлением смесей различных жидкостей, с целью получения пользующихся спросом готовых продуктов.

Представим себе фирму, торгующую различного рода химическими продуктами, каждый из которых является смесью нескольких компонентов. Предположим, что эта фирма планирует изготовление смесей m -видов. Обозначим подлежащее определению количество литров i –го химического компонента, используемого для получения j –го продукта через x_{ij} . Будем предполагать,

что

$$x_{ij} \geq 0, \quad i = \overline{1, n}; \quad j = \overline{1, m}.$$

Первая группа ограничений относится к объемам потребляемых химических компонентов:

$$\sum_{j=1}^m x_{ij} \leq S_i, \quad i = 1, 2, \dots, n,$$

где S_i - объем i –го химического компонента, которым располагает фирма в начале планируемого периода.

Вторая группа ограничений отражает требование, заключающееся в том, чтобы запланированный выпуск продукции хотя бы в минимальной степени удовлетворял имеющийся спрос на каждый из химических продуктов, т.е.

$$\sum_{i=1}^n x_{ij} \geq D_j, \quad j = 1, 2, \dots, m,$$

где D_j - минимальный спрос на продукцию j в течение планируемого периода.

Третья группа ограничений связана с технологическими особенностями, которые необходимо принимать во внимание при приготовлении смеси, например, простое ограничение, определяемое некоторыми минимально допустимыми значениями, отношения между объемами двух химических компонентов в процессе получения продукта j :

$$\frac{x_{ij}}{x_{i-1,j}} \geq r \quad \text{или} \quad x_{ij} - r x_{i-1,j} \geq 0.$$

где r - некоторая заданная константа.

Обозначив через P_{ij} доход с единицы продукции x_{ij} , запишем целевую функцию:

$$f = \sum_{i=1}^n \sum_{j=1}^m P_{ij} x_{ij} \rightarrow \max.$$

Пример 7. Задача о раскрое или о минимизации обрезков. Данная задача состоит в разработке таких технологических планов раскроя, при которых получается необходимый комплекс заготовок, а отходы (по длине, площади, объему, массе или стоимости) сводятся к минимуму.

Например, продукция бумажной фирмы выпускается в виде бумажных рулонов стандартной ширины L . По специальным заказам потребителей фирма поставляет рулоны других размеров, для этого производится разрезание стандартных рулонов. Типичные заказы на рулоны нестандартных размеров могут включать m видов шириной l_i , ($i = 1, 2, \dots, m$). Известна потребность

в нестандартных рулонах каждого вида, она равна b_i . Возможны n различных вариантов построения технологической карты раскроя рулонов стандартной ширины L на рулоны длиной l_j .

Обозначим через a_{ij} количество рулонов i -го вида, получаемых при раскрое единицы стандартного рулона по j -му варианту. При каждом варианте раскроя на каждый стандартный рулон возможны потери, равные P_j . К потерям следует относить также избыточные рулоны нестандартной длины l_i , получаемые при различных вариантах раскроя y_{ij} , $j = 1, 2, \dots, n$.

В качестве переменных следует идентифицировать количество стандартных рулонов, которые должны быть разрезаны при j -м варианте раскроя. Определим переменную следующим образом: x_j - количество стандартных рулонов, разрезаемых по варианту j , $j = 1, 2, \dots, n$.

Целевая функция - минимум отходов при раскрое

$$f = \left(\sum_{j=1}^n P_j x_j + \sum_{i=1}^m \sum_{j=1}^n y_{ij} \right) \rightarrow \min.$$

Ограничение на удовлетворение спроса потребителя

$$b_i = \sum_{j=1}^n a_{ij} x_j - \sum_{j=1}^n y_{ij}, \quad x_j \geq 0, \quad y_{ij} \geq 0.$$

Пример 8. Многосторонний коммерческий арбитраж (см. [5]). В сфере деятельности, связанной с валютными и биржевыми операциями, а также коммерческими сделками контрактного характера, возможны различного рода транзакции, позволяющие извлекать прибыль на разнице в курсе валют. Такого рода транзакции называются **коммерческим арбитражем**.

Представим себе коммерсанта (условно назовем его N), имеющего возможность реализовать многосторонний коммерческий арбитраж. Предположим, что число валютных рынков, вовлеченных в транзакционную деятельность коммерсанта N , равняется шести, а максимальное число воз-

можных транзакций равняется девяти. Подробные данные, характеризующие рассматриваемую задачу, приведены в табл. 4.

При транзакции x_1 продажа единицы валютного номинала (ценных бумаг) II позволяет приобрести r_{11} единиц валютного номинала I. При транзакции x_7 взамен единицы валютного номинала I можно получить r_{37} единиц валютного номинала III и r_{67} единиц валютного номинала VI. Остальные транзакции расшифровываются аналогично. Значения r_{ij} могут быть дробными. Заметим, что при любой транзакции x_i ($i = 1, 2, 3, 4, 5$) каждый из валютных номиналов можно обменять на валютный номинал I. Следует обратить внимание на правило выбора знака перед показателями в табл. 4. Чтобы отличать куплю от продажи, будем, соответственно, использовать знаки «плюс» и «минус» перед показателями, характеризующими данную транзакцию.

Таблица 4

Валютный номинал	Тип транзакции									Возможность рынка
	1	2	3	4	5	6	7	8	9	
I	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	-1	-1			≥ 0
II	-1					r_{26}			r_{29}	≥ 0
III		-1					r_{37}	r_{38}		≥ 0
IV			-1					-1		≥ 0
V				-1				r_{58}		≥ 0
VI					-1		r_{67}		-1	≥ 0
Размер транзакции	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	

Рассмотрим идеализированный случай, когда все транзакции коммерсанта N выполняются одновременно. Ограничения определяются единственным требованием - транзакция возможна лишь при условии, если коммерсант N располагает наличными ценными бумагами. Другими словами, количество

проданных ценных бумаг не должно превышать количество приобретенных.

Данные ограничения имеют вид

$$r_{11}x_1 + r_{12}x_2 + r_{13}x_3 + r_{14}x_4 + r_{15}x_5 - x_6 - x_7 \geq 0 ;$$

$$-x_1 + r_{26}x_6 + r_{29}x_9 \geq 0 ;$$

$$-x_2 + r_{37}x_7 + r_{38}x_8 \geq 0 ;$$

$$-x_3 - x_8 + r_{49}x_9 \geq 0 ;$$

$$-x_4 + r_{58}x_8 \geq 0 ;$$

$$-x_5 + r_{67}x_7 - x_9 \geq 0 ;$$

$$x_j \geq 0. \quad j = 1, 2, \dots, 9.$$

Пусть целевая функция представляет собой чистый доход, выраженный в единицах валютного номинала I, т.е. задача состоит в том, чтобы

$$f = r_{11}x_1 + r_{12}x_2 + r_{13}x_3 + r_{14}x_4 + r_{15}x_5 - x_6 - x_7 \rightarrow \max .$$

Упражнения

Составить математические модели следующих задач с экономическим содержанием:

1. В рационе животных используется два вида корма. Животные должны получать четыре вида питательных веществ. Составить рацион питания животных, обеспечивающий минимальные затраты, при исходных данных, заданных таблицей

Необходимое количество питательного вещества	Норма (ед. массы)	Содержание питательного вещества в единице корма	
		корм 1	корм 2
Пит. вещ. № 1	20	1	5
Пит. вещ. № 2	24	3	2
Пит. вещ. № 3	32	2	4
Пит. вещ. № 4	2	1	0
Стоимость единицы корма (ден. ед.)		4	6

2. Для изготовления изделий двух типов А и Б имеется 200 кг металла. На изготовление одного изделия типа А расходуется 2 кг металла, а одного

Вопросы для самопроверки

1. Что составляет содержание «Математического программирования»?
2. Из каких этапов состоит нахождение оптимальных решений экономических задач?
3. Что означает построение математической модели экономической задачи или экономического процесса?
4. Что включает в себя процесс построения математической модели?
5. Кто был выдающимся представителем зарождавшегося математического направления в экономической теории? На чем основана теория математического подхода к экономическим явлениям?
6. Сформулируйте задачу математического программирования в общем случае?
7. Всегда ли имеет решение задача математического программирования? Приведите какое-то достаточное условие существования решения задачи математического программирования.
8. Перечислите составные части математического программирования исходя из формулировки задачи математического программирования в общем виде.
9. Что называется детерминированной оптимизацией?
10. Что называется стохастической оптимизацией?
11. К какой оптимизации относится математическая теория игр?
12. Какие задачи математического программирования решаются методами скалярной оптимизации?
13. Какие задачи относятся к векторной оптимизации?
14. Дайте определение линейного программирования.
15. Какие Вы знаете еще математические модели экономических задач, кроме вышеприведенных математических моделей?
16. Разъясните, почему для нахождения решения задачи линейного программирования нельзя применить хорошо разработанные методы математического анализа исследования условного экстремума.

§11. Общая задача линейного программирования

Ключевые слова: задача в нормальной форме, задача в канонической форме, дополнительная переменная, план, опорный план, невырожденный опорный план, вырожденный опорный план, оптимальный план, выпуклая линейная комбинация точек, выпуклое множество (замкнутое, ограниченное), угловая точка, многогранник решений.

1. Задача в нормальной и канонической формах

Определение 1. Задачей линейного программирования в нормальной форме называется следующая задача:

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max. \quad (1)$$

[illegible]

$$x_j \geq 0, \quad (j=1,2,\dots,n), \quad (3)$$

где a_i, b_i, c_i - заданные постоянные величины. Соотношения (1)-(3) представляют координатную запись задачи.

Как отмечалось ранее (§10), в системе ограничений (2) все b_i , ($i=1, 2, 3, \dots, m$) можно считать неотрицательными.

Если ввести векторы

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, A_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}, A_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix}, \dots, A_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix}, A_0 = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix},$$

$$C = (c_1, c_2, \dots, c_n).$$

то можно привести векторную запись задачи (1)-(3).

Векторная запись

$$f = CX \rightarrow \max,$$

$$A_1x_1 + A_2x_2 + \dots + A_nx_n \leq A_0, \quad X \geq 0.$$

Формулировка задачи становится более компактной, если ввести матрицу $A = (a_{ij})$, $i = \overline{1, m}$; $j = \overline{1, n}$. В этом случае задачу (1)-(3) можно записать в матричной форме.

Матричная запись

$$f = CX \rightarrow \max,$$

$$AX \leq A_0, X \geq 0.$$

Приведем еще одну форму записи задачи (1)-(3).

Запись с помощью знаков суммирования

$$f = \sum_{j=1}^n c_j x_j \rightarrow \max,$$

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, 2, \dots, m; \quad x_j \geq 0, \quad j = 1, 2, \dots, n.$$

Матрица A называется *матрицей условий*.

Наряду с нормальной формой задачи линейного программирования широкое распространение получила каноническая форма.

Определение 2. Задачей линейного программирования в канонической форме называется следующая задача (в матричной записи):

$$f = CX \rightarrow \max,$$

$$AX = A_0, X \geq 0.$$

Таким образом, две формы задачи линейного программирования отличаются лишь типом основных ограничений: в нормальной форме ограничения заданы неравенствами, а в канонической они имеют вид равенств.

Отметим, что ограничения типа неравенств можно свести введением дополнительных переменных к ограничениям типа равенств.

Лемма 1. Каждая задача на максимум типа неравенств (1)-(3) эквивалентна задаче с ограничениями типа равенств

$$f(X) = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \rightarrow \max, \quad (4)$$

$$\sum_{j=1}^n a_{ij} x_j + x_{n+i} = b_i, \quad i = 1, 2, \dots, m, \quad (5)$$

$$x_j \geq 0, \quad j = 1, 2, \dots, n, n+1, \dots, n+m. \quad (6)$$

Переменные x_{n+i} , $i = \overline{1, m}$, в задаче (1)-(3) называются *дополнительными переменными*. Покажем, что точка X^0 будет решением задачи (1)-(3) тогда и только тогда, если

$$X_0, \quad x_{n+1}^0 = -\sum_{j=1}^n a_{1j} x_j^0 + b_1, \dots, \quad x_{n+m}^0 = -\sum_{j=1}^n a_{mj} x_j^0 + b_m \quad (7)$$

будет решением задачи (4)-(6).

Доказательство представляем читателю проводить самостоятельно в качестве упражнения.

Таким образом, ограничения типа неравенств можно свести введением дополнительных переменных к ограничениям типа равенств. С другой стороны, каждое ограничение типа равенства $AX = A_0$ можно записать в виде двух ограничений типа неравенств $AX \leq A_0$, $-AX \leq -A_0$. Поэтому задачи в нормальной и канонической формах эквивалентны между собой (с точностью до размерностей n и m).

2. Основные определения

В предыдущем пункте было показано, что систему ограничений любой задачи линейного программирования можно привести к системе m линейных уравнений с n неизвестными; значения переменных x_j , ($j = 1, 2, \dots, n$), при которых линейная функция достигает максимального значения. Следует заметить, что там, где необходимо найти минимальное значение линейной функции, достаточно заменить на противоположный знак линейной функции и найти максимальное значение последней функции. Заменяя на противоположный знак полученного максимального значения, определяем минимальное значение исходной линейной функции. Аналогичным образом, любую задачу максимизации можно свести к задаче минимизации.

Поэтому общую задачу линейного программирования можно определить следующим образом:

Определение 6. Опорный план называется *невыврожденным*, если он содержит m положительных компонент, в противном случае опорный план называется вырожденным.

Определение 7. *Оптимальным планом* или *оптимальным решением* задачи линейного программирования называется план, доставляющий наименьшее (наибольшее) значение линейной функции.

Свойства решений задачи линейного программирования тесно связаны со свойствами выпуклых множеств.

3. Выпуклые множества

Здесь приведем основные сведения о выпуклых множествах. Пусть на плоскости R^2 заданы две точки: $A_1(x_1^1, x_2^1)$ и $A_2(x_1^2, x_2^2)$, определяющие прямолинейный направленный отрезок $\overline{A_1A_2}$.

Точка A , для которой выполняются условия

$$A = \lambda_1 A_1 + \lambda_2 A_2, \quad (13)$$

$$\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1. \quad (14)$$

называется *выпуклой линейной комбинацией точек* A_1 и A_2 . При $\lambda_1 = 1$ и $\lambda_2 = 0$ точка A совпадает с концом отрезка A_1 и при $\lambda_1 = 0$ и $\lambda_2 = 1$ с концом отрезка A_2 . Точки A_1 и A_2 называются *угловыми или крайними точками отрезка* $\overline{A_1A_2}$.

Из определения выпуклой линейной комбинации точек видно, что угловая точка не может быть представлена как выпуклая линейная комбинация двух других точек отрезка. Соотношения (13) и (14) верны независимо от размерности пространства.

Пусть имеется n точек $A_1, A_2, \dots, A_n$. Точка A - выпуклая линейная комбинация, если выполняются условия

$$A = \lambda_1 A_1 + \lambda_2 A_2 + \dots + \lambda_n A_n,$$

$$\lambda_j \geq 0 \quad (j = 1, 2, \dots, n), \quad \sum_{j=1}^n \lambda_j = 1.$$

Множество точек называется выпуклым, если оно вместе с любыми двумя точками содержит и их произвольную выпуклую линейную комбинацию. Гео-

метрический смысл этого определения состоит в том, что множеству вместе с его двумя произвольными точками полностью принадлежит и прямолинейный отрезок, их соединяющий. Примерами выпуклых множеств служат прямолинейный отрезок, прямая, полуплоскость, круг, шар, куб, полупространство и др.

Угловыми точками выпуклого множества называются точки, не являющиеся выпуклой комбинацией двух различных точек множества. Например, угловыми точками треугольника являются его вершины, угловыми точками круга - точки окружности, которая его ограничивает. Таким образом, выпуклое множество может иметь конечное и бесконечное число угловых точек. Прямая, плоскость, полуплоскость, пространство, полупространство угловых точек не имеют.

Выпуклым многоугольником называется выпуклое замкнутое ограниченное множество на плоскости, имеющее конечное число угловых точек. Угловые точки многоугольника называются его вершинами, а отрезки, соединяющие две вершины и образующие его границу, - сторонами многоугольника. Опорной прямой выпуклого многоугольника называется прямая, имеющая с многоугольником по одну сторону от нее, хотя бы одну общую точку.

Выпуклым многогранником называется выпуклое замкнутое ограниченное множество трехмерного пространства, имеющее конечное число угловых точек. Угловые точки многогранника называются его вершинами; многоугольники, ограничивающие многогранник, - гранями; отрезки по которым они пересекаются, - ребрами. Опорной плоскостью многогранника называется плоскость, имеющая с многогранником, расположенным по одну от нее, хотя бы одну общую точку.

Теорема 1. Замкнутый, ограниченный, выпуклый многогранник является выпуклой линейной комбинацией своих угловых точек.

4. Геометрическая интерпретация задачи линейного программирования

Рассмотрим задачу линейного программирования, система ограничений которой задана в виде неравенств.

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min. \quad (15)$$

[illegible]

$$x_j \geq 0, \quad (j = 1, 2, \dots, n). \quad (17)$$

Совокупность чисел $x_1, x_2, \dots, x_n$, удовлетворяющих ограничениям (15)-(17) называется *решением*. Если система (16) при условии (17) имеет хотя бы одно решение, она называется *совместной*, в противном случае - *несовместной*.

Рассмотрим на плоскости R^2 совместную систему линейных нера-

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 \leq b_1, \\ a_{21}x_1 + a_{22}x_2 \leq b_2, \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m, \\ x_1 \geq 0, \quad x_2 \geq 0. \end{array}$$

Это все равно, что в системе (16)-(17) положить $n=2$. Каждое неравенство этой системы геометрически определяет полуплоскость с граничной прямой $a_i x_1 + a_i x_2 = b_i \ (i=1,2,\dots,m)$.

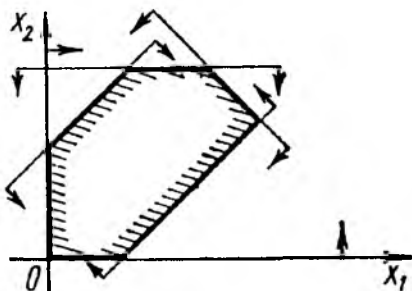


Рис.1

Условия неотрицательности определяют полуплоскости. соответственно, с граничными прямыми $x_1 = 0$, $x_2 = 0$. Система совместна, поэтому полуплоскости, как выпуклые множества, пересекаясь, образуют общую часть, которая является выпуклым множеством и представляет собой совокупность точек, координаты каждой из которых являются решением данной системы (рис. 1). Совокупность этих точек (решений) назовем многоугольником решений. Он может быть точкой, отрезком, лучом, многоугольником, неограниченной многоугольной областью.

Если в системе ограничений (16)-(17) $n = 3$, то каждое неравенство геометрически представляет полупространство трехмерного пространства, граничная плоскость которого $a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 = b_i$ ($i = 1, 2, \dots, m$), а условия неотрицательности – полупространства с граничными плоскостями, соответственно, $x_j = 0$ ($j = 1, 2, 3$). Если система ограничений совместна, то эти полупространства, как выпуклые множества, пересекаясь, образуют в трехмерном пространстве R^3 общую часть, которая называется многогранником решений. Многогранник решений может быть точкой, отрезком, лучом, многоугольником, многогранником, многогранной неограниченной областью.

Пусть в системе ограничений (16)-(17) $n > 3$; тогда каждое неравенство определяет полупространство пространства R^n с граничной гиперплоскостью $a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i$ ($i = 1, 2, \dots, m$), а условия неотрицательности – полупространства с граничными гиперплоскостями $x_j = 0$ ($j = 1, 2, \dots, n$).

Если система ограничений совместна, то по аналогии с R^3 пространством она образует общую часть пространства R^n , называемую многогранником решений, так как координаты каждой его точки являются решением.

Таким образом, геометрически задача линейного программирования представляет собой отыскание такой точки многогранника решений, координаты которой доставляют линейной функции минимальное значение, причем допустимыми решениями служат все точки многогранника решений.

5. Свойства решений задачи линейного программирования

Теорема 2. Множество всех планов задачи линейного программирования выпукло (если оно не пусто).

Теорема 3. Линейная функция задачи линейного программирования достигает своего минимального значения в угловой точке многогранника решений. Если линейная функция принимает минимальное значение более чем в одной угловой точке, то она достигает того же значения в любой точке являющейся выпуклой линейной комбинацией этих точек.

Теорема 4. Если известно, что система векторов $A_1, A_2, \dots, A_k$ ($k \leq n$) в разложении (12) линейно независима и такова, что

$$A_1 x_1 + A_2 x_2 + \dots + A_k x_k = A_0,$$

где все $x_i \geq 0$, то точка $X = (x_1, x_2, \dots, x_k, 0, \dots, 0)$ является угловой точкой многогранника решений.

Здесь X есть n -мерный вектор, последние $n-k$ компонент которого равны нулю.

Теорема 5. Если $X = (x_1, x_2, \dots, x_n)$ - угловая точка многогранника решений, то векторы в разложении (12), соответствующие положительным x_i , являются линейно независимыми.

Следствие 1. Так как векторы $A_1, A_2, \dots, A_n$ имеют размерность m , то угловая точка многогранника решений имеет не более чем m положительных компонент $x_i > 0$, ($i = 1, 2, \dots, m$).

Следствие 2. Каждой угловой точке многогранника решений соответствует $k \leq m$ линейно независимых векторов системы $A_1, A_2, \dots, A_n$.

Не теряя общности, можно предположить, что система векторов $A_1, A_2, \dots, A_n$ задачи линейного программирования (8)-(10) всегда содержит m линейно независимых векторов. Если при решении частной задачи это свойство не очевидно, то первоначальную систему векторов дополняют m линейно независимыми векторами, затем находят решение расширенной задачи.

Итак, если задача линейного программирования ограничена на многограннике решений, то:

1) существует такая угловая точка многогранника решений, в которой линейная функция задачи линейного программирования достигает своего оптимума;

2) каждый опорный план соответствует угловой точке многогранника решений. Поэтому для решения задачи линейного программирования необходимо исследовать угловые точки многогранника решений, т.е. только опорные планы.

Упражнения

Приведите каждую из следующих задач к канонической форме.

1. $f = 3x_1 - 4x_2 + x_3 \rightarrow \min.$

$$\begin{cases} 6x_1 + 5x_2 + 3x_3 = 8, \\ 12x_1 + 9x_2 + 3x_3 \leq 14, \\ 3x_1 + 5x_2 + 2x_3 \geq 8, \\ x_j \geq 0, (j = 1, 2, 3). \end{cases}$$

2. $f = 2x_1 - 4x_2 + 2x_3 \rightarrow \max.$

$$\begin{cases} x_1 + 2x_2 - x_3 \geq 6, \\ 4x_1 - x_2 + x_3 \leq 18, \\ 3x_1 + x_2 + 2x_3 \leq 6, \\ x_j \geq 0, (j = 1, 2, 3). \end{cases}$$

3. $f = 3x_1 + 2x_2 - x_3 \rightarrow \max.$

$$\begin{cases} x_1 + 2x_2 \leq 7, \\ 3x_1 + x_2 - 2x_3 = 8, \\ 2x_1 + x_3 \geq -5, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

4. $f = 3x_1 - 2x_2 - x_3 \rightarrow \min.$

$$\begin{cases} x_1 + 2x_2 = 3, \\ 3x_1 - x_2 + x_3 \leq 5, \\ x_1 + 3x_2 \geq -2, \\ x_1 \geq 0, x_3 \geq 0. \end{cases}$$

5. $f = -2x_1 + 3x_2 + 3x_3 \rightarrow \max.$

$$\begin{cases} x_1 + x_2 < 3, \\ x_2 + x_4 > -5, \\ 2x_1 + x_3 \geq -5, \\ x_1 \geq 0. \end{cases}$$

6. $f = 4x_2 + x_3 \rightarrow \min,$

$$\begin{cases} x_1 + x_2 + x_3 = 1, \\ x_2 - x_3 \geq 7, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{cases}$$

Вопросы для самопроверки

1. Дайте определение задачи линейного программирования в нормальной форме.

2. Дайте определение задачи линейного программирования в канонической форме.

3. Каким образом ограничения типа неравенств можно свести к ограничениям типа равенств? Доказать лемму. Какие переменные называются дополнительными?
4. Сформулируйте общую задачу линейного программирования.
5. Чем отличается каноническая форма задачи линейного программирования от общей?
6. Дайте определения плана: невырожденного и вырожденного плана, оптимального плана.
7. Какое множество называется выпуклым?
8. Какая точка выпуклого множества называется угловой?
9. В чем заключается важность теоремы 1?
10. Что называется многогранником решений?
11. Дайте геометрическое истолкование задачи линейного программирования.
12. В какой точке многогранника решений линейная функция задачи линейного программирования достигает своего оптимального решения?
13. Какой вид имеет угловая точка многогранника решений и какому плану она соответствует?
14. Какие планы необходимо исследовать, чтобы найти оптимальное значение линейной функции?
15. Дайте геометрическое истолкование задачи линейного программирования.
16. В какой точке многогранника решений линейная функция задачи линейного программирования достигает своего оптимального значения?
17. Какой вид имеет угловая точка многогранника решений и какому плану она соответствует?
18. Какие планы необходимо исследовать, чтобы найти оптимальное значение линейной функции?

§11 определяет полуплоскость с граничной прямой: $a_{i1}x_1 + a_{i2}x_2 = b_i$ ($i = 1, 2, \dots, m$), $x_1 = 0$, $x_2 = 0$.

Линейная функция (1) при фиксированных значениях f является уравнением прямой линии: $c_1x_1 + c_2x_2 = \text{const}$. Построим многоугольник решений системы ограничений (2)-(3) и график линейной функции (1) при $f = 0$ (рис. 1). Тогда поставленной задаче линейного программирования можно дать следующую интерпретацию. Найти точку многоугольника решений, в которой прямая $c_1x_1 + c_2x_2 = \text{const}$ - опорная и функция f при этом достигает минимума.

Значения $f = c_1x_1 + c_2x_2$ возрастают в направлении градиента

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (c_1, c_2)$$

линейной функции f . Поэтому прямую $f = 0$ передвигаем параллельно самой себе в направлении вектора ∇f . Из рис. 1 следует, что прямая дважды становится опорной по отношению к многоугольнику решений (в точках A и E), причем минимальное значение принимает в точке A . Координаты точки $A(x_1, x_2)$ находим, решая систему уравнений прямых AB и AE .

Если многоугольник решений представляет собой неограниченную область, то возможны два случая.

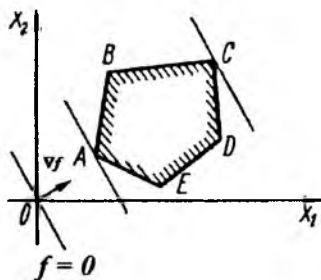


Рис. 1

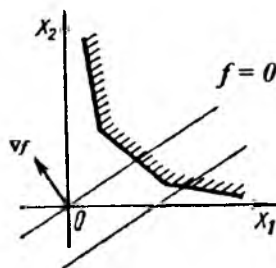


Рис. 2

Случай 1. Прямая $c_1x_1 + c_2x_2 = const$, передвигаясь в направлении вектора ∇f или противоположно ему, постоянно пересекает многоугольник решений и ни в какой точке не является опорной к нему. В этом случае линейная функция не ограничена на многоугольнике решений как сверху, так и снизу (рис. 2).

Случай 2. Прямая $c_1x_1 + c_2x_2 = const$, передвигаясь, все же становится опорной относительно многоугольника решений (рис. 3) Тогда в зависимости от вида области линейная функция может быть ограниченной сверху и неограниченной снизу (рис. 3а), ограниченной снизу и неограниченной сверху (рис. 3, б), либо ограниченной как снизу, так и сверху (рис. 3в).

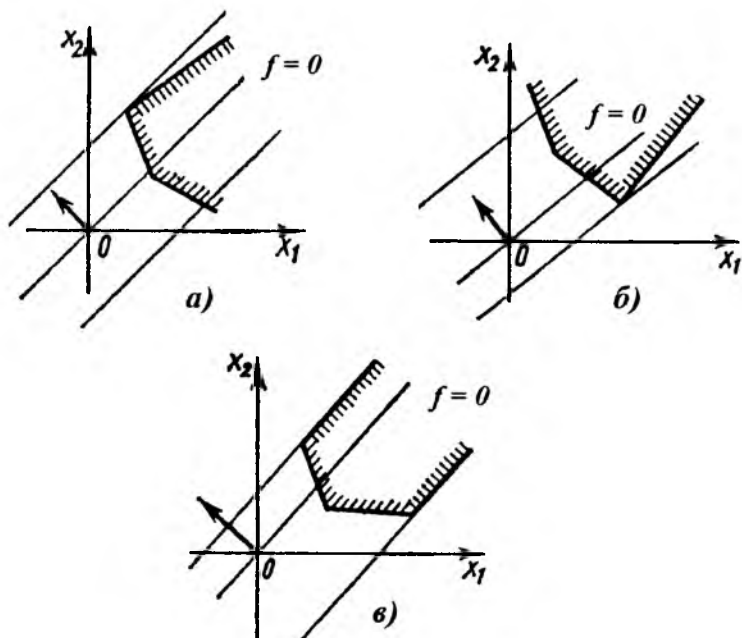


Рис. 3

2. Примеры задач, решаемых графическим методом

Решим графическим методом некоторые простейшие экономические задачи из §10.

Пример 1. Определение оптимального ассортимента продукции.

Предприятие изготавливает два вида продукции P_1 и P_2 , которая поступает в оптовую продажу. Для производства продукции используется два вида сырья A и B . Максимально возможные запасы сырья в сутки составляют 9 и 13 единиц, соответственно. Расход сырья на единицу продукции вида P_1 и вида P_2 дан в табл. 1.

Таблица 1

Сырье	Расход сырья на 1 ед. продукции P_1	Расход сырья на 1 ед. продукции P_2	Запас сырья, ед.
A	2	3	9
B	3	2	13

Опыт работы показал, что суточный спрос на продукцию P_1 никогда не превышает спроса на продукцию P_2 более чем на 1 ед. Кроме того, известно, что спрос на продукцию P_2 никогда не превышает 2 ед. в сутки.

Оптовые цены единицы продукции равны: 3 ден.ед. – для P_1 и 4 ден.ед. для P_2 .

Какое количество продукции каждого вида должно производить предприятие, чтобы доход от реализации продукции был максимальным?

Решение

Нами в п.3 §10 была построена математическая модель задачи. Математическая модель задачи об ассортименте продукции имеет вид:

$$Z = 3x_1 + 4x_2 \rightarrow \max,$$

$$\begin{cases} 2x_1 + 3x_2 \leq 9, \\ 3x_1 + 2x_2 \leq 13, \\ x_1 - x_2 \leq 1, \\ x_2 \leq 2, \end{cases}$$

$$x_1 \geq 0, x_2 \geq 0.$$

Эту задачу решим графическим методом.

Построим многоугольник решений (рис.4). Для этого на плоскости R^2 , где введена прямоугольная декартова система координат, изобразим граничные прямые

$$2x_1 + 3x_2 = 9 \quad (L_1),$$

$$3x_1 + 2x_2 = 13 \quad (L_2),$$

$$x_1 - x_2 = 1 \quad (L_3),$$

$$x_2 = 2 \quad (L_4),$$

$$x_1 = 0, \quad x_2 = 0,$$

и установим, какую полуплоскость определяет каждое неравенство относительно граничной прямой. В результате получим многоугольник $OABCD$.

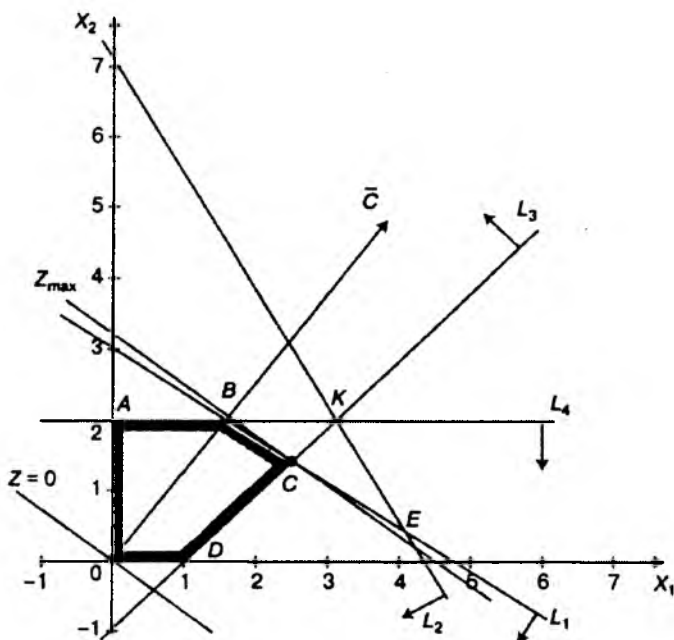


Рис. 4

Построим градиент $\bar{C} = \nabla Z = (3; 4)$ и прямую $3x_1 + 4x_2 = 0$. Перемещаем прямую $Z = 0$ параллельно самой себе в направлении вектора $\bar{C}$. Из рис. 4 следует, что по отношению к многоугольнику решений опорной эта прямая становится в точке C , где линейная функция принимает максимальное зна-

чение. Точка C лежит на пересечении прямых L_1 и L_3 . Для определения ее координат решим систему уравнений

$$\begin{cases} 2x_1 + 3x_2 = 9, \\ x_1 - x_2 = 1. \end{cases}$$

Оптимальный план задачи: $x_1 = 2,4$; $x_2 = 1,4$, т.е. $X_{opt} = (2,4; 1,4)$, и

$$Z_{\max} = Z(X_{opt}) = 3 \cdot 2,4 + 4 \cdot 1,4 = 12,8.$$

Полученное решение означает, что объем производства продукции P_1 должен быть равен 2,4 ед., а продукции P_2 - 1,4 ед. Максимальный доход, получаемый в этом случае, составит 12,8 ден.ед.

Пример 2. Задача использования сырья. Как нам известно (п.2 §10) математическая модель задачи имеет вид:

$$\begin{aligned} Z &= 50x_1 + 40x_2 \rightarrow \max, \\ \begin{cases} 2x_1 + 5x_2 \leq 20, \\ 8x_1 + 5x_2 \leq 40, \\ 5x_1 + 6x_2 \leq 30, \end{cases} \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

Решение

Построим многоугольник решений (рис. 5). Для этого на плоскости R^2 , где введена прямоугольная декартова система координат изобразим граничные прямые

$$\begin{cases} 2x_1 + 5x_2 = 20 & (L_1), \\ 8x_1 + 5x_2 = 40 & (L_2), \\ 5x_1 + 6x_2 = 30 & (L_3), \\ x_1 = 0, \quad x_2 = 0. \end{cases}$$

Взяв какую-нибудь точку, например, начало координат, установим, какую полуплоскость определяет соответствующее неравенство. Многоугольником решений данной задачи является ограниченный пятиугольник $OABCD$.

Для построения прямой $50x_1 + 40x_2 = 0$ строим градиент $\nabla Z = (50; 40) = 10 \cdot (5; 4)$ и через точку O проводим прямую, перпендикулярную ему. По-

строенную прямую $Z = 0$ перемещаем параллельно самой себе в направлении вектора ∇Z .

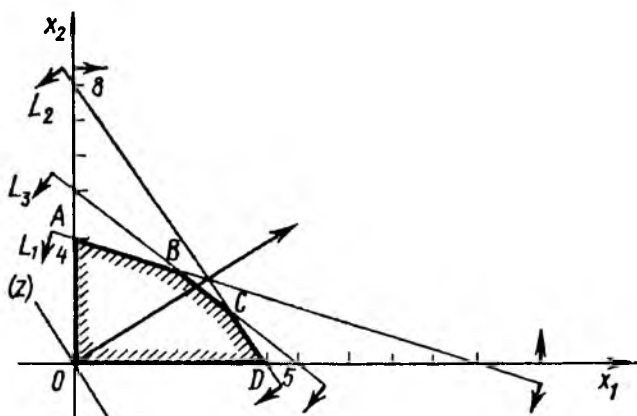


Рис. 5

Из рис. 5 следует, что опорной по отношению к многоугольнику решений эта прямая становится в точке C , где функция Z принимает максимальное значение. Точка C лежит на пересечении прямых L_2 и L_3 . Для определения ее координат решим систему уравнений

$$\begin{cases} 8x_1 + 5x_2 = 40, \\ 5x_1 + 6x_2 = 30. \end{cases}$$

Оптимальный план задачи: $x_1 = \frac{90}{23} \approx 3,9$; $x_2 = \frac{40}{23} \approx 1,7$, т.е. $X_{opt} = \left(\frac{90}{23}; \frac{40}{23} \right)$ и

$$Z_{\max} = Z(X_{opt}) = 50 \cdot 3,9 + 40 \cdot 1,7 \approx 260,3.$$

Таким образом для того чтобы получить максимальную прибыль в размере 260,3 ден. ед., необходимо запланировать производство 3,9 ед. продукции P_1 и 1,7 ед. продукции P_2 .

Вообще, с помощью графического метода может быть решена задача линейного программирования, система ограничений которой содержит n неизвестных и m линейно независимых уравнений, если n и m связаны соотношением $n - m = 2$.

Пример 3. Графическим методом решить следующую задачу линейного программирования

$$\begin{aligned} Z &= 2x_1 - x_2 + x_3 - 3x_4 + 4x_5 \rightarrow \max, \\ \begin{cases} x_1 - x_2 + 3x_3 - 18x_4 + 2x_5 = -4, \\ 2x_1 - x_2 + 4x_3 - 21x_4 + 4x_5 = 22, \\ 3x_1 - 2x_2 + 8x_3 - 43x_4 + 11x_5 = 38, \\ x_j \geq 0, \quad (j = 1, 2, \dots, 5). \end{cases} \end{aligned}$$

Решение

Используя метод Жордана-Гаусса, произведем три полных исключения неизвестных x_1, x_2, x_3 . В результате приходим к системе

$$\begin{cases} x_1 & +x_4 - 3x_5 & = 6, \\ x_2 & +7x_4 + 10x_5 & = 70, \\ x_3 & -4x_4 + 5x_5 & = 20. \end{cases} \quad (4)$$

откуда

$$x_1 = 6 - x_4 + 3x_5, \quad x_2 = 70 - 7x_4 - 10x_5, \quad x_3 = 20 + 4x_4 - 5x_5. \quad (5)$$

Подставляя эти значения в линейную функцию и отбрасывая в системе (4) базисные переменные, получаем задачу, выраженную только через свободные неизвестные x_4, x_5 :

$$\begin{aligned} Z &= 6x_4 + 15x_5 - 38 \rightarrow \max, \\ \begin{cases} x_4 - 3x_5 \leq 6, \\ 7x_4 + 10x_5 \leq 70, \\ -4x_4 + 5x_5 \leq 20, \\ x_4 \geq 0, \quad x_5 \geq 0. \end{cases} \end{aligned}$$

Построим многогранник решений в системе координат $x_4 O x_5$ (рис. 6).

Из рис. 6 заключаем, что линейная функция принимает максимальное значение в угловой точке B , которая лежит на пересечении прямых 2 и 3. В результате решения системы

$$\begin{cases} 7x_4 + 10x_5 = 70, \\ -4x_4 + 5x_5 = 20 \end{cases}$$

находим: $x_4 = 2$; $x_5 = \frac{28}{5}$. Максимальное значение функции

$$Z_{\max} = -38 + 12 + 84 = 58.$$

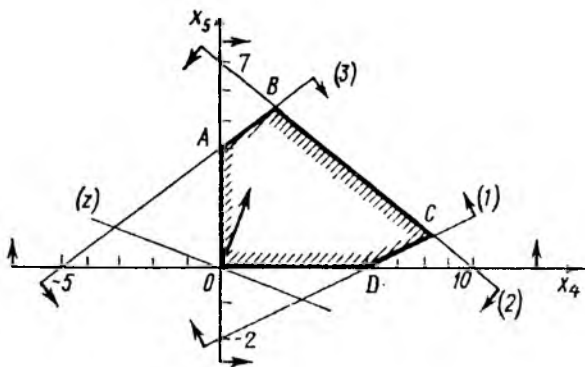


Рис. 6

Для отыскания оптимального плана исходной задачи подставляем в

(5) найденные значения $x_4 = 2$; $x_5 = \frac{28}{5}$. Окончательно получаем:

$$x_1 = \frac{104}{5}, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 2, \quad x_5 = \frac{28}{5}.$$

Таким образом,

$$X_{opt} = \left(\frac{104}{5}; \frac{40}{3}; 0; 2; \frac{28}{5} \right) \text{ и } Z_{\max} = Z(X_{opt}) = Z\left(\frac{104}{5}; \frac{40}{3}; 0; 2; \frac{28}{5} \right) = 58.$$

Упражнения

1. Решить графическим методом задачи с двумя переменными.

а) $Z = 3x_1 + 4x_2 \rightarrow \max.$

б) $Z = 3x_1 - x_2 \rightarrow \max.$

$$\begin{cases} 4x_1 - x_2 \geq 0, \\ -x_1 + x_2 \leq 3, \\ 3x_1 + 2x_2 \geq 6, \\ 2x_1 - 5x_2 \geq 0, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

$$\begin{cases} 2x_1 - x_2 \leq 4, \\ -x_1 + x_2 \leq 2, \\ 3x_1 - 2x_2 \geq 0, \\ x_1 - x_2 \leq 0, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

2. Решить графическим методом задачу с $n = 4$ переменными.

$$a) Z = 4x_1 + 12x_2 + 4x_3 + 8x_4 \rightarrow \max. \quad б) Z = -2x_1 + 6x_2 - 3x_3 + 6x_4 \rightarrow \min.$$

$$\begin{cases} x_1 + 4x_2 + x_3 + 2x_4 = 12, \\ 2x_1 + 3x_2 + 3x_3 + x_4 = 12. \end{cases}$$

$$x_j \geq 0, j = 1, 2, 3, 4.$$

$$\begin{cases} x_1 + 3x_2 + 3x_3 + x_4 = 6, \\ -2x_1 + 2x_2 - x_3 + 2x_4 = 6, \end{cases}$$

$$x_j \geq 0, j = 1, 2, 3, 4.$$

3. В рационе животных используется два вида корма. Животные должны получать четыре вида питательных веществ. Составить рацион питания животных, обеспечивающий минимальные затраты, при исходных данных, заданных таблицей

Необходимое количество питательного вещества	Норма (ед. массы)	Содержание питательного вещества в единице корма	
		Корм 1	Корм 2
Пит. вещ. № 1	20	1	5
Пит. вещ. № 2	24	3	2
Пит. вещ. № 3	32	2	4
Пит. вещ. № 4	2	1	0
Стоимость единицы корма (ден. ед.)		4	6

4. Для изготовления изделий двух типов А и Б имеется 200 кг металла. На изготовление одного изделия типа А расходуется 2 кг металла, а одного изделия типа Б – 4 кг. Составить план производства, обеспечивающий получение наибольшей выручки от продажи изготовленных изделий, если изделие типа А стоит 50 , а одно изделие типа Б стоит 70 , причём изделий типа А можно изготовить не более 60, и изделий типа Б – не более 30.

Вопросы для самопроверки

1. На чем основан графический метод решения задачи линейного программирования?
2. Как определить по рисунку имеет ли задача линейного программирования решение или ее оптимум находится в $\pm\infty$?
3. Какие задачи линейного программирования можно решать графическим методом?

§ 13. Симплексный метод решения задачи линейного программирования

Ключевые слова: симплексный метод, опорный план, первоначальный опорный план, оптимальный план, условие оптимальности.

Для решения задачи линейного программирования имеется много различных методов. В особенности это касается конкретных экономических задач. Систематическое исследование общих задач линейного программирования и разработка общих методов их решения были начаты в работах Л.В. Канторовича в 1939 г. В настоящее время основным общим методом решения общих задач линейного программирования является так называемый симплексный метод, предложенный Данцигом в 1949 году и многочисленные его модификации.

На основании материала, рассмотренного в §§11-12, можно сделать следующий вывод.

Существует такая угловая точка многогранника решений, в которой линейная функция достигает своего наименьшего (наибольшего) значения. Каждой угловой точке многогранника решений соответствует опорный план. Каждый опорный план определяется системой m линейно независимых векторов, содержащихся в данной системе из n векторов, $A_1, A_2, \dots, A_n$. Для отыскания оптимального плана необходимо исследовать только опорные планы. Верхняя граница количество опорных планов, содержащихся в данной задаче, определяется числом сочетаний C_n^m . При больших m и n найти оптимальный план, перебирая все опорные планы задачи, очень трудно. Поэтому необходимо, иметь схему, позволяющую осуществлять упорядоченный переход от одного опорного плана к другому. Такой схемой является симплексный метод, который позволяет, исходя из известного опорного плана задачи, за конечное число шагов получить ее оптимальный план. Каждый из шагов (или итераций) состоит в нахождении нового плана, которому соответствует

Меньшее значение линейной функции, чем значение этой же функции в предыдущем плане. Процесс продолжают до получения оптимального плана. Если задача не обладает планами или ее линейная функция не ограничена на многограннике решений, то симплексный метод позволяет установить это в процессе решения.

Отметим, что существуют много эквивалентных друг другу схем, позволяющих осуществлять упорядоченный переход от одного плана к другому. Ниже остановимся на одном из них.

1. Построение опорных планов

Пусть поставлена общая задача линейного программирования (см §11

и 2):

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2,$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m,$$

$$x_j \geq 0, \quad (j = 1, 2, \dots, n).$$

где a_{ij} , $b_i \geq 0$, c_j - заданные постоянные величины.

Предположим сначала, что система ограничений задачи содержит m единичных векторов, причем без ограничения общности можно предположить, что единичными являются первые m векторов. Тогда имеем задачу

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min, \quad (1)$$

$$x_1 + a_{1m+1}x_{m+1} + \dots + a_{1n}x_n = b_1,$$

$$x_2 + a_{2m+1}x_{m+1} + \dots + a_{2n}x_n = b_2, \quad (2)$$

.....

$$x_3 + a_{3m+1}x_{m+1} + \dots + a_{mn}x_n = b_m,$$

$$x_j \geq 0, \quad j = 1, 2, \dots, n. \quad (3)$$

Запишем систему (2) в векторной форме:

$$x_1 A_1 + x_2 A_2 + \dots + x_m A_m + x_{m+1} A_{m+1} + \dots + x_n A_n = A_0, \quad (4)$$

где

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}, \dots, A_m = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}, A_{m+1} = \begin{pmatrix} a_{1m+1} \\ a_{2m+1} \\ \dots \\ a_{mm+1} \end{pmatrix}, \dots, A_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{nn} \end{pmatrix}, A_0 = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}.$$

Поскольку векторы $A_1, A_2, \dots, A_m$ линейно независимы, они образуют базис в m -мерном пространстве. Поэтому в разложении (4) за базисные неизвестные выбираем $x_1, x_2, \dots, x_m$, свободные неизвестные $x_{m+1}, \dots, x_n$ приравняем нулю и учитывая, что $b_i \geq 0, i = 1, 2, \dots, m$, а векторы $A_1, A_2, \dots, A_m$ - единичные; получаем первоначальный план:

$$X_0 = (b_1, b_2, \dots, b_m, 0, \dots, 0). \quad (5)$$

Плану (5) соответствует разложение

$$x_1 A_1 + x_2 A_2 + \dots + x_m A_m = A_0, \quad (6)$$

где векторы $A_1, A_2, \dots, A_m$ линейно независимы, следовательно, построенный первоначальный план является и опорным.

Ниже рассмотрим, как исходя из первоначального опорного плана (5), можно построить второй опорный план. Векторы $A_1, A_2, \dots, A_m$ образуют базис в m -мерном пространстве, поэтому каждый из данных n векторов соотношения (4) можно разложить по векторам базиса, причем единственным образом:

$$A_j = \sum_{i=1}^m x_{ij} A_i, \quad j = 1, 2, \dots, n.$$

Предположим, что для некоторого вектора, не входящего в базис, например, для вектора A_{m+1} , положителен хотя бы один из коэффициентов x_{im+1} в разложении

$$x_{1m+1} A_1 + x_{2m+1} A_2 + \dots + x_{mm+1} A_m = A_{m+1}. \quad (7)$$

Выберем некоторую величину $\theta > 0$ (пока неизвестную), умножим на нее обе части равенства (7) и вычтем результат почленно из равенства (6).

Получаем

$$(x_1 - \theta x_{1m+1})A_1 + (x_2 - \theta x_{2m+1})A_2 + \dots + (x_m - \theta x_{mm+1})A_m + \theta A_{m+1} = A_0. \quad (8)$$

Таким образом, вектор

$$X_1 = (x_1 - \theta x_{1m+1}; x_2 - \theta x_{2m+1}; \dots; x_m - \theta x_{mm+1}; \theta; 0, \dots, 0)$$

является планом, если его компоненты неотрицательны.

Так как $\theta > 0$, то все компоненты вектора, X_1 в которые входят неположительные x_{im+1} , неотрицательны. Потому надо рассмотреть только компоненты, включающие положительные x_{im+1} , $i = 1, 2, \dots, m$, т.е. необходимо определить такое $\theta > 0$, при котором для всех $x_{im+1} > 0$

$$x_i - \theta x_{im+1} \geq 0. \quad (9)$$

Из (9) получаем $\theta \leq \frac{x_i}{x_{im+1}}$, следовательно, вектор X_1 - план задачи для любого θ , удовлетворяющего условию

$$0 < \theta \leq \min \frac{x_i}{x_{im+1}}, \quad (10)$$

где минимум берется по i , для которых $x_{im+1} > 0$.

Опорный план не может содержать $m+1$ положительных компонент, поэтому в плане X_1 необходимо обратить в нуль по крайней мере одну из компонент. Положим в (10), что

$$\theta = \theta_0 = \min_i \frac{x_i}{x_{im+1}}, \quad (11)$$

тогда компонента плана X_1 , для которой достигается минимум, обращается в нуль. Пусть эта компонента стоит на первом месте, т.е.

$$\theta_0 = \min \frac{x_i}{x_{im+1}} = \frac{x_1}{x_{1m+1}}.$$

Подставляя значение θ_0 в (8), имеем

$$\left(x_1 - \frac{x_1}{x_{1m+1}} x_{1m+1}\right)A_1 + \left(x_2 - \frac{x_1}{x_{1m+1}} x_{2m+1}\right) + \dots + \left(x_m - \frac{x_1}{x_{1m+1}} x_{mm+1}\right)A_m + \frac{x_1}{x_{1m+1}} A_{m+1} = A_0,$$

откуда получаем разложение

$$x'_2 A_2 + x'_3 A_3 + \dots + x'_m A_m + x'_{m+1} A_{m+1} = A_0,$$

которому соответствует новый опорный план:

$$X_1 = (0; x'_2, x'_3, \dots, x'_m, x'_{m+1}; 0; \dots; 0),$$

где $x'_i = \theta_0 x'_{im}$, $(i = 1, 2, \dots, m)$; $x'_{m+1} = \theta_0$.

Исключение одного вектора из базиса и включение вместо него другого с помощью θ_0 соответствуют переходу от одного базиса к другому с помощью метода Жордана - Гаусса. Поэтому система векторов $A_2, A_3, \dots, A_m, A_{m+1}$ линейно независима и является новым базисом.

Для определения следующего опорного плана необходимо любой вектор, не входящий в базис $A_2, A_3, \dots, A_m, A_{m+1}$, разложить по векторам этого базиса, а затем определить такое $\theta_0 > 0$, при котором исключался бы один из векторов этого базиса.

Таким образом, процесс получения новых опорных планов заключается в выборе вектора, который подлежит включению в базис, и определении вектора, подлежащего исключению из базиса. Критерий, используемый для определения вектора, который включается в базис, является одним из основных элементов симплексного метода. Заметим, что если вектор A_{m+1} подлежит включению в базис, а в его разложении (7) все $x_{im+1} \leq 0$, очевидно, нельзя выбрать такое $\theta > 0$, которое исключало бы один из векторов разложения (8). В этом случае план X_1 содержит $m+1$ положительных компонент, а система векторов $A_1, A_2, A_3, \dots, A_m, A_{m+1}$ линейно зависима и определяет не угловую, а внутреннюю точку многогранника решений, в которой линейная функция не может достигать минимального значения. Это указывает на то, что гиперплоскость соответствующая линейной функции, не может стать опорной к многограннику решений, как бы далеко ее ни перемещать в направлении, обратном вектору ∇f , т.е. линейная функция не ограничена на многограннике решений.

Таким образом, если система ограничений задачи линейного программирования при неотрицательных свободных членах содержит единичный базис, то без дополнительных вычислений нужно получить первоначальный опорный план, а также коэффициенты разложения векторов по векторам базиса.

Пример 1. Дана система

$$\begin{aligned} 3x_1 - x_2 + 2x_3 + x_4 &= 7, \\ 2x_1 - 4x_2 &+ x_5 = 12, \\ -4x_1 - 3x_2 + 8x_3 &+ x_6 = 10. \end{aligned}$$

или

$$x_1 A_1 + x_2 A_2 + x_3 A_3 + x_4 A_4 + x_5 A_5 + x_6 A_6 = A_0,$$

где

$$A_1 = \begin{pmatrix} 3 \\ 2 \\ -4 \end{pmatrix}, A_2 = \begin{pmatrix} -1 \\ -4 \\ -3 \end{pmatrix}, A_3 = \begin{pmatrix} 2 \\ 0 \\ 8 \end{pmatrix}, A_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, A_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, A_0 = \begin{pmatrix} 7 \\ 12 \\ 10 \end{pmatrix}.$$

За базис выбираем систему векторов A_4, A_5, A_6 , так как нам эти векторы единичные и линейно независимые. Базисными неизвестными являются неизвестные x_4, x_5, x_6 . Приравнявая нулю, свободные неизвестные $x_1 = x_2 = x_3$ получаем исходное базисное решение - первоначальный опорный план.

Этому плану соответствует разложение

$$7A_4 + 12A_5 + 10A_6 = A_0. \quad (12)$$

Чтобы перейти к другому опорному плану, возьмем любой вектор, не входящий в базис, но имеющий хотя бы одну положительную компоненту, например, A_1 и разложим его по базису. Так как базис единичный, то коэффициентами разложения вектора A_1 по векторам базиса являются компоненты вектора A_1 , т.е.

$$3A_4 + 2A_5 - 4A_6 = A_1. \quad (13)$$

В разложении вектора имеются два положительных коэффициента. Умножая последнее соотношение на $\theta > 0$ и вычитая из (12), получим

$$(7 - 3\theta)A_4 + (12 - 2\theta)A_5 + (10 + \theta)A_6 + \theta A_1 = A_0. \quad (14)$$

Для исключения какого-нибудь вектора из разложения (14) находим

$$\theta_0 = \min_i \frac{x_i}{x_{i1}} = \min\left(\frac{7}{3}; \frac{12}{3}\right) = \frac{7}{3}.$$

Подставляя значение $\theta_0 = \frac{7}{3}$ в (14), исключаем из разложения век-

тор A_4 .

Имеем:

$$\left(7 - 3 \cdot \frac{7}{3}\right)A_4 + \left(12 - 2 \cdot \frac{7}{3}\right)A_5 + \left(10 + 4 \cdot \frac{7}{3}\right)A_6 + \frac{7}{3}A_1 = A_0.$$

В результате получаем второе разложение:

$$\frac{22}{3}A_5 + \frac{58}{3}A_6 + \frac{7}{3}A_1 = A_0,$$

которому соответствует новое базисное решение – новый опорный план

$$X_1 = \left(\frac{7}{3}; 0; 0; 0; \frac{22}{3}; \frac{58}{3}\right).$$

Заметим, что не всякий вектор можно ввести в базис для получения нового базисного решения. В нашем примере, вводя в базис вектор A_2 , не удастся получить новое базисное решение – новый опорный план.

Действительно, возьмем вектор A_2 , который имеет следующее разложение в первоначальном базисе:

$$-A_4 - 4A_5 - 3A_6 = A_0. \quad (15)$$

Умножая (15) на $\theta > 0$ и вычитая из (12), получаем разложение

$$(7 + \theta)A_4 + (12 + 4\theta)A_5 + (10 + 3\theta)A_6 + \theta A_2 = A_0,$$

из которого ни при каком $\theta > 0$ нельзя исключить ни один из векторов, т.е. нет возможности получить новое базисное решение.

План $X_1 = (0; \theta; 0; 7 + \theta; 12 + \theta; 10 + 3\theta)$ не является опорным, так как содержит четыре положительные компоненты и соответствует внутренней точке многогранника решений.

2. Условия оптимальности

Предположим, что задача линейного программирования (1) - (3) об-
ладает планами. и каждый ее опорный план является невырожденным.
В этом случае для опорного плана (5) имеем:

$$x_1 A_1 + x_2 A_2 + \dots + x_m A_m = A_0, \quad (16)$$

$$x_1 c_1 + x_2 c_2 + \dots + x_m c_m = f(X_0), \quad (17)$$

где все $x_i > 0$, а $f(X_0)$ - значение линейной функции, соответствующее это-
му плану.

Разложение любого вектора A_j по векторам данного базиса $A_1, A_2, \dots, A_m$
единственное:

$$x_{1j} A_1 + x_{2j} A_2 + \dots + x_{mj} A_m = A_j, \quad j = 1, 2, \dots, n, \quad (18)$$

поэтому разложению вектора A_j в базисе соответствует и единственное, зна-
чение линейной функции

$$x_{1j} c_1 + x_{2j} c_2 + \dots + x_{mj} c_m = f_j, \quad j = 1, 2, \dots, n. \quad (19)$$

где f_j - значение линейной функции, если в нее вместо неизвестных подста-
вить соответствующие коэффициенты разложения j -го вектора по векторам
базиса.

Обозначим через c_j коэффициент линейной функции, соответствующе-
му вектору A_j . Тогда справедлива следующая теорема.

Теорема 1. Если для некоторого вектора A_j выполняется условие
 $f_j - c_j > 0$, то план X_0 не является оптимальным, и можно построить такой
план X , для которого выполняется неравенство $f(X) < f(X_0)$.

Следствие. Если для некоторого плана X_0 разложения всех векторов
 A_j , $j = 1, 2, \dots, n$ в данном базисе удовлетворяют условию

$$f_j - c_j \leq 0, \quad (20)$$

то план X_0 является оптимальным.

Неравенства (20) являются условием оптимальности плана задачи, решаемой на отыскание минимального значения линейной функции, а значения $\Delta_j = f_j - c_j$ называются оценками плана.

Таким образом, для того чтобы план задачи на отыскание минимального значения линейной функции был оптимальным, необходимо и достаточно, чтобы его оценки были неположительными.

Для задачи линейного программирования (1) - (3), заключающейся в отыскании максимального значения линейной функции, справедлива следующая теорема.

Теорема 2. Если для некоторого вектора A_j , выполняется условие $f_j - c_j < 0$, то план X_0 не является оптимальным и можно построить такой план X , для которого выполняется условие $f(X) > f(X_0)$.

Следствие. Если для некоторого плана X_0 разложения всех векторов A_j , $j=1,2,\dots,n$ в данном базисе удовлетворяют условию

$$f_j - c_j \leq 0, \quad (21)$$

то план X_0 является оптимальным.

Неравенство (21) - условие оптимальности плана задачи решаемой на отыскание максимального значения линейной функции.

Таким образом, для того чтобы план задачи на отыскание максимального значения линейной функции был оптимальным, необходимо и достаточно, чтобы его оценки были неотрицательными.

3. Алгоритм симплексного метода

Как следует из теорем 1 и 2 и следствий, начиная с исходного опорного плана задачи можно получить последовательность опорных планов, завершающихся оптимальным планом.

Продолжим рассмотрение задачи линейного программирования (1) - (3) на отыскание минимального значения линейной функции, опорный план ко-

торой $X_0 = (b_1, b_2, \dots, b_m, 0, \dots, 0)$, определяется системой m -мерных единичных векторов $A_1, A_2, \dots, A_m$. Для исследования этого опорного плана на оптимальность необходимо векторы A_j , $j = 1, 2, \dots, n$ системы (2) разложить по векторам базиса $A_1, A_2, \dots, A_m$ и подсчитать значения оценок $\Delta_j = f_j - c_j$. Базис является единичным, поэтому коэффициентами разложения вектора A_j по базису служат его компоненты, т. е. $x_{ij} = a_{ij}$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Дальнейшие вычисления удобнее проводить, если условия задачи и первоначальные данные, полученные после определения первого опорного плана, записать в симплексную таблицу (табл. 1). В столбце C базиса запишем коэффициенты линейной функции, соответствующие векторам базиса. В столбце A_0 - первоначальный опорный план X_0 , в нем же в результате вычислений получаем оптимальный план, в столбцах A_j , $j = 1, 2, \dots, n$ записываем коэффициенты разложения j -го вектора по базису, обозначаемые в дальнейшем через X_j .

В $m+1$ -й строке в столбце A_0 записываем значения линейной функции $f(X_0)$, которое она принимает при найденном опорном плане, а в столбцах A_j - значения оценок $\Delta_j = f_j - c_j$.

Значения $f(X_0)$ и $f_j = f(X_j)$ находим, подставляя в линейную функцию, соответственно, компоненты опорного плана и коэффициенты разложения j -го вектора по векторам базиса, поэтому эти значения в табл. 1 можно получить как скалярное произведение:

$$f(X_0) = C_0 X_0 = \sum_{i=1}^m c_i x_i,$$

$$f_j = C_0 X_j = \sum_{i=1}^m c_i x_{ij}, \quad j = 1, 2, \dots, n,$$

где c_j коэффициенты линейной функции, соответствующие векторам базиса.

Таблица 1

i	базис	C_0	A_0	c_1	c_2	...	c_l	...	c_m	c_{m+1}	...	c_j	...	c_k	...	c_n
				A_1	A_2	...	A_l	...	A_m	A_{m+1}	...	A_j	...	A_k	...	A_n
1	A_1	c_1	x_1	1	0	...	0	...	0	x_{1m+1}	...	x_{1j}	...	x_{1k}	...	x_{1n}
2	A_2	c_2	x_2	0	1	...	0	...	0	x_{2m+1}	...	x_{2j}	...	x_{2k}	...	x_{2n}
...	...	...	...	...	0	...	...	...	...	...	...	...	...	...	...	...
l	A_l	c_l	x_l	0	0	...	1	...	0	x_{lm+1}	...	x_{lj}	...	x_{lk}	...	x_{ln}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
m	A_m	c_m	x_m	0	0	...	0	...	1	x_{mm+1}	...	x_{mj}	...	x_{mk}	...	x_{mn}
$m+1$	Δ_j	f_0	0	0	...	0	...	0	$f_{m+1} - c_{m+1}$	...	$f_j - c_j$	...	$f_k - c_k$	...	$f_n - c_n$	

После составления табл. 1 просматриваем $m+1$ -ю строку. Если для всех $j=1,2,\dots,n$ разности $f_j - c_j \leq 0$, то опорный план X_0 оптимальный и минимальное значение линейной функции равно $f(X_0)$.

Предположим, что одна из оценок $f_j - c_j > 0$; тогда план X_0 не является оптимальным и, включая в базис вектор, соответствующий этой оценке, можно построить другой опорный план, которому соответствует меньшее значение линейной функции.

Если положительных оценок несколько, то в базис должен быть включен вектор, которому соответствует $\max[\theta_{0j}(f_j - c_j)]$, где максимум берется по тем j , для которых $f_j - c_j > 0$ и θ_{0j} определяется для каждого j . Это дает возможность на данном шаге перейти к вершине многогранника решений, связанной с наибольшим уменьшением линейной функции и в большинстве случаев приводящей к уменьшению количества итераций, что при решении задачи «вручную» позволяет быстрее получить опти-

мальное решение. При решении задачи на компьютере вектор, подлежащий исключению в базис, выбирается по $\max(f_j - c_j)$. Если имеется несколько одинаковых максимальных значений $\theta_{0j}(f_j - c_j)$, то из соответствующих им векторов включается в базис прежде всего вектор, которому соответствует $\min c_j$.

Если хотя бы для одной положительной оценки $f_j - c_j > 0$ коэффициенты разложения x_{ij} соответствующего вектора неположительны, то линейная функция не ограничена на многограннике решений и, выбирая θ , ее значение можно сделать сколь угодно малым; многогранник решений в этом случае (представляет собой неограниченную многогранную область).

Пусть $\max[\theta_{0j}(f_j - c_j)] = \theta_{0k}(f_k - c_k)$, т. е. максимальное значение достигается для k -го вектора, $m < k \leq n$. Тогда в базис включается вектор A_k и исключается вектор, которому соответствует $\theta_{0k} = \min \frac{x_i}{x_{ik}}$. ($x_{ik} > 0$).

Допустим, что $\theta_{0k} = \min \frac{x_i}{x_{ik}} = \frac{x_l}{x_{lk}}$ достигается для вектора базиса, состоящего в l -й строке; тогда вектор A_l исключается из базиса. Элемент x_{lk} называется разрешающим, а столбец и строка, на пересечении которых он находится, - направляющими. Новому опорному плану соответствует базис, состоящий из векторов $A_1, A_2, \dots, A_{l-1}, A_k, A_{l+1}, \dots, A_m$. Чтобы вычислить новый опорный план и проверить его на оптимальность, необходимо все векторы A_0, A_j , $j = 1, 2, \dots, n$ разложить по векторам базиса. Первоначальный базис был единичным, поэтому

$$A_0 = x_1 + \dots + x_l A_l + \dots + x_m A_m, \quad (22)$$

$$A_k = x_{1k} A_1 + \dots + x_{lk} A_l + \dots + x_{mk} A_m, \quad (23)$$

$$A_j = x_{1j} A_1 + \dots + x_{lj} A_l + \dots + x_{mj} A_m. \quad (24)$$

Из (23) имеем

$$A_l = \frac{1}{x_{lk}} (A_k - x_{lk} A_l - \dots - x_{mk} A_m). \quad (25)$$

Подставляя выражение A_l в (22), получаем

$$A_0 = x_1 A_1 + \dots + x_l \left[\frac{1}{x_{lk}} (A_k - x_{lk} A_l - \dots - x_{mk} A_m) \right] + \dots + x_m A_m,$$

или

$$A_0 = \left(x_1 - \frac{x_l}{x_{lk}} x_{lk} \right)_1 A_1 + \dots + \frac{x_l}{x_{lk}} A_k + \dots + \left(x_m - \frac{x_l}{x_{lk}} x_{mk} \right) A_m.$$

Таким образом, новый опорный план $X_1 = (x'_1; x'_2; \dots; x'_k; \dots; x'_m)$ вычисляется по формулам

$$\begin{cases} x'_i = x_i - \frac{x_l}{x_{lk}} x_{ik} & (i \neq l), \\ x'_k = \frac{x_l}{x_{lk}} & (i = l). \end{cases} \quad (26)$$

Подставляя в (25) в (27), получаем разложение вектора A_j по векторам нового базиса:

$$A_j = x'_{1j} A_1 + \dots + x'_{kj} A_k + \dots + x'_{mj} A_m,$$

где

$$\begin{cases} x'_{ij} = x_{ij} - \frac{x_{lj}}{x_{lk}} x_{ik} & (i \neq l), \\ x'_{kj} = \frac{x_{lj}}{x_{lk}} & (i = l). \end{cases} \quad (27)$$

Объединяя (26) и (27), находим, что новый опорный план и разложения векторов в новом базисе при $j = 1, 2, \dots, n$ определяются по формулам

$$\begin{cases} x'_{ij} = x_{ij} - \frac{x_{lj}}{x_{lk}} x_{ik} & (i \neq l), \\ x'_{kj} = \frac{x_{lj}}{x_{lk}} & (i = l), \end{cases} \quad (28)$$

которые являются формулами полных исключений Жордана-Гаусса. Действительно, полагая $j = k$, имеем

$$\begin{cases} x'_{ik} = x_{ik} - \frac{x_{jk}}{x_{ik}} x_{ik} & (i \neq l), \\ x'_{lk} = \frac{x_{jk}}{x_{ik}} = 1 & (i = l), \end{cases}$$

т.е. все коэффициенты разложения вектора, вводимого в базис, за исключением одного, обращаются в нуль, а коэффициент, взятый за разрешающий элемент, - в единицу. Вектору базиса соответствует оценка, равная нулю, поэтому для вычисления значений $m+1$ -й строки также используем формулы (47).

Таким образом, чтобы получить коэффициенты разложения векторов $A_0, A_j, j=1,2,\dots,n$ по векторам нового базиса, значения оценок нового опорного плана и значение линейной функции нужно разделить все элементы направляющей строки на разрешающий элемент и, производя одно полное исключение по методу Жордана - Гаусса с помощью этой преобразованной строки, составить симплексную таблицу (табл. 2).

Таблица 2

i	Basic	C_0	A_0	c_1	c_2	...	c_l	...	c_m	c_{m+1}	...	c_j	...	c_k	...	c_n
				A_1	A_2	...	A_l	...	A_m	A_{m+1}	...	A_j	...	A_k	...	A_n
1	A_1	c_1	x'_1	1	0	...	x'_{1l}	...	0	x'_{1m+1}	...	x'_{1j}	...	0	...	x'_{1n}
2	A_2	c_2	x'_2	0	1	...	x'_{2l}	...	0	x'_{2m+1}	...	x'_{2j}	...	0	...	x'_{2n}
...	...	...	...	...	0	...	...	...	...	...	...	...	...	...	...	...
l	A_k	c_k	x'_k	0	0	...	x'_{ll}	...	0	x'_{lm+1}	...	x'_{lj}	...	1	...	x'_{ln}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
m	A_m	c_m	x'_m	0	0	...	x'_{ml}	...	1	x'_{mm+1}	...	x'_{mj}	...	0	...	x'_{mn}
$m+1$	Δ_j		f'_0	0	0	...	$f'_l -$ $-c_l$	...	0	$f'_{m+1} -$ $-c_{m+1}$	...	$f'_j -$ $-c_j$	...	0	...	$f'_n -$ $-c_n$

$$f(X_0) = C_0 X_0; \quad f_j - c_j = C_0 X_j - c_j \quad (29)$$

используют для контроля за правильностью произведенных вычислений.

Если в табл. 2 в $m+1$ -й строке все оценки $f_j - c_j \leq 0$, то полученный план X_0 является оптимальным; если же имеются положительные оценки, то отыскивают следующий опорный план. Процесс продолжают либо до получения оптимального плана, либо до установления неограниченности линейной функции решаемой задачи.

Если среди оценок оптимального плана только нулевые оценки, соответствующие базисным векторам, то это говорит о единственности оптимального плана. Если же нулевая оценка соответствует вектору, не входящему в базис, то в общем случае это означает, что оптимальный план не единственный.

Действительно, пусть некоторому A_{m+1} , не входящему в базис, соответствует оценка $f_{m+1} - c_{m+1}$. Включим этот вектор в базис и по θ_{0m+1} какой-то вектор исключим из базиса. В результате получим новый опорный план, которому соответствует то же значение линейной функции, что и в первоначальном оптимальном плане, т.е. линейная функция достигает оптимума в двух угловых точках многогранника решений, но тогда по теореме 3 (см. §11) она достигает его в любой точке, являющейся выпуклой линейной комбинацией этих угловых точек. Таким образом, в этом случае задача линейного программирования обладает бесконечным множеством оптимальных планов.

Используя теорему 2 и ее следствие, решаем задачу линейного программирования на отыскание максимального значения линейной функции.

При невыполнении условия оптимальности (21) в базис включают в первую очередь тот вектор, которому соответствует $\min[\theta_{0j}(f_j - c_j)]$, где минимум берется по тем j , для которых $f_j - c_j < 0$. Если минимальных оценок несколько, то в базис прежде всего включают вектор, который соответ-

пуст $\max c_j$. В остальном симплексный процесс аналогичен процессу, имеющему место при отыскании минимального значения линейной функции.

Пример. На заводе изготавливаются изделия четырех типов. Стоимости единицы изделия каждого типа: $c_1 = 2$, $c_2 = 1$, $c_3 = 3$, $c_4 = 5$. На изготовление изделий расходуются энергия, материалы, труд. Размеры на единицу изделия каждого типа и заводские ресурсы на них приведены в следующей таблице:

	I тип	II тип	III тип	IV тип	Ресурсы
Энергия	2	3	1	2	30
Материалы	4	2	1	2	40
Трудовые затраты	1	2	3	1	25

Требуется найти такой план производства изделий, при котором стоимость всей продукции завода была максимальна.

Решение. Нетрудно составить математическую модель данной производственной задачи:

$$f = 2x_1 + x_2 + 3x_3 + 5x_4 \rightarrow \max,$$

$$2x_1 + 3x_2 + x_3 + 2x_4 \leq 30,$$

$$4x_1 + 2x_2 + x_3 + 2x_4 \leq 40,$$

$$x_1 + 2x_2 + x_3 + 2x_4 \leq 25,$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0.$$

От этой задачи в нормальной форме перейдем к задаче в канонической форме

$$f = 2x_1 + x_2 + 3x_3 + 5x_4 + 0 \cdot x_5 + 0 \cdot x_6 + 0 \cdot x_7 \rightarrow \max,$$

$$2x_1 + 3x_2 + x_3 + 2x_4 + x_5 = 30,$$

$$4x_1 + 2x_2 + x_3 + 2x_4 + x_6 = 40,$$

$$x_1 + 2x_2 + x_3 + 2x_4 + x_7 = 25,$$

$$x_1 \geq 0, \dots, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0, x_7 \geq 0,$$

где x_5, x_6, x_7 - дополнительные переменные, имеющие экономический смысл свободных ресурсов при рассматриваемом плане.

В качестве первоначального опорного плана можно взять точку $X = (0, 0, 0, 0, 30, 40, 25)$, которой соответствуют базисные векторы A_5, A_6, A_7 .

Составим первую симплексную таблицу.

Ба- зис	C_{θ}	A_0	2	1	3	5	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_5	0	30	2	3	1		1	0	0
A_6	0	40	4	2	1	2	0	1	0
A_7	0	25	1	2	3	1	0	0	1
Δ_j		0	-2	-1	-3	-5	0	0	0

В последней строке среди чисел Δ_j имеются отрицательные числа. Значит, первоначальный опорный план неоптимальный. Над отрицательными числами Δ_j нет столбцов из неположительных чисел. Значит пока нельзя сказать, что решение задачи не ограничено. Первоначальный план можно улучшить, включив в базис вектор, которому соответствует $\min\{\theta_{0j} \cdot \Delta_j\} < 0$.

Вычислим:

$$\theta_{01} = \min \left\{ \frac{30}{2}; \frac{40}{4}; \frac{25}{1} \right\} = 10, \quad \theta_{02} = \min \left\{ \frac{30}{3}; \frac{40}{2}; \frac{25}{2} \right\} = 10,$$

$$\theta_{03} = \min \left\{ \frac{30}{1}; \frac{40}{1}; \frac{25}{3} \right\} = \frac{25}{3}, \quad \theta_{04} = \min \left\{ \frac{30}{2}; \frac{40}{2}; \frac{25}{1} \right\} = 15;$$

$$\min\{\theta_{0j} \cdot \Delta_j\} = \min \left\{ (-2) \cdot 10; (-1) \cdot 10; (-3) \cdot \frac{25}{3}; (-5) \cdot 15 \right\} = (-5) \cdot 15 = -75.$$

Следовательно, разрешающим элементом служит число 2, стоящее на пересечении первой строки и четвертого столбца, первая строка и четвертый столбец являются направляющими; необходимо вектор A_4 включить в базис, а A_5 исключить.

Составим вторую симплексную таблицу.

Базис	C_6	A_0	2	1	3	5	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_4	5	15	1	3/2	1/2	1	1/2	0	0
A_6	0	10	2	-1	0	0	-1	1	0
A_7	0	10	0	-1/2		0	-1/2	0	1
Δ_j		55	3	13/2	-1/2	0	5/2	0	0

Аналогично строится следующая таблица.

Базис	C_6	A_0	2	1	3	5	0	0	0
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_4	5	13	1	8/5	0	1	3/5	0	-1/5
A_6	0	10	2	-9/10	0	0	-1	1	0
A_3	3	4	0	-1/5	1	0	0	0	2/5
Δ_j		77	3	32/5	0	0	3	0	3/5

Итак, имеем таблицу, у которой в последней строке все числа $\Delta_j \geq 0$.

Следовательно, опорный план $X = (0, 0, 4, 13, 10, 0, 0)$ является оптимальным планом исходной задачи, т.е.

$$X_{opt} = X = (0, 0, 4, 13, 10, 0, 0), \text{ и } f_{\max} = f(X_{opt}) = 77.$$

Упражнение

$$1. F = 6x_1 + 5x_2 + 9x_3 \rightarrow \max,$$

$$\begin{cases} 5x_1 + 2x_2 + 3x_3 \leq 25, \\ x_1 + 6x_2 + 2x_3 \leq 20, \\ 4x_1 + 3x_3 \leq 18. \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 3}.$$

$$2. F = 50x_1 + 40x_2 \rightarrow \max,$$

$$\begin{cases} 2x_1 + 5x_2 \leq 20, \\ 8x_1 + 5x_2 \leq 40, \\ 5x_1 + 6x_2 \leq 30. \end{cases}$$

$$x_1 \geq 0, x_2 \geq 0.$$

$$3. F = 3x_1 + 4x_2 \rightarrow \max.$$

$$\begin{cases} 2x_1 + 3x_2 \leq 9, \\ 3x_1 + 2x_2 \leq 13, \\ x_1 - x_2 \leq 1, \\ x_2 \leq 2, \end{cases}$$

$$x_1 \geq 0, x_2 \geq 0.$$

$$3. F = 2x_1 + x_2 - x_3 + x_4 - x_5 \rightarrow \max,$$

$$\begin{cases} x_1 + x_2 + x_3 = 50, \\ 2x_1 + x_2 + x_4 = 90, \\ x_1 + 2x_2 + x_5 = 70. \end{cases}$$

$$x_j \geq 0, j = \overline{1, 5}.$$

Вопросы для самопроверки

1. Перечислите условия оптимальности опорного плана задачи линейного программирования на отыскание минимального и максимального значений линейной функции.

2. Как определяется вектор для включения в базис, если первоначальный план не является оптимальным?

3. Когда линейная функция не ограничена на многограннике решений?

4. Как определить вектор, подлежащий исключению из базиса? Какой элемент называется разрешающим?

5. Выведите формулы разложения векторов по векторам базиса и покажите, что они являются формулами полного исключения.

6. Какой метод решения систем линейных уравнений лежит в основе симплексного метода?

§14. Метод искусственного базиса. Вырожденные задачи

Ключевые слова: Искусственный базис, двухфазный метод, М-метод, вырожденные задачи, запускание.

1. Решение задач линейного программирования методом искусственного базиса

В предыдущем параграфе было показано, что если ограничения задачи линейного программирования содержат единичную матрицу порядка m , то тем самым при неотрицательных правых частях уравнений определен первоначальный план, исходя из которого с помощью симплексного метода находится оптимальный план.

Если ограничения задачи линейного программирования можно преобразовать к виду $AX \leq A_0$ при $A_0 \geq 0$, то система ограничений всегда содержит единичную матрицу. Многие задачи линейного программирования не содержат единичной матрицы. В этом случае для решения задач применяют следующие методы:

- а) двухфазный метод решения задач линейного программирования;*
- б) решение задач линейного программирования с помощью М-метода*

Замечание. Фактически идейную основу этих методов мы демонстрировали при отыскании неотрицательного базисного решения систем линейных алгебраических уравнений (см. лек. №4.).

Двухфазный метод решения задач линейного программирования

Рассмотрим следующую задачу линейного программирования

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min, \quad (1)$$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\dots \dots \dots \end{aligned} \quad (2)$$

$$\begin{aligned} a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \\ x_j &\geq 0, \quad (j = 1, 2, \dots, n), \end{aligned} \quad (3)$$

$$X^* = \left(x_1^*, x_2^*, \dots, x_n^*, \overbrace{0, \dots, 0}^m \right).$$

Тогда

$$\varphi(X^*) = x_1^* + x_2^* + \dots + x_m^* < x_1 + x_2 + \dots + x_m = \varphi(X),$$

что противоречит тому, что на векторе $X = (x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+m})$ линейная функция (4) достигает минимума.

Значит, условие (7) является необходимым и достаточным для существования допустимых векторов задачи (1) - (3).

Замечание. Задача (4)-(6) всегда имеет решение, так как нижняя граница линейной функции на множестве планов равна нулю.

Вспомогательную задачу (4) - (6) можно решить симплексным методом, ибо первоначальный опорный план для нее известен:

$$X = (0, 0, \dots, 0, b_1, \dots, b_m).$$

При решении задачи (4)-(6) возможны три исхода:

1) в полученном решении не все искусственные переменные равны нулю;

2) все искусственные переменные равны нулю и среди базисных векторов нет векторов, соответствующих искусственным переменным;

3) все искусственные переменные равны нулю, но среди базисных векторов есть векторы, соответствующие искусственным переменным.

В случае 1) согласно доказанному задача (1)-(3) не разрешима из-за отсутствия планов, т.е. из-за того, что уравнения в (2) несовместны.

В случае 2) решение задачи (4)-(6) без искусственных переменных можно использовать в качестве первоначального опорного плана при решении задачи (1)-(3) симплексным методом.

В случае 3) решение задачи (4)-(6) нельзя использовать непосредственно, ибо в конечном базисе присутствуют векторы, соответствующие искусственным переменным. Поэтому, прежде чем переходить ко второй фазе, нужно из числа базисных исключить лишние векторы.

Для этого преобразуем последнюю таблицу первой фазы. В качестве разрешающей строки выберем строку с искусственной переменной x_{n-k} . В качестве разрешающего столбца берем столбец, в котором находится вектор условий A_j , $j \leq n$, не вошедший в базис, и в котором $x_{n-k} \neq 0$. После преобразования таблицы строка Δ_j не изменится. Не изменится и столбец A_0 , только теперь вместо переменной x_{n-k} будет стоять переменная $x_j = 0$, $j \leq n$. Этот процесс закончится или удалением из базиса всех векторов, соответствующих искусственным переменным, или получением равенств

$$x_{n-k_j} = 0, j = 1, 2, \dots, m; \quad k = \overline{k_1, k_2, \dots, k_s}, \quad (8)$$

для всех оставшихся искусственных базисных переменных x_{n-k} : $k = \overline{k_1, k_s}$. В первом случае переходим ко второй фазе. Второй случай по смыслу чисел x_{ij} означает, что среди уравнений в (2) есть линейно зависимые. Напомним, что x_{ij} - i -я координата вектора A_j по базисному вектору A_i . Если $x_{n-k_j} \neq 0$ для всех $j = 1, 2, \dots, m$, то это означает, что все векторы можно разложить по базисным, в которые не входит вектор A_{n-k} , т.е. на вектор A_{n-k} базисные векторы можно сократить. Поэтому при выполнении равенств (8) в симплексной таблице вычеркиваем s строк, содержащих искусственные переменные, и вторую фазу начнем с $m - s$ базисных векторов.

В терминах исходной задачи полученный результат означает, что из уравнений (2) исключается s линейно зависимых и процедура симплексного метода начинается при меньшем числе уравнений.

Проиллюстрируем двухфазный метод на следующем примере.

Пример 1.

$$\begin{aligned} f &= -x_1 + 2x_2 - x_3 \rightarrow \min, \\ x_1 + x_2 + x_3 &= 5, \\ 2x_1 + x_2 &= 3, \\ -2x_1 + 2x_2 &= 4, \\ x_1 \geq 0, x_2 \geq 0, x_3 &\geq 0. \end{aligned}$$

Решаем задачу первой фазы:

$$\varphi = x_4 + x_5 \rightarrow \min,$$

$$x_1 + x_2 + x_3 = 5,$$

$$2x_1 + x_2 + x_4 = 3,$$

$$-2x_1 + 2x_2 + x_5 = 4,$$

$$x_1 \geq 0, \dots, x_5 \geq 0.$$

Произведя вычисления, получим следующие таблицы:

Базис	C_6	A_0	0	0	0	1	1
			A_1	A_2	A_3	A_4	A_5
A_3	0	5	1	1	1	0	0
A_4	1	3	2	1	0	1	0
A_5	1	4	-2	2	0	0	1
Δ_1		7	0	3	0	0	0

Базис	C_6	A_0	0	0	0	1	1
			A_1	A_2	A_3	A_4	A_5
A_3	0	3	2	0	1	0	-1/2
A_4	1	1	3	0	0	1	-1/2
A_2	0	2	-1	1	0	0	1/2
Δ_1		1	3	0	0	0	-3/2

Базис	C_6	A_0	0	0	0	1	1
			A_1	A_2	A_3	A_4	A_5
A_3	0	7/3	0	0	1	-2/3	-1/6
A_1	0	1/3	1	0	0	1/3	-1/6
A_2	0	7/3	0	1	0	1/3	1/3
Δ_1		0	0	0	0	-1	-1

Далее переходим ко второй фазе, взяв за первоначальный опорный

план $X = \left(\frac{1}{3}, \frac{7}{3}, \frac{7}{3}\right)$ и базис A_1, A_2, A_3 . Для этого плана составлена следующая

таблица:

Базис	C_b	A_0	-1	2	-1
			A_1	A_2	A_3
A_3	-1	7/3	0	0	1
A_1	-1	1/3	1	0	0
A_2	2	7/3	0	1	0
Δ_j		2	0	0	0

Таким образом, $X = \left(\frac{1}{3}, \frac{7}{3}, \frac{7}{3}\right)$ доставляет минимум линейной функции

$f = -x_1 + 2x_2 - x_3$, т.е.

$$X_{opt} = X = \left(\frac{1}{3}, \frac{7}{3}, \frac{7}{3}\right), \text{ и } f_{min} = f(X_{opt}) = 2.$$

Решение задач линейного программирования с помощью М-метода

Этот метод предложен американским ученым Чарнесом. В этом методе обе фазы предыдущего метода объединены. Вместо исходной задачи заданной в общем виде

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \min.$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2,$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m,$$

$$x_j \geq 0, \quad (j = 1, 2, \dots, n).$$

где a_{ij} , $b_i \geq 0$, c_j - заданные постоянные величины и система ограничений не содержит единичной матрицы, рассматривается так называемая М-задача или расширенная задача:

$$\bar{f} = c_1x_1 + c_2x_2 + \dots + c_nx_n + Mx_{n+1} + \dots + Mx_{n+m} \rightarrow \min,$$

$$\begin{aligned} & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1, \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+2} = b_2, \\ & \dots\dots\dots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + x_{n+m} = b_m, \\ & x_j \geq 0, \quad j=1, 2, \dots, n+m. \end{aligned}$$

Замечание. Здесь для получения единичной матрицы к каждому равенству исходной задачи прибавили по одной переменной x_{n+i} , $i = 1, 2, \dots, m$, которые называют искусственными.

Величина M предполагается достаточно большим положительным числом, если задача решается на отыскание минимального значения линейной функции, и достаточно малым отрицательным числом, если находится максимальное значение линейной функции. Единичные векторы

$$A_{n+1}, A_{n+2}, \dots, A_{n+m}.$$

соответствующие искусственным переменным, образуют искусственный базис.

Для отыскания оптимального плана исходной задачи используют следующую теорему.

Теорема 4. Если в оптимальном плане $\bar{X} = (x_1, x_2, \dots, x_n, 0, \dots, 0)$ расширенной задачи искусственные переменные $x_{n+i} = 0, \quad i = 1, 2, \dots, m$, то план $X = (x_1, x_2, \dots, x_n)$ является оптимальным планом исходной задачи.

Применение симплексного метода к расширенной задаче обеспечивает построение плана, в котором каждое из искусственных переменных $x_{n+m} = 0$. Если первоначальная задача не обладает планами (т.е. она не совместна), то оптимальное решение расширенной задачи содержит по крайней мере одно $x_{n+m} > 0$.

Для отыскания оптимального плана расширенной задачи в случае, если заранее не задана величина M , применяется симплексный метод, с составлением симплексных таблиц, которые имеют на одну строку больше, чем обычная симплексная таблица. По этой $m + 2$ -й строке определяют вектор,

подлежащий включению в базис. Итерационный процесс по $m+2$ -й строке проводят для исключения из базиса всех искусственных векторов, затем процесс отыскания оптимального плана продолжают по $m+1$ -й строке.

Проиллюстрируем М-метод на следующем примере.

Пример 2.

$$f = -x_1 + 2x_2 - x_3 \rightarrow \min$$

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 + x_3 = 2,$$

$$x_1 + 2x_2 = 3,$$

$$-x_1 + 4x_2 = 3.$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.$$

Так как вектор условий A_3 является единичным, то его можно взять в качестве базисного. Поэтому вводим две искусственные переменные x_4, x_5 и составляем расширенную задачу (М-задачу):

$$\bar{f} = -x_1 + 2x_2 - x_3 + Mx_4 + Mx_5 \rightarrow \min$$

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 + x_3 = 2,$$

$$x_1 + 2x_2 + x_4 = 3,$$

$$-x_1 + 4x_2 + x_5 = 3.$$

$$x_1 \geq 0, \dots, x_5 \geq 0.$$

Произведя стандартные вычисления, получим следующие таблицы:

Базис	C_6	A_0	-1	2	-1	M	M
			A_1	A_2	A_3	A_4	A_5
A_3	-1	2	1/2	1/2	1	0	0
A_4	M	3	1	2	0	1	0
A_5	M	3	-1	4	0	0	0
Δ_j		-2	1/2	-5/2	0	0	0
		6	0	6	0	0	0

Базис	C_6	A_0	-1	2	-1	M	M
			A_1	A_2	A_3	A_4	A_5
A_3	-1	13/8	5/8	0	1	0	-1/8
A_4	M	3/2	3/2	0	0	1	-1/2
A_2	2	3/4	-1/4	1	0	0	1/4
Δ_j		-13/8	-1/8	0	0	0	5/8
		9/4	3/2	0	0	0	-3/2

Базис	C_6	A_0	-1	2	-1	M	M
			A_1	A_2	A_3	A_4	A_5
A_3	-1	1	0	0	1	-5/12	1/12
A_1	-1	1	1	0	0	2/3	-1/3
A_2	2	1	0	0	0	1/6	1/6
Δ_j		0	0	0	0	1/12-M	7/12-M

Таким образом, $\bar{X} = (1, 1, 1, 0, 0)$ является оптимальным планом расширенной задачи (М-задачи). В нем искусственные переменные $x_4 = 0, x_5 = 0$, следовательно, в силу теоремы 4 план $X = (1, 1, 1)$ является оптимальным планом исходной задачи, т.е.

$$X_{opt} = X = (1, 1, 1), \text{ и } f_{min} = f(X_{opt}) = 0.$$

Замечание. Отметим, что если исходная задача является задачей на отыскания максимального значения линейной функции, то для нее М-задача или расширенная задача имеет вид:

$$\overline{f} = c_1x_1 + c_2x_2 + \dots + c_nx_n - Mx_{n+1} - \dots - Mx_{n+m} \rightarrow \max.$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+2} = b_2,$$

$$\dots\dots\dots a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + x_{n+m} = b_m.$$

$$x_j \geq 0. \quad j = 1, 2, \dots, n + m.$$

Замечание. С помощью последней задачи довольно просто можно показать эвристическую основу М-метода. Она такова. Введением искусственных переменных $x_{n+1} \geq 0, x_{n+2} \geq 0, \dots, x_{n+m} \geq 0$ нарушается каждое из исходных уравнений. За нарушение условий платится «штраф» - $M \sum_{i=1}^m x_{n+i}$, где M - мера штрафа за единицу нарушения. Таким образом, общая прибыль от выбора вектора $X = (x_1, x_2, \dots, x_n)$, при котором условия задачи выполняются с заданными отклонениями $x_{n+1}, x_{n+2}, \dots, x_{n+m}$, будет равна $c_1x_1 + \dots + c_nx_n - M \sum_{i=1}^m x_{n+i}$. При больших относительных штрафах M эта величина может стать отрицательной, поэтому, чем больше M , тем менее выгодно нарушать условия задачи. В задачах линейного программирования существует конечный штраф M , при превышении которого ни одно из условий нельзя нарушать, если стремится получить максимальное значение линейной функции.

Упражнения

Используя вышерассмотренные методы, найдите решения следующих задач.

$$1. F = x_1 + 2x_2 + x_3 - x_4 - 6 \rightarrow \max.$$

$$\begin{cases} -x_1 + 5x_2 + x_3 + x_4 + x_5 = 10, \\ 2x_1 - x_2 + x_3 - 3x_4 = 6 \\ 10x_2 + x_3 + 2x_4 + 3x_5 = 25. \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 5}.$$

$$2. F = x_1 + 2x_2 + x_3 + x_4 \rightarrow \max,$$

$$\begin{cases} 2x_1 - x_2 - x_3 = -1, \\ x_2 + x_4 = 6, \\ x_1 + x_2 - x_5 = 25, \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 5}.$$

$$3. F = 5x_1 + x_2 - 3x_3 + 2x_4 \rightarrow \min,$$

$$\begin{cases} 3x_1 + 2x_2 + x_3 + x_4 = 7, \\ 5x_1 + 3x_2 + x_3 + 2x_4 = 11, \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 4}.$$

$$4. F = -2x_1 + 6x_2 + 2x_4 \rightarrow \max,$$

$$\begin{cases} 3x_1 - 3x_2 + 4x_3 - 2x_4 = 30, \\ -x_1 + 2x_2 - x_3 + x_4 = 5, \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 4}.$$

$$5. F = 2x_1 + 3x_2 + 2x_3 + 4x_4 \rightarrow \min,$$

$$\begin{cases} -x_1 + 2x_2 + x_3 + 6x_4 = 140, \\ x_1 + 2x_2 + 3x_3 + 2x_4 = 100, \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 4}.$$

$$6. F = x_1 - x_2 - 3x_3 \rightarrow \min,$$

$$\begin{cases} 2x_1 - x_2 + x_3 \leq 3, \\ 4x_1 - 2x_2 + x_3 \geq -6, \\ 3x_1 + x_3 \leq 15, \end{cases}$$

$$x_j \geq 0, \quad j = \overline{1, 3}.$$

2. Вырожденные задачи

Определение. Задача линейного программирования называется невырожденной, если у нее нет вырожденных опорных планов.

Определение. Задача линейного программирования называется вырожденной, если у нее имеется вырожденный опорный план.

Если проанализировать обоснование симплексного метода, приведенное в (п.2.) нетрудно обнаружить моменты, где используется предположение о невырожденности задачи. Главная трудность применения стандартной процедуры к вырожденным задачам заключается в том, что на некоторой итерации параметр θ_0 может оказаться равным нулю, в силу чего на ней линейная функция не возрастает. Если подобная ситуация повторится несколько раз, то появляется опасность заикливания, т.е. через несколько итераций возможно возвращение к тому же базису, при котором впервые оказалось, что $\theta_0 = 0$. Существуют примеры вырожденных задач, при решении которых получается заикливание. Однако пока не описано практических задач, где это явление наблюдалось. Поэтому целесообразно следовать во всех случаях стандартной процедуре симплексного метода. При этом, возможно, потребуются дополнительные итерации для вырожденных задач, но сохранится простота алгоритма. Существует несколько способов избежания заикливания.

Вопросы для самопроверки

1. Какая переменная называется искусственной, когда она вводится и какой коэффициент соответствует ей в линейной функции?
2. Зачем в системе ограничений необходим единичный базис?
3. Когда оптимальный план расширенной задачи является оптимальным планом исходной задачи?
4. Когда исходная задача несовместна, и как это определить с помощью решения расширенной задачи?
5. Как определяется вектор, подлежащий включению в базис при использовании искусственного базиса?
6. Что такое заикливание и в какой задаче линейного программирования оно может произойти?

§15. Теория двойственности в линейном программировании

Ключевые слова: двойственность, двойственная задача, симметрические двойственные задачи, несимметрические двойственные задачи, теоремы двойственности.

1. Понятие двойственности

Любой задаче линейного программирования можно определенным образом сопоставить некоторую другую связанную с ней задачу, которую называют **двойственной по отношению к первой**. Первоначальная задача называется **исходной** или **прямой**.

Связь исходной и двойственной задач состоит в том, что коэффициенты C_j функции цели исходной задачи являются свободными членами системы ограничений двойственной задачи, свободные члены b_i системы ограничений исходной задачи служат коэффициентами функции цели двойственной задачи, а матрица коэффициентов системы ограничений двойственной задачи является транспонированной матрицей коэффициентов системы ограничений исходной задачи. Решение двойственной задачи может быть получено из решения исходной и наоборот (см.п.2).

В качестве примера рассмотрим задачу использования ресурсов. Предприятие №1 имеет m видов ресурсов в количестве b_i ($i = 1, 2, \dots, m$) единиц, из которых производится n видов продукции. Для производства 1 ед. j -й продукции расходуется a_{ij} ед. i -го ресурса, а ее стоимость составляет C_j ед. Составить план выпуска продукции, обеспечивающий ее максимальный выпуск в стоимостном выражении. Обозначим через x_j ($j = 1, 2, \dots, n$) количество ед. j -й продукции. Тогда математическая модель исходной задачи имеет вид:

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max, \quad (1)$$

Исходная задача

$$\begin{aligned} f &= C'X \rightarrow \min, \\ AX &= A_0, \quad X \geq 0. \end{aligned} \quad (5)$$

Двойственная задача

$$\begin{aligned} g &= A'_0Y \rightarrow \max, \\ A'Y &\leq C. \end{aligned} \quad (6)$$

Замечание. Систему неравенств в симметрических двойственных задачах с помощью дополнительных переменных можно преобразовать в систему уравнений, поэтому всякую пару симметрических двойственных задач можно преобразовать в пару несимметрических. Поэтому ниже (в п.2.) теорему двойственности докажем для несимметрических двойственных задач.

Математические модели пары двойственных задач могут иметь один из следующих видов.

Таблица 1

№	Исходная задача	Двойственная задача
Несимметричные задачи		
I	$f = C'X \rightarrow \min,$ $AX = A_0, \quad X \geq 0$	$g = A'_0Y \rightarrow \max,$ $A'Y \leq C$
II	$f = C'X \rightarrow \max,$ $AX = A_0, \quad X \geq 0$	$g = A'_0Y \rightarrow \min,$ $A'Y \geq C$
Симметричные задачи		
III	$f = C'X \rightarrow \min,$ $AX \geq A_0, \quad X \geq 0$	$g = A'_0Y \rightarrow \max,$ $A'Y \leq C, \quad Y \geq 0$
IV	$f = C'X \rightarrow \max,$ $AX \leq A_0, \quad X \geq 0$	$g = A'_0Y \rightarrow \min,$ $A'Y \geq C, \quad Y \geq 0$

Следовательно, прежде чем записать двойственную задачу для данной исходной, систему ограничений исходной задачи необходимо привести к соответствующему виду.

Пример 1. Для следующей задачи

$$f = 2x_1 + x_2 + 3x_3 \rightarrow \max,$$

$$\begin{cases} -x_1 + 3x_2 - 5x_3 = 12, \\ 2x_1 - x_2 + 4x_3 = 24, \\ 3x_1 + x_2 + x_3 = 18, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{cases}$$

составить двойственную задачу.

Решение

Рассматриваемая задача относится к несимметричным задачам (II вид). Число переменных в двойственной задаче равно числу уравнений исходной задачи, т.е. равно трём. Коэффициентами целевой функции двойственной задачи являются свободные члены системы уравнений исходной задачи, т.е. 12, 24, 18.

Целевая функция исходной задачи исследуется на максимум, а система условий содержит только уравнения. Поэтому в двойственной задаче целевая функция исследуется на минимум, а её переменные могут принимать любые значения (в том числе и отрицательные). Так как все три переменные исходной задачи принимают только лишь неотрицательные значения, то в системе условий двойственной задачи должны быть три неравенства вида « $\geq$ ». Следовательно, для исходной задачи двойственная задача имеет вид:

$$\begin{aligned} g &= 12y_1 + 24y_2 + 18y_3 \rightarrow \min, \\ \begin{cases} -y_1 + 2y_2 + 3y_3 \geq 2, \\ 3y_1 - y_2 + y_3 \geq 1, \\ -5y_1 + 4y_2 + y_3 \geq 3, \\ y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{cases} \end{aligned}$$

Пример 2. Для следующей задачи

$$\begin{aligned} f &= 2x_1 + x_2 + 5x_3 \rightarrow \min, \\ \begin{cases} x_1 - x_2 - x_3 \leq 4, \\ x_1 - 5x_2 + x_3 \geq 5, \\ 2x_1 - x_2 + 3x_3 \geq 6, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{cases} \end{aligned}$$

составить двойственную задачу.

Решение

Эта задача в таком виде не соответствует ни одному из видов задач, приведённых в таблице. Однако, умножением первого неравенства на -1 , её можно привести к виду III:

$$\begin{aligned} f &= 2x_1 + x_2 + 5x_3 \rightarrow \min, \\ \begin{cases} -x_1 + x_2 + x_3 \geq -4, \\ x_1 - 5x_2 + x_3 \geq 5, \\ 2x_1 - x_2 + 3x_3 \geq 6, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{cases} \end{aligned}$$

Для этой задачи двойственной будет следующая задача:

$$\begin{aligned} g &= -4y_1 + 5y_2 + 6y_3 \rightarrow \max, \\ \begin{cases} -y_1 + y_2 + 2y_3 \leq 2, \\ y_1 - 5y_2 - y_3 \leq 1, \\ y_1 + y_2 + 3y_3 \leq 5, \\ y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{cases} \end{aligned}$$

Пример 3. Для следующей задачи

$$\begin{aligned} f &= 4x_1 + x_2 - 4x_3 \rightarrow \max, \\ \begin{cases} 2x_1 - x_2 + 4x_3 \leq 12, \\ x_1 + 3x_2 - 2x_3 = 13, \\ 2x_1 + 5x_2 - 6x_3 \leq 11, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{cases} \end{aligned}$$

составить двойственную задачу.

Решение

Эта задача тоже не соответствует ни одному из видов задач, приведённых в таблице. Рассматриваемую задачу можно переписать следующим образом:

$$f = 4x_1 + x_2 - 4x_3 \rightarrow \max,$$

$$\begin{cases} 2x_1 - x_2 + 4x_3 \leq 12, \\ x_1 + 3x_2 - 2x_3 - x_4 \leq 13, \\ 2x_1 + 5x_2 - 6x_3 \leq 11, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \end{cases}$$

Эта задача относится к симметричным задачам вида IV. Поэтому двойственная задача будет иметь вид:

$$\begin{aligned} g &= 12y_1 + 13y_2 + 11y_3 \rightarrow \min, \\ \begin{cases} 2y_1 + y_2 + 2y_3 \geq 4, \\ -y_1 + 3y_2 + 5y_3 \geq 1, \\ 4y_1 - 2y_2 - 6y_3 \leq -4, \\ y_1 \geq 0, y_2 \geq 0, y_3 \geq 0. \end{cases} \end{aligned}$$

2. Связь между решениями прямой и двойственной задач.

Рассмотрим пару двойственных задач (5) и (6). Существующие связи между оптимальными планами пары двойственных задач устанавливают следующие теоремы двойственности.

Теорема 1 (первая теорема двойственности). Если из пары двойственных задач одна обладает оптимальным планом, то и другая имеет решение, причём для экстремальных значений целевых функций выполняется соотношение

$$\min f = \max g, \quad X_{opt} = D^{-1}A_0, \quad Y_{opt} = C_B^* D^{-1},$$

где D^{-1} – матрица, составленная из компонент векторов окончательного базиса. A_0 – вектор, составленный из свободных членов исходной задачи, C_B^* – вектор, составленный из коэффициентов при окончательных базисных переменных в целевой функции.

Если целевая функция одной из задач не ограничена, то другая не имеет решения.

Приведенная теорема позволяет при решении одной из двойственных задач находить оптимальный план другой.

Пример 4. Для следующей задачи

$$f = x_2 - x_4 - 3x_5 \rightarrow \min.$$

$$x_1 + 2x_2 - x_4 + x_6 = 1,$$

$$-4x_2 + x_3 = 2x_4 - x_5 = 2,$$

$$3x_2 + x_5 + x_6 = 5,$$

$$x_j \geq 0, \quad j = \overline{1,6}$$

составить двойственную задачу и найти её решения.

Решение

Двойственная задача имеет вид:

$$g = y_1 + 2y_2 + 5y_3 \rightarrow \max.$$

$$\begin{cases} 2y_1 - 4y_2 + 3y_3 \leq 1 \\ -y_1 + 2y_2 \leq -1 \\ y_1 - y_2 + y_3 \leq -3 \\ y_1 \leq 0, y_2 \leq 0, y_3 \leq 0 \end{cases}$$

Решение исходной задачи находим симплексным методом.

Б	C _Б	A ₀	0	1	0	-1	-3	0
			A ₁	A ₂	A ₃	A ₄	A ₅	A ₆
A ₁	0	1	-1	2	0	-1	1	0
A ₃	0	2	0	-4	1	2	-1	0
A ₆	0	5	0	3	0	0	1	1
Δ_j		0	0	-1	0	1	3	0
A ₅	-3	1	1	2	0	-1	1	0
A ₃	0	3	1	-2	1	1	0	0
A ₆	0	4	-1	1	0	1	0	1
Δ_j		-3	-3	-7	0	4	0	0
A ₅	-3	4	2	2	1	0	1	0
A ₄	-1	3	1	-2	1	1	0	0
A ₆	0	1	-2	3	-1	0	0	1
Δ_j		-15	-7	1	-4	0	0	0
A ₅	-3	4	2	0	1	0	1	0
A ₄	-1	11/3	-1/3	0	1/3	1	0	2/3
A ₂	1	1/3	-2/3	1	-1/3	0	0	1/3
Δ_j		-46/3	-19/3	0	-11/3	0	0	-1/3

Оптимальный план исходной задачи $X_{opt} = (0; 1/3; 0; 11/3; 4; 0)$, при котором $f_{min} = -\frac{46}{3}$, получен в четвертой итерации. Используя эту итерацию,

найдем оптимальный план двойственной задачи. Согласно первой теоремы двойственности оптимальный план двойственной задачи находится из соотношения $Y^* = C^* D^{-1}$, где матрица D^{-1} - матрица, обратная матрице, составленной из компонент векторов, входящих в последний базис, при котором получен оптимальный план исходной задачи. В последний базис входят векторы A_5, A_4, A_2 ; значит,

$$D = (A_5, A_4, A_2) = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 2 & -4 \\ 1 & 0 & 3 \end{pmatrix}.$$

Обратная матрица D^{-1} образована из коэффициентов, стоящих в столбцах A_1, A_3, A_6 четвертой итерации:

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

Из этой же итерации следует $C^* = (-3; -1; -1)$. Таким образом,

$$\begin{aligned} Y^* = C^* D^{-1} &= (-3; -1; -1) \cdot \begin{pmatrix} 2 & 1 & 0 \\ -1/3 & 1/3 & 2/3 \\ -2/3 & -1/3 & 1/3 \end{pmatrix} = \\ &= (-19/3; -11/3; -1/3), \end{aligned}$$

т.е. $y_i = C^* X_i$, где X_i - коэффициенты разложения последней итерации, стоящие в столбцах векторов первоначального единичного базиса.

Итак, i -ю двойственную переменную можно получить из значения оценки последней строки симплексной таблицы, стоящей против соответствующего вектора, входившего в первоначальный единичный базис, если к ней прибавить соответствующее значение коэффициента целевой функции исходной задачи:

$$y_1 = -\frac{19}{3} + 0 = -\frac{19}{3}; \quad y_2 = -\frac{11}{3} + 0 = -\frac{11}{3}; \quad y_3 = -\frac{1}{3} + 0 = -\frac{1}{3}$$

При этом плане $g_{\max} = -\frac{46}{3}$.

Замечание. На практике при нахождении оптимальных планов двойственных задач, выбирают задачу более удобную для решения. Затем с помощью первой теоремы двойственности находят оптимальный план другой.

Пример 5. Найти оптимальный план следующей задачи

$$\begin{aligned} f &= x_1 + 2x_2 + 3x_3 \rightarrow \min, \\ \begin{cases} 2x_1 + 2x_2 - x_3 \geq 2, \\ x_1 - x_2 - 4x_3 \leq -3, \\ x_1 + x_2 - 2x_3 \geq 6, \\ 2x_1 + x_2 - 2x_3 \geq 3, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{cases} \end{aligned}$$

Решение

Для этой задачи двойственной будет следующая задача

$$\begin{aligned} g &= 2y_1 + 3y_2 + 6y_3 + 3y_4 \rightarrow \max, \\ \begin{cases} 2y_1 - y_2 + y_3 + 2y_4 \leq 1, \\ 2y_1 + y_2 + y_3 + y_4 \leq 2, \\ -y_1 + 4y_2 - 2y_3 - 2y_4 \leq 3, \\ y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_4 \geq 0. \end{cases} \end{aligned}$$

Для решения исходной задачи необходимо ввести четыре дополнительные переменные и одну искусственную. Таким образом, исходная симплексная таблица будет состоять из шести строк и девяти столбцов.

Для решения двойственной задачи необходимо ввести три дополнительные переменные. Её первая симплексная таблица содержит четыре строки и восемь столбцов.

Целесообразно решить двойственную задачу. Двойственную задачу решим симплексным методом.

Б	С _Б	А	2	3	6	3	0	0	0
			A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇
A ₅	0	1	2	-1	1	2	1	0	0
A ₆	0	2	2	1	1	-1	0	1	0
A ₇	0	3	-1	4	-2	-2	0	0	1
Δ_j		0	-2	-3	-6	-3	0	0	0
A ₃	6	1	2	-1	1	2	1	0	0
A ₆	0	1	0	2	0	-1	-1	1	0
A ₇	0	5	3	6	0	2	2	0	1
Δ_j		6	10	-9		9	6	0	0
A ₃	6	3/2	2	0	1	3/2	1/2	1/2	0
A ₂	3	1/2	0	1	0	-1/2	-1/2	1/2	0
A ₇	0	2	3	0	0	4	5	3	1
Δ_j		21/2	10	0	0	9/2	3/2	9/2	0

Оптимальный план двойственной задачи $Y^* = (0; 1/2; 3/2; 0)$, $g_{\max} = \frac{21}{2}$

Оптимальный план исходной задачи находим, используя оценки последней строки симплексной таблицы, стоящие в столбцах A_5, A_6, A_7 :

$$x_1 = 3/2 + 0 = 3/2; \quad x_2 = 9/2 + 0 = 9/2; \quad x_3 = 0 + 0 = 0;$$

При оптимальном плане исходной задачи

$$X^* = (3/2; 9/2; 0)$$

целевая функция достигает минимального значения $f_{\min} = \frac{21}{2}$.

Теорема 2 (вторая теорема двойственности). Если при подстановке компонент оптимального плана в систему ограничений исходной задачи i -е ограничение обращается в неравенство, то i -я компонента оптимального плана двойственной задачи равна нулю.

Если i -я компонента оптимального плана двойственной задачи по-

ложительна, то i -е ограничение исходной задачи удовлетворяется ее оптимальным решением как строгое равенство.

Упражнения

Построить двойственные задачи.

$$1. F = x_1 + 2x_2 - 3x_3 \rightarrow \max,$$

$$\begin{cases} 2x_1 - 3x_2 + 4x_3 + x_4 \leq 12, \\ x_1 - 2x_3 + 3x_4 \leq 10, \\ -x_1 + 2x_2 + 3x_3 \leq 9, \\ x_j \geq 0, \quad j = \overline{1, 4}. \end{cases}$$

$$2. F = 3x_1 - x_2 + 2x_3 \rightarrow \min,$$

$$\begin{cases} x_1 - x_2 + 2x_3 \geq 6, \\ 3x_1 + x_2 - 2x_3 - x_4 \geq -1, \\ 2x_1 + 3x_3 - 2x_4 \geq 0, \\ x_1 + x_2 + x_3 + x_4 \geq 10, \\ x_j \geq 0, \quad j = \overline{1, 4}. \end{cases}$$

Для каждой из следующих задач линейного программирования составить двойственную задачу и решить обе эти задачи.

$$3. F = 2x_1 + 3x_2 \rightarrow \max,$$

$$\begin{cases} 2x_1 - x_2 \leq -2, \\ -x_1 - x_2 \leq -6, \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

$$4. F = x_1 + 2x_2 - x_3 \rightarrow \max,$$

$$\begin{cases} -x_1 + 4x_2 - 2x_3 \leq 12, \\ x_1 + x_2 + 2x_3 \leq 17, \\ 2x_1 - x_2 + 2x_3 - x_4 = 4, \\ x_j \geq 0, \quad j = \overline{1, 3}. \end{cases}$$

Вопросы для самопроверки

1. В чем заключается сущность двойственности в линейном программировании?
2. Пусть исходная задача состоит в оптимальном использовании ресурсов. Дайте экономическую интерпретацию двойственной задачи.
3. Сформулируйте и доказите теорему двойственности.
4. Какие задачи линейного программирования относятся к несимметричным и симметричным, в чем их отличие?
5. Как по решению исходной (двойственной) найти решение двойственной (исходной) задачи?
6. Запишите возможные виды математических моделей двойственных задач.

§16. Анализ оптимального решения экономической задачи

Ключевые слова: оптимальное решение, статус ресурсов, дефицитный ресурс, недефицитный ресурс, ценность ресурса, вариация коэффициентов целевой функции, двойственные оценки ресурса, приращение ресурса.

1. Экономическая интерпретация решения задачи линейного программирования

Оказывается заключительная симплексная таблица «насыщена» весьма важными данными, лишь небольшую часть которых составляют оптимальные значения переменных. Из симплексной таблицы можно получить информацию относительно:

- оптимального решения;
- статуса ресурсов;
- ценности каждого ресурса;
- чувствительности оптимального решения к изменению запасов ресурсов, вариациям коэффициентов целевой функции и интенсивности потребления ресурсов.

Сведения, относящиеся к первым трем пунктам, можно извлечь непосредственно из итоговой симплексной таблицы. Получение информации, относящейся к четвертому пункту, требует дополнительных вычислений.

Для иллюстрации возможностей получения указанной выше информации из заключительной симплексной таблицы воспользуемся задачей об ассортименте продукции.

Определение оптимального ассортимента продукции

Предприятие изготавливает два вида продукции P_1 и P_2 , которая поступает в оптовую продажу. Для производства продукции используется два вида сырья - A и B . Максимально возможные запасы сырья в сутки составляют 9 и 13 единиц соответственно. Расход сырья на единицу продукции вида P_1 и вида P_2 дан в табл. 1.

Таблица 1

Сырье	Расход сырья на 1ед.продукции P_1	Расход сырья на 1ед.продукции P_2	Запас сырья, ед.
<i>A</i>	2	3	9
<i>B</i>	3	2	13

Опыт работы показал, что суточный спрос на продукцию P_1 никогда не превышает спроса на продукцию P_2 более чем на 1 ед. Кроме того, известно, что спрос на продукцию P_2 никогда не превышает 2 ед. в сутки.

Оптовые цены единицы продукции равны: 3 ден.ед. – для P_1 и 4 ден.ед. для P_2 .

Какое количество продукции каждого вида должно производить предприятие, чтобы доход от реализации продукции был максимальным?

Эта задача формулируется следующим образом:

$$\begin{aligned}
 Z &= 3x_1 + 4x_2 \rightarrow \max && (\text{доход}); \\
 2x_1 + 3x_2 &\leq 9 && (\text{сырье A}); \\
 3x_1 + 2x_2 &\leq 13 && (\text{сырье B}); \\
 x_1 - x_2 &\leq 1 && (\text{спрос}); \\
 x_2 &\leq 2 && (\text{спрос}); \\
 x_1 \geq 0, x_2 &\geq 0.
 \end{aligned}$$

Решим эту задачу методом симплекса. Симплексная таблица имеет приведена ниже (см. табл. 2). В этой таблице x_3, x_4, x_5, x_6 - выравнивающие (дополнительные) переменные.

Оптимальное решение

С точки зрения практического использования результатов решения задач линейного программирования классификация переменных на базисные и небазисные не имеет значения и при анализе оптимального решения могут не учитываться. Переменные, вектора которых отсутствуют в симплексной таблице в столбце *Б*- «базисные вектора», обязательно имеют нулевое зна-

чение. Значения остальных переменных приводятся в столбце A_0 - свободные члены».

Таблица 2

			3	4	0	0	0	0
B	C_0	A_0	A_1	A_2	A_3	A_4	A_5	A_6
A_3	0	9	2	3	1	0	0	0
A_4	0	13	3	2	0	1	0	0
A_5	0	1	1	-1	0	0	1	0
A_6	0	2	0	1	0	0	0	1
Δ_j		0	-3	-4	0	0	0	0
A_3	0	7	0	5	1	0	-2	0
A_4	0	10	0	5	0	1	-3	0
A_1	3	1	1	-1	0	0	1	0
A_6	0	2	0	1	0	0	0	1
Δ_j		3	0	-7	0	0	3	0
A_2	4	7/5	0	1	1/5	0	-2/5	0
A_4	0	3	0	0	-1	1	-1	0
A_1	3	12/5	1	0	1/5	0	3/5	0
A_6	0	3/5	0	0	-1/5	0	2/5	1
Δ_j		64/5	0	0	7/5	0	1/5	0

При интерпретации результатов оптимизации в задаче об ассортименте продукции нас прежде всего интересуют объемы производства продукции P_1 и P_2 , т. е. значения управляемых переменных x_1 и x_2 . Используя данные, содержащиеся в симплексной таблице для оптимального решения, основные

результаты можно представить в следующем виде:

Управляемые переменные	Оптимальные значения	Решение
x_1	2.4	Объем производства продукции P_1 должен быть равен 2.4 ед. в сутки
x_2	1,4	Объем производства продукции P_2 должен быть равен 1,4 ед. в сутки
$Z_{\max}$	12.8	Доход от реализации продукции будет равен 12,8 д. е. в сутки

Статус ресурсов

В линейном программировании ресурсы относят либо к дефицитным либо к недефицитным - в зависимости от того, полное или частичное их использование предусматривает оптимальное решение задачи. Сейчас цель состоит в том, чтобы получить соответствующую информацию непосредственно из оптимальной таблицы.

В модели, построенной для задачи об ассортименте продукции, фигурируют четыре ограничения со знаком « $\leq$ ». Первые два ограничения (определяющие допустимый расход исходного сырья) представляют собой истинные ограничения на ресурсы. Третье и четвертое ограничения относятся к спросу. Эти требования можно рассматривать как ограничения на соответствующие ресурсы, так как увеличение спроса на продукцию эквивалентно расширению представительства предприятия на рынке сбыта. В отношении финансовых средств такая ситуация имеет те же последствия, что и увеличение запасов ресурсов, требующее распределения дополнительных вложений.

Из вышеизложенного следует, что статус ресурсов (дефицитный или недефицитный) для любой модели линейного программирования можно уста-

получить непосредственно из результирующей симплексной таблицы, обращая внимание на значения выравнивающих переменных. Применительно к нашей задаче можно привести следующую сводную таблицу:

Ресурс	Выравнивающая переменная	Статус ресурса
Сырье A	$x_3 = 0$	Дефицитный
Сырье B	$x_4 = 3$	Недефицитный
Превышение объема производства продукции P_1 , по отношению к объему производства	$x_5 = 0$	Дефицитный
Спрос на продукцию P_2	$x_6 = 0.6$	Недефицитный

Положительное значение выравнивающей переменной указывает на неполное использование соответствующего ресурса, т.е. данный ресурс является недефицитным. Если же выравнивающая переменная равна 0, то это свидетельствует о полном потреблении соответствующего ресурса. Из сводной таблицы видно, что ресурсы 2 и 4 связаны с запасами сырья B и возможностями сбыта продукции P_2 . Поэтому любое увеличение их запасов сверх установленного максимального значения приведет лишь к тому, что они станут еще более недефицитными. Оптимальное решение задачи при этом останется неизменным.

Ресурсы, увеличение запасов которых позволяет улучшить решение (увеличить доход), — это сырье A и возможности по сбыту продукции P_1 , поскольку из оптимальной симплексной таблицы видно, что они дефицитные. В связи с этим логично поставить вопрос: какому из дефицитных ресурсов следует отдать предпочтение при вложении дополнительных средств

на увеличение их запасов, с тем чтобы получить от них максимальную отдачу? Ответ на этот вопрос будет дан в следующем разделе этой главы, где рассматривается ценность различных ресурсов.

Ценность ресурса

Ценность ресурса характеризуется величиной улучшения оптимального значения Z , приходящегося на единицу прироста объема данного ресурса. Графическая интерпретация этого определения применительно к условиям задачи об ассортименте продукции была дана в § 4 (вторая задача на чувствительность). Графический анализ показывает, что ценность ресурсов 1, 2, 3 и 4 равна:

$y_1 = 1,4$ д. е. на единицу прироста запасов ресурса сырья A :

$y_2 = 0, y_4 = 0$:

$y_3 = 0,2$ д. е. на единицу прироста превышения производства продукции P_1 по отношению к объему производства продукции P_2 .

Эта информация представлена в оптимальной таблице, а именно, в последней строке в столбцах, которые соответствуют векторам начального базиса A_3, A_4, A_5, A_6 : 1,4; 0; 0,2; 0. Указанные числа в точности соответствуют значениям y_1, y_2, y_3, y_4 .

Покажем, каким образом аналогичный результат можно получить непосредственно из симплексной таблицы.

Рассмотрим Z - уравнение оптимальной симплексной таблицы решения задачи об ассортименте продукции:

$$Z = 12,8 - (1,4 \cdot x_3 + 0 \cdot x_4 + 0,2 \cdot x_5 + 0 \cdot x_6) .$$

Положительное приращение переменной относительно x_3 текущего нулевого значения приводит к пропорциональному уменьшению Z , причем коэффициент пропорциональности равен 1,4 д. е. Однако из первого ограничения модели следует

$$2x_1 + 3x_2 + x_3 = 9,$$

т.е. увеличение x_2 эквивалентно снижению запаса ресурса 1 (сырья А). Отсюда следует, что уменьшение запаса первого ресурса вызывает пропорциональное уменьшение целевой функции с коэффициентом пропорциональности, равным 1.4 д. е. Аналогичные рассуждения справедливы и для ресурса 1.

В отношении ресурсов 2 и 4 было установлено, что их ценность равна 0 ($y_2 = y_4 = 0$). Этого и следовало ожидать, так как ресурсы 2 и 4 оказались недефицитными. Такой результат получается всякий раз, когда соответствующие выравнивающие переменные имеют положительное значение.

Несмотря на то, что ценность различных ресурсов, определяемая значениями переменных y_i , была представлена в стоимостном (д. е.) выражении, ее нельзя отождествлять с действительными ценами, по которым возможна закупка соответствующих ресурсов. На самом деле речь идет о некоторой мере, имеющей экономическую природу и количественно характеризующей ценность ресурса только относительно полученного оптимального значения Z . При изменении ограничений модели соответствующие экономические оценки будут меняться даже тогда, когда оптимизируемый процесс предполагает применение тех же ресурсов. Поэтому при характеристике ценности ресурсов экономисты предпочитают использовать такие термины, как теневая цена или двойственная оценка. Заметим, что теневая цена характеризует интенсивность улучшения оптимального решения Z . Однако при этом не фиксируется интервал значений увеличения запасов ресурсов, при которых интенсивность улучшения целевой функции остается постоянной. Для большинства практических ситуаций логично предположить наличие верхнего предела увеличения запасов, при превышении которого соответствующее ограничение становится избыточным, что, в свою очередь, приводит к новому базисному решению и соответствующим ему новым теневым ценам. Ниже определяется интервал значений запасов ресурса, при которых соответствующее ограничение не становится избыточным.

Максимальное изменение запаса ресурса

При решении вопроса о том, запас какого из ресурсов следует увеличивать в первую очередь, обычно используются двойственные оценки (теневые цены). Чтобы определить интервал значений изменения запаса ресурса, при которых двойственная оценка данного ресурса, фигурирующая в заключительной симплекс-таблице, остается неизменной, необходимо выполнить ряд дополнительных вычислений. Положим, что в задаче об ассортименте продукции запас первого ресурса (сырья A) изменился на Δ_1 , т.е. запас сырья A составит $(9 + \Delta_1)$ единиц. Введем это изменение в начальную симплекс-таблицу и затем выполним всю последовательность вычислений.

Поскольку элементы правых частей ограничений никогда не используются в качестве разрешающих, то, очевидно, что на каждой итерации вычислений Δ_1 будет оказывать влияние только на значения элементов столбца «свободные члены».

Результаты вычислений элементов столбца A_0 - «свободные члены» сведены в следующую таблицу:

Начальная симплексная таблица	Оптимальная симплексная таблица
$9 + \Delta_1$	$1.4 + 0.2 \cdot \Delta_1$
13	$3 - 1 \cdot \Delta_1$
1	$2.4 + 0.2 \cdot \Delta_1$
2	$0.6 - 0.2 \cdot \Delta_1$
0	$12.8 + 1.4 \cdot \Delta_1$

Все изменения элементов столбца «свободные члены» определяются непосредственно по данным, содержащимся в симплекс-таблицах. Каждый элемент столбца «свободные члены» представляет собой сумму двух величин: 1) постоянной и 2) члена, линейно зависящего от Δ_1 . Постоянные соответствуют числам, которые фигурируют в оптимальной симплекс-таблице до

введения Δ_1 в столбце «свободные члены». Коэффициенты при Δ_1 во вторых слагаемых равны коэффициентам при x_2 в оптимальной симплексной таблице.

Заметим, что при анализе изменений в правых частях второго, третьего и четвертого ограничений нужно пользоваться коэффициентами при переменных x_4 , x_5 , x_6 соответственно.

Так как введение Δ_1 сказывается лишь на правой части ограничений (на элементах столбца «свободные члены»), изменение запаса ресурса может повлиять только на допустимость решения. Поэтому Δ_1 не может принимать значений, при которых какая-либо из базисных переменных становится отрицательной. Из этого следует, что величина Δ_1 должна быть ограничена таким интервалом значений, при котором выполняется условие неотрицательности правых частей ограничений в результирующей симплекс-таблице, т.е.:

$$x_2 = 1,4 + 0,2 \cdot \Delta_1 \geq 0, \quad (I)$$

$$x_4 = 3 - 1 \cdot \Delta_1 \geq 0, \quad (II)$$

$$x_1 = 2,4 + 0,2 \cdot \Delta_1 \geq 0. \quad (III)$$

$$x_6 = 0,6 - 0,2 \cdot \Delta_1 \geq 0. \quad (IV)$$

Для определения допустимого интервала изменения Δ_1 рассмотрим два случая.

Случай 1: $\Delta_1 > 0$. Соотношения (I) и (III) всегда выполняются при $\Delta_1 \geq 0$. Соотношения (II) и (IV) определяют следующие предельные значения Δ_1 : $\Delta_1 \leq 3$; $\Delta_1 \leq 3$. Таким образом, все четыре соотношения выполняются при $\Delta_1 \leq 3$.

Случай 2: $\Delta_1 < 0$. Соотношения (II) и (IV) выполняются при $\Delta_1 < 0$. Соотношения (I) и (III) справедливы при $\Delta_1 \geq -12$; $\Delta_1 \geq -7$ соответственно.

Таким образом, оба соотношения справедливы при $\Delta_1 \geq -7$.

Объединяя результаты, полученные для обоих случаев, можно сделать

вывод, что при $-7 \leq \Delta_1 \leq 3$ решение рассматриваемой системы всегда будет допустимым. Любое значение Δ_1 , выходящее за предел указанного интервала (т.е. уменьшение запаса сырья A более чем на 7 единиц или увеличение более чем на 3 единицы), приведет к недопустимости решения и новой совокупности базисных переменных.

Анализ на чувствительность оптимального решения к вариации коэффициентов целевой функции

Иногда при определенных значениях изменения коэффициентов целевой функции оптимальные значения переменных остаются неизменными (хотя оптимальное значение Z при этом меняется). Возвращаясь к этому вопросу, покажем, каким образом интересующую нас информацию можно получить из данных, содержащихся в оптимальной симплексной таблице.

Следует отметить, что уравнение целевой функции также не используется в качестве ведущего уравнения. Поэтому любые изменения коэффициентов целевой функции окажут влияние только на Z -уравнение результирующей симплексной таблицы. Это означает, что такие изменения могут сделать полученное решение неоптимальным. Наша цель заключается в том, чтобы найти интервалы изменений коэффициентов целевой функции, при которых оптимальные значения переменных остаются неизменными.

Чтобы показать, как выполняются соответствующие вычисления, положим, что доход, получаемый с единицы продукции P_1 , изменяется от 3 до $3 + \delta_1$, где δ_1 может быть как положительным, так и отрицательным числом. Целевая функция в этом случае принимает следующий вид:

$$Z = (3 + \delta_1) \cdot x_1 + 4x_2 \rightarrow \max.$$

Если воспользоваться данными начальной симплексной таблицы и выполнить все вычисления, необходимые для получения оптимальной симплексной таблицы, то последнее Z -уравнение будет выглядеть следующим образом:

B	C_6	A_0	A_1	A_2	A_3	A_4	A_5	A_6
Δ_1		$12,8 + 2,4\delta_1$	0	0	$1,4 + 0,2\delta_1$	0	$0,2 + 0,6\delta_1$	0

Это уравнение (строка целевой функции) отличается от Z -уравнения до введения δ_1 , только наличием членов, содержащих δ_1 . Коэффициенты при δ_1 равны коэффициентам при соответствующих переменных в x_1 -уравнении (x_1 -строка) симплекс-таблицы для полученного ранее оптимального решения:

B	C_6	A_0	A_1	A_2	A_3	A_4	A_5	A_6
A_1	3	2,4	1	0	0,2	0	0,6	0

Мы рассматриваем x_1 -уравнение, так как коэффициент именно при этой переменной в выражении для целевой функции в начальной симплексной таблице изменился на δ_1 .

Оптимальные значения переменных будут оставаться неизменными при значениях δ_1 , удовлетворяющих условию неотрицательности (задача на отыскание максимума) всех коэффициентов при свободных переменных в Z -уравнении. Таким образом, должны выполняться следующие неравенства:

$$1,4 + 0,2 \cdot \delta_1 \geq 0;$$

$$0,2 + 0,6 \cdot \delta_1 \geq 0.$$

Из первого неравенства получаем, что $\delta_1 \geq -7$, а из второго следует, что $\delta_1 \geq -1/3$. Эти результаты определяют пределы изменения коэффициента

$$-1/3 \leq \delta_1 < +\infty.$$

Таким образом, при уменьшении коэффициента целевой функции при переменной x_1 до значения, равного $\left(3 + \left(\frac{1}{3}\right)\right) = 2\frac{2}{3}$, или при его увеличении до $+\infty$ оптимальные значения переменных остаются неизменными.

Следует отметить, что оптимальное значение Z будет изменяться в со-

ответствии с выражением $(12.8 + 2,4 \delta_1)$, где $-1/3 \leq \delta_1 < +\infty$.

Мы рассмотрели случай изменения коэффициента при базисной переменной x_1 . В случае изменения коэффициента при свободной переменной в целевой функции происходит изменение коэффициента только при данной переменной в оптимальной симплексной таблице. Рассмотрим в качестве иллюстрации случай, когда коэффициент при свободной переменной x_3 (первая выравнивающая переменная) изменяется от 0 до δ_2 . Выполнение преобразований, необходимых для получения заключительной симплекс-таблицы, приводит к следующему результирующему Z - уравнению:

B	C_6	A_0	A_1	A_2	A_3	A_4	A_5	A_6
Δ_j		12.8	0	0	$1.4 - \delta_2$	0	0,2	0

Из приведенного фрагмента заключительной симплексной таблицы видно, что единственное отличие от Z -уравнения до введения δ_2 состоит в том, что коэффициент при x_3 уменьшился на δ_2 . Таким образом, коэффициент при свободной переменной в результирующем Z - уравнении нужно уменьшить на ту же величину, на которую он увеличивался в исходном Z -уравнении.

2.Экономико-математический анализ полученных оптимальных решений

Оптимальное решение задачи линейного программирования существенно зависит от реальной экономической ситуации, складывающейся на предприятии. На решение задачи могут повлиять следующие экономические ситуации:

- 1) изменение запасов ресурсов;
- 2) внедрение нового технологического способа производства, позволяющего снизить расход сырья А и В;

- 3) произошедшие изменения в ценовой политике на предприятии;
- 4) предполагается выпуск нового вида продукции.

Результаты влияния данных экономических ситуаций на оптимальное решение можно получить в ходе проведения экономико-математического анализа модели на чувствительность.

Анализ на чувствительность оптимального решения базируется на второй теореме двойственности.

Теорема. Если при подстановке компонент оптимального плана в систему ограничений исходной задачи i -е ограничение обращается в неравенство, то i -я компонента оптимального плана двойственной задачи равна нулю.

Если i -я компонента оптимального плана двойственной задачи положительна, то i -е ограничение исходной задачи удовлетворяется ее оптимальным решением как строгое равенство.

Таким образом:

1. Двойственные оценки характеризуют дефицитность ресурсов. Величина y_i в оптимальном решении двойственной задачи является оценкой i -го ресурса: чем больше значение оценки y_i , тем выше дефицитность ресурса. Для недефицитного ресурса $y_i = 0$.

2. Двойственные оценки показывают, как влияют изменения правой части ограничений (запасов ресурсов) на значение целевой функции. Практический интерес представляют границы (нижняя и верхняя) изменения ресурсов, в которых величины оценок остаются неизменными.

3. Двойственные оценки являются показателем эффективности производства отдельных видов продукции с позиции критерия оптимальности. С этой точки зрения в оптимальный план может быть включена лишь та продукция j -го вида, для которой выполняется условие

$$\sum_{i=1}^m a_{ij} y_i \leq c_j.$$

4. Двойственные оценки позволяют провести сравнение суммарных

условных затрат и результатов.

Это свойство следует из принципа двойственности, в котором устанавливается связь между значениями функции прямой и двойственной задач, т. е. $Z_{\max} = L_{\min}$. Это означает, что оценка всех затрат производства должна равняться оценке произведенного продукта.

Используя данные свойства двойственных оценок, проведем анализ изменений исходной задачи, которые могут привести к недопустимости и неоптимальности решения.

Обратимся к конкретной задаче и проиллюстрируем применение анализа оптимального решения на чувствительность на примере задачи оптимизации ассортимента выпускаемой продукции.

Математическая модель исходной задачи имеет вид:

$$Z = 3x_1 + 4x_2 \rightarrow \max, \quad (1)$$

$$2x_1 + 3x_2 \leq 9 \quad (\text{сырье А});$$

$$3x_1 + 2x_2 \leq 13 \quad (\text{сырье В}); \quad (2)$$

$$x_1 - x_2 \leq 1 \quad (\text{спрос});$$

$$x_2 \leq 2 \quad (\text{спрос});$$

$$x_1 \geq 0, x_2 \geq 0.$$

Математическая модель двойственной задачи имеет вид:

$$L = 9y_1 + 13y_2 + y_3 + 2y_4 \rightarrow \min, \quad (3)$$

$$2y_1 + 3y_2 + y_3 \geq 3.$$

$$3y_1 + 2y_2 - y_3 + y_4 \geq 4, \quad (4)$$

$$y_1 \geq 0, \dots, y_4 \geq 0.$$

Как видно, задачи (1)-(2) и (3)-(4) образуют симметричную пару двойственных задач. Чтобы найти решение этих задач, следует сначала отыскать решение какой-либо одной из них. Так как система ограничений задачи (1) - (2) содержит лишь неравенства вида « $\leq$ », то лучше сначала найти решение этой задачи. Ее решение приведено в табл. 2:

$$X_{opt} = \left(\frac{12}{5}, \frac{7}{5} \right), \quad Z_{\max} = \frac{64}{5} = 12,8.$$

Из этой же таблицы получим решение двойственной задачи (3)-(4):

$$Y_{opt} = \left(\frac{7}{5}, 0, \frac{1}{5}, 0 \right), \quad L_{min} = \frac{64}{5} = 12.8.$$

Из этой таблицы видно, что оптимальным планом производства продукции является такой, при котором изготавливается $\frac{12}{5} = 2.4$ продукция P_1 и $\frac{7}{5} = 1.4$ продукции P_2 . При данном плане производства остается неиспользованным 3 ед. ресурса B и на $\frac{3}{5} = 0.6$ ед. продукции P_2 спроса не будет, а доход от реализации продукции равна $\frac{64}{5} = 12.8$ ден.ед.

Переменные y_1 и y_3 обозначают двойственные оценки ограничения 1 и 3. Первое ограничение определяет запасы сырья A , а третье ограничение определяет соотношение спроса на выпускаемую продукцию. Эти оценки отличны от нуля (см. оптимальное решение двойственной задачи), следовательно, ресурс 1 и 3 видов полностью используется при оптимальном плане производства продукции. Двойственная оценка единицы ресурса 2 и 4 видов равна нулю. Следовательно, ресурс B не полностью используется и спрос на продукцию P_2 не будет удовлетворен полностью.

Таким образом, положительную двойственную оценку имеют лишь те виды ресурсов, которые полностью используются при оптимальном плане производства продукции. Поэтому двойственные оценки определяют дефицитность используемого предприятием ресурсов. Более того, величина данной двойственной оценки показывает, на сколько возрастает максимальное значение целевой функции прямой задачи при увеличении количества сырья соответствующего вида на 1 ед. Так, увеличение количества сырья A вида на $\Delta b_1 = 1$ ед. приведет к тому, что появится возможность найти новый оптимальный план производства изделий, при котором общая стоимость изготавливаемой продукции возрастет на 1.4 ден.ед.: $\Delta Z = y_1 \cdot \Delta b_1 = \frac{7}{5} \cdot 1 = 1.4$ и станет равной

$$Z_{max} = 12.8 + 1.4 = 14.2 \text{ ден.ед.}$$

При этом числа, стоящие в столбце вектора A_3 табл.1, показывают, что указанное увеличение общей стоимости изготавливаемой продукции может быть достигнуто за счет увеличения выпуска продукции P_1 и P_2 на $1/5=0,2$ ед. Вследствие этого использование сырья A вида увеличится на 1 ед. Точно так же увеличение на $\Delta b_1 = 1$ ед. ресурса 3 вида (т.е. если предположить, что суточный спрос на продукцию P_1 никогда не превышает спроса на продукцию P_2 более чем на 2 ед.) позволит найти новый оптимальный план производства изделий, при котором общая стоимость изготавливаемой продукции возрастет на $\Delta Z = y_3 \cdot \Delta b_3 = \frac{1}{5} \cdot 1 = 0,2$ ден.ед. и составит $Z_{\max} = 12,8 + 0,2 = 13$ ден.ед.

Это будет (см. столбец вектора A_3) достигнуто в результате увеличения выпуска продукции P_1 на $3/5$ ед. и уменьшения изготовления продукции P_2 на $2/5$ ед.

Продолжим рассмотрение оптимальных двойственных оценок. При подстановке оптимальных двойственных оценок в систему ограничений двойственной задачи получаем

$$2 \cdot \frac{7}{5} + 3 \cdot 0 + \frac{1}{5} = 3, \quad 3 \cdot \frac{7}{5} + 2 \cdot 0 - \frac{1}{5} + 0 = 4.$$

Первое и второе ограничения двойственной задачи выполняются как строгие равенства. Это означает, что двойственные оценки сырья, используемого для производства единицы, соответственно, продукции P_1 и P_2 , равны в точности их ценам. Поэтому выпускать эти два вида продукции по двойственным оценкам экономически целесообразно. Их производство и предусмотрено оптимальным планом прямой задачи.

Таким образом, двойственные оценки тесным образом связаны с оптимальным планом прямой задачи. Всякое изменение исходных данных прямой задачи может оказать влияние как на ее оптимальный план, так и на систему оптимальных двойственных оценок. Поэтому, чтобы проводить

экономический анализ с использованием двойственных оценок, нужно знать их интервал устойчивости.

Упражнения

Решите задачи линейного программирования и проведите анализ на чувствительность.

1. $Z = 2x_1 - 5x_2 \rightarrow \min;$

$$\begin{cases} 3x_1 + 4x_2 \leq 6 \\ 2x_1 + 3x_2 \leq 4 \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

2. $Z = x_1 - 4x_2 \rightarrow \min;$

$$\begin{cases} 3x_1 + 5x_2 \geq 8 \\ -3x_1 + 10x_2 \leq 16 \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

3. $Z = x_1 + x_2 \rightarrow \max;$

$$\begin{cases} x_1 + 3x_2 \leq 30 \\ 2x_1 + x_2 \leq 20 \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

4. $Z = 2x_1 + 2x_2 \rightarrow \min;$

$$\begin{cases} x_1 + x_2 \geq 1 \\ -x_1 + x_2 \leq 1 \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

Вопросы для самопроверки

1. Каким образом из симплексной таблицы можно получить информацию относительно: оптимального решения; статуса ресурсов; ценности каждого ресурса; чувствительности оптимального решения к изменению запасов ресурсов, вариациям коэффициентов целевой функции и интенсивности потребления ресурсов?

2. Сведения, относящиеся к первым трем пунктам, можно извлечь непосредственно из итоговой симплексной таблицы. Получение информации, относящейся к четвертому пункту, требует дополнительных вычислений. Какие эти вычисления?

3. На какой теореме теории двойственности базируется анализ на чувствительность оптимального решения? Сформулируйте эту теорему.

4. Приведите схему проведения экономико-математического анализа полученного оптимального решения.

§17. Двойственный симплексный метод

Ключевые слова: двойственная задача, симплексный метод двойственный симплексный метод.

Иногда для получения оптимального решения исходной задачи переходят к двойственной и, используя оценки её оптимального плана, определяют оптимальное решение исходной задачи. Переход к двойственной задаче не обязателен, так как если рассмотреть первую симплексную таблицу с m единичным базисом, то легко заметить, что в столбцах записана исходная задача, а в строках - двойственная. Причём оценками плана исходной задачи являются c_j , а оценками плана двойственной задачи - b_i .

Решим двойственную задачу по симплексной таблице, в которой записана исходная задача. Найдём оптимальный план двойственной задачи, а вместе с ним и оптимальный план исходной задачи. Этот метод называется двойственным симплексным методом.

В обычном симплексном методе для преобразования симплексной таблицы (т.е. для перехода к новому допустимому базисному решению) сначала выбирают вектор подлежащий включению в базис, а затем определяют вектор, который необходимо исключить из базиса. В двойственном симплексном методе сначала по условию

$$\min_{b_i < 0} b_i = b_l$$

выбирают вектор, подлежащий исключению из базиса, затем определяют вектор, который необходимо включить в базис. Для этого находят

$$\min_{a_{ij} < 0} \frac{\Delta_i}{a_{ij}}$$

Если этот минимум достигается при $j=r$, тогда вектор A_r включают в базис, т.е. a_{lr} будет разрешающим элементом. Симплексная таблица преобразовывается как в обычном симплексном методе. Этот процесс продолжают до тех пор, пока из столбца вектора A_0 не будут исключены отрицательные

элементы. В результате находим оптимальный план исходной задачи, а следовательно, и двойственной.

Если на некотором шаге окажется, что в i -й строке симплексной таблицы в столбце вектора A_0 стоит отрицательное число b_i , а среди остальных элементов этой строки нет отрицательных, то исходная задача не имеет решения.

Пример. Следующую задачу

$$F = -x_1 + x_2 - 2x_3 \rightarrow \min.$$

$$x_1 + x_2 + x_3 = 8,$$

$$x_1 - x_2 \geq 4,$$

$$x_1 + 2x_2 \geq 6,$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$

решить двойственным симплексным методом.

Решение

Эту задачу не трудно привести к следующему виду:

$$F = -x_1 - x_2 - 2x_3 \rightarrow \min.$$

$$x_1 + x_2 + x_3 = 8.$$

$$-x_1 + x_2 + x_4 = -4.$$

$$-x_1 - 2x_2 + x_5 = -6.$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0.$$

Выбрав в качестве базиса векторы A_3 , A_4 и A_5 составим первоначальную симплексную таблицу для исходной задачи:

Б	C_6	A_0	-1	-1	-2	0	0
			A_1	A_2	A_3	A_4	A_5
A_3	0	8	1	1	1	0	0
A_4	0	-4	-1	1	0	1	0
A_5	0	-6	-1	-2	0	0	1
Δ_i		-16	-1	-1	0	0	0

В столбце вектора A_0 имеются два отрицательных числа (-4 и -6) и в строках векторов A_4 и A_5 имеются отрицательные числа (если бы они отсутствовали, то задача была бы неразрешима). Вектор, подлежащий исключению, выбираем из условия

$$\min\{-4; -6\} = -6,$$

т.е. из базиса исключаем вектор A_5 . Чтобы определить, какой вектор необходимо ввести в базис, находим

$$\min_{a_{3j} < 0} \frac{\Delta_i}{a_{3j}} = \min\left\{\frac{-1}{-1}; \frac{-1}{-2}\right\} = \frac{1}{2}.$$

Значит, в базис вводим вектор A_2 . Переходим к новой симплексной таблице.

Б	C_6	A_0	-1	-1	-2	0	0
			A_1	A_2	A_3	A_4	A_5
A_3	-2	5	1/2	0	1	0	1/2
A_4	0	-7	-3/2	0	0	1	1/2
A_2	1	3	1/2	1	0	0	-1/2
Δ_i		-13	-1/2	0	0	0	-1/2

В столбце вектора A_0 стоит отрицательное число -7. Поэтому рассмотрим элементы второй строки. Среди этих чисел есть одно отрицательное -3/2 (если бы такое число отсутствовало, то исходная задача была бы неразрешима). В данном случае переходим к новой симплексной таблице.

Б	C_6	A_0	-1	-1	-2	0	0
			A_1	A_2	A_3	A_4	A_5
A_3	-2	8/3	0	0	1	1/3	2/3
A_1	-1	14/3	1	0	0	-2/3	-1/3
A_3	-1	2/3	0	1	0	1/3	-1/3
Δ_i		-32/3	0	0	0	-1/3	-2/3

Как видно из таблицы, найдены оптимальные планы исходной и двойственной задач. Ими собственно являются:

$$X_{opt} = \left(\frac{14}{3}; \frac{2}{3}; \frac{8}{3}; 0 \right), \quad F_{\max} = -\frac{32}{3};$$

$$Y_{opt} = \left(2; \frac{1}{3}; \frac{2}{3} \right), \quad G_{\max} = -\frac{32}{3}.$$

Упражнения

Найдите решения следующих задач двойственным симплексным методом:

1. $F = 3x_1 + 2x_2 - 4x_3 \rightarrow \min$

$$x_1 + x_2 - 2x_3 \geq 4$$

$$3x_1 + x_2 - 4x_3 \geq 7$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$

2. $F = x_1 + x_2 + 2x_3 \rightarrow \max$

$$x_1 + x_2 + x_3 = 8$$

$$x_1 - x_2 \geq 4$$

$$x_1 + x_2 \geq 6$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$

3. $F = -4x_1 - 7x_2 - 8x_3 - 5x_4 \rightarrow \max$

$$x_1 + x_2 + 2x_4 \geq 4$$

$$2x_1 + x_2 + 2x_3 \geq 6$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0$$

4. $F = 5x_1 + 6x_2 + x_3 + x_4 \rightarrow \min$

$$1,5x_1 + 3x_2 - x_3 + x_4 \geq 18$$

$$3x_1 + 2x_3 - 4x_4 \geq 24$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0$$

5. $F = x_1 + 3x_2 + 4x_3 + 2x_4 \rightarrow \min$

$$x_1 - x_2 + 4x_3 + 5x_4 \geq 27$$

$$2x_1 + 3x_2 - x_3 + 4x_4 \geq 24$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0$$

6. $F = 2x_1 + 3x_2 + 5x_4 \rightarrow \max$

$$-2x_1 + x_2 - x_3 = 12$$

$$x_1 + 2x_2 + x_4 = 10$$

$$3x_1 - 2x_2 - x_5 = 18$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0$$

Вопросы для самопроверки

1. Описать простое правило нахождения решения двойственной задачи по известной симплексной таблице прямой задачи.

2. В чем состоит сущность двойственного симплексного метода?

§18. Общая постановка и математическая модель транспортной задачи

Ключевые слова: транспортная задача, условие баланса, закрытая модель транспортной задачи, открытая модель транспортной задачи, распределительный метод, метод северо-западного угла, метод минимальной стоимости

Транспортная задача - самая распространенная в линейном программировании. Она обладает рядом специфических свойств, в силу которых общие свойства и методы линейного программирования могут быть детализированы до весьма простых утверждений и операций.

1. Постановка задачи. Теорема существования

Имеется n пунктов производства (поставщики) $A_1, A_2, \dots, A_n$ и m пунктов потребления (потребители) $B_1, B_2, \dots, B_m$ некоторого продукта. В пункте A_i производится $a_i > 0$, $i = 1, 2, \dots, n$ (запасы), а в пункте B_j потребляется $b_j > 0$, $j = 1, 2, \dots, m$ единиц продукта (потребности). Стоимость перевозки единицы продукта из A_i в B_j равна $c_{ij} > 0$. Нужно найти оптимальный план перевозок, т.е. указать количество

$$x_{ij} \geq 0, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m \quad (1)$$

продукта, перевозимое из A_i в B_j при условии:

- весь продукт из пунктов производства будет вывезен:

$$\sum_{j=1}^m x_{ij} = a_i, \quad i = 1, 2, \dots, n; \quad (2)$$

- запросы всех пунктов потребления удовлетворены:

$$\sum_{i=1}^n x_{ij} = b_j, \quad j = 1, 2, \dots, m; \quad (3)$$

- стоимость перевозок окажется минимальной:

$$f = \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij} \rightarrow \min. \quad (4)$$

Задача линейного программирования, имеющая структуру (1)-(4), называется **транспортной задачей**.

В рассмотренной модели предполагается, что количество произведенного продукта равно его потреблению, т.е.

$$\sum_{i=1}^n a_i = \sum_{j=1}^m b_j. \quad (5)$$

Условие (5) называется **условием баланса**.

Транспортная задача для которой выполняется условие (5) называется **задача с закрытой моделью**; в противном случае – **задача с открытой моделью**.

Теорема. Для того чтобы закрытая транспортная задача имела решение, необходимо и достаточно, чтобы она удовлетворяла условию баланса.

2. Основные свойства матрицы условий транспортной задачи

Условия (2), (3) запишем в векторной форме

$$A_{11}x_{11} + \dots + A_{1m}x_{1m} + \dots + A_{n1}x_{1n} + \dots + A_{nm}x_{nm} = A_0. \quad (6)$$

Здесь через A_{ij} обозначен $(n+m)$ - мерный вектор, у которого все координаты равны нулю, за исключением i -й и $(n+j)$ -й, равных единице:

$$A_{ij}^i = 1, \quad A_{ij}^{n+j} = 1.$$

Другими словами, каждый вектор условий A_{ij} можно записать в форме

$$A_{ij} = e_i + e_{n+j},$$

где e_k - $(n+m)$ -вектор, все координаты которого равны нулю, кроме k -й, равной единице.

Вектор A_0 имеет размерность $n+m$:

$$A_0 = (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m).$$

Матрица A условий транспортной задачи (1) - (4) составлена из векторов A_{ij} :

$$A = (A_{11}, A_{12}, \dots, A_{1j}, \dots, A_{nm}). \quad (7)$$

у нее $n + m$ строк и nm столбцов.

Первое свойство. Ранг матрицы условий равен $n + m - 1$:

$$\text{rank} A = n + m - 1.$$

Доказательство. Ранг матрицы (7) меньше $n + m$. Действительно, если сложить первые n строк матрицы, то получим строку, состоящую из одних единиц. Сумма последних m строк также дает строку из одних единиц. Это значит, что если первые n строк умножить на $+1$, а последние m строк на -1 и сложить, то в результате получим нулевой вектор, что означает

$$\text{rank} A < n + m.$$

Осталось показать, что в матрице (7) можно найти отличный от нуля минор $(n + m - 1)$ -го порядка.

Составим матрицу из первых $n + m - 1$ компонент векторов $A_{1m}, A_{2m}, \dots, A_{nm}, A_{11}, A_{12}, \dots, A_{1m-1}$. Полученная матрица треугольная, на ее диагонали стоят единицы, значит, определитель равен единице. Утверждение доказано.

Замечание. Если у матрицы условий A вычеркнуть любую строку, то ранг оставшейся $(n + m - 1) \times nm$ -матрицы будет равен $n + m - 1$.

В силу доказанного свойства достаточно показать, что любое условие в (2), (3) есть следствие остальных.

Пусть план x_{ij} удовлетворяет всем условиям из (2), (3), кроме k -го, $1 \leq k \leq n$.

Тогда

$$\sum_{i=1}^m x_{ij} = a_i, \quad i = 1, 2, \dots, n; \quad i \neq k; \quad \sum_{i=1}^n x_{ij} = b_j, \quad j = 1, 2, \dots, m.$$

Поэтому

$$\sum_{j=1}^m x_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_{ij} - \sum_{\substack{i=1 \\ i \neq k}}^n \sum_{j=1}^m x_{ij} = \sum_{j=1}^m b_j - \sum_{\substack{i=1 \\ i \neq k}}^n a_i = a_k,$$

т.е. k -е условие также выполняется.

Следовательно, одно из условий в (2), (3) можно отбросить, однако ра-

ли симметрии все соотношения обычно сохраняются. Согласно результатам по задаче линейного программирования решение транспортной задачи можно искать среди угловых точек множества планов. Поскольку $\text{rank } A = n + m - 1$, то опорный план в транспортной задаче не может иметь более $n + m - 1$ положительных координат.

Второе свойство. Любой минор матрицы (7) равен одному из трех чисел: $-1, 0, +1$ (в силу этого свойства матрицу транспортной задачи называют унимодулярной).

Доказательство. Рассмотрим минор k -го порядка

$$A_k = \begin{pmatrix} i_1, \dots, i_k \\ j_1, \dots, j_k \end{pmatrix}.$$

В этом определителе каждый столбец содержит не более двух единиц. Если найдется столбец из нулей, то $A_k = 0$.

Пусть в каждом столбце по две единицы. Одна из них обязательно в строке, соответствующей ограничениям (2), другая - в строке, соответствующей ограничениям (3). Складывая строки с номерами из (2) и вычитая из них сумму строк по номерам из (3), получаем нуль. Это значит, что $A_k = 0$. Если найдется столбец с одной единицей, то, разлагая определитель по этому столбцу, получаем

$$A_k = \pm B_{k-1},$$

где B_{k-1} - некоторый минор $(k-1)$ -го порядка.

Повторив рассуждения с минором B_{k-1} , получим $B_{k-1} = 0$ или $B_{k-1} = \pm C_{k-2}$ и т.д. Поскольку миноры первого порядка матрицы A равны 0 или 1, то отсюда следует справедливость второго свойства.

Третье свойство. Если числа $a_i, i = 1, 2, \dots, n; b_j, j = 1, 2, \dots, m$ целые, то существует решение транспортной задачи из целочисленных компонент.

Доказательство. Пусть $x_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, m$ (угловая точка) - некоторое решение транспортной задачи, положительные координаты кото-

рой обозначим через $x_1, x_2, \dots, x_k$, $k \leq n + m - 1$. В §11 казано, что векторы условий $A_1, A_2, \dots, A_k$ с теми же номерами линейно независимы (см. теорему 5). В равенстве

$$A_1 x_1 + A_2 x_2 + \dots + A_k x_k = A_0$$

выделим k строк, соответствующих ненулевому минору k -го порядка. Остальные строки будут линейно выражаться через выделенные. Поскольку каждый отличный от нуля минор матрицы A равен ± 1 , то, решая выделенную систему уравнений по правилу Крамера, получаем

$$x_1 = \frac{\Delta_1}{\Delta} = \pm \Delta_1, \quad \dots, \quad x_k = \frac{\Delta_k}{\Delta} = \pm \Delta_k,$$

где Δ_i - определители из целочисленных элементов. Значит, и все $x_1, x_2, \dots, x_k$ целые числа.

3. Вычисление элементов симплексной таблицы. Распределительный метод

При решении задач линейного программирования симплексным методом наиболее трудоемким является вычисление элементов симплексной таблицы. Ситуация существенно упрощается для транспортной задачи.

Как известно, симплексная таблица состоит из координат разложения векторов условий по текущему базису, соответствующему рассматриваемому опорному плану. Пусть задача невырожденная. Тогда любой опорный план (x_{ij}) имеет ровно $n + m - 1$ положительных компонент, которым соответствуют $n + m - 1$ линейно независимых векторов условий $\{A_{ij}\}$. Множество индексов i, j , которые встречаются в этих векторах, обозначим через Ω . Любой вектор условий A_{kl} разлагается по базису A_{ij} , $(i, j) \in \Omega$:

$$A_{kl} = \sum_{(i,j) \in \Omega} x_{ij}^{kl} A_{ij}. \quad (8)$$

Числа x_{ij}^{kl} и есть элементы основной части симплексной таблицы.

Лемма. Элементы x_{ij}^{kl} , принимают одно из трех значений: $-1, 0, +1$.

Доказательство. В матричной форме соотношение (8) принимает

вид

$$A_{ij} = Qx^{kl}, \quad (9)$$

где Q - матрица, составленная из $n+m-1$ линейно независимых векторов условий A_{ij} , $(i, j) \in \Omega$. По доказанному второму свойству, если у матрицы Q вычеркнуть любую строку, то оставшаяся $(n+m-1) \times (n+m-1)$ -матрица будет неособой. В уравнении (9) вычеркнем k -ю строку. Тогда слева останется $n+m-1$ вектор e_{n+l-1} , а справа - Rx^{kl} , где R неособая матрица:

$$e_{n+l-1} = Rx^{kl}.$$

Отсюда $x^{kl} = R^{-1}e_{n+l-1} = r_{n+l-1}$, где r_{n+l-1} - $(n+l-1)$ -й столбец матрицы R^{-1} . Известно, что каждый элемент обратной матрицы R^{-1} равен некоторому минору, $(n+m-2)$ -го порядка, деленному на определитель матрицы R . Но $\det R$ и все миноры матрицы R являются минорами матрицы условий A транспортной задачи. Поэтому они равны одному из трех чисел: $-1, 0, 1$. Следовательно, $x_{ij}^{kl} = -1, 0, 1$. Лемма доказана.

В силу леммы каждый вектор условий транспортной задачи имеет представление

$$A_{kl} = \sum_{(i,j) \in \Omega^1} \pm A_{ij}, \quad \Omega^1 \subset \Omega. \quad (10)$$

Поскольку у вектора A_{ij} единицы стоят только на k -м и $(n+m)$ -м местах, то в базисе найдется вектор $A_{k j_1}$ первый индекс которого такой же, что и у разлагаемого вектора A_{kl} . Если $j_1 \neq l$, то в силу равенства (10) найдется вектор базиса $A_{i_1 j_1}$, который в (10) входит со знаком «-». Разложение на двух векторах не может закончиться, ибо слева $(n+l)$ -я координата равна 1; справа координаты с номерами, большими чем n , - нулю, i_1 -я координата равна -1. Поэтому найдется вектор базиса $A_{i_1 j_2}$, входящий в (10) со знаком «+». Если $j_2 = l$, то разложение вектора A_{kl} закончено. При $j_2 \neq l$ разложение

продолжается. В результате через конечное число шагов получаем разложение

$$A_{kl} = A_{k,l_1} - A_{l_1,l_2} + A_{l_2,l_3} - \dots - A_{l_{j-1},l_j} + A_{l_j,l}. \quad (11)$$

В этом разложении справа будет нечетное число слагаемых, знаки слагаемых чередуются, первое и последнее слагаемые имеют знак «+». Правило чередования индексов поясним на следующем примере. Пусть в табл. 1 точками обозначены клетки с номерами векторов базиса. Нужно разложить вектор A_{26} . Из клетки A_{26} движемся по строке до одной из точек, затем по столбцу до одной из точек, далее по строке и т.д. Наконец приходим по строке к точке, лежащей в том же столбце, что и раскладываемая точка. Если в очередную точку приходим по строке, то вектору базиса, соответствующему этой точке, приписываем знак «+», если по столбцу - знак «-». По доказанному, описанное построение всегда осуществимо и единственно. Для вектора A_{26} имеем

$$A_{26} = A_{23} - A_{33} + A_{34} - A_{44} + A_{46}.$$

Таблица 1

Поставщики	Потребители						Запасы
	В ₁	В ₂	В ₃	В ₄	В ₅	В ₆	
A ₁	•	•					a ₁
A ₂		•	•	→		← A ₂₆	a ₂
A ₃			•	→			a ₃
A ₄				•	→	•	a ₄
Потребности	b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	$\sum_{i=1}^n a_i = \sum_{j=1}^m b_j$

Последовательность построенных прямолинейных отрезков, исходя-

ших из клетки с A_{26} , назовем **цепочкой** этой клетки, отрезки - **звеньями** цепочки.

Теперь можно приступить к вычислению последней строки в симплексной таблице. При двуиндексной нумерации координат выражение для компонент вектора $f - c$ принимает вид

$$f_{kl} - c_{kl} = \sum_{(i,j) \in \Omega} c_{ij} x_{ij}^{kl} - c_{kl}.$$

Значения x_{ij}^{kl} определяются разложением (11) вектора A_{kl} . в соответствии с этим разложением получаем

$$f_{kl} - c_{kl} = c_{k_1} - c_{i_1 j_1} + c_{i_1 j_2} - \dots - c_{i_{l-1} j_{l-1}} + c_{i_l j_l} - c_{kl}. \quad (12)$$

Как следует из симплексного метода, величины (12) полностью определяют оптимальность текущего плана и вектор условий, который следует включить в базис на следующем этапе, если текущий план неоптимален.

Для подсчета чисел (12) составим новую таблицу, которую назовем транспортной таблицей или таблицей планирования (табл. 2). включим в нее данные о транспортной задаче.

Таблица 2

Поставщики	Потребители				Запасы
	B_1	B_2	...	B_n	
A_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1m} x_{1m}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}		c_{2m} x_{2m}	a_2
...					...
A_n	c_{n1} x_{n1}	c_{n2} x_{n2}		c_{nm} x_{nm}	a_n
Потребности	b_1	b_2	...	b_m	$\sum_{i=1}^n a_i = \sum_{j=1}^m b_j$

Вместо величин x_{ij} на каждом этапе вычислений заносим компоненты текущей крайней точки, причем, если $x_{ij} = 0$, то клетку оставляем пустой. Далее, для каждой пустой клетки с номером (k, l) составляем цепочку по правилу разложения вектора условия A_{kl} , соответствующего этой клетке. Числа c_{ij} , стоящие в клетках, где кончается горизонтальное звено цепочки, умножим на +1, где кончается вертикальное звено, - на -1 и результаты сложим. Из полученного результата вычитаем значение c_{kl} , стоящее в клетке, из которой начинается цепочка. Ясно, что эти операции приведут к величине (12).

Согласно критерию оптимальности, доказанному для симплексного метода, текущий опорный план будет оптимальным, если все числа (12) неположительны.

Если хотя бы для одной пустой клетки описанные операции приведут к положительному числу, то рассматриваемый план перевозок неоптимален. Номер (k, l) этой клетки указывает номер вектора условия A_{kl} , который можно ввести в базис. При рассмотрении симплексного метода брался тот номер (k, l) , при котором число (12) наибольшее. И здесь можно, просчитав числа (12) для пустых клеток, выбрать из результатов наибольший. Из физического смысла подобного выбора, отмеченного при исследовании симплексного метода, следует, что при такой замене скорость убывания линейной функции (4) будет наибольшей. Пусть индексы (s, t) таковы, что

$$f_{st} - c_{st} \max\{f_{kl} - c_{kl}\}.$$

Тогда вектор A_{st} войдет в базис и число x_{st} будет положительным на следующем этапе, т.е. клетка с номером (s, t) уже не будет пустой.

Замечание. Вопрос о неограниченности решения в транспортной задаче не возникает, ибо множество планов всегда ограничено (см. п. 1).

Значения $\Delta_{kl} = f_{kl} - c_{kl}$ помещаются также в клетках. Окончательный вид транспортной таблицы представлен табл.3.

Поставщики	Потребители				Запасы
	B_1	B_2	...	B_n	
A_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1m} x_{1m}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}		c_{2m} x_{2m}	a_2
...					...
A_n	c_{n1} x_{n1}	c_{n2} x_{n2}		c_{nm} x_{nm}	a_n
Потребности	b_1	b_2	...	b_m	$\sum_{i=1}^n a_i = \sum_{j=1}^m$

Найдем далее номер того вектора условий, который нужно исключить из базиса. Для этого, следуя симплексному методу, составим отношения

$$\theta_{ij} = \frac{x_{ij}}{x_{ij}^{st}}, \text{ для всех } x_{ij}^{st} > 0.$$

Но, по доказанному (см. лемму), все положительные x_{ij}^{st} равны единице, поэтому $\theta_{ij} = x_{ij}$. Согласно симплексному методу исключается из базиса вектор A_{pq} , индексы которого удовлетворяют соотношению

$$\theta_{pq} = \min \theta_{ij} = \max x_{ij} = x_{pq}. \quad (13)$$

На языке транспортной таблицы индексы исключаемого вектора находятся так: для клетки (s, t) строится цепочка. Среди чисел x_{ij} , стоящих на концах горизонтальных звеньев цепочки, находим наименьшее x_{pq} . Вектор A_{pq} на следующем этапе выводим из базиса, клетка с номером (p, q) становится пустой.

Осталось заполнить таблицу для нового набора базисных векторов. Общая формула перехода к новой угловой точке в симплексном методе име-

$$x'_{ij} = x_{ij} - \theta_{pq} x_{ij}^{pq}.$$

Пользуясь графической интерпретацией чисел x_{ij}^{pq} , формулой (13), приходим к следующим правилам заполнения таблицы в новом базисе:

$$1) \ x'_{st} = x_{st}, \quad x'_{ij} = x_{ij} - x_{pq},$$

если клетка (i, j) лежит на конце горизонтального звена цепочки, построенного для клетки (s, t) ;

$$2) \ x'_{ij} = x_{ij} + x_{pq},$$

если клетка (i, j) лежит на конце вертикального звена цепочки, построенного для клетки (s, t) ;

$$3) \ x'_{ij} = x_{ij},$$

если в непустой клетке (i, j) старой таблицы цепочка не терпит излома.

Описанная процедура перехода от старого базиса к новому, являющаяся интерпретацией симплексного метода для транспортной задачи, называется **распределительным методом решения транспортной задачи**.

4. Построение первоначального опорного плана

Известно, что нахождение оптимального решения любой задачи линейного программирования начинается с построения первоначального опорного плана. Первоначальный опорный план транспортной задачи как задачи линейного программирования можно построить ранее рассмотренными методами. При решении общей задачи линейного программирования симплексным методом для построения начальной крайней точки приходится, как правило, прибегать к искусственным переменным. В транспортной задаче введение искусственных переменных излишне, ибо существует допустимый план перевозок (см. п. 1) и имеются различные приемы построения угловой точки по этим планам. Остановимся на некоторых из них.

4.1. Метод северо-западного угла

Не учитывая стоимости перевозки единицы груза, удовлетворим потребности первого потребителя B_1 за счет запаса поставщика A_1 . Для этого $\min(a_1, b_1)$ записываем в левый нижний угол клетки A_1B_1 , т.е. $x_{11} = \min(a_1, b_1)$. Если $a_1 > b_1$, то излишек $a_1 - b_1$ завозим из A_1 в пункт B_1 , т.е. полагаем $x_{12} = \min(a_1 - b_1, b_2)$. Если $a_1 < b_1$, то остаток $b_1 - a_1$ завозим из пункта A_2 , т.е. полагаем $x_{12} = \min(b_1 - a_1, a_2)$. Процесс продолжим до тех пор, пока не удовлетворим всех потребителей за счет запасов поставщиков. На этом построение первоначального плана заканчивается.

Опорный план содержит не более $n+m-1$ положительных компонент (т.е. в таблице планирования количество занятых клеток не превосходит $n+m-1$). Если опорный план имеет ровно $n+m-1$ положительных компонент (т.е. в таблице планирования количество занятых клеток не превосходит $n+m-1$), тогда этот опорный план будет невырожденным.

Пример 1. Найти первоначальный план следующей задачи

Поставщики	Потребители				Запасы
	B_1	B_2	B_3	B_4	
A_1	3	5	7	11	100
A_2	1	4	6	2	130
A_3	5	8	12	7	170
Потребности	150	120	80	50	

Решение

Первоначальный опорный план найден методом северо-западного угла. Матрицу планирования заполним следующим образом:

$a_i \backslash b_j$	150	120	80	50
100	3 100	5	7	11
130	1 50	4 80	6	2
170	5	8 40	12 80	7 50

В этой таблице через a_i обозначено запасы поставщиков, а через b_j – потребности потребителей. Полученный опорный план имеет вид:

$$X = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 50 & 80 & 0 & 0 \\ 0 & 40 & 80 & 50 \end{pmatrix}.$$

Этот план невырожденный, так как содержит точно $n+m-1=3+4-1=6$ занятых клеток. Вычислим стоимость полученного плана:

$$f(X) = 100 \cdot 3 + 50 \cdot 1 + 80 \cdot 4 + 40 \cdot 8 + 80 \cdot 12 + 50 \cdot 7 = 2300.$$

Недостатком правила северо-западного угла является то, что оно не учитывает стоимости перевозок. В силу этого первоначальный опорный план может оказаться далеким от решения задачи и потребует большого числа вычислений по распределительному методу. Существуют другие методы построения первоначального опорного плана, учитывающие стоимость перевозок. Хотя они более трудоемки, но могут сократить общий объем вычислений при решении задачи, так как дают лучшее начальное приближение к решению.

4.2. Метод минимальной стоимости

Суть метода заключается в том, что из всей таблицы стоимостей выбирают наименьшую и в клетку, которая ей соответствует, помещают меньшее из чисел a_i или b_j . Затем из рассмотрения исключают либо строку, со-

ответствующую поставщику, запасы которого полностью израсходованы, либо и строку и столбец, если израсходованы запасы поставщика и удовлетворены потребности потребителя. Из оставшейся части таблицы стоимостей снова выбирают наименьшую стоимость, и процесс распределения запасов продолжают, пока все запасы не будут распределены, а потребности удовлетворены.

Пример 2. Составить с помощью метода минимальной стоимости опорный план выше рассмотренной транспортной задачи.

Решение

$b_j \backslash a_i$	150	120	80	50
100	3 20	5 80	7	11
130	1 130	4	6	2
170	5	8 40	12 80	7 50

Опорный план имеет вид:

$$X = \begin{pmatrix} 20 & 80 & 0 & 0 \\ 130 & 0 & 0 & 0 \\ 0 & 40 & 80 & 50 \end{pmatrix}$$

В этом случае также опорный план содержит точно 6 занятых клеток. Следовательно, опорный план является невырожденным. При составлении опорного плана учитывалась стоимость перевозки единицы груза. Поэтому план будет значительно ближе к оптимальному.

Действительно,

$$f(X) = 20 \cdot 3 + 80 \cdot 5 + 130 \cdot 1 + 40 \cdot 8 + 80 \cdot 12 + 50 \cdot 7 = 2200.$$

Существуют ещё и другие методы построения первоначального плана.

С помощью рассмотренных методов построения первоначального плана можно получить вырожденный и невырожденный опорный план.

Упражнения

Найти первоначальный опорный план транспортных задач 1-6 методами северо-западного угла и минимального элемента.

1.

b_j	40	60	80	60
a_i				
60	1	3	4	2
80	4	5	8	3
100	2	3	6	1

2.

b_j	150	140	190
a_i			
60	3	7	2
80	9	2	1
100	1	5	7
170	6	4	8

Вопросы для самопроверки

1. Сформулируйте транспортную задачу линейного программирования и напишите ее математическую модель.
2. Докажите теорему о существовании решения транспортной задачи.
3. Что называется распределительным методом решения транспортной задачи?
4. Какие существуют методы построения первоначального опорного плана? Постройте опорный план с помощью этих методов.
5. Сколько положительных перевозок должен содержать невырожденный опорный план и почему?
6. В чем заключается опорность плана транспортной задачи, условия которой записаны в виде таблицы?

§19. Нахождение оптимального плана транспортной задачи

Ключевые слова: транспортная задача, распределительный метод, дополнительные переменные (потенциалы), метод потенциалов, транспортная задача с открытой моделью.

1. Вводное замечание

Метод потенциалов является модификацией распределительного метода (поэтому иногда называется модифицированным распределительным методом), в которой введением дополнительных переменных (потенциалов) упрощается одна из процедур распределительного метода.

Как известно, в симплексном методе и его реализации - распределительном методе - одна из основных процедур связана с вычислением в опорном плане вектора $\Delta = f - c$, по которому можно судить об оптимальности опорного плана или находить новый вектор базиса. В методе потенциалов предлагается новый путь вычисления компонент этого вектора.

Введение потенциалов существенно уменьшает объем вычислений для транспортных задач больших размеров (числа n, m большие).

Метод потенциалов был предложен Л. В. Канторовичем вне связи с симплексным методом. Дж. Данциг получил его независимо исходя из симплексного метода.

С помощью рассмотренных методов построения первоначального плана можно получить вырожденный и невырожденный опорный план. Построенный опорный план транспортной задачи с помощью методов потенциалов можно довести до оптимального.

2. Метод потенциалов.

Теорема. Если план $X^* = (x_{ij}^*)$ транспортной задачи является оптимальным, то существует система $m + n$ чисел U_i^* и V_j^* удовлетворяющих условиям

$$U_i^* + V_j^* = c_{ij} \quad \text{для } x_{ij}^* > 0,$$

$$U_i^* + V_j^* \leq c_{ij} \quad \text{для } x_{ij}^* = 0,$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m$$

Числа U_i^* и V_j^* называются **потенциалами**, соответственно, **поставщиков** и **потребителей**.

Доказательство. Транспортную задачу минимизации линейной функции $f = \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij}$ при ограничениях

$$x_{i1} + x_{i2} + \dots + x_{im} = a_i, \quad i = 1, 2, \dots, n;$$

$$x_{1j} + x_{2j} + \dots + x_{nj} = b_j, \quad j = 1, 2, \dots, m;$$

$$x_{ij} \geq 0, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m.$$

можно рассматривать как двойственную некоторой исходной задачи линейного программирования, условия которой получают по общей схеме, рассмотренной в §15, если каждому ограничению вида $x_{i1} + x_{i2} + \dots + x_{im} = a_i$ в исходной задаче соответствует переменная U_i , $i = 1, 2, \dots, n$, а каждому ограничению вида $x_{1j} + x_{2j} + \dots + x_{nj} = b_j$ - переменная V_j , $j = 1, 2, \dots, m$, а именно, максимизировать линейную функцию

$$g = \sum_{i=1}^n a_i U_i + \sum_{j=1}^m b_j V_j$$

при ограничениях

$$U_i + V_j \leq c_{ij}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m.$$

План X^* - оптимальный план двойственной задачи, поэтому план $Y^* = (U_i^*, V_j^*)$ является планом исходной задачи и на основании теоремы двойственности

$$g_{\max} = f_{\min}$$

или

$$\sum_{i=1}^n a_i U_i^* + \sum_{j=1}^m b_j V_j^* = \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij}^*, \quad x_{ij}^* \geq 0,$$

На основании второй теоремы двойственности получаем, что ограничения исходной задачи, соответствующие положительным компонентам оптимального плана двойственной задачи, удовлетворяются как строгие равенства, а соответствующие компонентам, равным нулю, - как неравенства, т.е.

$$\begin{aligned} U_i^* + V_j^* &= c_{ij} & \text{для } x_{ij}^* > 0, \\ U_i^* + V_j^* &\leq c_{ij} & \text{для } x_{ij}^* = 0. \end{aligned}$$

Теорема доказана.

На основании доказанной теоремы для того чтобы первоначальный опорный план был оптимальным, необходимо выполнение следующих условий:

а) для каждой занятой клетки сумма потенциалов должна быть равна стоимости единицы перевозки, стоящей в этой клетке:

$$U_i^* + V_j^* = c_{ij}; \quad (1)$$

б) для каждой незанятой клетки сумма потенциалов должна быть меньше или равна стоимости единицы перевозки, стоящей в этой клетке:

$$U_i^* + V_j^* \leq c_{ij}. \quad (2)$$

Если хотя бы одна незанятая клетка не удовлетворяет условию (2), то опорный план является неоптимальным и его можно улучшить, вводя в базис вектор, соответствующий клетке, для которой нарушается условие оптимальности (т.е. в клетку надо переместить некоторое количество единиц груза).

Таким образом, для проверки плана на оптимальность необходимо сначала построить систему потенциалов.

3. Построение системы потенциалов

Для построения системы потенциалов используем условие

$$U_i^* + V_j^* = c_{ij}.$$

Систему потенциалов можно построить только для невырожденного опорного плана. Такой план содержит $n+m-1$ занятых клеток, поэтому для

него можно составить систему из $n+m-1$ линейно независимых уравнений с $n+m$ неизвестными. Уравнений на одно меньше, чем неизвестных, поэтому система является неопределенной и одному неизвестному (обычно U_1) придают нулевое значение. После этого остальные потенциалы определяются однозначно. Затем для каждой незанятой клетки проверяем условие $\Delta_{ij} = (U_i + V_j) - C_{ij} \leq 0$. Если для некоторых клеток $\Delta_{ij} \leq 0$, то план оптимальный. Если для некоторых клеток $\Delta_{ij} > 0$, то план неоптимальным. Тогда для каждой клетки, в которой не выполняется условие оптимальности, находим величину $\Delta_{ij} > 0$ и записываем её значение в левый нижний угол этой же клетки.

Для построения следующего опорного плана в первую очередь заполняется клетка, которой соответствует $\max \Delta_{ij}$. Для того чтобы эту клетку сделать занятой сначала необходимо определить, сколько единиц груза должно быть перераспределено в нее.

Для определения количества единиц груза, подлежащих перераспределению, отмечаем знаком «+» незанятую клетку, которую надо загрузить. Это означает, что клетка присоединяется к занятым клеткам. В таблице занятых клеток стало $n+m$, поэтому появляется цикл, все вершины которого, за исключением клетки, отмеченной знаком «+» находятся в занятых клетках, причем этот цикл единственный. Отыскиваем цикл и, начиная движение от клетки, отмеченной знаком «+», поочередно проставляем знаки «-» и «+». Затем находим $\theta_0 = \min x_{ij}$, где x_{ij} – перевозки, стоящее в вершине цикла, отмеченных знаком «-».

Величина θ_0 определяет сколько единиц груза можно перераспределить по найденному циклу. Значение θ_0 записываем в незанятую клетки, отмеченную знаком «+», двигаясь по циклу, вычитаем θ_0 из объемов перевозок, расположенных в клетках, которые отмечены знаком «-», и прибавляем к объёмам перевозок, находящихся в клетках, отмеченных знаком «+». Если θ_0 соответствуют несколько минимальных перевозок, то при вычитании оставляем в соответствующих клетках нулевые перевозки в таком количестве,

чтобы θ_0 во вновь полученном опорном плане занятых клеток было $n+m-1$.

В результате перераспределения θ_0 получается новый опорный план, который снова подлежит проверке на оптимальность.

Пример. С помощью метода потенциалов найти оптимальный план вышерассмотренной транспортной задачи.

Решение

Построим систему потенциалов для первоначального опорного невырожденного плана, найденного нами методом северо-западного угла.

$$\begin{aligned} U_1 + V_1 &= 3, & U_1 &= 0, & V_1 &= 3; \\ U_2 + V_1 &= 1, & U_2 &= 1 - V_1 = 1 - 3 = -2; \\ U_2 + V_2 &= 4, & V_2 &= 4 - U_2 = 4 - (-2) = 6; \\ U_3 + V_2 &= 8, & U_3 &= 8 - V_2 = 8 - 6 = 2; \\ U_3 + V_3 &= 12, & V_3 &= 12 - U_3 = 12 - 2 = 10; \\ U_3 + V_4 &= 7, & V_4 &= 7 - U_3 = 7 - 2 = 5 \end{aligned}$$

$b_j \backslash a_i$	150	120	80	50	U_i
100	- 3 100	5	7	11	0
130	+ 1 50	4 80	6	2	-2
170	5	8 40	12 80	7 50	2
V_j	3	6	10	5	

$$X_0 = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 50 & 80 & 0 & 0 \\ 0 & 40 & 80 & 50 \end{pmatrix}, \quad F(X_0) = 2300.$$

Теперь для незанятых клеток проверим условие оптимальности:

$$\Delta_{12} = (U_1 - V_2) - C_{12} = -1;$$

$$\Delta_{23} = (U_2 - V_3) - C_{23} = 2;$$

$$\Delta_{12} = (U_1 - V_2) - C_{13} = 3;$$

$$\Delta_{24} = (U_2 - V_4) - C_{24} = 1;$$

$$\Delta_{14} = (U_1 - V_4) - C_{14} = -6;$$

$$\Delta_{31} = (U_3 - V_1) - C_{31} = 0.$$

Видно, что среди Δ_{ij} имеются положительные. Поэтому этот план не является оптимальным. Переходим к новому опорному плану: $\max\{3; 2; 1\} = 3 = \Delta_{13}$.

Значит, в таблице клетки A_1B_3 отметим знаком «+» и построим цикл. Вычислим $\theta_0 = \min\{100; 80\} = 80$. Значение $\theta_0 = 80$ записываем в незанятую клетку со знаком «+»; двигаясь по циклу вычитаем $\theta_0 = 80$ из объёмов перевозок, расположенных в клетках со знаком «-» и прибавляем к объёмам перевозок, находящихся в клетках «+». В результате получим новый опорный план:

$a_i \backslash b_j$	150	120	80	50	U_i
100	3 20	5	7 80	11	0
130	- 1 130	4	6	+ 2 0	-2
170	5 +	8 120	12	- 7 50	3
V_j	3	5	7	4	

$$X_1 = \begin{pmatrix} 20 & 0 & 80 & 0 \\ 130 & 0 & 0 & 0 \\ 0 & 120 & 0 & 50 \end{pmatrix}, F(X_1) = 2060.$$

X_1 – вырожденный опорный план, ибо одной занятой клетки не хватает. Поэтому выбираем строку, которая содержит наибольшее количество занятых клеток (например, строка A_1) и полагаем $U_1 = 0$. Тогда $U_1 = 0$. $U_1 + V_1 = 3$; $V_1 = 3 - U_1 = 3 - 0 = 3$. $V_1 + U_1 = 3$; $U_1 + V_3 = 7$. $V_3 = 7 - U_1 = 7 - 0 = 7$; $V_3 + U_2 = 7$; $U_2 + V_1 = 4$; $U_2 = 4 - V_1 = 4 - 3 = 1$. А из соотношения $U_3 + V_2 = 8$. $U_3 + V_4 = 7$ невозможно определить U_3 , V_2 , V_4 .

Эта ситуация возникла из-за того что опорный план оказался невырожденным, т.е. количество занятых клеток равно 5. Чтобы выйти из этого положения и определение U_3 и V_4 сделать фиктивно занятой одну из незанятых клеток строки A_3 или столбца B_4 . **Фиктивно занятые клетки это те клетки, в которые введены нулевые перевозки.** Задача решается на минимизацию линейной функции, поэтому целесообразно сделать фиктивно занятой клетку, в которой стоит наименьшая стоимость. Такой клеткой является клетка A_2B_4 . Теперь последовательно находим

$$U_2 + V_4 = 2, V_4 = 2 - U_2 = 2 - (-2) = 4, V_4 = 4;$$

$$U_3 + V_4 = 7, U_3 = 7 - V_4 = 7 - 4 = 3, U_3 = 3;$$

$$U_3 + V_2 = 8, V_2 = 8 - U_3 = 8 - 3 = 5, V_2 = 5.$$

Теперь для не занятых клеток проверяем условие оптимальности:

$$\Delta_{12} = (U_1 + V_2) - C_{12} = 0; \quad \Delta_{31} = (U_3 + V_1) - C_{31} = 1;$$

$$\Delta_{14} = (U_1 + V_4) - C_{14} = -7; \quad \Delta_{33} = (U_3 + V_3) - C_{33} = 2.$$

$$\Delta_{23} = (U_2 + V_3) - C_{23} = -1;$$

План X_1 не является оптимальным, ибо $\Delta_{31} = 1 > 0$.

В таблице клетку A_3B_1 отметим знаком «+» и построим цикл; находим $\theta_0 = \min\{130; 50\} = 50$. Значение $\theta_0 = 50$ записываем в не занятую клетку со знаком «+»; двигаясь по циклу вычитаем $\theta_0 = 50$ из объемов перевозок, расположенных в клетках со знаком «-» и прибавляем к объемам перевозок, находящихся в клетках со знаком «+». В результате получим новый опорный

план

$a_i \backslash b_j$	150	120	80	50	U_i
100	- 3 20	5 +	7 80	11	0
130	80	1 4	6	2 50	-2
170	- 5 50	8 120	12 -	7	2
V_j	3	6	7	4	

$$X_2 = \begin{pmatrix} 20 & 0 & 80 & 0 \\ 80 & 0 & 0 & 50 \\ 50 & 120 & 0 & 0 \end{pmatrix}, \quad F(X_2) = 2010.$$

Полученный опорный план X_2 является невырожденным. Выполнив необходимые вычисления убеждаемся что план X_2 не является оптимальным. Построим очередной опорный план

$a_i \backslash b_j$	150	120	80	50	U_i
100	3	5 20	7 80	11	0
130	80	1 4	6	2 50	-2
170	5 70	8 100	12	7	2
V_j	3	6	7	4	

$$X_3 = \begin{pmatrix} 0 & 20 & 80 & 0 \\ 80 & 0 & 0 & 50 \\ 70 & 100 & 0 & 0 \end{pmatrix}, \quad F(X_3) = 1990.$$

X_3 является невырожденным опорным планом. Этот план является оптимальным. Для этого выполняются все условия оптимальности:

$$\Delta_{11} = (U_1 + V_1) - C_{11} = -1; \quad \Delta_{23} = (U_2 + V_3) - C_{23} = 0;$$

$$\Delta_{14} = (U_1 + V_4) - C_{14} = -12; \quad \Delta_{33} = (U_3 + V_3) - C_{33} = -2;$$

$$\Delta_{22} = (U_2 + V_2) - C_{12} = 0; \quad \Delta_{34} = (U_3 + V_4) - C_{34} = -5.$$

т.е. $X_3 = X_{opt}$ и $F_{min} = F(X_3) = 1990$.

4. Транспортная задача с «открытой» моделью

В ранее рассмотренных транспортных задачах суммарные запасы и потребности совпадали. Такие задачи называются транспортные задачи с «закрытой» моделью, в противном случае задачу называют **транспортная задача с «открытой» моделью**.

Транспортная задача с «открытой» моделью решается приведением к задаче с «закрытой» моделью.

Для приведения задачи с «открытой» модели к задаче с «закрытой» модели вводятся «фиктивный потребитель» или «фиктивный поставщик». А именно:

а) в случае, когда суммарные запасы превышают суммарные потребности, вводится «фиктивный потребитель» B_{n+1} , потребности которого

$$b_{n+1} = \sum_{i=1}^n a_i - \sum_{j=1}^m b_j.$$

б) в случае, когда суммарные потребности превышают суммарные запасы, вводится «фиктивный поставщик» A_{n+1} , запасы которого

$$a_{n+1} = \sum_{j=1}^m b_j - \sum_{i=1}^n a_i.$$

Стоимость перевозки как до «фиктивного потребителя», так и стоимость перевозки единицы груза от «фиктивного поставщика» полагают равными нулю, так как груз в обоих случаях не перевозится.

Пример. Приведите следующую транспортную задачу с «открытой» модели к задаче с «закрытой» модели и найдите ее оптимальное решение.

$b_i \backslash a_j$	3	3	3	2	2
4	3	2	1	2	3
5	5	4	3	1	1
7	0	2	3	4	5

В этой задаче

$$\sum_{i=1}^3 a_i = 16 > \sum_{j=1}^5 b_j = 13$$

Поэтому введем «фиктивного потребителя», потребности которого

$$b_6 = 16 - 13 = 3.$$

Теперь составим таблицу планирования:

$b_i \backslash a_j$	3	3	3	2	2	3
4	3	2	1	2	3	0
5	5	4	3	1	1	0
7	0	2	3	4	5	0

Полученную задачу с «закрытой» модели решаем методом потенциалов. После 7-ми итераций получим следующий оптимальный план задачи:

$$X_{opt} = \begin{pmatrix} 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 1 \\ 3 & 2 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad F_{min} = F(X_{opt}) = 13.$$

5. Усложненные задачи транспортного типа

Нами рассмотрена классическая транспортная задача, на которой показано, в частности, как используется метод потенциалов для нахождения оптимального плана. В экономике предприятия такие задачи встречаются крайне редко. Обычно при составлении математической модели задачи транспортного типа приходится вводить целый ряд дополнительных ограничений, и затем пользоваться методом потенциалов.

Ряд экономических задач легко сводимы к транспортной задаче. Рассмотрим наиболее часто встречающиеся ситуации в экономике предприятия.

1. Отдельные поставки от определенных поставщиков некоторым потребителям должны быть исключены (из-за отсутствия необходимых условий хранения, чрезмерной перегрузки коммуникаций и т. д.). Это ограничение требует, чтобы в матрице перевозок, содержащей оптимальный план, определенные клетки оставались свободными. Последнее достигается искусственным завышением затрат на перевозки c_{ij} в клетках, перевозки через которые следует запретить. При этом производят завышение величины c_{ij} до таких значений, которые будут заведомо больше всех и с которыми их придется сравнивать в процессе решения задачи.

2. На предприятии необходимо определить минимальные суммарные затраты на производство и транспортировку продукции. С подобной задачей сталкиваются при решении вопросов, связанных с оптимальным размещением производственных объектов. Здесь может оказаться экономически более выгодным доставлять сырье из более отдаленных пунктов, но зато при меньшей его себестоимости. В таких задачах за критерий оптимальности принимают сумму затрат на производство и транспортировку продукции.

3. Ряд транспортных маршрутов, по которым необходимо доставить грузы, имеют ограничения по пропускной способности. Если, например, по маршруту $A_i B_j$ можно провести не более q единиц груза, то B_j -й столбец матрицы разбивается на два столбца - B'_j и B''_j . В первом столбце спрос при-

нимается равным разности между действительным спросом b_j и ограничением q : $b'_j = b_j - q$, во втором - равным ограничению q , т.е. $b''_j = q$. Затраты c_{ij} в обоих столбцах одинаковы и равны данным, но в первом столбце B'_j , клетке, соответствующей ограничению i , вместо истинного тарифа c_{ij} ставится искусственно завышенный тариф M (клетка блокируется). Затем задача решается обычным способом.

4. Поставки по определенным маршрутам обязательны и должны войти в оптимальный план независимо от того, выгодно это или нет. В этом случае уменьшают запас груза у поставщиков и спрос потребителей и решают задачу относительно тех поставок, которые необязательны. Полученное решение корректируют с учетом обязательных поставок.

5. Экономическая задача не является транспортной, но в математическом отношении подобна транспортной, так как описывается аналогичной моделью, например, распределение производства изделий между предприятиями, оптимальное закрепление механизмов по определенным видам работы.

6. Необходимо максимизировать целевую функцию задачи транспортного типа. В этой ситуации при составлении опорного плана в первую очередь стараются заполнить клетки с наиболее высокими значениями показателей c_{ij} . Выбор клетки, подлежащей заполнению при переходе от одного допустимого плана к другому, должен производиться не по максимальной положительной разнице $\Delta_{ij} = U_i + V_j - c_{ij}$, а по минимальной отрицательной разнице $\Delta_{ij} = U_i + V_j - c_{ij}$. Оптимальным будет план, которому в последней таблице сопутствуют свободные клетки с неотрицательными элементами: все разности $\Delta_{ij} = U_i + V_j - c_{ij} \geq 0$.

7. Необходимо в одно время распределить груз различного рода по потребителям. Задачи данного типа называются многопродуктовыми транспортными задачами. В этих задачах поставщики n родов грузов разбиваются

на n условных поставщиков, а потребители m родов грузов разбиваются на m условных потребителей. С учетом этой разбивки составляют полную транспортную таблицу. При этом заметим, что некоторые маршруты $A_i B_j$, должны быть блокированы (закрыты), поскольку в данной постановке задачи грузы разного рода не могут заменять друг друга. Этим маршрутам $A_i B_j$ должна соответствовать очень высокая стоимость перевозки. Многопродуктовую задачу не всегда обязательно описывать одной моделью. Например, если поставки грузов различного рода независимы, то задачу можно представить в виде комплекса транспортных задач по каждому роду груза. Однако если между грузами различного рода существует связь (например, одни из грузов можно заменить другими), то в общем случае исходную модель (задачу) не удастся разбить на комплекс простых транспортных задач.

Рассмотрим примеры задач транспортного типа.

Пример 5.1. Одно фермерское хозяйство (A_1) имеет продовольственное зерно двух видов: 3 тыс. т - III класса и 4 тыс. т - IV класса. Второе фермерское хозяйство (A_2) также имеет зерно двух классов: 5 тыс. т - III класса и 2 тыс. т - IV класса. Зерно должно быть вывезено на два элеватора: на первый элеватор (B_1) необходимо поставить 2 тыс. т пшеницы III класса, 3 тыс. т пшеницы IV класса и остальные 2 тыс. т пшеницы любого класса.

Аналогично второй элеватор (B_2) должен получить 8,25 тыс. т, из них пшеницы - 1 тыс. т III класса и 1,5 тыс. т IV класса.

Стоимость перевозки в д. е. 1 т зерна составляет: из пункта A_1 в пункты B_1 и B_2 - 1 и 1,5 соответственно; из пункта A_2 в пункты B_1 и B_2 - 2 и 1 д. е. соответственно.

Составить оптимальный план перевозок.

Решение

Каждого поставщика условно разбиваем на две части согласно двум видам зерна (A_1^3 и A_1^4 ; A_2^3 и A_2^4), аналогично потребителей разбиваем на три

части (пшеница III класса, IV класса и любой класс): B_1^3 , B_1^4 и B_1^0 , а также B_2^3 , B_2^4 и B_2^0 . Потребности превышают запасы, поэтому вводим фиктивного поставщика A_3 . Часть клеток в таблице запираем большими числами M : например, в клетке (1; 2) стоит большое число. Это значит, что поставщик A_1 не может удовлетворить потребителя B_1^4 пшеницей IV класса за счет имеющейся пшеницы III класса.

С учетом сделанных замечаний составим первую таблицу (табл. 1).

Таблица 1

Исходные данные

Поставщики		Потребители						Запас, тыс. т
		B_1			B_2			
		B_1^3	B_1^4	B_1^0	B_2^3	B_2^4	B_2^0	
A_1	A_1^3	1	M	1	1,5	M	1,5	3
	A_1^4	M	1	1	M	1,5	1,5	4
A_2	A_2^3	2	M	2	1	M	1	5
	A_2^4	M	2	2	M	1	1	2
A_3		0	0	0	0	0	0	1,25
Спрос, тыс. т		2	3	2	1	1,5	5,75	15,25

Перевозки от фиктивного поставщика не производятся, поэтому $c_{s_1} = c_{s_2} = c_{s_3} = c_{s_4} = c_{s_5} = c_{s_6} = 0$. Величина M намного больше c_{ij} . Применяя метод потенциалов, в итоге получим таблицу с оптимальным решением (табл. 1).

Анализ решения

Первый поставщик поставит на первый элеватор B_1 пшеницу III класса ($x_{22} = 2$); пшеницу IV класса ($x_{23} = 3$), а также пшеницу любого класса (III или IV) ($x_{13} = 1; x_{23} = 1$).

Второй поставщик (A_2) поставит на второй элеватор (B_2) пшеницу III класса ($x_{31} = 1$), пшеницу IV класса ($x_{45} = 1,5$) и частично

Таблица 2

Оптимальное решение

Поставщики		Потребители						Запас, тыс. т
		B_1			B_2			
		B_1^3	B_1^4	B_1^0	B_2^3	B_2^4	B_2^0	
A_1	A_1^3	$\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}$	M	$\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}$	1,5	M	1,5	3
	A_1^4	M	$\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}$	$\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}$	M	1,5	1,5	4
A_2	A_2^3	2	M	2	$\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}$	M	$\begin{smallmatrix} 1 \\ 4 \end{smallmatrix}$	5
	A_2^4	M	2	2	M	$\begin{smallmatrix} 1 \\ 1,5 \end{smallmatrix}$	$\begin{smallmatrix} 1 \\ 0,5 \end{smallmatrix}$	2
A_3		0	0	0	0	0	$\begin{smallmatrix} 0 \\ 1,25 \end{smallmatrix}$	1,25
Спрос, тыс. т		2	3	2	1	1,5	5,75	15,25

любую пшеницу ($x_{36} = 4; x_{46} = 0,5$). Потребность элеватора в любой пшенице не удовлетворена на 1,25 тыс. т ($x_{56} = 1,25$). Минимальные затраты на перевозку составили: $F_{\min} = 14$ д.е.

Пример 5.2. Модель производства с запасами

Фирма переводит свой головной завод на производство определенного вида изделий, которые будут выпускаться в течение четырех месяцев. Вели-

чины спроса в течение этих четырех месяцев составляют 100, 200, 180 и 300 изделий соответственно. В каждый месяц спрос можно удовлетворить за счет:

- запасов изделий, произведенных в прошлом месяце, сохраняющихся для реализации в будущем;
- производства изделий в течение текущего месяца;
- избытка производства изделий в более поздние месяцы в счет невыполненных заказов.

Затраты на одно изделие в каждом месяце составляют 4 д.е. Изделие, произведенное для более поздней реализации, влечет за собой дополнительные издержки на хранение в 0,5 д. е. в месяц. С другой стороны, каждое изделие, выпускаемое в счет невыполненных заказов, облагается штрафом в размере 2 д. е. в месяц.

Объем производства изделий меняется от месяца к месяцу в зависимости от выпуска других изделий. В рассматриваемые четыре месяца предполагается выпуск 50, 180, 280 и 270 изделий соответственно.

Требуется составить план, имеющий минимальную стоимость производства и хранения изделий.

Решение

Задачу можно сформулировать как транспортную. Эквивалентность между элементами производственной и транспортной систем устанавливается следующим образом (табл. 3):

Таблица 3

Транспортная система	Производственная система
1. Исходный пункт i	1. Период производства i
2. Пункт назначения j	2. Период потребления j
3. Предложение в пункте i	3. Объем производства за период i
4. Спрос в пункте j	4. Реализация за периоду j
5. Стоимость перевозки из i в j	5. Стоимость производства и хранения за период i и j

Перед нами структура транспортной модели. Для рассматриваемой задачи стоимость «перевозки» изделия из периода i в период j выражается как:

$$c_{ij} = \begin{cases} \bullet \text{ стоимость производства в } i\text{-й период, } i = j; \\ \bullet \text{ стоимость производства в } i\text{-й период плюс стоимость} \\ \text{задержки от } i \text{ до } j, \quad i < j; \\ \bullet \text{ стоимость производства в } i\text{-й период плюс штраф} \\ \text{за нарушение срока, } i > j. \end{cases}$$

Из определения c_{ij} следует, что затраты в период i при реализации продукции в тот же период i ($i = j$) оцениваются только стоимостью производства. Если в период i производится продукция, которая будет потребляться позже ($i < j$), то имеют место дополнительные издержки, связанные с хранением. Аналогично производство в i -й период в счет невыполненных заказов $i > j$ влечет за собой дополнительные расходы в виде штрафа. Например,

$$c_{11} = 4 \text{ д. е.}; \quad c_{12} = 4 + (0,5 + 0,5) = 5 \text{ д. е.}; \quad c_{14} = 4 + (2 + 2 + 2) = 10 \text{ д. е.}$$

Исходная транспортная таблица выглядит следующим образом (табл. 4).

Таблица 4

Исходные данные

Период	Период				Объем произд-ва
	1	2	3	4	
1	4	4,5	5	5,5	50
2	6	4	4,5	5	180
3	8	6	4	4,5	280
4	10	8	6	4	270
Спрос	100	200	180	300	

Задача решается обычным методом потенциалов на минимум затрат по производству и хранению продукции.

Пример 5.3. Имеются три сорта бумаги в количестве 10, 8 и 5 т, которую можно использовать на издание четырех книг тиражом 8000, 6000, 10 000, 15 000 экземпляров. Расход бумаги на одну книгу составляет: 0,6; 0,4; 0,8; 0,5 кг, а себестоимость тиража книги при использовании i -го сорта бумаги задается следующей матрицей (д. е.):

$$c_{ij} = \begin{pmatrix} 24 & 16 & 32 & 25 \\ 18 & 24 & 24 & 20 \\ 30 & 24 & 16 & 20 \end{pmatrix}.$$

Определить оптимальное распределение бумажных резервов.

Решение

Задача по своему экономическому смыслу не является транспортной, в то же время можно построить математическую модель, аналогичную транспортной задаче.

Потребности в бумаге легко определить, зная тираж и расход на одну книгу (т):

$$8000 \cdot 0,6 = 4,8 \text{ т};$$

$$15000 \cdot 0,4 = 6 \text{ т};$$

$$6000 \cdot 0,8 = 4,8 \text{ т};$$

$$10000 \cdot 0,5 = 5 \text{ т}.$$

Общие запасы бумаги составляют 23 т, а общие потребности - 20,5 т, поэтому необходимо в таблицу ввести фиктивный тираж B_5 с нулевыми затратами. В связи с тем что мы составляем модель относительно бумаги, а матрица c_{ij} характеризует себестоимость печатания книги, необходимо исходную матрицу преобразовать относительно единицы бумаги (каждый столбец матрицы c_{ij} разделим на количество бумаги, приходящейся на одну книгу).

Согласно изложенному составим первую таблицу (табл. 5).

Исходные данные

Поставщики	Потребители					Запасы, т
	B_1	B_2	B_3	B_4	B_5^*	
A_1	40	20	80	50	0	10
A_2	30	30	60	40	0	8
A_3	50	30	40	40	0	5
Потребность, т	4,8	4,8	6	5	2,4	23

Используя метод потенциалов, получим оптимальное решение (табл.

6).

Таблица 6

Оптимальное решение

Поставщики	Потребители					Запасы, т
	B_1	B_2	B_3	B_4	B_5^*	
A_1	40	20 4,8	80	50 2,8	0 2,4	10
A_2	30 4,8	30	60 1	40 2,2	0	8
A_3	50	30	40 5	40	0	5
Потребность, т	4,8	4,8	6	5	2,4	23

Анализ решения.

Бумаги 1-го сорта в количестве 4,8 т затрачено на издание второй книги; 2,8 т - на издание четвертой книги; 2,4 т - не использовано. Бумаги 2-го сорта затрачено: на первую книгу - 4,8 т; на издание третьей книги - 1,0 т; на издание четвертой книги - 2,2 т; бумага 3-го сорта использована на издание третьей книги в количестве 5 т.

Упражнения

Решить следующие транспортные задачи, заданные следующими таблицами планирования.

1.

A_i	B_j	40	60	80	60
60		1	3	4	2
80		4	5	8	3
100		2	3	6	1

2.

A_j	B_j	150	140	190
60		3	7	2
80		9	2	1
100		1	5	7
170		6	4	8

3.

A_j	B_j	100	70	35	45	50
54		12	14	26	16	3
32		8	11	11	22	10
85		6	10	10	21	15
162		10	4	4	8	9

4.

A_j	B_j	20	10	60	30	70
60		18	2	8	3	2
36		8	2	3	12	4
90		4	3	5	7	14
84		9	4	16	5	8

5.

A_i	B_j	125	75	200	380	220
222		20	18	11	8	3
188		10	10	5	2	4
210		2	17	8	4	3
300		3	9	17	8	4

В задачах 6 – 9 решить транспортные задачи, в которых заданы векторы запасов A и потреблений B , а тарифы перевозок задаются матрицей C .

6.

$$A=(30,40,90), \quad B=(40,20,30,60).$$

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

8.

$$A=(400,300,900), \quad B=(400,200,300,600).$$

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

9.

$$A=(18,24,30), \quad B=(12,18,24,18).$$

$$C=\begin{pmatrix} 2 & 4 & 5 & 3 \\ 5 & 6 & 9 & 4 \\ 3 & 4 & 7 & 8 \end{pmatrix}.$$

Вопросы для самопроверки

1. Дайте определение системе потенциалов, расскажите, как она строится.
2. В каком случае опорный план транспортной задачи является оптимальным?
3. Какая модель транспортной задачи называется закрытой, а какая - открытой?
4. Как открытую модель преобразовать в закрытую?
5. Для решения, каких экономических задач применяется транспортная задача? Сформулируйте эти задачи и постройте их математические модели. Что такое усложненные задачи транспортного типа?

§20. Целочисленное линейное программирование

Ключевые слова: неделимые величины, полностью целочисленная задача линейного программирования, частичная целочисленная задача линейного программирования, метод Гомори, задачи с неделимостями, экстремальные комбинаторные задачи, задачи оптимизации капиталовложений, задачи оптимизации производственной программы, двумерная задачи раскрыя, задача о ранце.

1. Постановка задачи

Целочисленные задачи математического программирования могут возникать различными путями. Существуют задачи линейного программирования, которые формально к целочисленным не относятся (требование целочисленности переменных в них в явном виде не накладывается), но которые при целочисленных исходных данных всегда обладают целочисленным планом. Этим свойством обладают транспортная задача и различные ее варианты (задача о назначениях).

Первоначальным стимулом к изучению целочисленных задач явилось рассмотрение задач линейного программирования, в которых переменные представляли физически **неделимые** величины (скажем, количество единиц продукции разных видов). Для характеристики этого класса моделей используется термин «задачи с неделимостями».

Итак, во многих случаях на переменные задачи линейного программирования накладывается дополнительное требование целочисленности переменных. Если этому требованию должны удовлетворять все переменные, то получаем полностью целочисленную задачу линейного программирования, которая записывается следующим образом:

$$F(x) = \sum_{j=1}^n c_j x_j \rightarrow \min, \quad (1)$$

$$\sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \quad (2)$$

$$x_j \geq 0; \quad j = 1, 2, \dots, n, \quad (3)$$

$$x_j \in Z \quad j = 1, 2, \dots, n, \quad (4)$$

где Z – множество целых чисел.

2. Графический метод

Полностью целочисленную задачу с двумя переменными можно решить графически, учитывая, что множество допустимых решений $\bar{K}$ этой задачи состоит из точек целочисленной координатной сетки, принадлежащих множеству допустимых решений задачи (1) - (3), т.е. задачи линейного программирования без дополнительного требования (4).

Пример 1.

$$F = x_1 - 20x_2 \rightarrow \min,$$

$$-x_1 + 10x_2 \leq 40,$$

$$4x_1 + 2x_2 \leq 29,$$

$$x_j \geq 0, \quad x_j \in Z, \quad j = 1, 2$$

Решение

На плоскости R^2 построим множество допустимых решений K рассматриваемой задачи линейного программирования без требования целочисленности (многоугольник ABCD на рис. 1) и отметим точки множества K с целочисленными координатами. Совокупность этих точек представляет собой множество допустимых решений $\bar{K}$ полностью целочисленной задачи.

Перемещая линию уровня целевой функции $F(x)$ в направлении $\nabla F = (-1; 20)$ убывания $F(x)$, находим крайнее положение этой линии, в котором она ещё имеет непустое пересечение со множеством $\bar{K}$. В этом положении линия уровня проходит через точку B(0,4), поэтому решение задачи будет

$$\bar{X}_{opt} = (0, 4) \text{ и } F_{\min} = F(\bar{X}_{opt}) = -80.$$

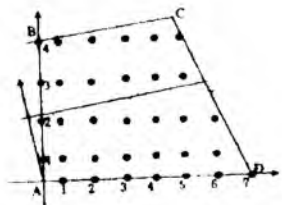


Рис.1

Отметим, что (как видно из рис. 1) точкой минимума $F(x)$ в данной задаче без требования целочисленности является точка $C(5; 4,5)$, т.е.

$$X_{opt} = (5; 4,5) \text{ и } F_{min} = F(X_{opt}) = -85.$$

Отсюда следует, что точкой минимума целевой функции на множестве допустимых решений $\tilde{K}$ целочисленной задачи не обязательно является ближайшая к решению X_{opt} обычной (не целочисленной) задачи точка множества K с целочисленными координатами.

Упражнения

Решить следующие полностью целочисленные задачи линейного программирования графическим методом.

$$1. \begin{cases} F(x) = -x_1 - x_2 \rightarrow \min, \\ 2x_1 + 3x_2 \leq 36, \\ x_1 \leq 13, \\ 3x_1 + x_2 \geq 6, \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2 \end{cases}$$

$$2. \begin{cases} F(x) = -9x_1 - 11x_2 \rightarrow \min. \\ 4x_1 + 3x_2 \leq 10, \\ x_1 \leq 5, \\ x_1 + 2x_2 \leq 8, \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2. \end{cases}$$

$$3. \begin{cases} F(x) = -x_1 - x_2 \rightarrow \min, \\ 2x_1 + 3x_2 \leq 5, \\ x_1 \leq 2, \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2. \end{cases}$$

$$4. \begin{cases} F(x) = -4x_1 - 3x_2 \rightarrow \min, \\ 4x_1 + x_2 \leq 10, \\ 2x_1 + 3x_2 \leq 8, \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2 \end{cases}$$

$$F(x) = 3x_1 + x_2 \rightarrow \min,$$

$$5. \begin{cases} -4x_1 + x_2 \leq 29, \\ 3x_1 - x_2 \leq 15, \\ 5x_1 + 2x_2 \geq 38. \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2. \end{cases}$$

$$F(x) = 5x_1 + 7x_2 \rightarrow \min,$$

$$6. \begin{cases} -3x_1 + 14x_2 \leq 78, \\ 5x_1 - 6x_2 \leq 26, \\ x_1 + 4x_2 \geq 25, \\ x_1 \geq 0, x_2 \geq 0, x_j \in Z, j = 1, 2. \end{cases}$$

3. Метод Гомори

Для решения полностью целочисленных задач линейного программирования с произвольным числом переменных используется метод Гомори. Он состоит в последовательном отсечении от множества допустимых решений K нецелочисленной задачи частей, не содержащих точек с целыми координатами. Эти отсечения производятся включением в задачу дополнительных ограничений на переменные x_j .

Опишем алгоритм метода Гомори:

1. С помощью симплекс-метода находится решение X_{opt} задачи линейного программирования без учета требования целочисленности (4). Если для X_{opt} условие (4) выполняется, то задача решена. В противном случае среди чисел β_i и столбца A_0 , определяющей решение X_{opt} , есть такие, что $\{\beta_i\} > 0$.

Замечание. Напомним, что любое $a \in R^1$ можно представить в виде $a = [a] + \{a\}$, где $[a]$ - целая часть числа a , а $\{a\} = a - [a]$ - его дробная часть.

Например,

$$\left[\frac{7}{3} \right] = 2; \left[-\frac{7}{3} \right] = -3, \quad \left\{ \frac{7}{3} \right\} = \frac{7}{3} - \left[\frac{7}{3} \right] = \frac{7}{3} - 2 = \frac{1}{3},$$

$$\left\{ -\frac{7}{3} \right\} = -\frac{7}{3} - \left[-\frac{7}{3} \right] = -\frac{7}{3} + 3 = \frac{2}{3}.$$

2. Среди нецелых элементов β_i выбирается произвольный элемент $\{\beta_r\}$ (например, исходя из условия $\{\beta_r\} = \max\{\beta_i\}$).

По r -й строке симплекс-таблицы составляется дополнительное ограничение вида

$$-\sum_{j=m+1}^n \{q_j\} x_j \leq -\{\beta_r\}.$$

С помощью вспомогательной переменной $x_{n+1} \geq 0$ это ограничение представляется в виде равенства

$$-\sum_{j=m+1}^n \{q_j\} x_j + x_{n+1} = -\{\beta_r\}$$

и вводится в симплекс-таблицу дополнительной строкой.

Так как $x_{n+1} = \{\beta_r\} < 0$, то теперь симплекс-таблица перестает соответствовать допустимому базисному решению задачи линейного программирования, которую она описывает (на столбце A_0 появляется $-\{\beta_r\} < 0$).

3. Для перехода к допустимому базисному решению производятся следующие операции:

а) строка с отрицательным свободным членом β_r считается разрешающей (на первом шаге, очевидно $k = n + 1$);

б) если все коэффициенты $q_{ij} > 0$, то задача **не имеет решения**, в противном случае номер l разрешающего столбца находится из условия

$$\frac{\Delta_l}{|q_{kl}|} = \min_{j: q_{kj} < 0} \frac{\Delta_j}{|q_{kj}|};$$

в) совершается преобразование симплекс-таблицы с опорным элементом q_{kl} .

4. Если найденное в разделе 3 решение задачи линейного программирования удовлетворяет условию целочисленности, то вычисления завершаются, а если нет, то продолжаются переходом к разделу 2 описания алгоритма. Описанный алгоритм позволяет найти решение полностью целочисленной задачи линейного программирования или установить отсутствие решений за конечное число итераций.

Пример 2. Решить задачу, рассмотренную в примере 1, методом Гомори.

Решение

Введя дополнительные переменные $x_3 \geq 0$, $x_4 \geq 0$, запишем эту задачу в каноническом виде:

$$\begin{aligned} F &= x_1 - 20x_2 \rightarrow \min, \\ -x_1 + 10x_2 + x_3 &= 40, \\ 4x_1 + 2x_2 + x_4 &= 29, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, \\ x_j &\in \mathbb{Z}, j = 1, 2, 3, 4. \end{aligned}$$

Отметим, что, так как все коэффициенты ограничений-равенств данной задачи целые, то целочисленность исходных переменных x_1, x_2 влечет целочисленность и дополнительных переменных x_3, x_4 . Поэтому и после перехода к каноническому виду можно рассматривать данную задачу как полностью целочисленную (что мы сделали выше) и применить для её решения метод Гомори.

Задачу решим симплексным методом.

Б	C _б	A ₀	1	-20	0	0
			A ₁	A ₂	A ₃	A ₄
A ₃	0	40	-1	10	1	0
A ₄	0	29	4	2	0	1
Δ_j		0	-1	20	0	0
A ₂	-20	4	$-\frac{1}{10}$	1	$\frac{1}{10}$	0
A ₄	0	21	$\frac{21}{5}$	0	$-\frac{1}{5}$	1
Δ_j		-80	2	0	-2	0
A ₂	-20	$\frac{9}{2}$	0	1	$\frac{2}{21}$	$\frac{1}{42}$
A ₁	1	5	1	0	$-\frac{1}{21}$	$\frac{5}{21}$
Δ_j		-85	0	0	$-\frac{41}{21}$	$-\frac{5}{21}$

Это решение

$$X_{opt} = \left(5; \frac{9}{2}; 0; 0\right), F_{min} = F(X_{opt}) = -85$$

не удовлетворяет условию целочисленности, поэтому дополняем последнюю симплекс-таблицу дополнительной строкой.

$$\{\beta_i\} = \frac{9}{2} - \left[\frac{9}{2}\right] = \frac{9}{2} - 4 = \frac{1}{2}; \quad q_{11} = q_{12} = 0; \quad q_{13} = \frac{2}{21} - 0 = \frac{2}{21}; \quad q_{14} = \frac{1}{42} - 0 = \frac{1}{42}.$$

Для перехода к допустимому базисному решению находим разрешающий а элемент по описанному правилу и преобразуем симплексную таблицу:

Б	C ₆	A	1	-20	0	0	0
			A ₁	A ₂	A ₃	A ₄	A ₅
A ₂	-20	9/2	0	1	2/21	1/42	0
A ₁	1	5	1	0	-1/21	5/21	0
Δ ₁		-85	0	0	-41/21	-5/21	0
A ₅	0	-1/2	0	0	-2/21	-1/42	1
A ₂	-20	4	0	1	0	0	1
A ₁	1	0	1	0	-1	0	10
A ₄	0	21	0	0	4	1	-42
Δ ₂		-80	0	0	-1	0	-10

Последняя симплекс-таблица дает решение рассматриваемой задачи:

$$\tilde{X}_{opt} = (0, 4) \text{ и } F_{min} = F(\tilde{X}_{opt}) = -80.$$

Отметим, что дополнительное ограничение, введенное в симплексную таблицу, имеет вид

$$-\frac{1}{42}x_4 - \frac{4}{42}x_5 \leq -\frac{1}{2}.$$

С помощью уравнений $x_1 = 40 + x_1 - 10x_2$, $x_4 = 29 - 4x_1 - 2x_2$ перепишем (5) для переменных x_1, x_2 : $x_2 \leq 4$. Отсюда видно, что дополнительное ограничение соответствует отсечению от множества K (многоугольника ABCD на рис.1) части, содержащей точку $X = (5; 9/2)$ (вершину C этого многоугольника).

Отметим, что переход к каноническому виду в полностью целочисленной задаче линейного программирования, содержащей ограничения-неравенства

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad (5)$$

не приводит, вообще говоря, к полностью целочисленной задаче в каноническом виде, так как в преобразованных ограничениях (5) вспомогательные переменные x_{n+i} не подчинены требованию целочисленности.

Однако, если все коэффициенты a_{ij}, b_i в (5) – целые числа, то условия целочисленности можно распространить и на x_{n+i} , как это сделано при решении примера 2.

Полностью целочисленную задачу в каноническом виде можно получить также, если (5) a_{ij}, b_i – рациональные числа. Для этого следует умножить (5) на общее кратное знаменателей коэффициентов a_{ij}, b_i (т.е. перейти к целым коэффициентам в (5)) и лишь после этого ввести вспомогательные переменные

x_{n+i} .

Упражнения

Решить следующие полностью целочисленные задачи линейного программирования методом Гомори.

$$F(x) = -x_1 + x_4 \rightarrow \min,$$

$$1. \begin{cases} -2x_1 + x_4 + x_5 = 1, \\ x_1 + x_2 - 2x_4 = 2, \\ x_1 + x_3 + 3x_4 = 3, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4. \end{cases}$$

$$F(x) = -x_1 + x_2 - x_3 + x_4 \rightarrow \min,$$

$$2. \begin{cases} x_1 + 2x_3 + x_4 = 8, \\ x_1 + x_2 - x_4 = 4, \\ -x_1 + 2x_2 + x_3 + 3x_4 = 6, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4. \end{cases}$$

$$F(x) = -x_1 - x_2 \rightarrow \min,$$

$$3. \begin{cases} 2x_1 + x_2 + x_3 = 5, \\ 2x_1 + 3x_2 + x_4 = 9, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4. \end{cases}$$

$$F(x) = x_1 + 2x_2 + x_5 \rightarrow \min,$$

$$5. \begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 5, \\ x_2 + x_3 + x_4 - x_5 = 2, \\ x_3 - x_4 + x_5 = 1, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4, 5. \end{cases}$$

$$F(x) = -4x_1 - 3x_2 \rightarrow \min,$$

$$6. \begin{cases} 2x_1 + 3x_2 + x_3 = 8, \\ 4x_1 + x_2 + x_4 = 10, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4. \end{cases}$$

$$F(x) = -x_1 - x_2 \rightarrow \min,$$

$$4. \begin{cases} x_1 + 2x_2 + x_3 = 6, \\ 3x_1 + 2x_2 + x_4 = 9, \\ x_j \geq 0, x_j \in Z, j = 1, 2, 3, 4. \end{cases}$$

7. Сформулируйте задачу оптимального раскроя материалов и составьте ее математическую модель.

8. В обработку поступила партия из 150 досок длиной 7,5 м каждая для изготовления комплектов из четырех деталей. Комплект состоит из одной детали длины 3 м, двух деталей размером по 2 м и одной детали размером 1,6 м. Как распилить все доски, чтобы получить возможно большее число комплектов?

Указание. Сначала составить таблицу возможных способов распила одной доски на детали заданных размеров.

4. Некоторые классы задач целочисленного линейного программирования

4.1. Задача с неделимостями

К данному классу принадлежат задачи распределения капиталовложений и задачи планирования производства.

Рассмотрим следующую задачу целочисленного линейного программирования:

проект A_j , либо проект A_k . Вместе они включены быть не могут. С помощью этой записи выражается отношение взаимоисключения между проектами.

В этих обозначениях взаимоисключение $A_j \vee A_k$ выражается неравенством $x_j + x_k \leq 1$:

2) *взаимобусловленность*. Запись $A_k \rightarrow A_j$ («проект A_k влечет за собой проект A_j ») означает, что проект A_k может быть включен в план только в том случае, если в план включен и проект A_j . С помощью этой записи выражается отношение между обуславливающими друг друга проектами, например, когда проект A_k - результат тиражирования проекта A_j на другом объекте или когда A_k базируется на результатах реализации проекта A_j .

В принятых обозначениях взаимобусловленность $A_k \rightarrow A_j$ выражается неравенством $x_k \leq x_j$.

4.2. Экстремальные комбинаторные задачи

Задачи данного класса, называемые также **задачами выбора**, состоят в отыскании среди конечного множества альтернатив одной, которой отвечает экстремальное значение принятой целевой функции.

Задача о коммивояжере - классический пример задачи выбора оптимального маршрута. Формулируется она следующим образом. Коммивояжер должен выехать из определенного города и вернуться в него, побывав в каждом из городов лишь по одному разу и проехав минимальное расстояние.

Пусть $x_{ij} = 1$, если коммивояжер переезжает из города i непосредственно в город j , и $x_{ij} = 0$ в противном случае. Обозначим через c_{ij} расстояние между городами i и j (чтобы избежать бессмысленных значений $x_{ij} = 1$, предполагается, что c_{ii} равны достаточно большому числу).

Тогда формальная модель имеет вид:

$$f = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min,$$

$$\sum_{j=1}^n x_{ij} = 1, \quad i = 1, 2, \dots, n,$$

$$\sum_{i=1}^n x_{ij} = 1, \quad j = 1, 2, \dots, n.$$

К приведенным ограничениям необходимо добавить условия на недопустимость подциклов, т.е. повторного посещения городов (за исключением исходного). Это ограничения вида

$$z_i - z_j + n x_{ij} \leq n - 1, \quad i = 2, \dots, n; \quad j = 2, \dots, n \quad (i \neq j),$$

где на переменные z_i и z_j не требуется накладывать никаких ограничений.

Общая задача календарного планирования формулируется следующим образом. Имеется n станков (машин), на которых требуется обработать m деталей. Заданы маршруты (в общем случае различные) обработки каждой детали на каждом из станков или группе станков. Задана также продолжительность операций обработки деталей. Предполагается, что одновременно на станке можно обрабатывать не более одной детали. Требуется определить оптимальную последовательность обработки. Критерием оптимальности могут выступать продолжительность обработки всех деталей, суммарные затраты на обработку, общее время простоя станков и др. Существует огромное число постановок данной задачи, учитывающих конкретные условия производства.

Один из представителей задач данного типа - так называемая *задача о ранце*. Имеется n предметов. Предмет j ($j = 1, \dots, n$) обладает весом w_j и полезностью c_j . Пусть b - общий максимально допустимый вес предметов, которые можно положить в ранец. Требуется выбрать предметы таким образом, чтобы их общий вес не превышал максимально допустимый и при этом суммарная полезность (ценность) содержимого ранца была максимальной. Пусть $x_j = 1$, если предмет положен в ранец, и $x_j = 0$ в противном случае. Математическая формулировка задачи имеет вид:

$$\begin{aligned} f &= \sum_{j=1}^n c_j x_j \rightarrow \max, \\ \sum_{j=1}^n w_j x_j &\leq b, \\ x_j &\in \{0; 1\}, \quad j = 1, 2, \dots, n. \end{aligned}$$

К классу экстремальных комбинаторных задач принадлежит также линейный и нелинейный варианты задачи о назначениях.

Большинство целочисленных и комбинаторных типов задач таких, как задача с неделимостями, задача коммивояжера, задача календарного планирования, принадлежит к разряду так называемых *трудно решаемых*. Это означает, что вычислительная сложность алгоритма их точного решения - зависимость числа элементарных операций (операций сложения или сравнения), необходимых для получения точного решения, от размерности задачи n - является экспоненциальной (порядка 2^n), т.е. сравнимой по трудоемкости с полным перебором вариантов. В качестве n , например, для задачи с неделимостями служит число целочисленных переменных и число ограничений, для задачи коммивояжера - число городов (или узлов графа маршрутов), для задачи календарного планирования - число деталей и число станков. Такие задачи называют еще *NP-трудными* или *NP-полными*. Получение их точного решения не может быть гарантировано, хотя для некоторых задач данного типа существуют эффективные методы, позволяющие находить точное решение даже при больших размерностях. Примером таких задач служит задача о ранце с булевыми переменными.

Задачи с вычислительной сложностью, определяемой полиномиальной зависимостью от n , называются *эффективно решаемыми*. К такого типа задачам принадлежат задачи транспортного типа и линейные задачи о назначениях.

Пример 3. Оптимизация капиталовложений. Имеется 10 работ, каждая из которых характеризуется тремя технико-экономическими показателями: a_i - трудозатраты; b_i - размер, необходимых капиталовложений; c_i - ожидаемый экономический эффект.

Исходные данные приведены в следующей таблице:

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
a_i	3	3	3	4	4	2	4	3	6	5
b_i	4	3	2	4	6	4	3	5	3	4
c_i	3	7	5	6	8	4	7	4	7	6

Общие трудозатраты не должны превышать 20. Общий объем капиталовложений не должен превышать 20. Определите, какие из 10 работ следует выполнить, чтобы максимизировать ожидаемый экономический эффект, учитывая следующие условия взаимообусловленности и взаимоисключения:

$$\begin{array}{ll}
 \text{a)} & \begin{array}{c} A_1 \rightarrow A_7 \vee A_4 \\ \downarrow \\ A_{10} \vee A_2 \end{array} \\
 \text{б)} & \begin{array}{c} A_3 \\ \downarrow \\ A_5 \vee A_8 \vee A_9 \rightarrow A_6 \end{array}
 \end{array}$$

Решение

Помимо целевой функции и двух ограничений по общему объему трудозатрат и капиталовложений, данную задачу характеризует следующая система неравенств:

$$\begin{aligned}
 x_1 &\leq x_7, & x_3 &\leq x_8, & x_7 &\leq x_{10}, & x_9 &\leq x_8, \\
 x_{10} + x_2 &\leq 1, & x_7 + x_8 &\leq 1, & x_7 + x_4 &\leq 1, & x_8 + x_9 &\leq 1.
 \end{aligned}$$

В результате расчетов получаем $X^* = (0101110010)$.

Пример 4. Оптимизация производственной программы. Автомобилестроительный завод выпускает три модели автомобилей, которые изготавливаются последовательно в трех цехах. Мощность цехов составляет 300, 250 и 200 человеко-дней в декаду. В первом цехе для сборки одного автомобиля первой модели требуется 6 человеко-дней, второй модели - 4 и третьей модели - 2 человеко-дня в декаду, соответственно. Во втором цехе трудоемкость равна 3, 4 и 5 человеко-дней, соответственно, в третьем - по 3 человеко-дня на каждую модель. Прибыль, получаемая заводом от продажи одного автомобиля каждой модели, составляет, соответственно, 15, 13 и 10 тыс.

Постройте модель для определения оптимального плана.

Решение

Пусть x_i - количество выпускаемых автомобилей i -й модели в течение декады ($i = 1, 2, \dots, n$).

В принятых обозначениях модель имеет вид:

$$F = 15x_1 + 13x_2 + 10x_3 \rightarrow \max$$

$$6x_1 + 4x_2 + 2x_3 \leq 300,$$

$$3x_1 + 4x_2 + 5x_3 \leq 250,$$

$$3(x_1 + x_2 + x_3) \leq 200.$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0,$$

$$x_j - \text{целые}, j = 1, 2, 3.$$

Пример 5. Двумерная задача раскроя. Из минимального количества листов стекла размером $8 \times 6 \text{ м}^2$ требуется вырезать 10 оконных стекол размером $4 \times 4 \text{ м}^2$, 20 оконных стекол размером $4 \times 5 \text{ м}^2$ и 30 оконных стекол размером $3 \times 3 \text{ м}^2$. Множество вариантов раскроя показано в следующей таблице:

Размер, м^2	4x4	4x5	3x3
Вариант			
1	2	-	-
2	1	1	-
3	1	-	2
4	-	2	-
5	-	-	4
6	1	1	2

Постройте модель для определения плана раскроя, требующего минимального количества материала.

Решение

Пусть x_i - количество листов стекла размером $8 \times 6 \text{ м}^2$, которые следует раскроить по варианту i .

Тогда модель имеет вид:

$$F = \sum_{i=1}^n x_i \rightarrow \min,$$

$$2x_1 + x_2 + x_3 \geq 10,$$

$$x_2 + 2x_4 + x_6 \geq 20,$$

$$2x_3 + 4x_5 + 2x_6 \geq 30,$$

$$x_1 \geq 0, x_2 \geq 0, \dots, x_6 \geq 0,$$

$$x_j - \text{целые}, j = 1, 2, \dots, 6.$$

Пример 6. Задача о рации. Некая торговая компания имеет свои универсамы в Ташкенте, Самарканде, Бухаре, Карши, Термезе, Ургендже и Нукусе. В результате ошибок менеджмента экономическое положение компании стало ухудшаться, ей пришлось взять кредит в размере 13 млн. и в конечном счете, чтобы вовремя его погасить, срочно продавать некоторые из своих универсамов. Средства, которые компания могла бы получить от продажи универсамов в Ташкенте, Самарканде, Бухаре, Карши, Термезе, Ургендже и Нукусе составляют, соответственно, 5,2; 4,9; 4,5; 3,6; 3,4; 3,2 и 3,1 млн. Однако продажа универсамов сопряжена с необходимостью увольнения персонала. Его численность составляет, соответственно, 200,190,180,170, 150,130 и 110 человек. По требованию объединенного профсоюза работников торговли компания должна минимизировать численность увольняемого персонала.

Постройте модель для нахождения оптимального решения.

Решение

Пронумеруем города в соответствии с порядком их перечисления. Пусть $x_i = 1$, если универсам, расположенный в городе, продается, и $x_i = 0$ в противном случае. Тогда оптимизационная модель имеет вид:

$$F = 200x_1 + 190x_2 + 180x_3 + 170x_4 + 150x_5 + 130x_6 + 110x_7 \rightarrow \min.$$

$$5,2x_1 + 4,9x_2 + 4,5x_3 + 3,6x_4 + 3,4x_5 + 3,2x_6 + 3,1x_7 \geq 13$$

$$x_i \in \{0; 1\}, \quad i = 1, 2, \dots, 7.$$

Вопросы для самопроверки

1. Как возникают целочисленные задачи математического программирования?
2. Прокомментируйте термин «задачи с неделимостями».
3. Какую задачу называют полностью целочисленной задачей линейного программирования?
4. Запишите полностью целочисленную задачу линейного программирования?

5. Опишите графический метод решения полностью целочисленной задачи с двумя переменными.
6. Опишите метод Гомори решения полностью целочисленной задачи с произвольным числом переменных.
7. Переход к каноническому виду в полностью целочисленной задаче линейного программирования, содержащей ограничения-неравенства $\sum_{j=1}^n a_{ij} x_j \leq b_i$, не приводит, вообще говоря, к полностью целочисленной задаче в каноническом виде, так как в преобразованных ограничениях $\sum_{j=1}^n a_{ij} x_j \leq b_i$ вспомогательные переменные x_{n+i} не подчинены требованию целочисленности. В каких случаях условия целочисленности можно распространить и на дополнительные переменные x_{n+i} и получить полностью целочисленную задачу в каноническом виде?
8. Приведите задачи, принадлежащие к классу задач с неделимостями, и прокомментируйте их. Какие из них допускают экономическую интерпретацию?
9. Приведите задачи, принадлежащие к классу экстремальные комбинаторные задачи, и прокомментируйте их.
10. Приведите постановку задачи оптимизации капиталовложений и прокомментируйте ее.
11. Приведите постановку задачи оптимизации, производственной программы и прокомментируйте ее.
12. Приведите постановку двумерной задачи раскроя и прокомментируйте ее.
13. Приведите постановку задачи о ранце и прокомментируйте ее.

Нелинейным программированием называют раздел математического программирования, в котором либо целевая функция, либо ограничения, либо то и другое нелинейные.

§21. Постановка задачи нелинейного программирования.

Задачи на безусловный минимум

Ключевые слова: задача нелинейного программирования, сепарабельная функция, задача на безусловный минимум, задача на условный минимум, необходимые и достаточные условия безусловного минимума, морсовское I-седло, информативная функция, обезьянье седло.

1. Введение

1.1. Необходимость в методах нелинейного программирования

Даже самого краткого обзора простых, но все же поучительных моделей линейного программирования, приведенных ранее, будет достаточно для того, чтобы вызвать у читателя сомнения в действительной адекватности строго линейных моделей многим реальным ситуациям.

Легко может создаться впечатление, что при линейном подходе попросту игнорируются такие явления как: эффективность или неэффективность укрупнения операций в многопродуктовых моделях; отсутствие аддитивности объемных показателей при составлении химических смесей; влияние объема реализации на цену реализации, а следовательно, на выручку от реализации.

Отметим, что при более глубоком исследовании в ряде задач появляются связи нелинейного характера, когда с изменением одного элемента другие элементы непропорциональны первому. Например, даже простейшая

транспортная задача принимает нелинейный вид, если стоимость перевозки единицы груза зависит от их общего количества.

Если у читателя и создалось такое впечатление, приведенное ниже обсуждение позволит устранить возникшее недоразумение.

Имеется много данных об очень успешном применении моделей линейного программирования в условиях нелинейности. Поскольку любая модель неизбежно оказывается лишь приближением к реальности, возникает важный вопрос: «*в каких случаях* линеаризованный вариант является адекватным отображением нелинейного явления?». Следовательно, читатель должен научиться отличать условия, в которых непосредственное применение линейного программирования приемлемо, от условий, где оно неприемлемо.

Далее следует два примера, на основе которых можно судить о вероятной неадекватности линеаризованного варианта в некоторых реальных ситуациях.

Пример 1. Рассмотрим фирму, находящуюся на начальном этапе разработки модели перспективного планирования в масштабах всей фирмы. Специалист по управлению обычно знает, что даже опытному бизнесмену трудно дать точный и детальный прогноз оптимальных объемов производства фирмы и распространения ее контроля на рынки сбыта на последующие 10 и более лет. В самом деле, применение администратором такой модели в основном обусловлено именно тем, что он понимает, как легко ошибиться при использовании только лишь интуитивных соображений в стремлении оценить влияние экономических факторов в последующие периоды. Если затраты производства и выручка от реализации зависят от объема операции нелинейно, линеаризованные догадки могут оказаться недостаточными для получения надежных ответов (вместе с тем при многократном применении линейная модель может оказаться применимой, если только параметры модели значительно не меняются во времени).

Пример 2. Этот пример относится к фирме, составляющей производственную программу с помощью динамической многопродуктовой модели, отображающей существенные затраты времени на наладку станков, ограниченную мощность отдельных групп оборудования, колеблющийся спрос. Обычно в таких случаях *сущность* оптимизационной задачи состоит в варьировании различных нелинейных факторов, влияющих на принимаемые решения, относительно производственной программы. Если только специалист, разрабатывающий программу, не имеет очень хорошего представления о характере оптимального решения, любая простая линейаризация задачи, вероятно, приведет к нарушению фундаментальных принципов оптимизации.

Однако даже если в данном конкретном случае может потребоваться построение нелинейной модели, иногда удается использовать метод решения, лишь немногим отличающийся от *метода решения* для линейной модели. Наряду с этим при существенной нелинейности в связи с ее спецификой или влиянием ее на характер модели приходится применять методы оптимизации, гораздо более сложные, чем симплексный алгоритм.

1.2. Значение нелинейных моделей для управления

В настоящее время применение математического программирования в преобладающем большинстве реальных ситуаций сводится к моделям линейной аппроксимации, а не к линейным моделям в явном их виде. Однако значимость нелинейного программирования и его использования постоянно возрастает. Это обусловлено быстро растущими познаниями руководителей и специалистов в части использования математических моделей, предназначенных для подготовки решений, а также все большей доступностью компьютерных программ решения нелинейных задач большой размерности.

В большинстве случаев нелинейности, которые необходимо отобразить в моделях, относятся к одной из двух категорий:

1. эмпирически наблюдаемые соотношения, такие, как непропорциональные изменения затрат, выхода продукции, показателей качества;

2. структурно полученные соотношения, к которым относятся постулируемые экономические явления, а также выведенные математически или установленные руководством правила поведения.

Очевидно, четкое разграничение этих двух категорий невозможно, поскольку при наличии достаточных данных можно вывести структурное соотношение, лежащее в основе эмпирически наблюдаемого явления.

Пример 3. Первым примером может служить тот случай, когда на предприятии в течение ряда лет прирост выпуска продукции отстает от роста затрат труда, тогда как темпы роста количества отходов его обгоняют.

Пример 4. Вторым примером является фирма которая должна оплатить счет за электроэнергию в случае, когда расчеты ведутся по нелинейной формуле, учитывающей как среднесуточный расход, так и сведения о нелинейном характере затрат из договора о ставках оплаты, заключенного с компанией, обеспечивающей энергоснабжение.

Нелинейность «встраивается» в модели программирования и в других случаях, например, в следующих:

Пример 5. Приготовление бензиновых смесей. В модели приготовления бензина определенного состава из отдельных фракций, полученных в результате перегонки нефти, обычно имеется нелинейное ограничение на октановое число смеси, поскольку эта характеристика качества нелинейно зависит от количества добавляемого к смеси тетра – этилового свинца.

Пример 6. Выручка от реализации продукции. Спрос на продукцию компании может существенно зависеть от цен реализации: чем ниже цена продукта, тем больше объем реализации, несмотря на аналогичное снижение цен, производимое конкурентами. Следовательно, выручка от реализации продукции не изменяется пропорционально цене, и это обстоятельство должно быть отражено в целевой функции многопродуктовой модели с помощью нелинейного слагаемого. Для иллюстрации примем, что $x(p)$ есть объем реализации, зависящий от цены p ; тогда выручка от реализации равна

$p \cdot x(p)$. Пусть на представляющем для нас интерес интервале изменения p функция объема реализации от цены линейна, т.е. имеет вид $x(p) = ap + b$. Тогда слагаемые в целевой функции, относящиеся к выручке от реализации, являются квадратичным относительно управляющей переменной p и имеют вид $(ap^2 + bp)$.

Пример 7. Размер многопродуктового заказа. В моделях управления запаса обычно число продуктов бывает равно одному. Однако нередко оптовый покупатель пополняет свои запасы, заказывая у одного и того же поставщика одновременно несколько видов продукции. Тем самым достигается экономия на транспортных затратах, расходах по оформлению документации и скидке на размер заказа, представляемой поставщиком. Эта ситуация может рассматриваться на основе использования модели математического программирования большой размерности, в которой затраты на пополнения запасов являются нелинейной функцией нескольких переменных – размеров заказов отдельных продуктов.

Пример 8. Уровень страховых запасов. В большинстве моделей математического программирования, используемых для общеприемного планирования, длительность отрезков планового периода редко составляет менее трех месяцев и часто превышает год и более. В таких динамических, «многопериодных» моделях обычно предусматривается условие наличия страховых запасов, которые должны выполнять компенсатора колебаний еженедельного объема реализации. В этих моделях применяется, в частности, следующий подход: уровень страхового запаса предполагается зависимым как от прогнозируемого объема реализации, так и от степени использования производственных мощностей, обусловленной этим прогнозом. Так, например, пусть c - максимально возможный недельный объем производства рассматриваемого продукта, s - прогнозируемый *средненедельный* объем реализации этого продукта и $n \cdot s$ - уровень страхового запаса продукта, где n - число недель, зависимое от коэффициента использования производственных мощностей

s/c . Для примера предположим, что администрация приняла следующую формулу расчета n : $n = m + f(s/c)$. Тогда уровень страхового запаса представляет собой квадратичную функцию прогнозируемого средненедельного уровня реализации, имеющую следующий вид: $[ms + (j/c)s^2]$. Этот уровень может входить как в ряд ограничений модели, так и в целевую функцию.

Как видно из сказанного, множество разнообразных обстоятельств приводит к нелинейной формулировке ограничений или целевых функций задач математического программирования.

2. Постановка задачи нелинейного программирования

В общем случае, задачу нелинейного программирования можно поставить следующим образом:

Пусть в R^n заданы функции $f(X), g_1(X), g_2(X), \dots, g_m(X)$, из которых хотя бы одна является нелинейной. Определим в R^n каким-нибудь образом скалярное произведение векторов X и Y , например, следующим образом:

$$Y'X = \sum_{i=1}^n x_i y_i = XY.$$

Символом $\|X\|$ будем обозначать норму вектора X , определенную $\|X\|^2 = XX$.

Требуется найти точку, $X^0 \in R^n$ удовлетворяющую условиям:

$$g_1(X) = 0, g_2(X) = 0, \dots, g_m(X) = 0,$$

такую, что

$$f(X^0) = \min_{\substack{g_i(X)=0 \\ i=1, \dots, m}} f(X).$$

Кратко задачу нелинейного программирования в общем случае можно записать так:

$$f(X) \rightarrow \min, \quad (1)$$

$$g_i(X) = 0, \quad i = 1, 2, \dots, m. \quad (2)$$

Очевидно, что задача нелинейного программирования имеет смысл лишь в том случае, когда допустимое множество K :

$$K = \{X : g_1(X) = 0, g_2(X) = 0, \dots, g_m(X) = 0\}$$

не пусто (система ограничений (2) совместна).

Нетрудно понять, что задача нелинейного программирования не всегда имеет решение.

Во многих случаях для доказательства существования решения задачи нелинейного программирования достаточно воспользоваться теоремой Вейерштрасса о минимуме на компакте непрерывной функции.

Задачу (1), (2) называют **задачей на глобальный условный минимум** в отличие от задачи на локальный условный минимум, которая состоит в следующем: найти точку X^0 , удовлетворяющую ограничениям (2), такую, что при некотором достаточно малом ε , $\varepsilon > 0$, для всех допустимых векторов (точек) X из ε -окрестности точки X^0 , $\|X - X^0\| \leq \varepsilon$, выполняется неравенство

$$f(X^0) \leq f(X).$$

В других обозначениях точка X^0 - локального условного минимума - определяется равенством:

$$f(X^0) = \min_{\|X - X^0\| \leq \varepsilon, X \in K} f(X).$$

Замечание. Некоторые авторы используют термины абсолютный и относительный, а не глобальный и локальный условный минимум. Мы ниже будем использовать те и другие термины.

Замечание. Так как любое уравнение равносильно системе двух неравенств

$$\begin{cases} \varphi(X) \leq 0, \\ -\varphi(X) \leq 0, \end{cases}$$

то можно считать, что допустимое множество K задачи нелинейного программирования задается только неравенствами.

Замечание. Отметим также, что если ограничения задачи нелинейного программирования задаются ограничениями типа неравенств, то их можно свести введением дополнительных переменных к ограничениям типа равенств.

Задачи нелинейного программирования образуют широкий класс оптимизационных задач. В частности, к этому классу принадлежит задача минимизации функции n переменных на всем пространстве R^n , которую называют **задачей безусловного минимума**.

Отметим, что основные результаты в нелинейном программировании получены при рассмотрении задач, в которых система ограничений линейная, а целевая функция нелинейная. Даже в таких задачах оптимальное решение может быть найдено только для узкого класса целевых функций. Рассматривают частные случаи, когда целевая функция сепарабельная (является суммой n функций $f_i(X_i)$) или квадратичная. Отметим также, что в некоторых специальных случаях эти задачи можно решить графическим методом. Мы на них остановимся ниже в п. 4.1.

3. Задачи на безусловный минимум (минимизация функции многих переменных)

Приступим к исследованию функций, определенных на n -мерном пространстве R^n . Точки пространства R^n будем обозначать символом X . При операциях с вектором X будем считать его записанным в виде вектора-столбца, хотя часто для экономии места компоненты вектора будут записываться в строку. Для обозначения вектора-строки используется символ $(\cdot)$ -транспонирование. Поэтому

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad X' = (x_1, x_2, \dots, x_n).$$

где $x_i, \quad i = 1, 2, \dots, n$ - компоненты вектора X (координаты точки X).

Для дважды непрерывно дифференцируемой скалярной функции

$$f(X), \quad X \in R^n, \quad (\text{т.е. } f(X) \in C^{(2)}),$$

символы $\nabla f \equiv \frac{\partial f}{\partial X}, \quad \frac{\partial^2 f}{\partial X^2}$ означают

$$\nabla f \equiv \frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad \frac{\partial^2 f}{\partial X^2} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

Если функция $g(X)$ - m -мерная, то символ $\frac{\partial g}{\partial X}$ означает $n \times m$ -мерную матрицу $(\nabla g_1, \nabla g_2, \dots, \nabla g_m)$.

3.1. Постановка задачи на безусловный минимум

Пусть в R^n задана скалярная функция $f(X)$. Требуется найти точку $X^0, \quad X^0 \in R^n$, такую, что

$$f(X^0) = \min_{X \in R^n} f(X).$$

Эта задача называется **задачей на глобальный (абсолютный) минимум**, точка X^0 - точкой глобального (абсолютного) минимума. **Задача на локальный (относительный) минимум** состоит в поиске точки $X^0, \quad X^0 \in R^n$, такой, что при некотором достаточно малом $\varepsilon, \quad \varepsilon > 0$, выполняется соотношение

$$f(X^0) = \min_{\|X - X^0\| \leq \varepsilon, \quad X \in R^n} f(X).$$

Вопрос о существовании решения поставленных задач во многих случаях решается с помощью теоремы Вейерштрасса (п. 2).

3.2. Вспомогательные сведения из теории квадратичных форм

Пусть $A = (a_{ij})$ - симметричная $n \times n$ матрица. Выражение

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = X'AX$$

называется **квадратичной формой**. Эта форма называется **знакоположительной**, если неравенство

$$X'AX \geq 0$$

выполняется для всех точек $X \in R^n$, и **определенно положительной**, если неравенство

$$X'AX > 0$$

выполняется для всех $X \in R^n$, $X \neq 0$. Аналогично определяются **знакоотрицательные** и **определенно отрицательные квадратичные формы**. С введенными квадратичными формами связаны понятия положительных и неотрицательных матриц. Симметричная матрица A называется **положительной (неотрицательной)** и обозначается $A > 0$ (соответственно, $A \geq 0$), если она служит матрицей коэффициентов определено положительной (знакоположительной) квадратичной формы.

Рассмотрим симметрическую матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Минор матрицы A , составленный из строк с номерами $i_1, \dots, i_p$ и столбцов $j_1, \dots, j_p$, обозначим через

$$A \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix} = \begin{vmatrix} a_{i_1 j_1} & \dots & a_{i_1 j_p} \\ \dots & \dots & \dots \\ a_{i_p j_1} & \dots & a_{i_p j_p} \end{vmatrix}.$$

Минор называется **главным**, если $i_1 = j_1, \dots, i_p = j_p$, т.е. он составлен из строк и столбцов с одинаковыми номерами. Миноры

$$D_1 = a_{11}, \quad D_{22} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots, \quad D_n = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

называются **последовательными главными**.

Справедливы следующие утверждения (**критерии Сильвестра**):

1) для того чтобы матрица была положительной, необходимо и достаточно, чтобы ее последовательные главные миноры были положительны:

$$D_1 > 0 \quad \dots, \quad D_n > 0; \quad (3)$$

2) для того чтобы матрица была неотрицательной, необходимо и достаточно, чтобы все ее главные миноры были неотрицательны:

$$A \begin{pmatrix} i_1, \dots, i_p \\ i_1, \dots, i_p \end{pmatrix} \geq 0, \quad 1 \leq i_1 \leq i_2 \leq \dots \leq i_p \leq n, \quad p = 1, 2, \dots, n. \quad (4)$$

Условий $D_1 \geq 0 \quad \dots \quad D_n \geq 0$ недостаточно, чтобы матрица была неотрицательной. Действительно, у матрицы

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

последовательные главные миноры

$$D_1 = a_{11}, \quad D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2.$$

При $a_{11} = 0$, $a_{12} = 0$, $a_{22} < 0$, получаем $D_1 = 0$, $D_2 = 0$, однако в этом случае соответствующая форма $a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = a_{22}x_2^2$ не является знакоположительной (она знакоотрицательна).

Замечание. Применяя условия (3), (4) к матрице A , получаем критерии:

а) отрицательности матрицы: $(-1)^p D_p > 0$, $p = 1, 2, \dots, n$;

б) неположительности матрицы: $(-1)^p A \begin{pmatrix} i_1, \dots, i_p \\ i_1, \dots, i_p \end{pmatrix} \geq 0$.

Замечание. Проверку условий (4) следует начинать с построения последовательных главных миноров, ибо из неравенств (3) следует, что все главные миноры положительны.

3.3. Необходимые условия минимума

Пусть $X^0, X^0 \in R^n$ - точка локального (относительного) минимума, т.е. существует такое $\varepsilon, \varepsilon > 0$, что для всех $X, \|X - X^0\| \leq \varepsilon$, выполняется неравенство $f(X^0) \leq f(X)$.

Теорема 1. В точке минимума X^0 гладкой функции ($f(X) \in C^{(1)}$) выполняется условие

$$\nabla f(X^0) = 0. \quad (3)$$

Доказательство мы опускаем, так как его можно найти в любом учебнике по математическому анализу. но по поводу гладкости (т.е. непрерывно дифференцируемости) функции хотим отметить, что каковая не излишня в данном контексте. Функция

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

имеет в нуле строгий минимум и равную нулю производную, но в сколь угодно малой окрестности нуля принимает, как положительные, так и отрицательные значения (см. рис. 1).

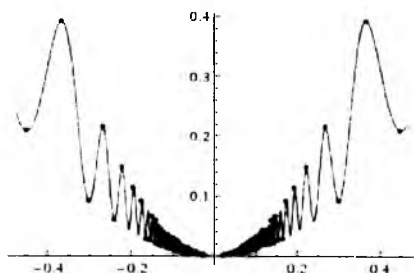


Рис. 1

Условие (5) называется **необходимым условием минимума** первого порядка. Вектор $\nabla f(X^0) = \frac{\partial f(X^0)}{\partial X}$ - принято называть **градиентом функции** $f(X)$ в точке X^0 .

В новых терминах теорема 1 утверждает: в точке локального (относительного) минимума градиент функции равен нулю.

Теорема 1 сводит поиск относительного минимума к решению уравнения

$$\nabla f(X) = 0. \quad (6)$$

Решения уравнения (6) называют стационарными точками функции $f(X)$. Смысл теоремы 1 можно выразить и таким образом: точка минимума функции является стационарной точкой функции. Обратное утверждение, конечно, неверно, ибо уравнению могут удовлетворять и точки максимума и другие точки.

Для того чтобы среди стационарных точек выделить точки минимума, нужно использовать дополнительные условия, например, необходимое условие второго порядка, которое содержится в следующем утверждении.

Теорема 2. Пусть функция $f(X)$ определена, непрерывна вместе с производными первого и второго порядков во всех точках n - мерного пространства R^n . Если X^0 - точка относительного минимума, то в этой точке матрица вторых производных минимизируемой функции неотрицательна:

$$\frac{\partial^2 f(X^0)}{\partial X^2} \geq 0. \quad (5)$$

Доказательство этой теоремы также опускаем.

3.4. Достаточное условие относительного минимума

Из курса математического анализа без доказательства приведем следующее утверждение.

Теорема 3. Для того чтобы стационарная точка X^* была точкой относительного минимума дважды непрерывно дифференцируемой функции

$f(X)$ ($f(X) \in C^{(2)}$), достаточно, чтобы матрица вторых производных функции $f(X)$ в точке X^* была положительной

$$\frac{\partial^2 f(X^*)}{\partial X^2} > 0. \quad (8)$$

Пример 9. Исследовать на экстремум функцию

$$f(x, y, z) = x^2 + 2xy + 4xz + 8yz + 5y^2 + 9z^2.$$

Решение

Найдем стационарные точки функции $f(x, y, z)$. Они определяются из системы уравнений

$$\begin{cases} \frac{\partial f}{\partial x} = 2x + 2y + 4z = 0, \\ \frac{\partial f}{\partial y} = 2x + 10y + 8z = 0, \\ \frac{\partial f}{\partial z} = 4x + 8y + 18z = 0. \end{cases}$$

Определитель этой системы

$$\begin{vmatrix} 2 & 2 & 4 \\ 2 & 10 & 8 \\ 4 & 8 & 18 \end{vmatrix} = 8 \cdot \begin{vmatrix} 1 & 1 & 2 \\ 1 & 5 & 4 \\ 2 & 4 & 9 \end{vmatrix} = 8 \cdot 16 > 0.$$

Поэтому единственное решение однородной системы есть $x = y = z = 0$.

Итак, функция $f(x, y, z)$ имеет единственную стационарную точку $(0, 0, 0)$. Найдем

$$\frac{\partial^2 f(0,0,0)}{\partial X^2} = \begin{pmatrix} \frac{\partial^2 f(0,0,0)}{\partial x^2} & \frac{\partial^2 f(0,0,0)}{\partial x \partial y} & \frac{\partial^2 f(0,0,0)}{\partial x \partial z} \\ \frac{\partial^2 f(0,0,0)}{\partial y \partial x} & \frac{\partial^2 f(0,0,0)}{\partial y^2} & \frac{\partial^2 f(0,0,0)}{\partial y \partial z} \\ \frac{\partial^2 f(0,0,0)}{\partial z \partial x} & \frac{\partial^2 f(0,0,0)}{\partial z \partial y} & \frac{\partial^2 f(0,0,0)}{\partial z^2} \end{pmatrix} = \begin{pmatrix} 2 & 2 & 4 \\ 2 & 10 & 8 \\ 4 & 8 & 18 \end{pmatrix}.$$

Так как

$$D_1 = 2 > 0, \quad D_2 = \begin{vmatrix} 2 & 2 \\ 2 & 10 \end{vmatrix} > 0, \quad D_3 = \begin{vmatrix} 2 & 2 & 4 \\ 2 & 10 & 8 \\ 4 & 8 & 18 \end{vmatrix} = 8 \cdot 16 > 0,$$

по в силу критерия Сильвестра

$$\frac{\partial^2 f(0,0,0)}{\partial X^2} = \begin{pmatrix} 2 & 2 & 4 \\ 2 & 10 & 8 \\ 4 & 8 & 18 \end{pmatrix} > 0.$$

Точка $(0,0,0)$ является точкой локального минимума функции $f(x, y, z)$, причем $f_{\min} = f(0,0,0) = 0$.

Пример 10. Исследовать на экстремум функцию

$$f(x, y, z) = 3x^3 + 6xy + y^2 + z^2 - 2z + 1.$$

Решение

Найдем стационарные точки функции $f(x, y, z)$. Они определяются из системы уравнений

$$\begin{cases} \frac{\partial f}{\partial x} = 9x^2 + 6y = 0, \\ \frac{\partial f}{\partial y} = 6x + 2y = 0, \\ \frac{\partial f}{\partial z} = 2z - 2 = 0. \end{cases}$$

Решив эту систему, найдем стационарные точки $(2, -6, 1)$ и $(0, 0, 1)$. Вычислив частные производные построим матрицу

$$\frac{\partial^2 f(x, y, z)}{\partial X^2} = \begin{pmatrix} \frac{\partial^2 f(x, y, z)}{\partial x^2} & \frac{\partial^2 f(x, y, z)}{\partial x \partial y} & \frac{\partial^2 f(x, y, z)}{\partial x \partial z} \\ \frac{\partial^2 f(x, y, z)}{\partial y \partial x} & \frac{\partial^2 f(x, y, z)}{\partial y^2} & \frac{\partial^2 f(x, y, z)}{\partial y \partial z} \\ \frac{\partial^2 f(x, y, z)}{\partial z \partial x} & \frac{\partial^2 f(x, y, z)}{\partial z \partial y} & \frac{\partial^2 f(x, y, z)}{\partial z^2} \end{pmatrix} = \begin{pmatrix} 18x & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

В точке $(2, -6, 1)$ имеем

$$\frac{\partial^2 f(2, -6, 1)}{\partial X^2} = \begin{pmatrix} 32 & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Так как

$$D_1 = 32 > 0, \quad D_2 = \begin{vmatrix} 32 & 6 \\ 6 & 2 \end{vmatrix} = 36 > 0, \quad D_3 = \begin{vmatrix} 32 & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 72 > 0.$$

то, в силу критерия Сильвестра

$$\frac{\partial^2 f(2, -6, 1)}{\partial X^2} = \begin{pmatrix} 32 & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} > 0.$$

Точка $(2, -6, 1)$ является точкой локального минимума функции $f(x, y, z)$, причем $f_{\min} = f(2, -6, 1) = -12$.

В точке $(0, 0, 1)$ имеем

$$\frac{\partial^2 f(0, 0, 1)}{\partial X^2} = \begin{pmatrix} 0 & 6 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Для исследования функции в точке $(0, 0, 1)$ нельзя использовать критерий Сильвестра, так как $D_1 = 0$. Легко видеть, что в этой точке экстремума нет. В самом деле, $f(0, 0, 1) = 0$, а в столь угодно малой окрестности точки $(0, 0, 1)$ функция принимает как положительные, так и отрицательные значения. Например, $f(0, 0, \varepsilon) > 0$, если $\varepsilon > 0$ и $f(0, 0, \varepsilon) < 0$, если $\varepsilon < 0$.

Если матрица $\frac{\partial^2 f}{\partial X^2}$ вырождена, ситуация принципиально усложняется. В скалярном случае о характере критической точки можно судить по первому ненулевому члену ряда Тейлора. Для функций n переменных дело обстоит совершенно иным образом по особой схеме. Эта схема позволяет свести анализ экстремали функции к исследованию стационарной точки функции меньшего числа переменных – информативной функции. Рас-

смотрение этой схемы исследования вырожденных стационарных точек выходит за рамки данного курса.

Предвидеть трудности (но не их масштаб) довольно легко. Понятно, что в отсутствие знакоопределенности дифференциалов порядка выше второго - поверхности их вырождения могут накладываться друг на друга весьма разнообразно, в результате чего аномалии становятся нормой.

Островок порядка характеризуется невырожденными критическими точками, в которых невырождены соответствующие матрицы Гессе. Классическая лемма Морса гарантирует существование в некоторой окрестности критической точки - локальной системы координат $x_1, x_2, \dots, x_n$ такой что

$$f = x_1^2 + \dots + x_l^2 - x_{l+1}^2 - \dots - x_n^2.$$

Функция такого вида называется **морсовским l -седлом**. В случае $l = 0$ имеем максимум, при $l = n$ - минимум.

Лемма Морса дает по существу полную классификацию невырожденных критических точек. Для **вырожденных критических точек такой классификации нет**. Простой пример вырожденной критической точки - **обезьянье седло** (рис. 2), которое описывается функцией

$$f(x, y) = x^3 - 3xy^2.$$

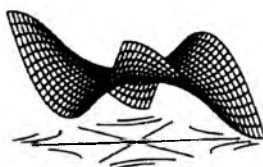


Рис. 2

Упражнения

Исследовать на экстремум функции двух переменных (1-17).

1. $f(x, y) = x^3 + 3xy^2 - 39x - 36y + 26.$
2. $f(x, y) = x^2 + xy + y^2 - 12x - 3y.$
3. $f(x, y) = 3 + 2x - y - x^2 + xy - y^2.$
4. $f(x, y) = 3x + 6y - x^2 - xy + y^2.$
5. $f = 4x^2 - 4xy + y^2 + 4x - 2y + 1.$
6. $f(x, y) = (x + y^2)e^{\frac{x}{2}}.$

7. $f(x, y) = (x^2 - 2y^2)e^{x-y}$.

8. $f(x, y) = (8x^2 - 6xy + 3y^2)e^{2x+3y}$.

9. $f(x, y) = (5 - 2x + y)e^{x^2-y}$.

10. $f(x, y) = \frac{x^3}{3} + 3x^2e^y - e^{-x^2}$.

11. $f(x, y) = (25 - 5x - 7y)e^{-(x^2-y-x^2)}$.

12. $f(x, y) = 108 \ln x - xy^2 + \frac{y^3}{3}$.

13. $f(x, y) = x^2 + y^2 - 32 \ln(xy)$.

14. $f(x, y) = xy \ln(x^2 + y^2)$.

15. $f(x, y) = x + y + 4 \sin x \sin y$.

16. $f(x, y) = (1 + e^y) \cos x - ye^x$.

17. Найти все стационарные точки функции $f(x, y) = x^4 + y^4 - 2x^2$ и исследовать ее на экстремум. Можно ли использовать при этом достаточное условие экстремума?

18. Может ли непрерывная дифференцируемая функция $f(x, y)$ иметь бесконечное множество максимумов и ни одного минимума?

19. Верно ли утверждение: если непрерывно дифференцируемая функция $f(x, y)$, $(x, y) \in R^2$, имеет только одну стационарную точку (x_0, y_0) , в которой у нее локальный минимум, то справедливо неравенство $f(x, y) \geq f(x_0, y_0)$, $(x, y) \in R^2$?

Исследовать на экстремум функции трех переменных (21-30).

21. $f(x, y, z) = 3x^3 + y^2 + z^2 + 6xy - 2z + 1$.

22. $f(x, y, z) = x^2 + y^2 + (z+1)^2 - xy + x$.

23. $f(x, y, z) = 8 - 6x + 4y - 2z - x^2 - y^2 - z^2$.

24. $f(x, y, z) = x^2 + y^2 - z^2 - 4x + 6y - 2z$.

25. $f(x, y, z) = x^3 + y^2 + z^2 + 6xy - 4z$.

26. $f(x, y, z) = zyz(16 - x - y - 2z)$.

3.5. Замечание к вопросу о глобальных экстремумах

Существенный интерес нередко представляет вопрос о глобальных экстремумах. В одномерном случае локальный минимум при отсутствии других стационарных точек является одновременно глобальным минимумом. В общем случае это не так. Вот соответствующий контрпример.

Пример 11. Пусть X_0 - единственная стационарная точка дважды непрерывно дифференцируемой функции $f(X)$, $X \in R^n$, и X_0 реализует его локальный минимум. Будет ли эта точка точкой его глобального минимума? Ответ на этот вопрос отрицателен. Соответствующий пример дает функция

$$f(x, y) = \frac{3x^2 - 2x^3 - 1}{1 + y^2} + (3x^2 - 2x^3)e^{-y}.$$

Эта функция имеет лишь нулевую стационарную точку, которая реализует ее локальный минимум, поскольку матрица

$$\frac{\partial^2 f(0,0)}{\partial X^2} = \begin{pmatrix} 12 & 0 \\ 0 & 2 \end{pmatrix} > 0.$$

Однако $f(0,0) = -1$, $f(2,0) = -9$.

Существуют различные признаки существования глобального минимума (максимума). Они связаны с понятиями, введения которых выходят за рамки данного курса.

Упражнение

Покажите, что если функция $f(X)$ имеет в точке X^* локальный минимум, и $\lim_{\|X\| \rightarrow \infty} \|F(X)\| = \infty$, то в X^* достигается глобальный минимум $f(X)$.

Вопросы для самопроверки

1. Сформулируйте постановку задачи нелинейного программирования в общем случае.
2. Приведите необходимые условия минимума в задаче на безусловный минимум.
3. Приведите достаточное условие минимума в задаче на безусловный минимум.
4. В каком случае для исследования функции на минимум нельзя использовать критерий Сильвестра?
5. Что дает полную классификацию невырожденных критических точек?
6. Существует ли классификация вырожденных критических точек?

§22. Задачи на условный минимум

Ключевые слова: *графический метод, метод исключения, функция Лагранжа, нормальная функция Лагранжа, нормальные задачи, правило множителей Лагранжа, экономическая интерпретация множителей Лагранжа, достаточное условие условного минимума.*

1. Решение задачи нелинейного программирования графическим методом

Если число переменных в задаче нелинейного программирования равно двум, то она может быть решена графическим методом.

Математическая формулировка такой задачи имеет вид

$$z = f(x, y) \rightarrow \max (\min), \quad (*)$$

$$\begin{aligned} g_1(x, y) &\leq b_1, \\ g_2(x, y) &\leq b_2, \\ &\dots\dots\dots \\ g_n(x, y) &\leq b_n, \end{aligned} \quad (**)$$

При решении задачи нелинейного программирования графическим методом сначала надо построить область допустимых решений, т.е. множество точек на плоскости, удовлетворяющих неравенствам (**). Экстремум функции может достигаться как внутри области, так и на границе. Затем следует записать уравнения линий уровня целевой функции, построить графики этих уровней на рисунке и определить направление возрастания (убывания) целевой функции. Далее, перемещая линии уровня в нужном направлении в области допустимых решений, найти точки области, в которых целевая функция принимает оптимальные значения.

Пример 1. Решить задачу графическим методом

$$\begin{aligned} z &= (x-2)^2 + (y-3)^2 \rightarrow \text{extr}, \\ x + y &= 7, \\ 0 &\leq x \leq 5, \\ 0 &\leq y \leq 10. \end{aligned}$$

Решение

Область определения - замкнутая ограниченная (прямоугольник $OABC$). Поэтому глобальные экстремумы, в том числе и условные существуют. Уравнение $x + y = 7$ есть прямая, отрезок которой DE (рис. 1) располагается внутри области.

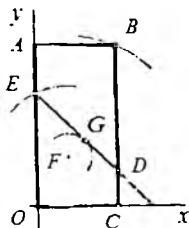


Рис. 1

Следовательно, значения функции должны сравниваться не во всей области $OABC$, а только вдоль этого отрезка DE . На рисунке показаны линии уровня, представляющие собой концентрические окружности с центром в точке $F(2;3)$. Как видно из рисунка, безусловные экстремумы достигаются в точках F (где $z_{\min} = 0$) и B (где $z_{\max} = 58$). При этом первый является одновременно локальным и глобальным минимумом, а второй только глобальным максимумом. Если же рассматривать только точки, лежащие на отрезке DE , то из того же рисунка видно, что условный глобальный максимум достигается в точке $E(0;7)$, где $z_{\max} = 20$, а условный глобальный минимум (он же и локальный) достигается в точке G , в которой окружность касается отрезка DE . Определив координаты этой точки из равенства угловых коэффициентов, получим $x_0 = 3$, $y_0 = 4$ и $z_{\min} = 2$. Ниже решим эту задачу аналитически.

Упражнения

Решить следующие задачи графическим методом:

1. $z = (x-4)^2 + (y-3)^2 \rightarrow \text{extr.}$

$2x + 3y \geq 6,$

$3x - 2y \leq 18,$

$-x + 2y \leq 8,$

$x \geq 0, y \geq 0.$

2. $z = 2x^2 - y^2 \rightarrow \max,$

$x - y \geq 6,$

$y \leq 4,$

$x + y - xy \geq 0,$

$x \geq 0, y \geq 0.$

$$\begin{aligned}
3. \quad & z = y - x^2 \rightarrow \max, \\
& 2x + 3y \leq 24, \\
& x + 2y \leq 15, \\
& 3x + 2y \leq 24, \\
& y \leq 4, \\
& x \geq 0, y \geq 0.
\end{aligned}$$

$$\begin{aligned}
4. \quad & z = (x-3)^2 + (y-4)^2 \rightarrow \text{extr}, \\
& 3x + 2y \geq 7, \\
& 10x - y \leq 8, \\
& -18x + 4y \leq 12, \\
& x \geq 0, y \geq 0.
\end{aligned}$$

$$\begin{aligned}
5. \quad & z = 3x + 4y^2 \rightarrow \max, \\
& x^2 + y^2 \leq 25, \\
& xy \geq 4, \\
& x \geq 0, y \geq 0.
\end{aligned}$$

$$\begin{aligned}
6. \quad & z = 2x + y \rightarrow \text{extr}, \\
& x^2 + y^2 \leq 16, \\
& x \geq 0, y \geq 0.
\end{aligned}$$

2. Решение задачи нелинейного программирования методом исключения

Рассмотрим задачу нелинейного программирования в общем виде

$$f(X) \rightarrow \min, \quad (1)$$

$$g_i(X) = 0, \quad i = 1, 2, \dots, m. \quad (2)$$

В некоторых задачах условия (2) составлены из несложных функций $g_1(X), g_2(X), \dots, g_m(X)$, позволяющих исключить m переменных. Пусть ими будут $x_1, x_2, \dots, x_m$. Тогда подставив значения

$$x_1 = h_1(x_{m+1}, \dots, x_n), x_2 = h_2(x_{m+1}, \dots, x_n), \dots, x_m = h_m(x_{m+1}, \dots, x_n)$$

в $f(X)$, получим новую функцию

$$\psi(x_{m+1}, \dots, x_n) = f(h_1(x_{m+1}, \dots, x_n), \dots, h_m(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n) \quad (3)$$

$n - m$ переменных.

Лемма 1. Если $X^0 = (x_1^0, \dots, x_n^0)$ - точка условного минимума функции $f(X)$, то $(x_{m+1}^0, \dots, x_n^0)$ - точка безусловного минимума функции $\psi(x_{m+1}, \dots, x_n)$ и, наоборот, если $(x_{m+1}^0, \dots, x_n^0)$ - точка безусловного минимума функции $\psi(x_{m+1}, \dots, x_n)$, то $(h_1(x_{m+1}^0, \dots, x_n^0), \dots, h_m(x_{m+1}^0, \dots, x_n^0), x_{m+1}^0, \dots, x_n^0)$ - точка условного минимума функции $f(X)$.

Доказательство опускаем.

Хотя метод исключения часто затруднителен для реализации, все же в тех случаях, где он осуществим, следует им пользоваться. Например, метод исключения реализуется в симплексном методе.

Теоретическая возможность применения метода исключения основана на теоремах о неявных функциях. Приведем простейшую теорему такого рода, которая будет использоваться и в дальнейшем.

Теорема 1 (о неявных функциях). Пусть $A, B, A \in R^m, B \in R^k$ - некоторые точки. Если m -мерная функция $g(Y, Z), Y \in R^m, Z \in R^k$ определена в окрестности точки $(Y, Z) \in R^m \times R^k$, дифференцируема там по Y и удовлетворяет условиям:

$$g(A, B) = 0, \quad \left| \frac{\partial g(A, B)}{\partial Y} \right| \neq 0,$$

то найдется m -мерная функция $Y = h(Z)$, определенная, непрерывная в окрестности точки $Z = B$ такая, что

1) $g(h(Z), Z) \equiv 0$ в окрестности точки $Z = B$;

2) $h(B) = A$;

3) функция $h(Z)$ имеет в окрестности точки $Z = B$ непрерывные производные по Z того же порядка, что и функция $g(Y, Z)$.

Используем теорему о неявных функциях для обоснования метода исключения в случае гладких функций $g_1(X), g_2(X), \dots, g_m(X)$.

Справедливо следующее утверждение:

Лемма 2. Метод исключения в задаче на относительный минимум применим, если в точке $X = X^0$

$$\text{rank} \left(\frac{\partial g_1(X^0)}{\partial X}, \dots, \frac{\partial g_m(X^0)}{\partial X} \right) = m.$$

Пример 2. Решить методом исключения задачу из примера 1.

Решение

Разрешим относительно y уравнения $x + y = 7$: $y = 7 - x$. Подставляя найденное значение в выражения задачи получаем, следующую задачу на на-

хождения наибольшего и наименьшего значения для функции одной переменной

$$\psi(x) = (x-2)^2 + (4-x)^2 = 2x^2 - 12x + 20, \quad 0 \leq x \leq 5.$$

Так как $\psi'(x) = 4x - 12$, $\psi''(x) = 4 > 0$, то функция ψ в точке $x_0 = 3$ имеет минимум $\psi_{\min} = 2$. На границах отрезка $0 \leq x \leq 5$ имеем следующие значения: $\psi(0) = 20$, $\psi(5) = 10$. Сравнивая найденные значения находим, что наименьшее значения функции ψ равно 2. Следовательно, $x_0 = 3$, $y_0 = 4$ и $z_{\min} = 2$, что совпадает с результатом графического решения (см. п.1).

Упражнения

Решить методом исключения следующие задачи:

$$1. f(x, y) = x^2 + y^2 \rightarrow \text{extr},$$

$$x + y = 1.$$

$$2. f(x, y, z) = xyz \rightarrow \text{extr},$$

$$x + y + z = 6,$$

$$x + 2y + 3z = 6.$$

3. Метод множителей Лагранжа

По условиям задачи (1), (2) составим функцию

$$F(X, \bar{\Lambda}) = \lambda_0 f(X) + \Lambda' g(X), \quad (4)$$

где $\bar{\Lambda} = (\lambda_0, \Lambda)$ есть $(m+1)$ - мерный вектор, состоящий из скаляра λ_0 и m - мерного вектора Λ .

Функцию $F(X, \bar{\Lambda})$ называют **функцией Лагранжа** задачи (1), (2), а компоненты $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_m$ вектора $\bar{\Lambda}$ - множителями Лагранжа.

Теорема 2 (правило множителей Лагранжа). Пусть функции

$$f(X), g_1(X), g_2(X), \dots, g_m(X)$$

определены, непрерывны и дифференцируемы на R^n . Если X^0 - точка локального условного минимума, то найдется ненулевой вектор $\bar{\Lambda}$ такой, что

$$\frac{\partial F(X^0, \bar{\Lambda})}{\partial X} = 0.$$

Замечание. Если $\bar{\Lambda}$, удовлетворяет правилу множителей Лагранжа, то, в силу однородности по $\bar{\Lambda}$ функции Лагранжа, правилу множителей удовлетворяет и вектор $-\bar{\Lambda}$. Поэтому знак одной компоненты вектора $\bar{\Lambda}$ можно выбрать заранее. Для определенности задают знак множителя $\lambda_0 : \lambda_0 \geq 0$. В уточненной формулировке правило множителей утверждает о существовании чисел $\lambda_0 > 0, \lambda_1, \dots, \lambda_m$, не равных одновременно нулю, таких что $\frac{\partial F(X^0, \bar{\Lambda})}{\partial X} = 0$.

В приложениях часто используется функция Лагранжа,

$$F(X, \Lambda) = f(X) + \Lambda'g(X), \quad (5)$$

которую будем называть **нормальной функцией Лагранжа**. Функция (5) получается из (4) при $\lambda_0 = 1$. Нормальная функция Лагранжа проще функции Лагранжа, удобнее для вычислений. Однако правило множителей для нее не всегда верно.

Пример 3. $n = 2, m = 1, f(x) = x_1, g(x) = x_1^3 - x_2^2$. Допустимые точки удовлетворяют уравнению $x_1^3 - x_2^2 = 0$ и лежат на полукубической параболе.

Ясно, что точка $x_1^0 = x_2^0 = 0$ есть точка условного минимума. Составим нормальную функцию Лагранжа

$$F(X, \lambda) = f(X) + \lambda g(X) = x_1 + \lambda(x_1^3 - x_2^2).$$

Правило множителей приводит к уравнениям

$$\frac{\partial F}{\partial x_1} = 1 + 3\lambda x_1^2 = 0, \quad \frac{\partial F}{\partial x_2} = -2\lambda x_2 = 0.$$

Точка минимума $x_1^0 = x_2^0 = 0$ не удовлетворяет этим уравнениям, т.е. правило множителей с нормальной функцией Лагранжа в данной задаче не справедливо.

Множители λ_0, Λ , при которых выполняется правило множителей в точке X^0 , назовем множителями, соответствующими точке X^0 . Точке X^0 может соответствовать несколько систем множителей Лагранжа.

4. Нормальные задачи на условный минимум

Рассматриваются задачи на условный минимум, которые не сводятся к перебору конечного числа допустимых точек.

4.1. Нормальные точки минимума и обыкновенные допустимые точки

Если среди систем множителей Лагранжа, соответствующих точке условного минимума X^0 задачи

$$f(X) \rightarrow \min, \quad g(X) = 0,$$

нет множителей $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_m$ с $\lambda_0 = 0$, то точка называется **нормальной точкой минимума**; а сама задача - **нормальной задачей на условный минимум**.

Если X^0 - нормальная точка минимума, то соответствующие ей множители можно разделить на $\lambda_0 \neq 0$ и считать, что множители Лагранжа имеют вид $\lambda_0 = 1, \lambda_1, \lambda_2, \dots, \lambda_m$.

Для нормальной точки соответствующие множители $1, \lambda_1, \lambda_2, \dots, \lambda_m$ существуют и определяются единственным образом.

Действительно, если допустить еще одну систему множителей $1, \bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_m$, ($\bar{\lambda}_i \neq \lambda_i$), то получаем

$$\begin{aligned} \frac{\partial f(X^0)}{\partial X} + \lambda_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \lambda_m \frac{\partial g_m(X^0)}{\partial X} &= 0, \\ \frac{\partial f(X^0)}{\partial X} + \bar{\lambda}_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \bar{\lambda}_m \frac{\partial g_m(X^0)}{\partial X} &= 0. \end{aligned}$$

Вычитая одно уравнение из другого, приходим к равенству

$$\mu_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \mu_m \frac{\partial g_m(X^0)}{\partial X} = 0,$$

где $\mu_i = \lambda_i - \bar{\lambda}_i \neq 0$. Полученное равенство означает, что точке X^0 соответствует еще одна система множителей $\mu_0 = 0, \mu_1, \dots, \mu_m$, и это противоречит определению нормальности точки X^0 . Следовательно, система множителей Лагранжа $1, \lambda_1, \lambda_2, \dots, \lambda_m$ единственна. Оказывается, характер точки минимума (нормальна она или нет) определяется полностью ограничениями задачи.

Точку $\tilde{X}$ назовем **обыкновенной допустимой точкой**, если $g(\tilde{X})=0$ и векторы

$$\frac{\partial g_1(\tilde{X})}{\partial X}, \frac{\partial g_2(\tilde{X})}{\partial X}, \dots, \frac{\partial g_m(\tilde{X})}{\partial X},$$

вычисленные в этой точке, линейно независимы.

Справедлива следующая теорема.

Теорема 3. Точка X^0 является нормальной в том и только в том случае, когда она обыкновенная допустимая точка.

Доказательство. Необходимость. Пусть точка X^0 - нормальная:

$$\frac{\partial f(X^0)}{\partial X} + \lambda_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \lambda_m \frac{\partial g_m(X^0)}{\partial X} = 0,$$

но не является обыкновенной, т.е. для некоторого вектора $M = (\mu_1, \dots, \mu_m)$, $M \neq 0$, выполняется равенство

$$\mu_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \mu_m \frac{\partial g_m(X^0)}{\partial X} = 0.$$

Отсюда следует, что точке X^0 соответствует система множителей Лагранжа $\mu_0 = 0, \mu_1, \dots, \mu_m = 0$, чего не может быть по определению нормальной точки X^0 .

Необходимость доказана.

Достаточность. Пусть допустимая точка X^0 обыкновенная. Если предположить, что она не является нормальной, то найдутся множители $\lambda_0 = 0, \lambda_1, \lambda_2, \dots, \lambda_m$; $\Lambda \neq 0$ такие, что

$$\lambda_0 \frac{\partial f(X^0)}{\partial X} + \lambda_1 \frac{\partial g_1(X^0)}{\partial X} + \dots + \lambda_m \frac{\partial g_m(X^0)}{\partial X} = 0.$$

Поскольку первое слагаемое слева равно нулю, то оставшаяся часть равенства означает, что векторы $\frac{\partial g_1(X^0)}{\partial X}, \frac{\partial g_2(X^0)}{\partial X}, \dots, \frac{\partial g_m(X^0)}{\partial X}$ линейно зависимы.

Это противоречит определению обыкновенной точки. Теорема доказана.

4.2. Правило множителей для нормальной задачи на условный минимум

Как следует из определения нормальных задач на условный минимум, при их исследовании достаточно пользоваться нормальными функциями Лагранжа. В этом случае правило множителей верно. Более того, оно упрощается и может быть записано в симметричном виде.

Теорема 4. Если точка X^0 доставляет решение нормальной задаче на условный минимум, то существует m -мерный вектор Λ такой, что выполняются равенства

$$\frac{\partial F(X^0, \Lambda)}{\partial X} = 0, \quad \frac{\partial F(X^0, \Lambda)}{\partial \Lambda} = 0, \quad (6)$$

где $F(X, \Lambda) = f(X) + \Lambda'g(X)$.

Для доказательства утверждения достаточно заметить, что первое равенство в (6) есть следствие правила множителей для рассматриваемого нормального случая, а второе эквивалентно условию $g(X^0) = 0$, которое должно, очевидно, выполняться для допустимой точки X^0 .

Соотношения (6) представляют собой систему из $n + m$ уравнений относительно $n + m$ неизвестных X^0, Λ . В аномальных задачах правило множителей сводит исходную задачу к решению $n + m$ уравнений относительно $n + m + 1$ неизвестных X^0, λ_0, Λ .

Если сравнить метод исключения с правилом множителей, то видно, что в первом методе исходная задача на условный минимум функции n переменных сводится к задаче безусловной минимизации функции $n - m$ переменных, во втором методе исходная задача сводится к условиям стационарности функции Лагранжа от $n + m$ аргументов. Следует подчеркнуть, что точка условного минимума X^0 , вообще говоря, не является точкой минимума функции Лагранжа.

Пример 4. $n = 2, m = 1, f(X) = x_2, g(X) = -x_1^2 + x_2$. В обыкновенной точке $x_1^0 = x_2^0 = 0$ достигается условный минимум. Правило множителей приводит к уравнениям

$$\frac{\partial F}{\partial x_1} = -2\lambda x_1 = 0, \quad \frac{\partial F}{\partial x_2} = 1 + \lambda = 0,$$

откуда $\lambda = -1$. Функция Лагранжа $F(X, -1) = x_1^2$ в точке $x_1 = 0$, соответствующей точке условного минимума, также достигает минимума.

Пример 5. $n = 2, \quad m = 1. \quad f(X) = x_1^3, \quad g(X) = -x_1^2 + x_2$. Поскольку

$$\frac{\partial g}{\partial x_1} = -2x_1, \quad \frac{\partial g}{\partial x_2} = 1,$$

то

$$\frac{\partial g}{\partial X} \neq 0$$

и любая допустимая точка – обыкновенная, т.е. задача нормальная и

$$F(X, \lambda) = x_2^3 + \lambda(x_2 - x_1^2).$$

Правило множителей приводит к уравнениям

$$-2\lambda x_1 = 0, \quad 3x_2^2 + \lambda = 0, \quad x_2 - x_1^2 = 0.$$

Точке условного минимума $x_1^0 = x_2^0 = 0$ соответствует $\lambda = 0$. Функция Лагранжа с этим значением $F(X, 0) = x_2^3$ в точке $x_2^0 = 0$ не достигает минимума.

Точка $x_2^0 = 0$ является точкой перегиба функции.

Пример 6. $n = 2, \quad m = 1, \quad f(X) = -x_2^2, \quad g(X) = x_2$. Точка $x_1^0 = x_2^0 = 0$ является нормальной точкой условного минимума. Правило множителей для функции $F(X, \lambda) = -x_2^2 + \lambda x_2$ приводит к уравнениям

$$\frac{\partial F}{\partial x_1} = 0, \quad \frac{\partial F}{\partial x_2} = -2x_2^0 + \lambda = 0,$$

из которых следует, что $\lambda = 0$. При этом значении множителя функция Лагранжа $F(X, 0) = -x_2^2$ достигает не минимума, а максимума.

4.3. Лемма о включении

Место нормальных задач в теории минимизации функций с дополнительными ограничениями определяется тем, что нормальные задачи на условный минимум не могут быть тривиальными, т.е. множество их допустимых векторов состоит не из конечного числа точек, а из бесконечного. Это следует из ме-

тогда исключения. Такой же смысл можно придать, и следующей лемме, ярко отражающей существо классических методов исследования задач на минимум.

Лемма 3. Пусть X^* - обыкновенная допустимая точка. $g(X^*) = 0$, где $g(X)$ - гладкая функция. Тогда для n - мерного вектора Y , лежащего в гиперплоскости

$$\frac{\partial g'(X^*)}{\partial X} Y = 0, \quad (7)$$

найдется n - мерная функция $h(\beta)$ такая, что в окрестности точки $\beta = 0$ выполняется тождество

$$g(h(\beta)) \equiv 0$$

и

$$h(0) = X^*, h'(0) = Y \quad \left(h'(0) = \frac{dh(0)}{d\beta} \right).$$

Замечание. Найденная функция $h(\beta)$ по теореме о неявных функциях дифференцируема столько раз, сколько раз дифференцируема функция $g(X)$.

Геометрический смысл леммы можно пояснить следующим образом. Совокупность уравнений $g_1(X) = 0, \dots, g_m(X) = 0$, $m < n$, задает $(m - n)$ -мерное многообразие. В частности, при $m - n = 2$ это будет поверхность.

Уравнение

$$\frac{\partial g'(X^*)}{\partial X} Y = 0$$

задает касательную гиперплоскость к этой поверхности в точке $X = X^*$. Лемма о включении утверждает, что какой бы вектор Y ни взять из касательной плоскости, на поверхности можно провести линию, которая исходит из точки $X = X^*$ с касательной, содержащей вектор Y .

По доказанной лемме, точку минимума нормальной задачи можно погрузить в семейство допустимых точек, зависящее от параметров β, Y . Поэтому нормальная задача на условный минимум не может оказаться тривиальной. Анормальные же задачи могут быть и тривиальными.

Пример 7. $n = 4$, $m = 1$, $f(X) = x_1$, $g(X) = x_1^2 + x_2^2 = 0$. Множество допустимых точек состоит лишь из одной точки $x_1 = x_2 = 0$, поэтому задача минимизации вырождается: минимизировать функцию $f(X)$ на других точках не имеет смысла.

4.4. Экономическая интерпретация множителей Лагранжа

В конкретных прикладных вопросах множители Лагранжа имеют содержательную интерпретацию. Так, в механике множители Лагранжа задают реакции связей, а в экономике — **цены на продукты производства**.

Ниже будем обсуждать экономическую интерпретацию множителей Лагранжа.

Пример 8. Пусть производственные функции каждого из двух продуктов зависят от двух (одних и тех же) факторов. Суммарное количество каждого фактора фиксировано. Пусть заданы цены продуктов. При каких условиях доход от выпуска будет максимальным?

Обозначим через x_1 , x_2 объемы выпуска первого и второго продуктов, соответственно. Пусть x_3 , x_4 — объемы факторов, использованные при производстве первого продукта с производственной функцией $x_1 = \varphi(x_3, x_4)$. Аналогично, $x_2 = \psi(x_5, x_6)$. Предполагается, что x_3 и x_5 представляют один и тот же фактор, так же как и x_4 и x_6 . Задача состоит в следующем:

$$\begin{aligned} f(X) &= p_1 x_1 + p_2 x_2 \rightarrow \max, \\ x_1 - \varphi(x_3, x_4) &= 0, \quad x_3 + x_5 - k_1 = 0, \\ x_2 - \psi(x_5, x_6) &= 0, \quad x_4 + x_6 - k_2 = 0. \end{aligned}$$

Функция Лагранжа в этом случае запишется в виде

$$\begin{aligned} F(X, \Lambda) &= \\ &= p_1 x_1 + p_2 x_2 - \lambda_1 (x_1 - \varphi(x_3, x_4)) - \lambda_2 (x_2 - \psi(x_5, x_6)) - \lambda_3 (x_3 + x_5 - k_1) - \lambda_4 (x_4 + x_6 - k_2). \end{aligned}$$

Приравнявая частные производные нулю, получим шесть уравнений:

$$p_1 = x_1, \quad p_2 = x_2, \quad \lambda_1 \frac{\partial \varphi}{\partial x_3} = \lambda_3, \quad \lambda_1 \frac{\partial \varphi}{\partial x_4} = \lambda_4, \quad \lambda_2 \frac{\partial \psi}{\partial x_5} = \lambda_3, \quad \lambda_2 \frac{\partial \psi}{\partial x_6} = \lambda_4.$$

Отсюда легко получить условия на выпуск, максимизирующий доход:

$$p_1 \frac{\partial \varphi}{\partial x_3} = p_2 \frac{\partial \psi}{\partial x_3} = \lambda_3, \quad p_1 \frac{\partial \varphi}{\partial x_4} = p_2 \frac{\partial \psi}{\partial x_4} = \lambda_4.$$

Следовательно, стоимость предельного продукта по каждому фактору будет одна и та же в обеих отраслях. Заметим, что все множители Лагранжа оказались равными ценам. Если в приведенном случае имеет место конкурентное ценообразование, то λ_1, λ_2 - цены продуктов, а λ_3, λ_4 - цены факторов.

Таким образом, в примере, приведенном выше, множители Лагранжа оказались равными ценам. Такими же свойствами, как мы помним, обладают двойственные переменные в теории линейного программирования.

Ниже будет предложена формальная интерпретация множителей Лагранжа.

Рассмотрим стандартную задачу условной оптимизации. Решим ее с помощью метода множителей Лагранжа, который дает оптимальные векторы X^*, Λ^* . Пусть i -е ограничение имеет вид $g_i(x) = b_i$.

Первоначально полагалось, что $b_i = 0$. Исследуем здесь влияние малого ослабления ограничения.

Обозначим через V^* оптимальное значение целевой функции ($V^* = f(X^*)$). Малое ослабление i -го ограничения приводит к малым изменениям оптимальных значений переменных. Однако предполагается, что условия оптимальности по-прежнему удовлетворяются, так что новое состояние, достигаемое в результате ослабления ограничений, также оптимально. Влияние ослабления на оптимальное значение целевой функции определяется формулой

$$\frac{\partial V^*}{\partial b_i} = \sum_j \frac{\partial f(x^*)}{\partial x_j} \frac{\partial x_j}{\partial b_i}. \quad (8)$$

Из ограничений имеем

$$\sum_j \frac{\partial g_i(x^*)}{\partial x_j} \frac{\partial x_j}{\partial b_i} = \begin{cases} 0, & k \neq i, \\ 1, & k = i. \end{cases} \quad (9)$$

Умножим k -е равенство в (9) на λ_k^* и просуммируем по k . Получим

$$\sum_k \sum_j \lambda_k^* \frac{\partial g_j(x^*)}{\partial x_j} \frac{\partial x_j}{\partial b_i} = \lambda_i^*.$$

Вычтем это выражение из (8). Получим

$$\frac{\partial V^*}{\partial b_i} = \lambda_i^* + \sum_j \left[\frac{\partial f(x^*)}{\partial x_j} - \sum_k \lambda_k^* \frac{\partial g_j(x^*)}{\partial x_j} \right] \frac{\partial x_j}{\partial b_i}.$$

Выражение справа в скобках в силу условия оптимальности равно нулю, поэтому

$$\frac{\partial V^*}{\partial b_i} = \lambda_i^*.$$

Таким образом, λ_i^* соответствует маргинальной (предельной) скорости изменения целевой функции относительно малого ослабления i -го ограничения при условии, что все остальные ограничения неизменны. Эта интерпретация аналогична интерпретации двойственных переменных в теории линейного программирования.

В типичных экономических приложениях ограничения могут задаваться лимитами на ресурсы, а целевая функция - некоторым индексом общественного благосостояния. Тогда оптимальные множители Лагранжа соответствовали бы маргинальным (предельным) общественным оценкам ресурсов. В примере, рассмотренном выше, множители соответствуют ценам на продукты и на факторы. Цены на факторы соответствуют предельным оценкам для фиксированного предложения факторов. Чему же соответствуют цены на продукты? Оказывается, они соответствуют маргинальным оценкам производственных функций, выступающих в качестве ограничений, или параметрам эффективности производственных функций.

Таким образом, множители Лагранжа в общей задаче оптимизации, играют роль двойственных переменных в линейном программировании, и сводится к ним, если общая задача линейна.

В экономических задачах можно, поэтому интерпретировать множители Лагранжа так же, как двойственные переменные.

5. Необходимое условие второго порядка. Достаточное условие минимума

До сих пор в задачах нелинейного программирования рассматривались необходимые условия первого порядка, которые используют лишь значения первых производных. Этой информации может оказаться недостаточно для различения точек, не являющихся точками минимума. Ниже приводятся более тонкие условия минимума, основанные на свойствах вторых производных.

5.1. Необходимое условие второго порядка

Теорема 5. Пусть функции $f(X)$, $g(X)$ определены, непрерывны и дважды дифференцируемы в окрестности точки $X = X^0$. Если X^0 - нормальная точка условного минимума, а Λ - соответствующий ей множитель Лагранжа, то квадратичная форма

$$Y' \frac{\partial^2 f(X^0, \Lambda)}{\partial X^2} Y$$

знакоположительна на гиперплоскости

$$\frac{\partial g'(X^0)}{\partial X} Y = 0, \quad (10)$$

т.е. для всех Y , удовлетворяющих уравнению (16), выполняется неравенство

$$Y' \frac{\partial^2 f(X^0, \Lambda)}{\partial X^2} Y \geq 0.$$

Доказательство опускаем.

Без предположения о нормальности точки X^0 теорема 7 может вырождаться.

Пример 9. $n = 2$, $m = 1$, $f(X) = x_1^4 + x_2^4$, $g(X) = x_1 x_2$. Допустимые точки заполняют оси координат $x_1 = 0$, $x_2 = 0$. Точка $x_1^0 = 0$, $x_2^0 = 0$ является точкой условного минимума. Правило множителей приводит к уравнениям

$$\frac{\partial F(X^0, \bar{\Lambda})}{\partial x_1} = 4\lambda_0 (x_1^0)^3 + \lambda x_2^3 = 0, \quad \frac{\partial F(X^0, \bar{\Lambda})}{\partial x_2} = 4\lambda_0 (x_2^0)^3 + \lambda x_1^3 = 0,$$

из которых следует, что любые числа λ_0, λ являются множителями, соответствующими точке минимума. Матрица вторых производных функции Лагранжа имеет вид

$$\frac{\partial^2 F(X^0, \Lambda)}{\partial X^2} = \begin{pmatrix} 12\lambda_0(x_1^0)^2 & \lambda \\ \lambda & 12\lambda_0(x_2^0)^2 \end{pmatrix} = \begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix}.$$

Уравнению

$$\frac{\partial g'(X^0)}{\partial X} Y = 0 \cdot y_1 + 0 \cdot y_2 = 0$$

удовлетворяют все точки плоскости. Квадратичная форма $2\lambda y_1 y_2$ неотрицательна лишь при $\lambda = 0$.

5.2. Достаточное условие минимума

Допустимую точку X^* назовем условно - стационарной точкой функции $f(X)$, если найдется m - мерный вектор Λ такой, что

$$\frac{\partial F(X^*, \Lambda)}{\partial X} = 0,$$

где $F(X, \Lambda) = f(X) + \Lambda'g(X)$.

Теорема 6. Пусть функции $f(X), g(X)$ определены, непрерывны и дважды дифференцируемы в окрестности точки $X = X^*$. Для того чтобы условно - стационарная точка была точкой относительного условного минимума, достаточно, чтобы квадратичная форма

$$Y' \frac{\partial^2 f(X^*, \Lambda)}{\partial X^2} Y$$

была определенно положительной на гиперплоскости

$$\frac{\partial g'(X^*)}{\partial X} Y = 0. \quad (17)$$

Замечание. Из доказательства теоремы следует, что X^* - точка строгого относительного условного минимума, т.е. найдется такое ε_2 , что для всех точек $X, X \neq X^*, X \in B_\varepsilon, 0 < \varepsilon < \varepsilon_2$, выполняется неравенство $f(X^*) < f(X)$.

Пример 10. Найти параметры цилиндрической цистерны, которая при заданной площади поверхности S_0 имеет объем.

Решение. Площадь поверхности цистерны равна $S = 2\pi x_2^2 + 2\pi x_1 x_2$. Объем цистерны равен $V = -\pi x_1 x_2^2$. Введя функцию

$$g(x_1, x_2) = \pi x_2^2 + 2\pi x_1 x_2 - S_0,$$

исходную задачу сведем к задаче

$$f(X) \rightarrow \min, \quad g(X) = 0. \quad (18)$$

Полученная задача не эквивалентна исходной, ибо не учтены дополнительные ограничения

$$x_1 \geq 0, x_2 \geq 0. \quad (19)$$

вытекающие из физической сущности этих переменных.

Однако будем решать задачу в форме (18), в случае необходимости отбрасывая решения, не удовлетворяющие ограничениям (19). Если задача (18) имеет решение, то искомое, очевидно, будет среди локальных минимумов задачи (18). Но без учета (19) задача (18) не имеет решения, ибо при $x_1 \rightarrow \infty, x_2 \rightarrow -\infty$ имеем $f(X) \rightarrow -\infty$. Значит, доказанные необходимые условия минимума нельзя применить. И все же имеющегося материала достаточно для решения поставленной задачи. Для этого нужно найти условно-стационарные точки и к ним применить достаточное условие минимума.

Составляем нормальную функцию Лагранжа

$$F(X, \lambda) = -\pi x_1 x_2^2 + \lambda[2\pi x_2^2 + 2\pi x_1 x_2 - S_0].$$

Условно-стационарные точки удовлетворяют уравнениям:

$$\frac{\partial F}{\partial x_1} = -2\pi x_2^2 + 2\pi \lambda x_2 = 0,$$

$$\frac{\partial F}{\partial x_2} = -2\pi x_1 x_2 + 4\pi \lambda x_2 + 2\pi \lambda x_1 = 0,$$

$$2\pi x_2^2 + 2\pi x_1 x_2 - S_0 = 0.$$

Отсюда

$$x_1 = 4\lambda, \quad x_2 = 2\lambda, \quad \lambda = \pm \sqrt{\frac{S_0}{24\pi}}.$$

Выберем $\lambda > 0$. Подсчитаем матрицу

$$\frac{\partial^2 F}{\partial X^2} \begin{pmatrix} 0 & -2\pi x_2 + 2\pi\lambda \\ -2\pi x_2 + 2\pi\lambda & -2\pi x_1 + 4\pi\lambda \end{pmatrix} = \begin{pmatrix} 0 & -2\pi\lambda \\ -2\pi\lambda & -4\pi\lambda \end{pmatrix}.$$

Уравнение гиперплоскости, на которой проверяется квадратичная форма:

$$\frac{\partial g}{\partial x_1} y_1 + \frac{\partial g}{\partial x_2} y_2 = 2\pi x_2 y_1 + [4\pi x_2 + 2\pi x_1] y_2 = 4\pi\lambda y_1 + 16\pi\lambda y_2 = 0. \quad (20)$$

Квадратичная форма с матрицей $\frac{\partial^2 F}{\partial X^2}$ имеет вид $-4\pi\lambda y_1 y_2 - 4\pi\lambda y_2^2$

и на гиперплоскости (20) переходит в выражение $12\pi\lambda y_2^2$, которое положительно при $y_2 \neq 0$. Таким образом, условия теоремы 6 для точки

$$x_1 = 2\sqrt{\frac{S_0}{6\pi}}, \quad x_2 = \sqrt{\frac{S_0}{6\pi}}$$

выполнены. Поэтому цистерна с такими параметрами, по крайней мере, среди цистерн с близкими параметрами имеет наибольший объем.

Упражнения

Решить следующие задачи:

1. $f(x_1, x_2, x_3) = x_1 - 2x_2 + 2x_3 \rightarrow \min,$

$$x_1^2 + x_2^2 + x_3^2 = 1.$$

2. $f(x_1, x_2, x_3) = x_1 x_2 x_3 \rightarrow \min,$

$$x_1^2 + x_2^2 + x_3^2 = 1.$$

$$x_1 + x_2 + x_3 = 0.$$

3. $f(x_1, x_2) = x_1^2 + 12x_1 x_2 + 2x_2^2 \rightarrow \min.$

$$4x_1^2 + x_2^2 = 25.$$

4. $f(x_1, x_2) = e^{ax_1 x_2} \rightarrow \min, \quad (a < 0),$

$$x_1^5 + x_2^5 + x_1 + x_2 - 4 = 0.$$

5. Мукомольный комбинат реализует муку двумя способами: в розницу через магазин и оптом через торговых агентов. При продаже x кг муки через

магазин расходы на реализацию составляют x^2 ден. ед., а при продаже y кг муки через торговых агентов - y^2 ден. ед. В сутки для продажи выделяется 5000 кг муки. Сколько кг муки надо продавать каждым способом при минимальных затратах? Задачу решить методом Лагранжа.

6. **Задача Шварца.** В остроугольный треугольник надо вписать треугольник минимального периметра с вершинами, лежащими на сторонах исходного.

7. **Задача Евклида.** В треугольник вписать параллелограмм наибольшей площади.

Вопросы для самопроверки

1. В чем состоит суть решения задачи на условный минимум методом исключения?
2. На чем основана возможность применения метода исключения?
3. Приведите определение функции Лагранжа. Сформулируйте правило множителей Лагранжа.
4. Приведите определение нормальной функции Лагранжа. В чем преимущества нормальной функции Лагранжа от функции Лагранжа?
5. Дайте определения нормальной точки минимума и нормальной задачи на условный минимум.
6. Дайте определения обыкновенной допустимой точки.
7. Приведите критерий нормальности задачи на условный минимум.
8. Сформулируйте правило множителей Лагранжа для нормальной задачи на условный минимум.
9. Дайте экономическую интерпретацию множителей Лагранжа.
10. Приведите необходимое условие второго порядка минимума в задаче на условный минимум.
11. Приведите достаточное условие второго порядка минимума в задаче на условный минимум.

§23. Выпуклое программирование

Ключевые слова: выпуклое программирование, теорема Куна-Таккера, условие Слейтера, сепарабельная функция, сепарабельное программирование.

Раздел нелинейного программирования, в котором задачи минимизации формулируются в терминах выпуклых функций и выпуклых множеств, называется **выпуклым программированием**.

1. Необходимые сведения из анализа.

1.1. Выпуклые множества

Основные сведения о выпуклых множествах приведены нами ранее в §11. Здесь приведем лишь определения выпуклого множества.

Определение. Множество $K \subset R^n$ называется выпуклым, если оно наряду с любыми двумя своими точками содержит соединяющий их отрезок. Другими словами, если $X^1, X^2 \in K$, то $X(\lambda) = \lambda X^1 + (1 - \lambda)X^2 \in K$ для любого λ , $0 \leq \lambda \leq 1$.

1.2. Выпуклые функции

Определение. Функция $f(X)$, определенная на выпуклом множестве K , называется **выпуклой**, если для любых $X^1, X^2 \in K$ и любого λ , $0 \leq \lambda \leq 1$, выполняется неравенство

$$f(\lambda X^1 + (1 - \lambda)X^2) \leq \lambda f(X^1) + (1 - \lambda)f(X^2). \quad (1)$$

Выпуклая функция $f(X)$, $X \in K$, называется **строго выпуклой**, если неравенство

$$f(\lambda X^1 + (1 - \lambda)X^2) < \lambda f(X^1) + (1 - \lambda)f(X^2)$$

выполняется для каждой двух точек $X^1, X^2 \in K$, $X^1 \neq X^2$ и числа λ , $0 < \lambda < 1$. Иными словами, график строго выпуклой функции не содержит прямолинейных отрезков.

Упражнение

Показать, что линейная функция $f(X) = c_1x_1 + c_2x_2 + \dots + c_nx_n + c_0$ являясь выпуклой функцией в пространстве R^n .

Лемма 1. Если $f(X)$ - выпуклая функция, то множество

$$K = \{X : f(X) \leq c\}$$

при любом c выпукло (или пусто).

Доказательство. Действительно, пусть $X^1, X^2 \in K$. λ , $0 \leq \lambda \leq 1$. Тогда $f(X^1) \leq c$, $f(X^2) \leq c$ и в силу выпуклости функции $f(X)$

$$f(\lambda X^1 + (1-\lambda)X^2) \leq \lambda f(X^1) + (1-\lambda)f(X^2) \leq \lambda c + (1-\lambda)c = c,$$

т.е. точка $\lambda X^1 + (1-\lambda)X^2$ принадлежит множеству K . Поскольку X^1, X^2 , λ выбирались произвольными в пределах ограничений, то полученный факт означает выпуклость множества K . Доказанное свойство не полностью характеризует связь между выпуклыми множествами и выпуклыми функциями. Она полностью раскрывается в следующем утверждении.

Теорема 1. Для того чтобы функция $f(X)$, определенная на выпуклом множестве K , была выпуклой, необходимо и достаточно, чтобы было выпуклым множество

$$\bar{K} = \{X, Y : X \in K, Y \geq f(X)\}.$$

На геометрическом языке: функция $f(X)$, определенная на выпуклом множестве K , выпукла тогда и только тогда, когда ее надграфик представлял выпуклое множество.

Доказательство. Необходимость. Пусть $f(X)$, $X \in K$ - выпуклая функция. Возьмем две точки X^1, X^2 множества $\bar{K}$:

$$\bar{X}^1 = \begin{pmatrix} X^1 \\ Y^1 \end{pmatrix}, \quad \bar{X}^2 = \begin{pmatrix} X^2 \\ Y^2 \end{pmatrix},$$

где $X^1, X^2 \in K$, $Y^1 \geq f(X^1)$, $Y^2 \geq f(X^2)$.

Составим отрезок

$$\bar{X}(\lambda) = \lambda \bar{X}^1 + (1-\lambda) \bar{X}^2 = \lambda \begin{pmatrix} X^1 \\ Y^1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} X^2 \\ Y^2 \end{pmatrix} = \begin{pmatrix} \lambda X^1 + (1-\lambda) X^2 \\ \lambda Y^1 + (1-\lambda) Y^2 \end{pmatrix}, \quad 0 < \lambda < 1.$$

Поскольку K - выпуклое множество, то $\lambda X^1 + (1-\lambda) X^2 \in K$. Далее, в силу определения множества $\bar{K}$ и выпуклости функции $f(X)$

$$\lambda Y^1 + (1-\lambda) Y^2 \geq \lambda f(X^1) + (1-\lambda) f(X^2) \geq f(\lambda X^1 + (1-\lambda) X^2).$$

т.е. $\bar{X}(\lambda)$ принадлежит множеству $\bar{K}$. Необходимость доказана.

Достаточность. Пусть множество $\bar{K}$ выпукло. Рассмотрим две его точки:

$$\bar{X}^1 = \begin{pmatrix} X^1 \\ f(X^1) \end{pmatrix}, \quad \bar{X}^2 = \begin{pmatrix} X^2 \\ f(X^2) \end{pmatrix}.$$

При любом λ , $0 \leq \lambda \leq 1$,

$$\lambda \bar{X}^1 + (1-\lambda) \bar{X}^2 = \lambda \begin{pmatrix} X^1 \\ Y^1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} X^2 \\ Y^2 \end{pmatrix} = \begin{pmatrix} \lambda X^1 + (1-\lambda) X^2 \\ \lambda Y^1 + (1-\lambda) Y^2 \end{pmatrix}, \quad 0 < \lambda < 1.$$

По определению множества $\bar{K}$ это означает, что

$$f(\lambda X^1 + (1-\lambda) X^2) \leq \lambda f(X^1) + (1-\lambda) f(X^2).$$

Но это неравенство было положено в определение выпуклой функции. Теорема доказана.

Функция $f(X)$, определенная на выпуклом множестве K , называется вогнутой, если функция $g(X) = -f(X)$, $X \in K$ выпукла. Основные свойства вогнутых функций непосредственно следуют из свойств выпуклых функций.

1.2. Свойства выпуклых функций

Перечислим свойства, которые понадобятся в дальнейшем. Большинство из них доказывается в курсах анализа.

1. Выпуклая функция $f(X)$, непрерывна во всех относительно внутренних точках множества K . Это значит, что выпуклая функция может иметь разрыв лишь на границе множества K .

2. Если функции $f_i(X)$, $i = 1, 2, \dots, m$; $X \in K$, выпуклы, то при любых $\alpha_i \geq 0$ выпуклы и функции

$$f(X) = \sum_{i=1}^m \alpha_i f_i(X), \quad (2)$$

$$f(X) = \max_{1 \leq i \leq m} f_i(X). \quad (3)$$

Докажем второе свойство. Для функции (2)

$$\begin{aligned} f(\lambda X^1 + (1-\lambda)X^2) &= \sum_{i=1}^m \alpha_i f_i(\lambda X^1 + (1-\lambda)X^2) \leq \sum_{i=1}^m \alpha_i [\lambda f_i(X^1) + (1-\lambda)f_i(X^2)] = \\ &= \lambda \sum_{i=1}^m \alpha_i f_i(X^1) + (1-\lambda) \sum_{i=1}^m \alpha_i f_i(X^2) = \lambda f(X^1) + (1-\lambda)f(X^2), \end{aligned}$$

т. е. функция (2) выпукла. Аналогично для (3)

$$\begin{aligned} f(\lambda X^1 + (1-\lambda)X^2) &= \max_{1 \leq i \leq m} f_i(\lambda X^1 + (1-\lambda)X^2) \leq \max_{1 \leq i \leq m} [\lambda f_i(X^1) + (1-\lambda)f_i(X^2)] \leq \\ &\leq \lambda \max_{1 \leq i \leq m} f_i(X^1) + (1-\lambda) \max_{1 \leq i \leq m} f_i(X^2) = \lambda f(X^1) + (1-\lambda)f(X^2). \end{aligned}$$

3. Выпуклая функция $f(X)$ не может иметь двух различных локальных минимумов (локальный минимум является и глобальным); у строго выпуклой функции минимум может достигаться в единственной точке.

Действительно, если $X^1, X^2 \in K$ - точки локального минимума выпуклой функции $f(X)$ и $f(X^1) > f(X^2)$, то для точки $X(\lambda) = \lambda X^1 + (1-\lambda)X^2$, $0 < \lambda < 1$

$$f(\lambda X^1 + (1-\lambda)X^2) \leq \lambda f(X^1) + (1-\lambda)f(X^2) \leq \lambda f(X^1) + (1-\lambda)f(X^1) = f(X^1).$$

(4)

При достаточно малых $\lambda, \lambda > 0$, точка $X(\lambda)$ попадает в сколь угодно малую окрестность точки X^1 и поэтому соотношение (4) противоречит тому, что X^1 - точка локального минимума. Значит, $f(X^1) = f(X^2)$.

Предположим, что у строго выпуклой функции $f(X)$ имеются две разные точки минимума $X^1, X^2 \in K$, $X^1 \neq X^2$, $f(X^1) \neq f(X^2)$. Тогда при $0 < \lambda < 1$ имеем соотношение

$$f(\lambda X^1 + (1-\lambda)X^2) < \lambda f(X^1) + (1-\lambda)f(X^2) = \min_{X \in K} f(X).$$

Значит, в точке $\lambda X^1 + (1-\lambda)X^2 \in K$ функция $f(X)$ принимает значение, меньшее, чем $\min_{X \in K} f(X)$. Противоречие доказывает утверждение: $X^1 = X^2$.

Доказанные свойства означают, что график выпуклой функции может быть «плоским» в окрестности точки минимума, у строго выпуклой функции этого не может быть. Более того, множество точек минимума выпуклой функции выпукло и состоит из единственной точки, если функция строго выпуклая.

4. Если $f(X) \in C^{(2)}$, то для выпуклости функции $f(X)$ на открытом выпуклом множестве необходимо и достаточно, чтобы матрица

$$\frac{\partial^2 f(X)}{\partial X^2}, \quad X \in K. \quad (5)$$

была неотрицательной. Если матрица (5) положительна, то функция $f(X)$ строго выпукла.

Пример 1. Рассмотрим функцию

$$f(X) = 2x_1^2 - 4x_1x_3 + 4x_2^2 - 8x_2x_3 + 9x_3^2 + 4x_1 + 8x_2 - 20x_3.$$

Эта функция дважды дифференцируема и ее частные производные

$$\frac{\partial f}{\partial x_1} = 4x_1 - 4x_3 + 4; \quad \frac{\partial f}{\partial x_2} = 8x_2 - 8x_3 + 8; \quad \frac{\partial f}{\partial x_3} = -4x_1 - 8x_2 + 18x_3 - 20;$$

$$\frac{\partial^2 f}{\partial x_1^2} = 4; \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0; \quad \frac{\partial^2 f}{\partial x_1 \partial x_3} = -4; \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = -8; \quad \frac{\partial^2 f}{\partial x_2^2} = 8; \quad \frac{\partial^2 f}{\partial x_2 \partial x_3} = 18.$$

Матрица вторых производных

$$\frac{\partial^2 f(X)}{\partial X^2} = \begin{pmatrix} 4 & 0 & -4 \\ 0 & 8 & -8 \\ -4 & -8 & 18 \end{pmatrix} > 0, \quad X \in R^n,$$

так как

$$\Delta_1 = 4 > 0, \Delta_2 = \begin{vmatrix} 4 & 0 \\ 0 & 8 \end{vmatrix} = 32 > 0, \Delta_3 = \begin{vmatrix} 4 & 0 & -4 \\ 0 & 8 & -8 \\ -4 & -8 & 18 \end{vmatrix} = 192 > 0.$$

Следовательно, функция строго выпукла.

Упражнение

Исследуйте на выпуклость производственную функцию

$$Q(K, L) = \alpha K^\beta L^{1-\beta}, \quad \alpha > 0, \quad 0 < \beta < 1,$$

выражающую объем производства Q через используемого капитала K и затраты труда L .

5. В каждой точке $X^* \in K$ выпуклая функция имеет опорную функцию, т. е. существует линейная функция $C'(X - X^0)$ такая, что для всех $X \in K$ выполняется неравенство

$$f(X) - f(X^*) \geq C'(X - X^*)$$

Если $f(X) \in C^{(1)}$, то $C = \nabla f(X^*)$.

6. Выпуклая функция $f(X)$ в любой относительно внутренней точке множества K дифференцируема по всем направлениям l , $\|l\| = 1$, и

$$\frac{\partial f(X)}{\partial l} = \max C'l, \quad C \in \Omega(X),$$

где $\Omega(X)$ - множество нормалей опорных функций к $f(X)$ в точке X .

7. Пусть K - выпуклое множество. Вектор l , $l \neq 0$, назовем допустимым направлением в точке X , $X \in K$, относительно множества K , если найдется число ε_1 , $\varepsilon_1 > 0$ такое, что $X + \varepsilon l \in K$ при всех ε , $0 \leq \varepsilon \leq \varepsilon_1$.

Для того чтобы точка $X^0 \in K$ была точкой минимума выпуклой дифференцируемой по всем направлениям функции $f(X)$ на множестве K , необходимо и достаточно, чтобы производная функции $f(X)$ по любому допустимому направлению в точке $X^0 \in K$ была неотрицательной.

Следствие. Из этого свойства, в частности, следует, что каждая стационарная точка X^* $\left(\frac{\partial f(X^*)}{\partial X} = 0 \right)$ выпуклой гладкой функции $f(X)$ является ее точкой минимума.

2. Теорема Куна – Таккера

В этом параграфе основная теорема для гладких задач выпуклого программирования получается как следствие теорем нелинейного программирования.

2.1. Постановка задачи

Пусть в пространстве R^n заданы выпуклые функции $f(X), g_1(X), \dots, g_m(X)$ и выпуклое множество Q . Задача выпуклого программирования: найти точку $X^0, X^0 \in Q, g(X^0) \leq 0$, удовлетворяющую соотношению

$$f(X^0) = \min_{\substack{g(X) \leq 0, \\ X \in Q}} f(X). \quad (6)$$

Допустимые точки в задаче выпуклого программирования задаются соотношениями

$$g(X) \leq 0, \quad X \in Q.$$

Поскольку функции $g_i(X), i=1,2,\dots,m$, выпуклы, то в силу приведенных в п.1 данной лекции свойств выпуклых множеств и функций множество K допустимых точек представляет выпуклое множество. Как доказано в том же параграфе, выпуклая функция $f(X)$ на множестве K может иметь один локальный минимум, который является и глобальным.

Лемма 2. Множество решений задачи выпуклого программирования - выпуклое множество.

Доказательство. Действительно, если

$$f(X^1) = \min_{X \in K} f(X), \quad f(X^2) = \min_{X \in K} f(X), \quad f(X^1) \neq f(X^2) = \alpha,$$

то для всех $\lambda, 0 \leq \lambda \leq 1$,

$$f(\lambda X^1 + (1-\lambda)X^2) \leq \lambda f(X^1) + (1-\lambda)f(X^2) \leq \lambda \alpha + (1-\lambda)\alpha = \min_{X \in K} f(X).$$

Поскольку строгое неравенство в последнем выражении противоречит определению минимума, то получаем равенство

$$f(\lambda X^1 + (1-\lambda)X^2) = \min_{X \in K} f(X),$$

в силу которого точка $\lambda X^1 + (1-\lambda)X^2$ есть решение задачи (6). Лемма 2. доказана.

Следствие. Из леммы следует, что если задача - (6) имеет два решения, то она имеет и континуум решений.

2.2. Седловая точка функции Лагранжа и решение задачи выпуклого программирования

Определение. Функция

$$F(X, \Lambda) = f(X) + \Lambda'g(X), \quad \Lambda \in R^m, \quad \Lambda \geq 0 \quad (7)$$

называется **нормальной функцией Лагранжа** для задачи выпуклого программирования.

Определение. Совокупность (X^*, Λ^*) векторов X^* , $\Lambda^* \geq 0$, называется **седловой точкой функции Лагранжа**, если для всех $X \in Q$, $\Lambda \geq 0$, выполняются неравенства

$$F(X^*, \Lambda) \leq F(X^*, \Lambda^*) \leq F(X, \Lambda^*). \quad (8)$$

Теорема 2. Если (X^*, Λ^*) , $X^* \in Q$, $\Lambda^* \geq 0$ - седловая точка функции Лагранжа (7), то X^* - решение задачи выпуклого программирования (6).

Доказательство. Запишем неравенства (8) более подробно (в исходных функциях):

$$f(X^*) + \Lambda^*g(X^*) \leq f(X^*) + \Lambda^*g(X^*) \leq f(X) + \Lambda^*g(X), \quad X \in Q, \Lambda \geq 0. \quad (9)$$

Из левого неравенства следует

$$\Lambda^*g(X^*) \leq \Lambda^*g(X^*). \quad (10)$$

Для выполнения (10) необходимо, чтобы

$$g(X^*) \leq 0$$

ибо если при некотором i , $1 \leq i \leq m$, $g_i(X^*) > 0$, то, полагая $\lambda_i = 0$, $i \neq i$, $j = 1, 2, \dots, m$, и выбирая λ_j , достаточно большим, получаем слева в (10) сколь угодно большое положительное число.

Из (10) следует условие дополняющей нежесткости:

$$\Lambda^* g(X^*) = 0. \quad (11)$$

Действительно, если допустить, что $\Lambda^* g(X^*) \leq 0$, то, положив $\Lambda = \frac{\Lambda^*}{2}$ из (10) имеем противоречивые неравенства $0 \leq \frac{1}{2} \Lambda^* g(X^*) < 0$. С учетом (11) из правого неравенства в (9) получаем

$$f(X^*) \leq f(X) + \Lambda^* g(X).$$

Отсюда следует, что для всех $X \in Q$, удовлетворяющих неравенству $g(X) \leq 0$, должно выполняться неравенство

$$f(X^*) \leq f(X).$$

Это значит, что X^* - решение задачи выпуклого программирования. Теорема доказана.

Замечание. В доказательстве не использовалась выпуклость функций $f(X)$, $g(X)$ множества Q . Эта теорема верна и для невыпуклых функций и множеств.

Согласно теореме 2 решение задачи выпуклого программирования сводится к нахождению седловых точек функции Лагранжа. Теоремы о существовании седловых точек функции Лагранжа называют **теоремами Куна - Таккера**, по имени ученых, получивших первые результаты такого типа.

2.3. Теорема Куна - Таккера для частной задачи

Пусть выпуклая скалярная функция $f(X)$, $X \in R^n$, и выпуклые компоненты m -мерной векторной функции $g(X)$, $X \in R^n$ определены и непрерывны вместе с производными до второго порядка включительно. Рассмотрим задачу

$$f(X) \rightarrow \min, \quad (12)$$

$$g(X) \leq 0. \quad (13)$$

Данная задача отличается от общей задачи выпуклого программирования (6) предположением гладкости функций и отсутствием ограничения $X \in Q$.

Пусть вектор X^0 (решение задачи (12), (13)) является обыкновенной допустимой точкой по ограничениям типа неравенств (13). Тогда в силу теоремы о необходимых условиях минимума в задачах нелинейного программирования найдется m - мерный вектор Λ^0 , $\Lambda^0 \geq 0$ такой, что выполняются условия:

$$\Lambda^0 g(X^0) = 0, \quad (14)$$

$$\frac{\partial F(X^0, \Lambda^0)}{\partial X} = 0, \quad (15)$$

где

$$F(X, \Lambda) = f(X) + \Lambda' g(X). \quad (16)$$

Поскольку $f(X), g_1(X), \dots, g_m(X)$ - выпуклые функции и $\Lambda^0 \geq 0$, то функция $F(X, \Lambda^0)$ по свойству 2 из п. 1 также выпукла по X .

Следовательно, точка X^0 , удовлетворяющая условию (15), является по свойству 7 из п. 1 точкой минимума функции Лагранжа (16)

$$F(X^0, \Lambda^0) \leq F(X, \Lambda^0) \quad \text{для всех } X \in R^n. \quad (17)$$

По предположению $g(X^0) \leq 0$, поэтому для всех $\Lambda \geq 0$ выполняется неравенство $\Lambda' g(X^0) \leq 0$. С другой стороны, выполняется равенство (14), таким образом,

$$F(X^0, \Lambda) = f(X^0) + \Lambda' g(X^0) \leq f(X^0) + \Lambda^0 g(X^0) = F(X^0, \Lambda^0). \quad (18)$$

Объединяя (17) и (18), приходим к следующему утверждению.

Теорема 3. Если обыкновенная допустимая, точка X^0 является решением задачи выпуклого программирования (12), (13), то существует вектор

Λ'' , $\Lambda'' \geq 0$ такой, что пара (X^0, Λ^0) составляет седловую точку функции Лагранжа, т.е. для всех $\Lambda \geq 0$ и $X \in R^n$ выполняются неравенства

$$F(X^0, \Lambda) \leq F(X^0, \Lambda^0) \leq F(X, \Lambda^0).$$

Замечание. В силу доказанной теоремы точка условного минимума в классе выпуклого программирования является и точкой безусловного минимума функции Лагранжа. Известно, что в общем случае задач нелинейного программирования это утверждение неверно.

Следствие. Если ограничения (13) линейны

$$A_i' X + b_i \leq 0, \quad i = 1, 2, \dots, m,$$

то теорема 3 верна без предположения о том, что X^0 - обыкновенная допустимая точка.

2.4. Задача выпуклого программирования с дополнительным ограничением

Сохраняя предположение о гладкости функций $F(X)$, $g(X)$, вместо (12), (13) рассмотрим задачу

$$f(X) \rightarrow \min, \quad (19)$$

$$g(X) \leq 0, \quad X \geq 0. \quad (20)$$

Точка X^* называется **обыкновенной допустимой по ограничениям** (20), если она обыкновенная допустимая по ограничениям $g(X) \leq 0$, т.е. в точке X^* векторы

$$\frac{\partial g_1(X^*)}{\partial X}, \dots, \frac{\partial g_k(X^*)}{\partial X}, \quad (21)$$

составленные из функций ограничений, активных в точке X^* , линейно независимы.

Каждое из дополнительных ограничений

$$x_i \geq 0, \quad i = 1, 2, \dots, m,$$

с помощью функций $g_{m+i} = -x_i$ можно записать в виде неравенства

$$g_{m+i} \leq 0. \quad (22)$$

Поэтому вводя $(n + m)$ - мерную векторную функцию задачи (19), (20) запишем в виде

$$f(X) \rightarrow \min, \quad (23)$$

$$\bar{g}(X) \leq 0. \quad (24)$$

Если точка X^* - обыкновенная допустимая по ограничениям (20), то она обыкновенная допустимая по ограничениям (24), так как все дополнительные ограничения линейны.

Компоненты функции $\bar{g}(X)$ являются выпуклыми функциями. Поэтому, в силу теоремы 3, найдется неотрицательный $(n + m)$ - мерный вектор $\bar{\Lambda}^0 = (\Lambda^0, \lambda_{m+1}^0, \dots, \lambda_{m+n}^0)$ такой, что пара (X^0, Λ^0) , где X^0 - решение задачи (23). (24) составляет седловую точку функции Лагранжа

$$F(X, \bar{\Lambda}) = f(X) + \Lambda' g(X) - \sum_{i=1}^m \lambda_{m+i} x_i.$$

Точнее, для всех $X \in R^n$, $\Lambda, \lambda_{m+i} \geq 0$, $i = 1, 2, \dots, n$ выполняются неравенства

$$\begin{aligned} f(X^0) + \Lambda' g(X^0) - \sum_{i=1}^n \lambda_{m+i} x_i^0 &\leq f(X^0) + \Lambda^0' g(X^0) - \sum_{i=1}^n \lambda_{m+i}^0 x_i^0 \leq \\ &\leq f(X) + \Lambda^0' g(X) - \sum_{i=1}^n \lambda_{m+i}^0 x_i. \end{aligned} \quad (25)$$

Докажем, что из (25) следует неравенство

$$f(X^0) + \Lambda' g(X^0) \leq f(X^0) + \Lambda^0' g(X^0) \leq f(X) + \Lambda^0' g(X) \quad (26)$$

для всех $\Lambda \geq 0$, $X \geq 0$.

Действительно, если при некотором $\tilde{\Lambda}$, $\tilde{\Lambda} \geq 0$, имеем

$$f(X^0) + \tilde{\Lambda}' g(X^0) > f(X^0) + \Lambda^0' g(X^0),$$

то в (25) левое неравенство при $\Lambda = \tilde{\Lambda}$ и достаточно малых λ_{m+i} , $i = 1, 2, \dots, n$, нарушится. Отсюда следует, что $\Lambda^0' g(X^0) = 0$. По той же причине из (25) следует

$$\Lambda^0 g(X^0) - \sum_{i=1}^m \lambda_{m+i}^0 x_i^0 = 0.$$

Поэтому, из правого неравенства в (25) получаем

$$f(X^0) + \Lambda' g(X^0) \leq f(X^0) + \Lambda^0 g(X^0) - \sum_{i=1}^n \lambda_{m+i}^0 x_i^0 \leq f(X) + \Lambda^0 g(X)$$

для всех $x_i, x_i \geq 0$. Неравенства (26) доказаны. Следовательно, справедлива теорема 4.

Теорема 4. Пусть $f(X), g(X) \in C^{(2)}$. Для того чтобы обыкновенная допустимая точка X^0 была решением задачи (19), (20) необходимо и достаточно существование неотрицательного вектора $\Lambda^0, \Lambda^0 \geq 0$ такого, что для всех $X \geq 0, \Lambda \geq 0$ выполнялись неравенства

$$F(X^0, \Lambda) \leq F(X^0, \Lambda^0) \leq F(X, \Lambda^0),$$

где $F(X, \Lambda) = f(X) + \Lambda' g(X)$.

Таким образом, в задаче выпуклого программирования (14), (15) дополнительные ограничения $X \geq 0$ можно не вносить в функцию Лагранжа, а перенести в формулировку результата.

Пример 2. Найдите объемы ресурсов K и L , при которых затраты на производство не менее 80 единиц продукции минимальны, если производственная функция Кобба - Дугласа $Q(K, L) = K^{3/4} L^{1/4}$, цены на ресурсы $p_K = 6, p_L = 2$.

Решение

Поскольку целевая функция имеет вид $f(K, L) = p_K K + p_L L$, то имеем задачу

$$f(K, L) = 6K + 2L \rightarrow \min,$$

$$K^{3/4} L^{1/4} \geq 80.$$

$$K \geq 0, L \geq 0.$$

Заметим, что функция $f(K, L)$ линейна, т.е. выпукла и любая точка, удовлетворяющая условию задачи, является обыкновенной допустимой, так как вектор

$$\frac{\partial g(K, L)}{\partial X} = \begin{pmatrix} \frac{3}{4K^{1/4}} \\ \frac{1}{4L^{3/4}} \end{pmatrix}$$

для указанных точек является ненулевым - линейно независимым (достаточно проверить точку $(81, 81)$). Теперь для применения теоремы Куна - Таккера достаточно записать постановку задачи в виде

$$\begin{aligned} f(K, L) &= 6K + 2L \rightarrow \min, \\ -K^{3/4}L^{1/4} + 80 &\leq 0, \\ K &\geq 0, \quad L \geq 0. \end{aligned}$$

Составим функцию Лагранжа

$$F(K, L, \lambda) = 6K + 2L + \lambda(-K^{3/4}L^{1/4} + 80).$$

Тогда условия теоремы Куна - Таккера записываются следующим образом:

$$\begin{aligned} \frac{\partial F}{\partial K} &= 6 - \frac{3}{4}\lambda K^{-1/4}L^{1/4} = 0, \\ \frac{\partial F}{\partial L} &= 2 - \frac{1}{4}\lambda K^{3/4}L^{-3/4} = 0, \\ \lambda(-K^{3/4}L^{1/4} + 80) &= 0, \\ \lambda &\geq 0. \end{aligned}$$

Отсюда приходим к системе

$$\begin{cases} \frac{3}{4}\lambda K^{-1/4}L^{1/4} = 6, \\ \frac{1}{4}\lambda K^{3/4}L^{-3/4} = 2, \\ -K^{3/4}L^{1/4} + 80 = 0, \\ \lambda > 0. \end{cases} \Rightarrow \begin{cases} 8K^{1/4}L^{-1/4} = 8K^{-1/4}L^{1/4}, \\ K^{3/4}L^{1/4} = 80. \end{cases}$$

Из последней системы заключаем, что $K = L = 80$, причем $\lambda = 8 > 0$.
 Таким образом, все условия теоремы Куна – Таккера выполнены и $K = L = 80$
 производственный план с минимальными издержками, причем

$$f_{\min} = f(80, 80) = 640.$$

Упражнение

Найдите объемы ресурсов K и L , при которых затраты на производство не менее 140 единиц продукции минимальны, если производственная функция Кобба - Дугласа $Q(K, L) = K^{2/3} L^{1/3}$, цены на ресурсы $p_K = 12$, $p_L = 3$.

Теорема 5. Теорема 4 для случая линейных ограничений верна без предположения о линейной независимости векторов (21).

3. Основная теорема выпуклого программирования

Ниже приведем основную теорему выпуклого программирования без предположения о гладкости функций. Метод доказательства основан на теореме об отделимости выпуклых множеств, а это выходит за рамки математической подготовки студента - экономиста. Поэтому эту теорему приведем без доказательства.

3.1. Условие Слейтера. Основная теорема

Рассмотрим задачу выпуклого программирования

$$f(X) \rightarrow \min, \quad (27)$$

$$g(X) \leq 0, \quad X \in Q, \quad (28)$$

где $f(X)$ и компоненты m -мерной функции $g(X)$ - выпуклые функции при $X \in Q$; Q - выпуклое множество.

Говорят, что ограничения (28) удовлетворяют условию Слейтера, если существует вектор X^* , $X^* \in Q$ такой, что

$$g(X^*) < 0. \quad (29)$$

Теорема 6. Пусть ограничение (28) задачи (27), (28) удовлетворяет условию Слейтера. Тогда для того чтобы вектор X^0 , $X^0 \in Q$ был решением задачи (27), (28) необходимо и достаточно, чтобы существовал вектор

$\Lambda^0, \lambda^0 \geq 0$ такой, что пара (X^0, Λ^0) составляла седловую точку функции Лагранжа $F(X, \Lambda) = f(X) + \Lambda'g(X)$.

Другими словами, теорема утверждает, что для всех $X \in Q, \Lambda \geq 0$, выполняются неравенства

$$F(X^0, \Lambda) \leq F(X^0, \Lambda^0) \leq F(X, \Lambda^0). \quad (30)$$

3.2. Обсуждение

Сравнивая теорему 6 с результатами п.2, видим, что теорема верна для выпуклых функций без дополнительных предположений типа гладкости. Уже отмечалось, что для невыпуклых функций теорема неверна. Однако она справедлива и не для всех выпуклых функций $f(X), g(X)$.

Предположение о том, что ограничения удовлетворяют условию Слейтера, существенно.

Пример 3. Действительно, пусть $f(x) = -x, g(x) = x^2$ и точка $x^0 = 0$ является решением задачи (27), (28). Составим функцию Лагранжа

$$F(x, \lambda) = -x + \lambda x^2.$$

Допустим, что теорема Куна - Таккера верна. Тогда существует число $\lambda^0, \lambda^0 \geq 0$ такое, что для всех $x, x \geq 0$ выполняется неравенство

$$F(x^0, \lambda^0) \leq F(x^0, \lambda), \text{ т.е. } 0 \leq -x + \lambda^0 x^2.$$

Ясно, что при любом $\lambda^0 \geq 0$ найдется достаточное малое число $x > 0$ такое, что $-x + \lambda^0 x^2 < 0$. Значит, теорема Куна - Таккера для этого примера неверна. В примере условие Слейтера не выполняется, ибо $g(x) = x^2$ при всех $x \geq 0$.

Упражнения

Проверить указанные точки на оптимальность в задачах выпуклого программирования:

$$1) \quad f(X) = -2x_1^2 - 3x_2^2 + x_1 - 6 \rightarrow \max,$$

$$W = \{X: x_1^2 + x_1 - 3 \leq 0, 2x_1 + x_2 - 5 \leq 0, x_2 \geq 0\},$$

$$X^1 = (1, 1), X^2 = (2, 1), X^3 = (1/4, 0), X^4 = (0, 0);$$

$$2) f(X) = x_1^2 + 3x_1 \rightarrow \min,$$

$$W = \{X: x_1^2 + x_2^2 - 2x_1 + 8x_2 + 16 \leq 0, \quad x_1 - x_2 \leq 5\}.$$

$$X^1 = (1, -4), \quad X^2 = (0, -4), \quad X^3 = (2, -4);$$

$$3) f(X) = 7x_1^2 + 2x_2^2 - x_1x_2 + x_1 - x_2 \rightarrow \min.$$

$$W = \{X: x_1 + x_2 \leq 2, \quad x_1 - 3x_2 \leq 4, \quad -2x_1 + x_2 \leq -3\},$$

$$X^1 = (5/2, -1/2), \quad X^2 = (1, -1), \quad X^3 = (2, 0);$$

$$4) f(X) = x_1^2/2 + x_2^2 - 5x_1 + x_2 \rightarrow \min,$$

$$W = \{X: x_1 + x_2 - 2x_3 \leq 3, \quad 2x_1 - x_2 - 3x_3 \geq -11, \quad x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0\},$$

$$X^1 = (1, 0, 2), \quad X^2 = (0, 0, -1), \quad X^3 = (1, 3, 0), \quad X^4 = (2, 1, 1), \quad X^5 = (5, 0, 1);$$

$$5) f(X) = e^{x_1 + x_2} + x_1^2 - 2x_2 \rightarrow \min.$$

$$W = \{X: x_2 \leq \ln x_1, \quad x_1 \geq 1, \quad x_2 \geq 0\}.$$

$$X^1 = (2, \ln 2), \quad X^2 = (2, 0), \quad X^3 = (1, 0);$$

$$6) f(X) = e^{x_1 + x_2} + x_3^2 + 2x_1 + 2x_2 \rightarrow \min,$$

$$W = \{X: x_1^2 + x_2^2 + x_3^2 \leq 18, \quad 2x_1 + x_2 - x_3 + 3 \leq 0\},$$

$$X^1 = (-3, 3, 0), \quad X^2 = (1, -1, 4), \quad X^3 = (-3, -3, 0), \quad X^4 = (0, 0, 3);$$

$$7) f(X) = -5x_1^2 - 6x_2^2 - x_3^2 + 8x_1x_2 + x_1 \rightarrow \max,$$

$$W = \{X: x_1^2 - x_2 + x_3 \leq 5, \quad x_1 + 5x_2 \leq 8, \quad x_1 \geq 0, \quad x_2 \geq 0\};$$

$$X^1 = (0, 0, 0), \quad X^2 = (1, 0, 4), \quad X^3 = (3/14, 1/7, 0), \quad X^4 = (4, 0, -11);$$

$$8) f(X) = -x_1^2 - 2x_2^2 + x_1x_2 - 26 \rightarrow \max,$$

$$W = \{X: x_1^2 \leq 25, \quad x_1 + 2x_2 - 5 \leq 0, \quad x_2 \geq 0\}.$$

$$X^1 = (0, 0), \quad X^2 = (-1, 2), \quad X^3 = (0, 6), \quad X^4 = (3, 0);$$

$$9) f(X) = 10x_1^2 + 5x_2^2 - x_1 + 2x_2 - 10 \rightarrow \min,$$

$$W = \{X: 2x_1^2 - x_2 \leq 4, \quad x_1 + x_2 \leq 8, \quad x_1 \geq 0\}$$

$$X^1 = (0, 0), \quad X^2 = (1, 1), \quad X^3 = (0, -1);$$

$$10) f(X) = 4x_1^2 + 3x_2^2 + 4x_1x_2 - x_1 + 6x_2 - 5 \rightarrow \min,$$

$$W = \{X: x_1^2 - x_2^2 \geq -3, \quad 3x_1^2 + x_2 \leq 4, \quad x_1 \geq 0, \quad x_2 \geq 0\},$$

$$X^1 = (0, 0), \quad X^2 = (5, 0), \quad X^3 = (1, -1), \quad X^4 = (1, 1);$$

$$11) f(X) = x_1^2 + 5/2x_2^2 - x_1x_2 \rightarrow \min,$$

$$\mathcal{W} = \{ X : x_1^2 - 4x_1 - x_2 \leq -5, \quad -x_1^2 + 6x_1 - x_2 \geq 7 \}.$$

$$X^1 = (2, 1), \quad X^2 = (3, 2);$$

$$12) \quad f(X) = x_1^2 + 2x_2^2 + 30x_1 - 16x_3 \rightarrow \min.$$

$$\mathcal{W} = \{ X : 5x_1 + 3x_2 - 4x_3 = -20, \quad x_1 - 6x_2 + 3x_3 \leq 0, \quad x_3 \geq 0.$$

$$X^1 = (-3, 1, 2), \quad X^2 = (-5, 3, 1).$$

Условие Слейтера является довольно стеснительным. Оно не выполняется при любой функции $g(X)$ для ограничений

$$g^1(X) \leq 0, \quad g^2(X) \leq 0, \quad (g^1(X) = g(X), \quad g^2(X) = -g(X)),$$

полученных из ограничения $g(X) = 0$. Однако можно показать, что если функция $g(X)$ линейна, то теорема верна без условия Слейтера. Это расширяет область ее применения.

Как и в случае гладких задач, смысл основного результата, содержащегося в теореме Куна - Таккера, состоит в том, что точка условного минимума X^0 задачи выпуклого программирования при некотором Λ^0 , $\Lambda^0 \geq 0$, является точкой безусловного минимума функции Лагранжа на множестве $\mathcal{Q}$.

4. Метод кусочно-линейной аппроксимации для решения некоторого класса задач выпуклого программирования

4.1. Введение

Рассмотрим задачу нелинейного программирования, в которой целевая функция и функции в системе ограничений являются сепарабельными.

Определение. Функция $f(x_1, x_2, \dots, x_n)$ называется **сепарабельной**, если она может быть представлена в виде суммы функций, каждая из которых является функцией одной переменной, т.е. если

$$f(x_1, x_2, \dots, x_n) = \sum_{j=1}^n f_j(x_j).$$

Определение. Если целевая функция и функции в системе ограничений задачи нелинейного программирования являются сепарабельными, то задача называется **задачей сепарабельного программирования**.

Приближенное решение такой задачи можно найти с использованием метода кусочно-линейной аппроксимации. Однако его применение в общем случае позволяет получить приближенный локальный экстремум. Если множество допустимых решений выпукло, а $f_j(x_j)$ – вогнутые функции, то локальный максимум является одновременно и глобальным. Поэтому рассмотрим использование метода кусочно-линейной аппроксимации для решения задачи выпуклого программирования.

4.2. Метод кусочно-линейной аппроксимации

Пусть требуется определить максимальное значение вогнутой функции

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n f_i(x_i) \rightarrow \max \quad (31)$$

$$\sum_{j=1}^n g_{ij}(x_j) \leq b_i, \quad i = \overline{1, m}, \quad (32)$$

$$x_j \geq 0, \quad j = \overline{1, n}. \quad (33)$$

Чтобы найти решение задачи (31) - (33), заменим функции $f_i(x_i)$, $g_{ij}(x_j)$ кусочно-линейными функциями $\tilde{f}_i(x_i)$, $\tilde{g}_{ij}(x_j)$ и перейдем от задачи (31) - (33) к задаче, состоящей в определении максимального значения функции

$$F = \sum_{j=1}^n \tilde{f}_j(x_j) \rightarrow \max \quad (34)$$

при условии

$$\sum_{j=1}^n \tilde{g}_{ij}(x_j) \leq b_i, \quad i = \overline{1, m}, \quad (35)$$

$$x_j \geq 0, \quad j = \overline{1, n}. \quad (36)$$

В задаче (34) - (36) пока не определен вид функций. Чтобы сделать это, будем считать, что переменная x_j может принимать значения из промежутка $[0, \alpha_j]$ (α_j - максимальное значение переменной x_j). Разобьем промежутки

$[0, \alpha_j]$ на r_j промежутков с помощью $r_j + 1$ точек так, что $x_{0j} = 0$, $x_{r_j j} = \alpha_j$.

Тогда функции $\tilde{f}_j(x_j)$, $\tilde{g}_{ij}(x_j)$ можно записать в виде

$$\tilde{f}_j(x_j) = \sum_{k=0}^{r_j} \lambda_{kj} f_{kj}; \quad \tilde{g}_{ij}(x_j) = \sum_{k=0}^{r_j} \lambda_{kj} g_{kij}, \quad (37)$$

где

$$f_{kj} = f_j(x_k); \quad g_{kij} = g_{ij}(x_k), \quad i = \overline{1, m}; \quad (38)$$

$$x_j = \sum_{k=0}^{r_j} \lambda_{kj} x_{kj}.$$

$\sum_{k=0}^{r_j} \lambda_{kj} = 1$, $\lambda_{kj} > 0$ для всех k и j , причем для данного x_j не более двух чисел

λ_{kj} могут быть положительными и должны быть соседними.

Подставляя теперь в равенства (34) и (35) выражения функций $\tilde{f}_j(x_j)$ и $\tilde{g}_{ij}(x_j)$ в соответствии с формулами (37), приходим к следующей задаче:

$$\tilde{F} = \sum_{j=1}^n \sum_{k=0}^{r_j} f_{kj} \lambda_{kj} \rightarrow \max \quad (39)$$

$$\sum_{j=1}^n \sum_{k=0}^{r_j} g_{kij} \lambda_{kj} \leq b_i, \quad i = \overline{1, m}, \quad (40)$$

$$\sum_{k=0}^{r_j} \lambda_{kj} = 1, \quad j = \overline{1, n}, \quad (41)$$

$$\lambda_{kj} \geq 0, \quad \text{для всех } k \text{ и } j. \quad (42)$$

Полученная задача отличается от обычной задачи линейного программирования тем, что наложено дополнительное ограничение на переменную λ_{kj} , состоящее в том, чтобы для каждого j не более двух были положительными и эти положительные λ_{kj} были соседними. Выполнение этих условий может быть соблюдено при решении задачи (39) - (42) симплексным методом за счет соответствующего выбора базиса, определяющего как каждый опорный, так и оптимальный план данной задачи. При этом в общем случае точ-

ность полученного решения зависит от принятого шага разбиения промежутка $[0, \alpha_j]$. Чем меньше шаг, тем более точным является полученное решение.

Таким образом, процесс нахождения решения задачи нелинейного программирования методом кусочно-линейной аппроксимации включает следующие этапы:

1. Каждую из сепарабельных функций заменяют кусочно-линейной функцией.
2. Строят задачу линейного программирования (39) - (42).
3. С помощью симплексного метода находят решение задачи (39) - (42).
4. Определяют оптимальный план задачи (31) - (33) и находят значение целевой функции при этом плане.

Пример 4. Используя метод кусочно-линейной аппроксимации, решить следующую задачу

$$F = x_2 - x_1^2 + 6x_1 - 9 \rightarrow \max,$$
$$\begin{cases} 2x_1 + 3x_2 \leq 24, \\ x_1 + 2x_2 \leq 15, \\ 3x_1 + 2x_2 \leq 24, \\ x_2 \leq 4, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Решение

В данном случае целевую функцию $F = x_2 - x_1^2 + 6x_1 - 9$ можно представить как сумму двух функций $f_1(x_1) = -x_1^2 + 6x_1 - 9$ и $f_2(x_2) = x_2$, каждая из которых есть функция одной переменной. Следовательно, функция F является сепарабельной. Кроме того, она является вогнутой (как сумма двух вогнутых функций), а область допустимых решений - выпуклой. Значит, используя метод кусочно-линейной аппроксимации, можно найти приближенно глобальный максимум целевой функции.

Здесь нелинейной функцией является только целевая функция. Значит, кусочно-линейной функцией следует заменить только ее. При этом так как функция $f_2(x_2)$ линейная, то аппроксимировать будем функцию $f(x_1)$.

Если построить область допустимых решений задачи, то видно будет, что переменная x_1 может принимать значения в промежутке $[0;8]$. Разобьем этот промежуток на восемь частей точками

$$x_{01} = 0, \quad x_{11} = 1, \quad x_{21} = 2, \quad x_{31} = 3, \quad x_{41} = 4, \quad x_{51} = 5, \quad x_{61} = 6, \quad x_{71} = 7, \quad x_{81} = 8$$

Вычислим теперь значения функции $f(x_1)$ в этих точках (табл.1).

Таблица 1

x_{kj}	0	1	2	3	4	5	6	7	8
$f_1(x_{k1})$	-9	-4	-1	0	-1	-4	-9	-16	-25

Используя формулы (37) – (38) находим

$$\bar{f}_1(x_1) = -9\lambda_{01} - 4\lambda_{11} - \lambda_{21} - \lambda_{31} - 4\lambda_{51} - 9\lambda_{61} - 16\lambda_{71} - 25\lambda_{81},$$

$$x_1 = 0\lambda_{01} + 1\lambda_{11} + 2\lambda_{21} + 3\lambda_{31} + 4\lambda_{41} + 5\lambda_{51} + 6\lambda_{61} + 7\lambda_{71} + 8\lambda_{81}.$$

Подставляя найденные значения $f(x_1)$ и x_1 в исходные данные, получим:

$$F = -9\lambda_{01} - 4\lambda_{11} - \lambda_{21} - \lambda_{31} - 4\lambda_{51} - 9\lambda_{61} - 16\lambda_{71} - 25\lambda_{81} + x_2 \rightarrow \max,$$

$$\left\{ \begin{array}{l} 2\lambda_{11} + 4\lambda_{21} + 6\lambda_{31} + 8\lambda_{41} + 10\lambda_{51} + 12\lambda_{61} + 14\lambda_{71} + 16\lambda_{81} + 3x_2 + x_3 = 24, \\ \lambda_{11} + 2\lambda_{21} + 3\lambda_{31} + 4\lambda_{41} + 5\lambda_{51} + 6\lambda_{61} + 7\lambda_{71} + 8\lambda_{81} + 2x_2 + x_4 = 15, \\ 3\lambda_{11} + 6\lambda_{21} + 9\lambda_{31} + 12\lambda_{41} + 15\lambda_{51} + 18\lambda_{61} + 21\lambda_{71} + 24\lambda_{81} + 2x_2 + x_5 = 24, \\ x_2 + x_6 = 4, \\ \lambda_{01} + \lambda_{11} + \lambda_{21} + \lambda_{31} + \lambda_{41} + \lambda_{51} + \lambda_{61} + \lambda_{71} + \lambda_{81} = 1 \\ x_2 \geq 0, \dots, x_6 \geq 0, \quad \lambda_k \geq 0, \quad k = \overline{1,8}. \end{array} \right.$$

Для полученной задачи пять векторов $P_{01}, P_3, P_4, P_5, P_6$ являются единичными. Значит, ее решение может быть найдено симплексным методом. Определяем его (табл. 2).

Таблица 2

P_k	c_k	P_0	-9	-4	-1	0	-1	-4	-9	-16	-25	1	0	0	0	0
			P_{01}	P_{11}	P_{21}	P_{31}	P_{41}	P_{51}	P_{61}	P_{71}	P_{81}	P_2	P_3	P_4	P_5	P_6
P_1	0	24	0	2	4	6	8	10	12	14	16	3	1	0	0	0
P_4	0	15	0	1	2	3	4	5	6	7	8	2	0	1	0	0
P_5	0	24	0	3	6	9	12	15	18	21	24	2	0	0	1	0
P_6	0	4	0	0	0	0	0	0	0	0	0	1	0	0	0	1
P_{01}	-9	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Δ		-9	0	-5	-8	-9	-8	-5	0	7	16	-1	0	0	0	0
P_3	0	18	-6	-4	-2	0	2	4	6	8	10	3	1	0	0	0
P_4	0	12	-3	-2	-1	0	1	2	3	4	5	2	0	1	0	0
P_5	0	15	-9	-6	-3	0	3	6	9	12	15	2	0	0	1	0
P_6	0	4	0	0	0	0	0	0	0	0	0	1	0	0	0	1
P_{31}	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Δ		0	9	4	1	0	1	4	9	16	25	-1	0	0	0	0
P_3	0	6	-6	-4	-2	0	2	4	6	8	10	0	1	0	0	-3
P_4	0	4	-3	-2	-1	0	1	2	3	4	5	0	0	1	0	-2
P_5	0	7	-9	-6	-3	0	3	6	9	12	15	0	0	0	1	-2
P_2	1	4	0	0	0	0	0	0	0	0	0	1	0	0	0	1
P_{31}	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Δ		4	9	4	1	0	1	4	9	2	25	0	0	0	0	1

По найденным в табл. 2 значениям λ_j находим $x_1^* = 3$. Далее, из табл. 2 видим, что $x_1^* = 4$. Значит, $X^* = (3; 4)$ является приближенным оптимальным

решением, которое случайно совпало с точным решением. При данном решении $F_{\max} = 4$.

Замечание. Рассмотренный способ приближенного нахождения локального максимума весьма полезен, так как не зависит от аналитических свойств функций, входящих в условие задачи. Однако его нужно применять с большой осторожностью, так как локальные экстремумы приближенной задачи могут существенно отличаться от глобального экстремума.

Упражнения

Используя приближенную задачу, найти оптимальное решение следующих задач.

1. $F = x_1 - x_1^2 + x_2 \rightarrow \max$

$$\begin{cases} x_1^2 + 2x_1 + 3x_2^2 \leq 3, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

2. $F = 3x_1 + 2x_2 \rightarrow \max$

$$\begin{cases} 4x_1 + x_2^2 \leq 16, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

3. $F = -(x_1 - 4)^2 - (x_2 - 3)^2 \rightarrow \max$

$$\begin{cases} x_1^2 + x_2^2 \leq 6, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

4. $F = 4(x_1 - 1)^2 + 16(x_2 - 5)^2 \rightarrow \min$

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

5. $F = 4x_1^2 + 3x_2^2 \rightarrow \min$

$$\begin{cases} x_1 + 3x_2 \geq 5, \\ x_1 + 4x_2 \geq 4. \end{cases}$$

Вопросы для самопроверки

1. Дайте определения: а) выпуклой функции; б) строго выпуклой функции.

2. Какая связь имеется между выпуклыми множествами и выпуклыми функциями. Приведите утверждения, в которых эта связь раскрывается.

3. Приведите свойства выпуклых функций.

4. Сформулируйте постановки задачи выпуклого программирования.
5. Что можно утверждать о множестве решений задачи выпуклого программирования?
6. Дайте определения седловой точки функции Лагранжа.
7. Как решается задача выпуклого программирования. Сформулируйте теорему Куна-Таккера – основную теорему задачи выпуклого программирования в частном случае.
8. Как решается задача выпуклого программирования с дополнительным ограничением?
9. Что представляет условие Слейтера?
10. Сформулируйте теорему Куна-Таккера – основную теорему задачи выпуклого программирования без предположения о гладкости функций.
11. Почему условие Слейтера является стеснительным?
12. В чем состоит смысл основного результата содержащего в теореме Куна - Таккера как в случае гладких, так и в случае негладких задач?
13. Что называется сепарабельной функцией?
14. Какая задача называется задачей сепарабельного программирования?
15. Известно, что приближенное решение задачи сепарабельного программирования можно найти с использованием метода кусочно-линейной аппроксимации. Всегда ли его применение позволяет получить приближенный локальный экстремум?
16. Когда полученный приближенный локальный максимум задачи сепарабельного программирования является одновременно и глобальным?
17. Почему в выражении $x_j = \sum_{k=0}^{r_j} \lambda_k x_{kj}$ не более двух λ_k могут быть отличными от нуля?
18. Как определить, является ли решение задачи, приведенное в симплексной таблице, оптимальным?
19. При решении каких задач можно не учитывать ограничения на выбор базиса?

§24. Элементы динамического программирования

Ключевые слова: Динамическое программирование, принцип оптимальности, оптимальное управление, условно оптимальное управление, метод рекуррентных соотношений, оптимальное распределение инвестиций, уравнение состояний, оптимальное управление, оптимальная прибыль.

Динамическое программирование — раздел математического программирования, в котором процесс принятия решения и управления может быть разбит на отдельные шаги (этапы).

1. Введение

Динамическое программирование позволяет свести одну сложную задачу со многими переменными ко многим задачам с малым числом переменных. В отличие от линейного программирования, в котором симплексный метод является универсальным методом решения, в динамическом программировании такого метода не существует. Одним из основных методов динамического программирования является метод рекуррентных соотношений, который основывается на использовании принципа оптимальности, разработанного американским математиком Р.Беллманом в 1953 г.

Экономический процесс является управляемым, если можно влиять на ход его развития. Под управлением понимается совокупность решений, принимаемых на каждом этапе для влияния на ход развития процесса. В качестве примеров управляемых процессов можно назвать, например, выпуск продукции предприятием, обеспечение предприятия сырьем, замена оборудования и т.д.

Принцип оптимальности Р. Беллмана: каково бы ни было состояние системы в результате какого-либо числа шагов, на ближайшем шаге нужно выбирать управление так, чтобы оно в совокупности с оптимальным управлением на всех последующих шагах приводило к оптимальному выигрышу на всех оставшихся шагах, включая данный.

Основным условием, при котором принцип оптимальности является верным, состоит в том, что процесс управления должен быть без обратной связи. Это означает, что управление на данном шаге не должно оказывать влияния на предшествующие шаги. Из принципа оптимальности следует, что для любого процесса без обратной связи оптимальное управление таково, что оно является оптимальным для любого подпроцесса по отношению к его исходному состоянию. Поэтому решение на каждом шаге оказывается наилучшим с точки зрения управления в целом.

Принцип оптимальности можно представить в виде формулы Беллмана:

$$F_k^*(s_{k-1}) = \max_{\{X_k\}} \{f_k(s_k, X_k) + F_{k+1}^*(s_k)\}, \quad (1)$$

где $k = 1, 2, \dots, n$ - номер шага;

n - общее количество шагов (расчет по формуле (1) ведется начиная с последнего шага, т.е. $k = n-1, n-2, \dots, 2, 1$);

s_{k-1}, s_k - состояние системы в конце шага под номером k и шага под номером $k-1$, соответственно;

$F_k^*(s_{k-1})$ - условный максимум целевой функции, при котором достигается максимум в (1), полученный при оптимальном управлении на $n-k+1$ шагах, начиная с k -го до конца, при условии, что к концу k -го шага система находилась в состоянии s_{k-1} ;

X_k - произвольное управление на k -м шаге и оптимальном управлении на последующих $n-k$ шагах;

$f_k(s_k, X_k)$ - показатель эффективности k -го шага.

Управление X_k на k -м шаге, при котором достигается максимум (1), обозначается через $X_k^*(s_{k-1})$ и называется *условным оптимальным управлением на k -м шаге*.

Состояние s_k системы в конце k -го шага зависит только от предшествующего состояния s_{k-1} и управления X_k , и не зависит от предшествующих

состояний и управлений. Это требование называется «отсутствием последствий». Состояние s_k системы в конце k -го шага определяется уравнениями состояния

$$s_k = \varphi_k(s_{k-1}, X_k), \quad k = 1, 2, \dots, n. \quad (2)$$

В п. 3 и в следующем рассмотрим применение принципа оптимальности Р. Беллмана к ряду экономических задач.

2. Значение динамического программирования

Отметим, что модели линейного программирования в большинстве случаев используются в промышленности для принятия крупномасштабных плановых решений в сложных ситуациях. В то же время модели динамического программирования обычно применяются при решении задач значительно меньшего масштаба. Вот некоторые типичные области применения моделей динамического программирования при принятии решений:

- Разработка правил управления запасами, устанавливающих момент пополнения запасов и размер пополняющего заказа.
- Разработка принципов календарного планирования производства и выравнивания занятости в условиях колеблющегося спроса на продукцию.
- Определение необходимого объема запасных частей, гарантирующего эффективное использование дорогостоящего оборудования.
- Распределение дефицитных капитальных вложений между возможными новыми направлениями их использования.
- Выбор методов проведения рекламной кампании, знакомящей покупателя с продукцией фирмы.
- Систематизация методов поиска ценного вида ресурсов.
- Составление календарных планов текущего и капитального ремонта сложного оборудования.
- Разработка долгосрочных правил замены выбывающих из эксплуатации основных фондов.

Процессы принятия решений, к которым относится ряд упомянутых выше моделей, часто относятся к числу *микроэкономических*. Однако во многих реально функционирующих системах еженедельно требуется принимать тысячи таких решений. В связи с этим, модели динамического программирования ценны именно тем, что они позволяют принимать миллионы решений на основе стандартного подхода (часто с использованием компьютеров) при минимальном вмешательстве человека. Без слов понятно, что даже если каждое из таких решений, взятое в отдельности, не является существенным, то в совокупности они могут оказывать большое влияние на прибыли фирм. Многие фирмы, воспользовавшиеся моделями динамического программирования, были неплохо вознаграждены - им удалось добиться сокращения затрат на обслуживание и ремонт или же уровня запасов на 25% и более, причем без какого-либо ухудшения качества обслуживания.

3. Оптимальная стратегия замены оборудования

Задача состоит в определении оптимальных сроков замены старого оборудования (станков, производственных зданий и т.д.) на новое. Старение оборудования включает его физический и моральный износ, в результате чего растут затраты на ремонт и обслуживание, снижается производительность труда, уменьшается остаточная стоимость. Наступает момент, когда старое оборудование выгоднее продать и заменить новым, чем эксплуатировать с большими затратами. Критерием оптимизации при решении задачи о замене оборудования может служить максимальная прибыль от эксплуатации оборудования или минимальные суммарные затраты на его эксплуатацию.

Возраст оборудования t будем отсчитывать в направлении течения процесса (рис. 1). При $t = 0$ имеем случай использования нового оборудования. Длительность процесса равна n лет.

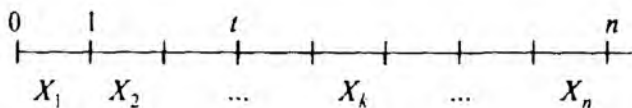


Рис. 1

На рис. 1 представлены произвольные управления X_k на k -м шаге и оптимальном управлении на последующих $n - k$ шагах.

Пример. Оборудование эксплуатируется в течение пяти лет, после чего продается. В начале каждого года может быть принято решение о сохранении оборудования или его замене. Стоимость нового оборудования $P = 4000$ ден. ед. После t лет эксплуатации оборудования может быть продано за $q(t) = \frac{P}{2^t}$ ден. ед. Затраты на содержание оборудования в течение года зависят от его возраста и равны $r(t) = 600(t + 1)$. Определить оптимальную стратегию эксплуатации оборудования при условии минимизации затрат с учетом начальной покупки и заключительной продажи.

Решение

Параметром состояния (2) в данном случае является возраст машины, т.е. $s_{k-1} = t$. Для новой машины в начале первого года эксплуатации $s_0 = 0$. Управление на каждом шаге определяется двумя переменными $X_k = X_c$ (сохранить) и $X_k = X_z$ (заменить).

Тогда

$$s_k = \varphi_k(s_{k-1}, X_k) = \begin{cases} t + 1, & \text{если } X_k = X_c, \\ 1, & \text{если } X_k = X_z, \end{cases} \quad k = 1, 2, 3, 4, 5.$$

Значение, определяемое второй строчкой данного равенства, связано с тем, что интерпретируется следующим образом: если оборудование заменяется, то это означает, что к началу k -го шага его возраст будет равен нулю, а после года эксплуатации $t = 1$, т.е. $s_k = 1$.

Показатель эффективности k -го шага

$$f_k(s_k, X_k) = \begin{cases} 600(t+1), & \text{если } X_k = X_c \rightarrow \text{сохранение,} \\ 4000 + 600 - \frac{4000}{2^t}, & \text{если } X_k = X_z \rightarrow \text{замена.} \end{cases} \quad (3)$$

При сохранении оборудования возраста t лет (первая строка) имеют место только затраты на его эксплуатацию, равные $600(t+1)$ ден. ед. При покупке нового оборудования за 4000 ден. ед., оплате его эксплуатации в течение года в размере 600 ден. ед. и продаже старого оборудования за $\frac{4000}{2^t}$ ден. ед. общие затраты (вторая строка) составят $4600 - \frac{4000}{2^t}$ ден. ед. (здесь и далее затратам присваивается знак «+», а доходам знак «-»).

Для условий примера функция Беллмана приобретает вид

$$F_k^*(s_{k-1}) \min \begin{cases} 600(t+1) + F_{k+1}^*, & \text{если } X_k = X_c \rightarrow \text{сохранение,} \\ 4000 + 600 - \frac{4000}{2^t} + F_{k+1}^*(1), & \text{если } X_k = X_z \rightarrow \text{замена,} \end{cases} \quad (4)$$

где $s_k = t+1$. Во втором уравнении в качестве аргумента условного максимума целевой функции взята единица, так как возраст оборудования после замены в конце года равен одному году.

Решим данную задачу геометрически. Для этих целей построим график состояний и управлений (рис. 2). На этом графике по оси абсцисс будем откладывать номер шага k , а по оси ординат – возраст оборудования t .

Состояние оборудования обозначены на рисунке прямоугольниками. Начальное и конечное состояния оборудования соответствуют точкам $(0;0)$ и $(5;1)$ соответственно. Обозначим их $s(0;0)$ и $s(5;1)$. Каждое состояние машины, кроме состояний из последнего столбца, соединяется стрелками с двумя последующими. Верхняя стрелка соответствует сохранению оборудования, а нижняя – замене. Над каждой стрелкой записаны соответствующие затраты.

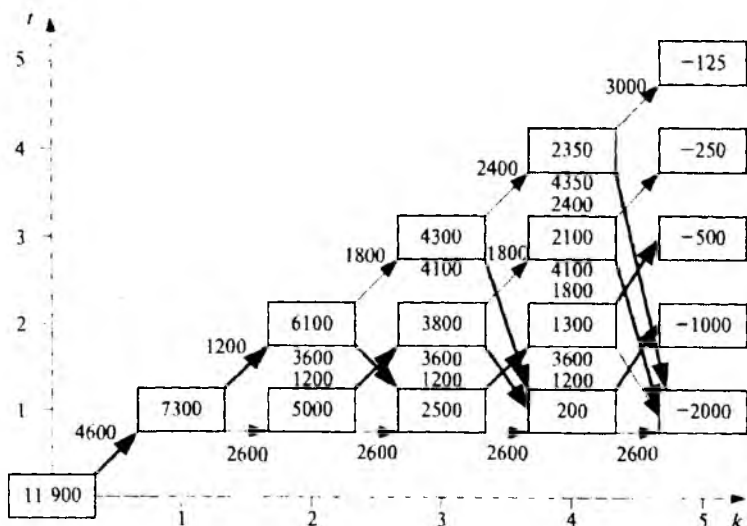


Рис. 2

Рядом со стрелкой, соединяющей начальное событие $s(0;0)$ и событие $s(1;1)$, стоит цифра 4600, получившаяся из следующих соображений. В начале процесса куплено новое оборудование за 4000 ден. ед. и оплачена его эксплуатация в течение первого года, равная $600 \cdot (t+1) = 600 \cdot (0+1) = 600$ ден. ед. Поэтому общие затраты первого шага составили $4000 + 600 = 4600$ ден. ед. (вторая формула в (3)). Так как старого оборудования в начале процесса не было, то и доход от его продажи равен нулю.

Рядом со стрелкой, соединяющей событие $s(1;1)$ и событие $s(2;2)$ стоит цифра 1200, так как затраты при сохранении оборудования после первого шага равны $600 \cdot (t+1) = 600 \cdot (1+1) = 1200$ ден. ед. (строка «сохранение» в (3)). Число, которое необходимо написать рядом со стрелкой, соединяющий событие $s(1;1)$ и $s(2;1)$, вычисляется по формуле (3), т.е.

$$4600 - \frac{4000}{2'} = 4600 - \frac{4000}{2} = 2600.$$

Рядом со стрелкой, соединяющей событие $s(3;3)$ и событие $s(3;1)$, стоит цифра 3600, так как затраты при замене оборудования равны ден. ед. (строка «сохранение» в (3)).

Используя формулы (3), можно записать числа затрат для всех событий (стрелок) на схеме рис. 2.

Расчет условного максимума целевой функции $F_k^*(s_{k+1})$ ведется по формуле (4), начиная с последнего, пятого, шага. В этом случае за начальное состояние принимаются точки $(4;t)$, а за конечное – точки $(5;t)$. В состоянии $(5;t)$ оборудование продается. Доход от продажи рассчитывается по формуле

$$q(t) = \frac{4000}{2^t}. \text{ Напомним, что доходам присваиваем знак «-». Полученные результаты занесены в прямоугольники последнего столбца рис. 2.}$$

Определим для любого исхода пятого шага оптимальное управление и отметим эти управления жирной стрелкой на рис. 2.

Из состояния $(4;1)$ можно попасть в состояния $(5;1)$ и $(5;2)$. Для случая $(4;1) \rightarrow (5;1)$ затраты определяются по второй формуле (3);

$$4000 + 600 - \frac{4000}{2^1} + F_{k+1}^*(1) = 4600 - \frac{4000}{2} - 2000 = 600 \text{ ден. ед.}$$

Это число также можно получить, используя данные рис 2. Для этих целей надо сложить значение, стоящее у стрелки, соединяющей $(4;1)$ и $(5;1)$, со значением, стоящем в прямоугольнике $(5;1)$:

$$2600 - 2000 = 600 \text{ ден. ед.}$$

Для случая $(4;1) \rightarrow (5;2)$ затраты определяются по первой формуле (4):

$$600(t+1) + F_{k+1}^*(t+1) = 600(1+1) - 1000 = 200 \text{ ден. ед.}$$

Используя данные рис. 2, получим:

$$1200 - 1000 = 200 \text{ ден. ед.}$$

Таким образом, оптимальным управлением для состояния $(4;1)$ является

$X_k = X_c$ (сохранить), т.е. случай $(4;1) \rightarrow (5;2)$.

Условный минимум целевой функции при этом

$$F_s^*(s_{k-1}) = F_s^*(1) = 200 \text{ ден. ед.},$$

так как возраст оборудования $s_{k-1} = t = 1$. Поместим условный минимум целевой функции в квадрат точки (4;1), а управления (4;1) → (5;2) отметим жирной стрелкой на рис. 2.

Из состояния (4;2) можно попасть в состояния (5;1) и (5;3). Для случая (4;2) → (5;1) затраты определяются как сумма числа, стоящего у стрелки, соединяющей (4;2) и (5;1), с числом, стоящим в прямоугольнике (5;1), т.е. 3600-2000=1600 ден. ед. Для случая (4;2) → (5;3) затраты равны 1800-500=1300 ден. ед. Оптимальным управлением для состояния (4;2) является $X_k = X_*$ (сохранить), т.е. случай (4;2) → (5;3). Условный минимум целевой функции при этом

$$F_s^*(s_{k-1}) = F_s^*(2) = 1300 \text{ ден. ед.},$$

так как возраст оборудования $s_{k-1} = t = 2$. Поместим этот условный минимум целевой функции в квадрат точки (4;2), а управления (4;2) → (5;3) отметим жирной стрелкой.

Из состояния (4;3) можно попасть в состояние (5;1) и (5;4). Для случая (4;3) → (5;1) затраты 4100-2000=2100 ден. ед., а для случая

$$(4;3) \rightarrow (5;4) - 2400 - 250 = 2150 \text{ ден. ед.}$$

Оптимальным управлением для состояния (4;3) является $X_k = X_*$ (заменить), т.е. случай (4;3) → (5;1). Условный минимум целевой функции при этом

$$F_s^*(s_{k-1}) = F_s^*(3) = 2100 \text{ ден. ед.}$$

Аналогичным образом определяются затраты на любом шаге. Все результаты расчета представлены на рис. 2.

После проведения условной оптимизации в точке (0;0) получим минимальные затраты на эксплуатацию оборудования в течение пяти лет с последующей продажей. Эта величина

$$F_1^*(0) = F_{\min} = 11900 \text{ ден. ед.}$$

Далее строим оптимальную траекторию, перемещаясь из точки (0;0) по жирным стрелкам. Оптимальному уравнению соответствует набор точек

$$(0;0) \rightarrow (1;1) \rightarrow (2;2) \rightarrow (3;1) \rightarrow (4;2) \rightarrow (5;3).$$

Таким образом, оптимальное управление замены оборудования состоит в том, чтобы заменить его на новое в начале третьего года.

Ниже в качестве упражнения приведем две задачи на замену оборудования. В первой задаче в качестве критерия оптимизации выберем прибыль, которую следует максимизировать. Напомним, что в вышерассмотренной задаче, в качестве критерия оптимизации выбраны минимизации затрат с учетом начальной покупки и заключительной продажи. В качестве второй задачи приведем бесконечношаговую модель задачи о замене (или тоже самое задача определения оптимальной стратегии замены оборудования при бесконечном плановом периоде).

Упражнения

1. Предположим, что ежегодные затраты на эксплуатацию, ликвидная и начальная стоимость зависят не только от возраста оборудования t , но и от времени, прошедшего с начала процесса. Пусть $r_k(t)$ -затраты на эксплуатацию в течение k -го года, если со времени последней замены прошло t лет; $\varphi_k(t)$ -ликвидная стоимость оборудования возраста t лет, если оно продается в начале k -го – года; p_k - начальная стоимость оборудования, если оно куплено в начале k -го года.

Требуется определить оптимальные сроки замены старого оборудования новым в течение n лет с тем, чтобы минимизировать затраты на его содержание.

Числовой пример. Решить задачу 1 при заданных конкретных условиях. Автомашина эксплуатируется в течение шести лет. В начале каждого года может быть принято решение о замене машины новой. Стоимость новой машины на k -м году эксплуатации составляет $p_k = 5000 + 500(k-1)$. После t

лет эксплуатации машину на k -м году можно продать за $\varphi(t) = p_k \cdot 2^{-t}$. Стоимость содержания машины в течение k -го года составляет $r_k(t) = 0,1 p_k(t+1)$. Найти оптимальный способ эксплуатации машины: когда нужно заменить машину новой, чтобы суммарные затраты (с учетом затрат на покупку новой машины в начале срока эксплуатации и компенсации за счет заключительной продажи) были минимальны.

2. Бесконечно шаговая модель задачи о замене оборудования (задача определения оптимальной стратегии замены оборудования при бесконечном плановом периоде). Определить оптимальные сроки замены оборудования при неограниченном времени его использования, если известны:

p - начальная стоимость оборудования;

$f(t)$ - стоимость производимой продукции на оборудовании возраста t лет;

$r(t)$ - ежегодные затраты на эксплуатацию оборудования возраста t лет;

$\varphi(t)$ - ликвидная стоимость оборудования возраста t лет.

В задаче требуется минимизировать затраты.

4. Оптимальное распределение инвестиций. Постановка задачи и ее решения

В п.3 рассмотрели применение принципа оптимальности Р. Беллмана к задаче замены оборудования. В настоящем пункте продолжим применение принципа оптимальности Р. Беллмана к задачам экономики, а именно, к задаче оптимального распределения инвестиций.

Постановка задачи. Распределить выделенные средства в размере $s_0 = 5$ ден. ед. между четырьмя предприятиями на очередной год. Вложения в каждое предприятие кратно одной ден. ед. Средства X_k , выделенные предприятию под номером k ($k = 1, 2, 3, 4$), приносят в конце года прибыль $f_k(X_k)$. Функция $f_k(X_k)$ задана таблично и представлена в табл. 1.

X_k	$f_1(X_k)$	$f_2(X_k)$	$f_3(X_k)$	$f_4(X_k)$
1	8	6	3	4
2	10	9	4	6
3	11	11	7	8
4	12	13	11	13
5	18	15	18	16

Определить, какое количество средств нужно выделить каждому предприятию, чтобы суммарная прибыль была максимальной.

Решение

Перепишем задачу в виде

$$F = \sum_{k=1}^4 f_k(X_k) \rightarrow \max$$

при условиях

$$\sum_{k=1}^4 X_k = 5.$$

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_4 \geq 0.$$

В задаче заданы целочисленные переменные, но функция $f_k(X_k)$ задана таблично, поэтому для ее решения нельзя применить методы целочисленного линейного программирования. Используем для решения этой задачи метод динамического программирования.

Требуется найти оптимальные переменные (оптимальное управление) $X_1^*, X_2^*, X_3^*, X_4^*$, удовлетворяющие условиям задачи и обращающие в максимум целевую функцию.

Уравнение состояния (2) из п. 1 для данной задачи имеет вид

$$s_k = \varphi(s_{k-1}, X_k) = s_{k-1} - X_k, k = 1, 2, 3, 4, \quad (5)$$

где s_k - количество средств, оставшихся после k -го шага; X_k - произвольное управление на k -м шаге и при оптимальном управлении на последующих $4-k$ шагах.

Схема решения задачи методом динамического программирования представлена на рис 3. По оси абсцисс будем откладывать номер шага (номер

предприятия). произвольные управления X_k на k -м шаге и оптимальном управлении на последующих $4-k$ шаге, а также уравнения состояния после каждого шага.

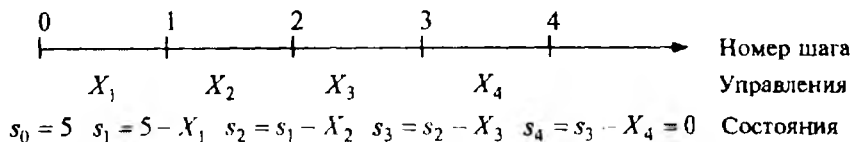


Рис. 3

Последнее уравнение состояния $s_4 = 0$, так как конечное состояние процесса распределения равно нулю, т.е. все средства вложены в производство. Допустимое управление на k -м шаге удовлетворяет условию $0 \leq X_k \leq s_{k-1}$.

Результаты расчетов возможных размещений средств приведены в табл. 2.

В первый столбец табл. 2 внесены возможные распределяемые между предприятиями средства s_{k-1} в конце шага под номером $k-1$, изменяющиеся от нуля до пяти.

Таблица 2

s_{k-1}	X_k	s_k	$k=3$			$k=2$			$k=1$		
			$f_3(s_1, X_1) + F_4^*(s_3)$	$F_3^*(s_2)$	$X_3^*(s_2)$	$f_2(s_2, X_2) + F_3^*(s_2)$	$F_2^*(s_1)$	$X_2^*(s_1)$	$f_1(s_1, X_1) + F_2^*(s_1)$	$F_1^*(s_0)$	$X_1^*(s_0)$
1	2	3	4	5	6	7	8	9	10	11	12
0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0+4=4	4	0	0+4=4	6	1	0+6=6	8	1
	1	0	3+0=3			6+0=6			8+0=8		
2	0	2	0+6=6	7	1	0+7=7	10	1	0+10=10	14	1
	1	1	3+4=7			6+4=10			8+6=14		
	2	0	4+0=4			9+0=9			10+0=10		
3	0	3	0+8=8	9	1	0+9=9	13	1	0+13=13	18	1
	1	2	3+6=9			6+7=13		2	8+10=18		
	2	1	4+4=8			9+4=13			10+6=16		
	3	0	7+0=7			11+0=11			11+0=11		
4	0	4	0+13=13	13	0	0+13=13	16	2	0+16=16	21	1
	1	3	3+8=11			6+9=15			8+13=21		
	2	2	4+6=10			9+7=16			10+10=20		
	3	1	7+4=11			11+4=15			0		

	4	0	11+0=11			13+0=13			11+6=17 12+0=12		
5	0	5	0+16=16	18	5	0+18=18	19	1	0+19=19	24	1
	1	4	3+13=16			6+13=19			8+16=24		
	2	3	4+8=12			9+9=18			10+13=2		
	3	2	7+6=13			11+7=18			3		
	4	1	11+4=15			13+4=17			11+10=2		
	5	0	18+0=18			15+0=15			1		
									12+6=18		
									18+0=18		

Во второй столбец внесены возможные управления X_k для заданного количества s_{k-1} . Например, если средств выделено 0 ден. ед., то и управление может быть только нулевое, либо 1 ден. ед. (вторая строка). Если выделены 2 ден. ед., управление может быть либо нулевое, либо 1 ден. ед., либо 2 ден. ед. (третья строка) и т.д.

В третий столбец внесены возможные оставшиеся после k -го шага количества средств s_k , рассчитанные по уравнению состояния (5). При расчете в эту формулу подставляют соответствующие данные из первого и второго столбцов табл. 2.

Далее, используя формулу для расчета условной оптимальной прибыли (1), заполняем оставшиеся графы табл. 2, проводя условную оптимизацию каждого шага.

Четвертый шаг. Все средства, оставшиеся к четвертому шагу, следует вложить в четвертое предприятие, так как прибыль $f_4(X_k)$ в соответствии с данными табл. 1 возрастает при увеличении вложений. Формула для расчета условной оптимальной прибыли (1) на четвертом шаге принимает вид

$$F_4^*(s_1) = \max_{\{x_4\}} \{f_4(s_1, X_4)\},$$

где $X_4 = s_1$ - средства, оставшиеся к четвертому шагу.

Третий шаг. Рассмотрим все возможные варианты финансирования третьего предприятия. Формулу для расчета условной оптимальной прибыли (1) для третьего шага запишем в виде

$$F_3^*(s_2) = \max_{\{X_3\}} \{f_3(s_3, X_3) + F_4^*(s_3)\}. \quad (6)$$

В четвертый столбец внесены результаты расчета показателя эффективности третьего шага $f_3(s_3, X_3)$, сложенного с условной оптимальной прибылью четвертого шага $F_4^*(s_3)$. Пусть возможная сумма средств, распределяемых между предприятиями в конце второго шага, равна $s_{k-1} = s_2 = 0$. Тогда прибыль, полученная от третьего предприятия, равна нулю (первая строка четвертого столбца).

Если возможная сумма средств, распределяемых между предприятиями в конце второго шага, равна $s_{k-1} = s_2 = 1$, то возможны два варианта (вторая строчка четвертого столбца).

1. В первом случае при $X_3 = 0$ имеем

$$f_3(s_3 = 1, X_3 = 0) = 0; \quad F_4^*(s_3 = 1) = f_4(s_4 = 1, X_4 = 1) = 4$$

(см. табл. 1). В сумме получим:

$$f_3(s_3 = 1, X_3 = 0) + F_4^*(s_3 = 1) = 0 + 4 = 4.$$

2. Во втором случае при $X_3 = 1$ имеем

$$f_3(s_3 = 0, X_3 = 1) = 3; \quad F_4^*(s_3 = 0) = f_4(s_4 = 0, X_4 = 0) = 0.$$

В сумме получим:

$$f_3(s_3 = 0, X_3 = 1) + F_4^*(s_3 = 0) = 3 + 0 = 3.$$

Эти результаты приведены во второй строке четвертого столбца.

Далее рассмотрим возможную сумму средств, распределяемых между предприятиями в конце второго шага, равную $s_{k-1} = s_2 = 2$. В этом случае возможны три варианта (третья строка четвертого столбца).

В первом варианте при $X_3 = 0$ имеем

$$f_3(s_3 = 2, X_3 = 0) = 0; \quad F_4^*(s_3 = 2) = f_4(s_4 = 2, X_4 = 2) = 6$$

(см. табл. 1). В сумме получим:

$$f_3(s_3 = 2, X_3 = 0) + F_4^*(s_3 = 2) = 0 + 6 = 6.$$

Во втором варианте при $X_3 = 1$ имеем

$$f_3(s_1 = 1, X_3 = 1) = 3; F_4^*(s_3 = 1) = f_4(s_4 = 0, X_4 = 1) = 4.$$

В сумме получим:

$$f_3(s_3 = 1, X_3 = 1) + F_4^*(s_3 = 3) = 3 + 4 = 7.$$

В третьем варианте при $X_3 = 2$ имеем

$$f_3(s_3 = 0, X_3 = 2) = 4; F_4^*(s_3 = 0) = f_4(s_4 = 0, X_4 = 0) = 0.$$

В сумме получим:

$$f_3(s_3 = 0, X_3 = 2) + F_4^*(s_3 = 0) = 4 + 0 = 4.$$

Эти результаты приведены в третьей строке четвертого столбца.

Аналогично проводятся расчеты для $s_{k-1} = s_2 = 3$, $s_{k-1} = s_2 = 4$ и $s_{k-1} = s_2 = 5$. Результаты расчетов приведены в четвертой, пятой и шестой строках четвертого столбца.

В пятом столбце табл. 2 в соответствии с формулой (6), предназначенной для расчета условной оптимальной прибыли для третьего шага, запишем максимальные значения из четвертого столбца. В шестом столбце запишем управления, пришедшие к максимальным условным оптимальным прибылям.

Второй шаг. Запишем формулу для расчетов условной оптимальной прибыли (1) для второго шага:

$$F_2^*(s_1) = \max_{\{X_2\}} \{f_2(s_2, X_2) + F_3^*(s_2)\}. \quad (7)$$

Показатели эффективности второго шага $f_2(s_2, X_2)$, сложенные с условными оптимальными прибылями третьего шага $F_3^*(s_2)$, внесены в седьмой столбец табл. 2. Пусть возможная сумма средств, распределяемых между предприятиями в конце первого шага, равна $s_{k-1} = s_1 = 1$, то возможны два варианта (вторая строка седьмого столбца). В первом варианте при $X_2 = 0$ имеем

$$f_2(s_2 = 1, X_2 = 0) = 0; F_3^*(s_2 = 1) = 4$$

(см. вторую строку пятого столбца табл. 1). В сумме получим

$$f_2(s_2 = 1, X_2 = 0) + F_3^*(s_2 = 1) = 0 + 4 = 4.$$

Во втором варианте при $X_2 = 1$ имеем

$$f_2(s_2 = 0, X_2 = 1) = 6 \text{ и } F_3^*(s_2 = 0) = 0$$

(так как средств для третьего и четвертого предприятий не осталось).

В сумме получим:

$$f_2(s_2 = 0, X_2 = 1) + F_3^*(s_2 = 0) = 6 + 0 = 6.$$

Эти результаты приведены во второй строке седьмого столбца.

Далее рассмотрим возможную сумму средств, распределяемых между предприятиями в конце второго шага, равную $s_{k-1} = s_1 = 2$. В этом случае возможны три варианта (третья строка седьмого столбца).

В первом варианте при $X_2 = 0$ имеем

$$f_2(s_2 = 2, X_2 = 0) = 0; F_3^*(s_2 = 2) = 7.$$

(см. третью строку пятого столбца табл. 1). В сумме получим:

$$f_2(s_2 = 2, X_2 = 0) + F_3^*(s_2 = 2) = 0 + 7 = 7.$$

Во втором варианте при $X_2 = 1$ имеем

$$f_2(s_2 = 1, X_2 = 1) = 6; F_3^*(s_2 = 1) = 4,$$

Так как последнее значение соответствует условным оптимальным прибылям третьего шага $F_3^*(s_2)$ при $s_2 = 1$. Это вторая строка пятого столбца.

В сумме получим:

$$f_2(s_2 = 1, X_2 = 1) + F_3^*(s_2 = 1) = 6 + 4 = 10.$$

В третьем варианте при $X_2 = 2$ имеем

$$f_2(s_2 = 0, X_2 = 2) = 9; F_3^*(s_2 = 0) = 0$$

(так как средств для третьего и четвертого предприятий не осталось). В сумме получим:

$$f_2(s_2 = 0, X_2 = 2) + F_3^*(s_2 = 0) = 9 + 0 = 9.$$

Эти результаты приведены в третьей строке седьмого столбца.

По этой же методике проводятся расчеты для $s_{k-1} = s_1 = 3, s_{k-1} = s_1 = 5$. Результаты расчетов приведены в четвертой, пятой и шестой строках седьмого столбца.

В восьмом столбце табл. 2 приведены максимальные значения из седьмого столбца, в девятом – управления, пришедшие к максимальным условным оптимальным прибылям.

Первый шаг. Формула для расчета условной оптимальной прибыли (1) для первого шага имеет вид

$$F_1^*(s_0) = \max_{\{x_1\}} \{f_1(s_1, X_1) + F_2^*(s_1)\}.$$

Расчет для первого шага проведен аналогично предыдущему. Результаты занесены в табл. 2.

Анализ результатов первого шага показывает, что максимальная прибыль от вложения денег в четыре предприятия

$$F_{\max} = F_1^*(s_0 = 5) = 24 \text{ ден. ед.}$$

при оптимальном вложении в первое предприятие $X_1^* = X_1^*(s_0 = 5) = 1 \text{ ден. ед.}$

Как следует из уравнения состояния (1), оптимальное количество средств, оставшихся в конце первого шага,

$$s_1 = s_{1-1} - X_1^* = 5 - 1 = 4 \text{ ден. ед.}$$

Это соответствует пятой строке табл. 2. Оставшиеся три предприятия принесут прибыль 16 ден. ед. при вложении во второе предприятие 2 ден. ед., т.е. $X_2^* = X_2^*(s_1 = 4) = 2 \text{ ден. ед.}$ Далее находим $s_2 = s_{2-1} = X_2^* = 4 - 2 = 2 \text{ ден. ед.}$ Это соответствует второй строке табл. 2. Оставшиеся два предприятия принесут прибыль 7 ден. ед. при вложении в третье предприятие 1 ден. ед., т.е.

$$X_3^* = X_3^*(s_2 = 1) = 1 \text{ ден. ед.}$$

Оставшаяся 1 ден. ед. вкладывается в четвертое предприятие, которое принесет прибыль 4 ден. ед.

Таким образом, оптимальное решение задачи имеет вид

$$X^* = X^*(1; 2; 1; 1) \text{ при } F_{\max} = 24 \text{ ден. ед.}$$

Достоинством рассмотренного метода является возможность анализа на чувствительность к изменению количества выделенных средств и количества предприятий.

Пример 2. В условиях примера 1 положить количество выделенных средств в размере $s_0 = 4$ ден. ед.

Решение

Приведем анализ табл. 2, исключив при этом последнюю строку. Максимальная прибыль от вложения денег в четыре предприятия в данном случае

$$F_{\max} = F_1(s_0 = 4) = 21 \text{ ден. ед.}$$

при оптимальном вложении в первое предприятие $X_1^* = X_1^*(s_0 = 4) = 1$ ден. ед.

Оптимальное количество средств, оставшихся в конце первого шага,

$$s_1 = s_{1-1} - X_1^* = 4 - 1 = 3 \text{ ден. ед.}$$

Это соответствует четвертой строке табл. 2. Оставшиеся три предприятия принесут прибыль 13 ден. ед. при вложении во второе предприятие 1 или 2 ден. ед.:

$$X_{2,1}^* = X_{2,1}^*(s = 3) = 1 \text{ ден. ед. или } X_{2,2}^* = X_{2,2}^*(s_1 = 3) = 2 \text{ ден. ед.}$$

Далее находим

$$s_{2,1} = s_{2-1} - X_{2,1}^* = 3 - 1 = 2 \text{ ден. ед. или } s_{2,2} = s_{2-1} - X_{2,2}^* = 3 - 2 = 1.$$

Первое значение $s_{2,1} = 2$ оставшиеся два предприятия принесут прибыль 7 ден. ед. при вложении в третье предприятие 1 ден. ед., т.е. $X_{3,1}^* = X_{3,1}^*(s_{2,1} = 1) = 1$ ден. ед. Оставшаяся 1 ден. ед. вкладывается в четвертое решение задачи для $s_{2,1} = 2$ имеет вид

$$X^* = X^*(1; 1; 1; 1) \text{ при } F_{\max} = 21 \text{ ден. ед.}$$

Для $s_{2,1} = 1$ оставшиеся два предприятия принесут прибыль 4 ден. ед., причем в третье предприятие вкладывается 0 ден. ед., т.е.

$$X_{3,2} = X_{3,2}^* = (s_{2,2} = 0) = 0 \text{ ден. ед.}$$

Оставшаяся 1 ден. ед. вкладывается в четвертое предприятие, которое принесет прибыль 4 ден. ед. Оптимальное решение задачи для $s_{2,2} = 1$ имеет вид

$$X^* = X^*(1; 2; 0; 1) \text{ при } F_{\max} = 21 \text{ ден. ед.}$$

Пример 3. В условиях примера 1 положить количество выделенных средств в размере $s_0 = 6$ ден. ед.

Решение

При увеличении средств на 1 ден. ед. в табл. 2 достаточно добавить строку для $s_0 = s_{1-1} = 6$. Эта строка представлена в табл. 3. Заполним эту таблицу по рассмотренной выше методике.

Таблица 3

s_{k-1}	X_k	S_k	$k = 3$		
			$f_3(s_3, X_3) + F_4^*(s_3)$	$F_3^*(s_2)$	$X_3^*(s_2)$
1	2	3	4	5	6
6	0	6	0+16=16	22	5
	1	5	3+16=19		
	2	4	4+13=17		
	3	3	7+8=15		
	4	2	11+6=17		
	5	1	18+4=22		

Продолжение таблицы 3

s_{k-1}	X_k	S_k	$k = 2$			$k = 1$		
			$f_2(s_2, X_2) + F_3^*(s_2)$	$F_2^*(s_1)$	$X_2^*(s_1)$	$f_1(s_1, X_1) + F_2^*(s_1)$	$F_1^*(s_0)$	$X_1^*(s_0)$
1	2	3	7	8	9	10	11	12
6	0	6	0+22=22	24	1	0+24=24	27	1
	1	5	6+18=24			8+19=27		
	2	4	9+13=22			10+16=26		
	3	3	11+9=20			11+13=24		
	4	2	13+7=20			12+10=22		
	5	1	15+4=19			18+6=24		

Анализ результатов первого шага показывает, что максимальная прибыль от вложения денег в четыре предприятия

$$F_{\max} = F_1^*(s_0 = 6) = 27 \text{ ден. ед.}$$

Это соответствует шестой строке табл. 2. Оставшиеся три предприятия принесут прибыль 19 ден. ед. при вложении во второе предприятие 1 ден. ед., т.е. $X_2^* = X_2^*(s_1 = 5) = 1$ ден. ед. Далее находим $s_2 = s_{2-1} - X_2 = 5 - 1 = 4$ ден. ед. Это соответствует пятой строке табл. 2. Оставшиеся два предприятия принесут прибыль 13 ден. ед. При этом вложение в третье предприятие равно 0 ден. ед., т.е. $X_3^* = X_3^*(s_2 = 4) = 0$ ден. ед. Оставшиеся 4 ден. ед. вкладываются в четвертое предприятие, которое принесет прибыль 13 ден. ед.

Таким образом, оптимальное решение задачи имеет вид

$$X^* = X^*(1; 1; 0; 4) \text{ при } F_{\max} = 27 \text{ ден. ед.}$$

Упражнение

В условиях примера 1 положить количество выделенных средств в размере а) $s_0 = 3$ ден.ед., б) $s_0 = 7$ ден.ед.

Вопросы для самопроверки

1. Что называют динамическим программированием?
2. Чем отличается динамическое программирование от линейного программирования?
4. Какой метод является одним из основных методов динамического программирования?
5. Когда экономический процесс является управляемым?
6. Что понимается под управлением?
7. Приведите примеры управляемых процессов.
8. Сформулируйте принцип оптимальности Р. Беллмана.
9. Что является основным условием, при котором принцип оптимальности является верным?
10. Что означает, что процесс управления должен быть без обратной связи?
11. Представьте принцип оптимальности в виде формулы Беллмана и прокомментируйте эту формулу.

12. Что называется условным оптимальным управлением на k – м шаге?
13. Какое требование называется «отсутствием последствия»?
14. Расскажите значение динамического программирования.
15. Приведите постановку задачи оптимальной стратегии замены оборудования и прокомментируйте ее.
16. Как решается задача оптимального распределения инвестиций между предприятиями?

§25. Элементы теории игр. Матричные игры

Ключевые слова: игра, матричная игра, равновесная ситуация, седловая точка в матричной игре, смешанная стратегия, правило доминирования

1. Предмет и задачи теории игр

В экономике сталкиваются с ситуациями, когда эффективность решений одних участников экономического процесса зависит от решений, принятых другими участниками. Например, объем выпускаемой предприятием продукции зависит от объема аналогичной продукции, выпускаемой другими предприятиями. При этом каждое предприятие сознательно стремится добиться наилучших результатов за счет других. Такие ситуации называются *конфликтными*. При этом ни одна из сторон не может полностью контролировать положение, так как все участники процесса принимают решения в условиях неопределенности. В реальных условиях нередко возникают ситуации, когда антагонизм отсутствует, но существуют противоречия. Например, для нормальной работы производства необходимо наличие запасов, но их чрезмерное увеличение приводит к дополнительным затратам на содержание и хранение. Конфликтные ситуации возникают как в результате сознательной деятельности людей, так из-за их недостаточной информированности об условиях проведения планируемой операции.

Построением математических моделей конфликтных ситуаций и разработкой методов решения возникающих в этих ситуациях задач занимается теория и др.

Методы и рекомендации теории игр применимы к многократно повторяющимся конфликтным ситуациям. Если конфликтная ситуация реализуется ограниченное число раз, то рекомендации теории игр теряют смысл.

Игра – это упрощенная математическая модель конфликтной ситуации.

Игра ведется по определенным правилам. Суть игры состоит в том, что каждый участник принимает такое решение, которое, как он полагает, обеспечит ему наилучший исход. Исходом игры называется значение некоторой функции, называемой *функцией выигрыша (платежной функцией)*, которая может задаваться в матричном или аналитическом виде.

Величина выигрыша зависит от стратегии, принимаемой игроком.

Стратегия – это совокупность правил, однозначно определяющих последовательность действий игрока в каждой конкретной ситуации, складывающиеся в процессе игры.

Оптимальной называется стратегия, которая при многократном повторении игры обеспечивает данному игроку максимально возможный средний выигрыш. Количество стратегий у каждого игрока может быть конечным или бесконечным. В зависимости от этого игры подразделяются на *конечные* и *бесконечные*.

Игра состоит из отдельных партий.

Партия – это каждый вариант реализации игры.

Ход – это выбор и реализация игроком одного из допустимых вариантов поведения.

В игре могут сталкиваться интересы двух или нескольких игроков, поэтому игры бывают парные и множественные. В зависимости от взаимоотношений участников различают игры *некооперативные* (бескоалиционные), когда участники не имеют права заключать соглашения, и *кооперативные* (коалиционные).

В экономике нередко приходится сталкиваться с ситуацией, когда один из участников безразличен к результату игры. Такие игры называются *играми с природой*. Под термином «природа» понимают всю совокупность внеш-

них обстоятельств, в которых сознательному игроку приходится принимать решения. Например, определение объема выпуска сезонной продукции в ожидании наиболее выгодного для ее реализации уровня спроса. В качестве второго игрока здесь выступает уровень спроса. В играх с природой степень неопределенности для сознательного игрока возрастает. Поскольку «природа» безразлична к результатам игры, то она может реализовать такие состояния, которые невыгодны ей, а выгодны сознательному игроку.

2. Матричные игры

Пусть в игре два игрока. Игрок A имеет m стратегий, а игрок B – n стратегий.

Обозначим стратегии игрока A как $A_1, A_2, \dots, A_m$, а стратегии игрока B – как $B_1, B_2, \dots, B_n$.

Если игрок A выбрал стратегию A_i , а игрок B – стратегию B_k , то выигрыш игрока A составит a_{ik} , а игрока B – b_{ik} , причем

$$a_{ik} = -b_{ik}. \quad (1)$$

Поэтому при анализе такой игры достаточно рассмотреть выигрыш только одного игрока, например, выигрыш a_{ik} игрока A . Зная выигрыш a_{ik} по формуле (1) легко определить выигрыш b_{ik} .

Матричные игры называются *парными играми с нулевой суммой*, в которых выигрыш одного игрока равен проигрышу другого.

Если известны значения a_{ik} для каждой пары стратегий $\{A_i, B_k\}$, $i = 1, 2, \dots, m$, $k = 1, 2, \dots, n$, то их удобно записать в виде прямоугольной таблицы, строки которой соответствуют стратегиями игрока A , а столбцы – стратегиями игрока B (табл. 1).

Таблица 1

	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}

Чаще эти выигрыши записывают в виде платежной матрицы (матрицы игры) размера $m \times n$, поэтому такие игры называют *матричными играми* $m \times n$:

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

Матричные игры относятся к разряду антагонистических, т.е. игр, в которых интересы игроков противоположны.

Пример 1. Игроки A и B одновременно и независимо друг от друга записывают числа 1 или 2. Если записанные числа одинаковы, то игрок A выигрывает одно очко, если разные, то одно очко выигрывает игрок B . Построить платежную матрицу данной игры.

Решение

У игрока A две стратегии: A_1 - записать число 1 и A_2 - записать число 2. У игрока B также две стратегии: B_1 - записать число 1 и B_2 - записать число 2.

Матрица выигрышей игрока A в этой игре 2×2 имеет вид

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

В данной матрице строки соответствуют стратегиям игрока A , а столбцы – стратегиям игрока B .

Действительно, если игрок A делает ход A_1 , а игрок B - ход B_1 , то одно очко выигрывает игрок A и $a_{11} = 1$. Если игрок A делает ход A_1 , а игрок B - ход B_2 , то одно очко выигрывает игрок B , соответственно, игрок A это очко проигрывает, т.е. $a_{12} = -1$ и т.д.

В примере 1 рассмотрена простейшая математическая модель конечной конфликтной ситуации, когда имеются два участника и выигрыш одного ра-

вен проигрышу другого. Такая модель называется *антагонистической игрой двух лиц с нулевой суммой*.

3. Равновесная ситуация

Пусть матричная игра $m \times n$ задана платежной матрицей

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2k} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ik} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mk} & \dots & a_{mn} \end{pmatrix}. \quad (2)$$

Строки этой матрицы соответствуют стратегиям игрока A , а столбцы – стратегиям игрока B . В теории игр предполагается, что оба игрока действуют разумно, т.е. стремятся к получению максимального выигрыша, считая, что соперник действует наилучшим для себя образом.

Определим оптимальные стратегии каждого из игроков. Начнем с анализа стратегий игрока A . На стратегию A_i игрока A игрок B отметит такой стратегией B_k , при которой выигрыш игрока A будет минимальным. Аналогично игрок B будет отвечать на все m стратегий игрока A . Другими словами, найдем в каждой строке матрицы минимальный элемент (минимальные выигрыши игрока A):

$$\alpha_i = \min_k a_{ik}, i = 1, 2, \dots, m$$

и запишем их в правом столбце табл. 2.

Таблица 2

	B_1	B_2	...	B_k	...	B_n	Минимальные выигрыши игрока A
A_1	a_{11}	a_{12}	...	a_{1k}	...	a_{1n}	α_1
A_2	a_{21}	a_{22}	...	a_{2k}	...	a_{2n}	α_2
...	...	...	...	...	...	...	...
A_i	a_{i1}	a_{i2}	...	a_{ik}	...	a_{in}	α_n
...	...	...	...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mk}	...	a_{mn}	α_m
Максимальные выигрыши игрока A	β_1	β_2	...	β_k	...	β_n	

Действуя разумно игрок A остановится на той стратегии A_i , для которой a_i окажется максимальным. Поэтому среди чисел $\alpha_i = \max_k a_{ik}$ выбираем максимальное число

$$\alpha = \max_i \alpha_i = \max_i \max_k a_{ik}. \quad (3)$$

Число α называется *нижней ценой игры*.

Принцип построения стратегии игрока A , основанный на максимизации минимальных выигрышей, называется принципом максимина ($\max \min$).

Проведем анализ стратегий игрока B . Для этого найдем в каждом столбце матрицы максимальный элемент (максимальные выигрыши игрока A):

$$\beta_k = \max_i a_{ik}, \quad k = 1, 2, \dots, n$$

и запишем их в нижней строке табл. 2. Действуя разумно, игрок B остановится на той стратегии B_k , для которых β_k окажется минимальным. Поэтому среди чисел $\beta_k = \max_i a_{ik}$ выбираем минимальное число

$$\beta = \min_k \beta_k = \min_k \max_i a_{ik}. \quad (4)$$

Число β называется *верхней ценой игры*.

Принцип построения стратегии игрок B , основанные на минимизации максимальных выигрышей. Называется принципом минимакса ($\min \max$).

Нижняя цена игры α и верхняя цена игры β связаны неравенством

$$\alpha \leq \beta. \quad (5)$$

Если $\alpha = \beta = a_{i, \text{opt}, k, \text{opt}}$ или

$$\max_k a_{ik} = \min_k \max_i a_{ik} = a_{i, \text{opt}, k, \text{opt}}, \quad (6)$$

то ситуация $(A_{i, \text{opt}}, B_{k, \text{opt}})$ оказывается равновесной, и ни один игрок не заинтересован в том, чтобы ее нарушить. В том случае, когда верхняя цена игры равна нижней, их называют просто *ценой игры*.

Если $\alpha = \beta$, то такую игру называют также *игрой с седловой точкой*, а пара оптимальных стратегий $(A_{i, \text{opt}}, B_{k, \text{opt}})$ - *седловой точкой матрицы*. Цена игры обозначается буквой v . Тогда $v = a_{i, \text{opt}, k, \text{opt}}$.

Седловых точек в матричной игре может быть несколько, но все они имеют одно и то же значение.

Заметим, что Седловый элемент $a_{i, \text{opt}, k, \text{opt}}$ является наименьшим в i -й строке i_{opt} , содержащей Седловый элемент, и наибольшим в k -м столбце k_{opt} , также содержащем Седловый элемент:

$$a_{i, k, \text{opt}} \leq a_{i, \text{opt}, k, \text{opt}} \leq a_{i, \text{opt}, k} \quad (7)$$

Пример 2. Игроки A и B записывают одну из трех цифр: 1, 2 или 3. Затем сравнивают эти цифры и расплачиваются друг с другом так, как показано в платежной матрице размера 3×3 :

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

Определить оптимальные стратегии.

Решение

Здесь строки соответствуют стратегиям игрока A , а столбцы – стратегиям игрока B .

Стратегии игрока A : $A_1 = 1$; $A_2 = 2$; $A_3 = 3$.

Стратегии игрока B : $B_1 = 1$; $B_2 = 2$; $B_3 = 3$.

Определим оптимальные стратегии каждого из игроков. Начнем с анализа стратегий игрока A . Игрок A анализирует свою стратегию на максимум (max min), т.е. на максимальный из своих минимальных выигрышей. На стратегию A_1 игрока A игрок B ответит стратегией B_1 , т.е. той стратегией, при которой выигрыш игрока A будет минимальным (выигрыш – 2 игрока A означает его проигрыш +2). На стратегию игрока A игрок B ответит стратегией B_2 (минимальный выигрыш игрока A равен 1), на стратегию A_3 - стра-

гией B_2 (минимальный выигрыш игрока A равен -3). Запишем минимальные выигрыши игрока A в правом столбце табл. 3.

Таблица 3

	B_1	B_2	B_3	Мин.-й выигрыши игрока A
A_1	-2	2	-1	-2
A_2	2	1	1	1
A_3	3	-3	1	-3
Макс. -й выигрыши игрока	3	2	1	

Игрок A , проанализировав правый столбец таблицы, остановит свой выбор на стратегии A_2 , при которой его минимальный выигрыш максимален. Это максимальное значение называется максимин ($\max \min$). В данном случае $\max \min = 1$. Если игрок A будет придерживаться этой стратегии, то ему гарантирован выигрыш, не меньший 1, при любом поведении противника.

Аналогично проведем анализ стратегий игрока B . Игрок B анализирует свою стратегию на минимакс ($\min \max$), т.е. на минимальный из максимальных выигрышей игрока A . На стратегию B_1 игрока B игрок A ответит стратегией A_3 , т.е. той стратегией, при которой его выигрыш будет максимальным и равным 3. На стратегию B_2 игрока B игрок A ответит стратегией A_1 (его максимальный выигрыш равен 2), на стратегию B_3 – стратегией A_3 (его максимальный выигрыш равен 1). Запишем максимальные выигрыши игрока A в нижней строке табл. 3. В данном случае $\min \max = 1$. Если игрок B будет придерживаться этой стратегии, то он проиграет не больше 1 при любом поведении противника.

Таким образом, в рассматриваемой игре $\max \min$ и $\min \max$ совпали:

$$\max \min = \min \max = 1.$$

Таким образом, ситуация оказывается равновесной, матрица игры имеет седловую точку

$$\alpha = \beta = 1.$$

Упражнение

Найдите нижнюю и верхнюю цены и определите седловую точку игры, представленной матрицей:

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

4. Смешанные стратегии

Если платежная матрица не имеет седловой точки, т.е. $\alpha < \beta$, то поиск решения игры приводит к применению сложной стратегии с определенными частотами. Такая сложная стратегия называется *смешанной*.

В табл. 4 приведен пример, когда нижняя цена игры α не совпадает с верхней ценой игры β .

Таблица 4

	B_1	B_2	B_3	Минимальные выигрыши игрока A
A_1	4	1	-3	-3
A_2	-2	1	3	-2
A_3	0	2	-3	-3
Максимальные выигрыши игрока A	4	2	3	

Здесь $\alpha = -2$, а $\beta = 2$. Таким образом, игрок A может выиграть не менее -2 (знак «-» перед цифрой означает проигрыш игроком B двух единиц), а игрок B может ограничить свой проигрыш (выигрыш игрока A) двумя единицами. Область между числами -2 и 2 остается нейтральной, и каждый игрок может попытаться улучшить свой результат за счет области. Для компромиссного распределения разности $\beta - \alpha$ при многократном повторении игры игроками используется случайное применение своих чистых стратегий. Это обеспечивает наибольшую скрытность выбора стратегии, поскольку ре-

ультат выбора не может быть известен противнику, так как он неизвестен самому игроку.

Обратимся к общему случаю матричной игры, представленной в табл. 2. Обозначим через $p_1, p_2, \dots, p_m$ вероятности, с которыми игрок A использует в ходе игры свои чистые стратегии $A_1, A_2, \dots, A_m$. Для этих вероятностей выполняются условия

$$\sum_{i=1}^m p_i = 1; p_1 \geq 0, p_2 \geq 0, \dots, p_m \geq 0. \quad (8)$$

Вектор $\bar{p} = \bar{p}(p_1, p_2, \dots, p_m)$, проекция которого удовлетворяет условиям (8), полностью определяет характер игры игрока A и называется его *смешанной стратегией*. Механизм случайного выбора чистых стратегий, которым пользуется игрок A , обеспечивает ему бесконечное множество смешанных стратегий.

Аналогично, вектор $\bar{q} = \bar{q}(q_1, q_2, \dots, q_n \geq 0)$, проекция которого удовлетворяет условиям (9),

$$\sum_{k=1}^n q_k = 1; q_1 \geq 0, q_2 \geq 0, \dots, q_n \geq 0 \quad (9)$$

полностью определяет характер игры игрока B и называется *смешанной стратегией игрока B* . Игрок B , как и игрок A , располагает бесконечным множеством смешанных стратегий.

Пусть игроки A и B применяют смешанные стратегии $\bar{p}$ и $\bar{q}$, соответственно, т.е. игрок A использует стратегию A_i с вероятностью p_i , а игрок B - стратегию B_k с вероятностью q_k . Поскольку события A_i и B_k независимы, то вероятность появления комбинации (A_i, B_k) равна произведению вероятностей p_i и q_k , т.е. $p_i q_k$. При использовании смешанных стратегий игра приобретает случайный характер, случайными становятся и величины выигрышей игроков. Поэтому выигрыш игрока A (проигрыш игрока B) определяют его математическим ожиданием, рассчитываемым по формуле

$$E(A, \bar{p}, \bar{q}) = \sum_{i=1}^m \sum_{k=1}^n a_{ik} p_i q_k. \quad (10)$$

Функция (10) называется *платежной функцией игры с матрицей*, заданной табл.5.

Нижней ценой игры называется число α , рассчитываемое по формуле:

$$\alpha = \max_q \max_p E(A, \bar{p}, \bar{q}). \quad (11)$$

Верхней ценой игры называется число β , рассчитываемое по формуле:

$$\alpha = \max_q \max_p E(A, \bar{p}, \bar{q}). \quad (12)$$

Таблица 5

	B_1	B_2	...	B_k	...	B_n	
A_1	a_{11}	a_{12}	...	a_{1k}	...	a_{1n}	p_1
A_2	a_{21}	a_{22}	...	a_{2k}	...	a_{2n}	...
...	...	...	...	...	...	...	...
A_i	a_{i1}	a_{i2}	...	a_{ik}	...	a_{in}	p_i
...	...	...	...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mk}	...	a_{mn}	
Вероятности использования чистых стратегий игроком В	q_1	q_2	...	q_k	...	q_n	

Оптимальными смешанными стратегиями называются стратегии, удовлетворяющие соотношению (сравнить с формулой (6))

$$\max_q \min_p E(A, \bar{p}, \bar{q}) = \min_q \max_p E(A, \bar{p}, \bar{q}) = E(A, \bar{p}_{opt}, \bar{q}_{opt}). \quad (13)$$

Величину

$$v = E(A, \bar{p}_{opt}, \bar{q}_{opt}),$$

Определенную соотношением (13), называют *ценой игры*.

Дадим другое определение оптимальных смешанных стратегий.

Векторы $\bar{p}_{opt}$ и $\bar{q}_{opt}$ называются **оптимальными смешанными стратегиями**, если они образуют седловую точку платежной функции игры $E(A, \bar{p}, \bar{q})$, т.е. удовлетворяют неравенству (сравнить с неравенством (7))

$$E(A, \bar{p}, \bar{q}) \leq E(\bar{p}_{opt}, \bar{q}_{opt}) \leq E(A, \bar{p}_{opt}, \bar{q}). \quad (14)$$

Из отношения (14) следует, что в Седловой точке $(\bar{p}_{opt}, \bar{q}_{opt})$ платежная функция $E(A, \bar{p}, \bar{q})$ достигает максимума по смешанным стратегиям $\bar{p}$ игрока A и минимума по смешанным стратегиям $\bar{q}$ игрока B . Возникают два вопроса:

1. Какие матричные игры имеют решение в смешанных стратегиях?
2. Как находить решение матричной игры, если оно существует?

Ответы на данные вопросы дают две следующие теоремы.

Теорема 1. Основная теорема теории матричных игр (Дж. фон Нейман). Для матричной игры с любой матрицей A величины $\max_p \min_q E(A, p, q)$ и $\min_q \max_p E(A, p, q)$ существуют и равны между

$$\max_p \min_q E(A, p, q) = \min_q \max_p E(A, p, q).$$

Более того, существует хотя бы одна ситуация в смешанных стратегиях $(\bar{p}_{opt}, \bar{q}_{opt})$, для которой выполняется соотношение

$$E(A, \bar{p}_{opt}, \bar{q}_{opt}) = \max_p \min_q E(A, p, q) = \min_q \max_p E(A, p, q).$$

Теорема 2. Основные свойства оптимальных смешанных стратегий. Пусть

$\bar{p}_{opt} = \bar{p}_{opt}(p_{1,opt}, p_{2,opt}, \dots, p_{m,opt})$, $\bar{q}_{opt} = \bar{q}_{opt}(q_{1,opt}, q_{2,opt}, \dots, q_{m,opt})$ - оптимальные смешанные стратегии и $v = E(A, \bar{p}_{opt}, \bar{q}_{opt})$ - цена игры.

Оптимальная смешанная стратегия $\bar{p}_{opt}$ игрока A складывается только из тех чистых стратегий $A_i, i = 1, 2, \dots, m$ (т.е. только те вероятности $p_i, i = 1, 2, \dots, m$) могут отличаться от нуля), для которых

$$\sum_{k=1}^n a_{ik} q_{k,opt} = v.$$

Аналогично, только те вероятности q_k , $k = 1, 2, \dots, n$, могут отличаться от нуля, для которых

$$\sum_{i=1}^m a_{ik} p_{i, \text{opt}} = v.$$

Имеют место соотношения

$$v = \min_{1 \leq k \leq n} \sum_{i=1}^m a_{ik} p_{i, \text{opt}} = \max_p \min_{1 \leq k \leq n} \sum_{i=1}^m a_{ik} p_i = \min_q \max_{1 \leq i \leq m} \sum_{k=1}^n a_{ik} q_k = \max_{1 \leq i \leq m} \sum_{k=1}^n a_{ik} q_{k, \text{opt}} = v$$

Рассмотрим методы решения некоторых матричных игр.

5. Графические решения матричных игр

Графический метод применим к тем играм, в которых хотя бы один игрок имеет две стратегии.

Рассмотрим игру $2 \times n$, представленную в табл. 6. Эта игра не имеет седловой точки. Согласно теореме 2. имеем

$$v = \min_{1 \leq k \leq n} \sum_{i=1}^2 a_{ik} p_{i, \text{opt}} = \min_{1 \leq k \leq n} (a_{1k} p_{\text{opt}} + a_{2k} (1 - p_{\text{opt}})) = \max_p \min_{1 \leq k \leq n} (a_{1k} p + a_{2k} (1 - p)). \quad (15)$$

Таблица 6

	B_1	B_1	...	B_l	...	B_l	Вероятности использования чистых стратегий игроком A
A_1	a_{11}	a_{12}	...	a_{1k}	...	a_{1n}	p
A_1	a_{21}	a_{22}	...	a_{2k}	...	a_{2n}	$1 - p$
Вероятности использования чистых стратегий игроком B	q_1	q_2	...	q_k	...	q_n	

Максимум функции

$$\min_{1 \leq k \leq n} (a_{1k} p + a_{2k} (1 - p)) \quad (16)$$

найдем, построив ее график. Для этого поступаем следующим образом. Построим графики прямых

$$w_k = a_{1k}p + a_{2k}(1-p) = (a_{1k} - a_{2k})p + a_{2k} \quad (17)$$

для каждого $k = 1, 2, \dots, n$ в системе координат pO (рис. 1). В соответствии с требованием (16) на каждой из построенных прямых определяются и отмечаются наименьшие значения. На рис. 2 эти значения выделены полужирной ломаной линией. Эта ломанная огибает снизу все семейство построенных прямых и называется *нижней огибающей семейства*.

В соответствии с (15) цену игры v определяет верхняя точка построенной нижней огибающей. Координаты этой точки являются оптимальной стратегией игрока A :

$$\bar{p}_{opt} = \bar{p}_{opt}(p_{opt}; (1 - p_{opt})).$$

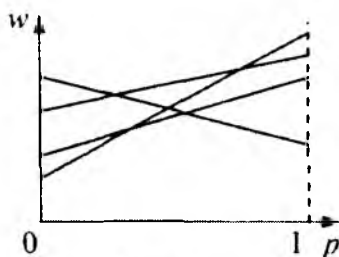


Рис. 1

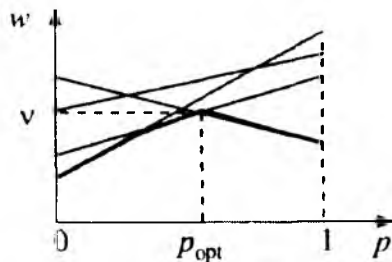


Рис. 2

Пример 3. Найти решение игры вида $2 \times n$, приведенной в табл. 7.

Таблица 7

	B_1	B_1	B_1	B_1	B_1	B_1	Вероятности использования чистых стратегий игроком A
A_1	6	4	3	1	-1	0	p
A_1	-2	-1	1	0	5	4	$1-p$
Вероятности использования чистых стратегий игроком B	q_1	q_2	q_3	q_4	q_5	q_6	

Решение

Проведем анализ игры на наличие седловой точки. Нижняя цена игры равна -1 , верхняя равна 1 . Седловой точки нет. Решение надо искать в смешанных стратегиях.

Построим график нижней огибающей (16). Предварительно запишем уравнения прямых (17):

$$w_1 = 6p - 2(1 - p) = 8p - 2 ;$$

$$w_2 = 4p - 1 \cdot (1 - p) = 5p - 1 ;$$

$$w_3 = 3p + (1 - p) = 2p + 1 ;$$

$$w_4 = p + 0 \cdot (1 - p) = p ;$$

$$w_5 = -p + 5(1 - p) = -6p + 5 ;$$

$$w_6 = 0 \cdot p + 4(1 - p) = -4p + 4.$$

Графики данных прямых, построенных в системе координат $p(w)$, представлены на рис. 3.

Нижняя огибающая выделена на рис. 3 полужирной ломаной линией. Точка максимума нижней огибающей лежит на пересечении прямых w_4 и w_5 .

Решая уравнение $p = -6p + 5$, получим $p_{opt} = \frac{5}{7}$. Цена игры, являющаяся математическим ожиданием выигрыша игрока A , равна $v = E(A, \bar{p}_{opt}, \bar{q}_{opt}) = \frac{5}{7}$.

Таким образом, цена игры и оптимальная стратегия игрока A равны:

$$v = \frac{5}{7}; \bar{p}_{opt} = \left(\frac{5}{7}; \frac{2}{7} \right).$$

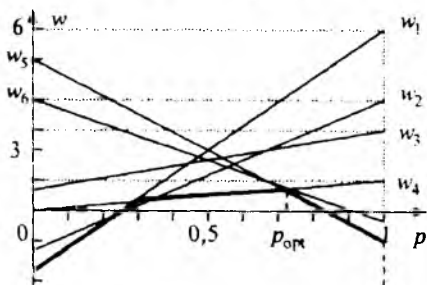


Рис. 3

Иногда решение матричной игры сводится только к поиску оптимальных смешанных стратегий игрока A . При этом стратегии противника могут

не интересовать исследователя. Однако в целом ряде случаев необходимо знать оптимальные смешанные стратегии обоих игроков.

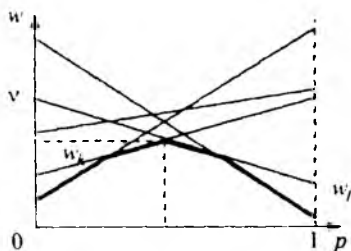


Рис. 4

Пусть в наивысшей точке нижней огибающей пересекаются прямые w_k и w_l (рис. 4), при этом прямая w_k - имеет положительный наклон, а прямая w_l - отрицательный. Оптимальная смешанная стратегия игрока B получается, если положить

$$q_k = q, q_l = 1 - q, q_j = 0 \text{ при } j \neq k, l,$$

где q находят из уравнения

$$a_{1k}q + a_{1l}(1 - q) = a_{2k}q + a_{2l}(1 - q).$$

Таким образом, игрок B применяет стратегию B_k с вероятностью $q_k = q$, а стратегию B_l - с вероятностью $q_l = 1 - q$.

Упражнение

Найдите решения следующих матричных игр:

а) $C = \begin{pmatrix} -5 & 8 \\ 4 & -7 \end{pmatrix}$.

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

в) $C = \begin{pmatrix} 4 & 2 & 3 & -1 \\ -4 & 0 & -2 & 2 \end{pmatrix}$.

г) $C = \begin{pmatrix} 1 & 4 \\ 3 & -2 \\ 0 & 5 \end{pmatrix}$.

Пример 4. Для условий примера 3 определить смешанные стратегии строки B .

Решение

В наивысшей точке нижней огибающей пересекаются прямые w_4 и w_5 (рис. 3), при этом прямая w_4 имеет положительный наклон, а прямая w_5 - отрицательный. Составим уравнение:

$$q - (1 - q) = 0 \cdot q + 5(1 - q) \text{ или } 7q = 6.$$

Отсюда находим

$$q_{opt} = \frac{6}{7}.$$

Таким образом, цены игры и оптимальная стратегия игрока B равны

$$v = \frac{5}{7}; \bar{q}_{opt} = \left(0; 0; 0; \frac{6}{7}; \frac{1}{7}; 0 \right).$$

Для решения игры $m \times 2$ (табл. 8) также может быть применен графический метод.

Таблица 8

	B_1	B_2	Вероятности использования чистых стратегий игроком B
A_1	a_{11}	a_{12}	p_1
A_2	a_{21}	a_{22}	p_2
...	...	...	...
A_i	a_{i1}	a_{i2}	p_i
...	...	...	...
A_m	a_{m1}	a_{m2}	p_m
Вероятности использования чистых стратегий игроком B	q	$q - 1$	

Эта игра не имеет седловой точки. Согласно теореме 2 имеем

$$v = \max_{1 \leq i \leq m} \sum_{k=1}^2 a_{ik} q_{k,opt} = \max_{1 \leq i \leq m} (a_{i1} q_{opt} + a_{i2} (1 - q)) = \min_q \max_{1 \leq i \leq m} (a_{i1} q + a_{i2} (1 - q)). \quad (18)$$

Графиком функции

$$\max_{1 \leq i \leq m} (a_{i1} q + a_{i2} (1 - q)) \quad (19)$$

является верхняя огибающая семейства прямых (15.20)

$$w_i = a_{i1}q + a_{i2}(1-q) = (a_{i1} - a_{i2})q + a_{i2}, \quad (20)$$

соответствующих чистым стратегиям игрока A (рис. 5).

Абсцисса нижней точки верхней огибающей семейства прямых (20) является оптимальной стратегией игрока B :

$$\bar{q}_{opt} = \bar{q}_{opt}(q_{opt}; (1 - q_{opt})),$$

а ордината – ценой игры v .

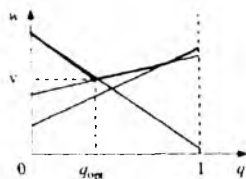


Рис. 5

Отыскание оптимальной смешанной стратегии игрока A проводится по той же схеме, которая позволяет находить оптимальную смешанную стратегию игрока B в игре $2 \times n$.

Пусть в самой нижней точке верхней огибающей пересекаются прямые w_i и w_j , при этом прямая w_i имеет отрицательный наклон, а прямая w_j – положительный. Оптимальная смешанная стратегия игрока A получается, если положить

$$p_i = p, \quad p_j = 1 - p, \quad p_r = 0 \quad \text{при } r \neq i, j,$$

где p находят из уравнения

$$a_{i1}p + a_{i1}(1-p) = a_{j2}p + a_{j2}(1-p).$$

Пример 5. Найти решение игры вида $m \times 2$, приведенной в табл. 9.

Таблица 9

	B_1	B_2	Вероятности использования чистых стратегий игроком A
A_1	3	-1	p_1
A_2	-1	3	p_2
A_3	1	0	p_3
Вероятности использования чистых стратегий игроком B	q	$1 - q$	

Решение

Проведем анализ игры на наличие седловой точки. Нижняя цена игры равна 0, верхняя цена игры равна 3. Седловой точки нет. Решение надо искать в смешанных стратегиях.

Построим график верхней огибающей (19), предварительно запишем уравнения прямых (20):

$$\begin{aligned}w_1 &= 3q - (1 - q) = 4q - 1; \\w_2 &= -q + 3(1 - q) = -4q + 3; \\w_3 &= q + 0 \cdot (1 - q) = q.\end{aligned}$$

Графики этих прямых, построенных в системе координат qwO , представлены на рис. 6.

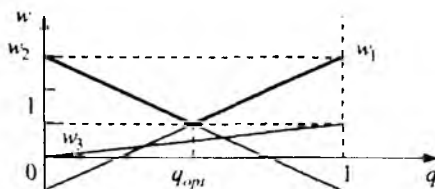


Рис. 6

Верхняя огибающая выделена на рис. 6 полужирной ломанной линией. Точка минимума верхней огибающей лежит на пересечении прямых w_1 и w_2 .

Решая уравнение $4q - 1 = -4 + 3$, получим $q_{opt} = \frac{1}{2}$. Цена игры являющаяся математическим ожиданием выигрыша игрока B , равна

$$v = E\left(A, \bar{p}_{opt}, \bar{q}_{opt}\right) = 4q_{opt} - 1 = 4 \cdot \frac{1}{2} - 1 = 1.$$

Найдем оптимальную стратегию игрока A . В самой нижней точке верхней огибающей пересекаются прямые w_1 и w_2 (рис. 6), при этом прямая w_1 имеет положительный наклон, а прямая w_2 — отрицательный. Составим уравнение:

$$-p + 3(1-p) = 3p - (1-p) \text{ или } 8p = 4.$$

Отсюда находим

$$p_{opt} = \frac{1}{2}.$$

Таким образом, цена игры и оптимальные стратегии игроков A и B равны

$$v = 1; \quad \bar{p}_{opt} = \left(\frac{1}{2}; \frac{1}{2}; 0\right); \quad \bar{q}_{opt} = \left(\frac{1}{2}; \frac{1}{2}\right).$$

Пример 6 (локальный конфликт). Рассмотрим войну между двумя государствами A и B , которая ведется в течение 30 дней.

Для бомбардировки моста, принадлежащего стране B , страна A использует два имеющихся у нее самолета. Разрушенный мост восстанавливается в течение суток, а каждый самолет совершает один полет по одному из двух возможных маршрутов, соединяющих эти страны. У страны B есть два зенитных орудия, при помощи которых можно сбивать самолеты страны A . Если самолет сбит, то страна A закупает новый самолет в течение суток у некоторой третьей страны.

Страна A может посылать самолеты либо по одному маршруту, либо по разным.

Страна B может поместить обе зенитки либо 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.

Таким образом, у каждой из стран есть две стратегии:

A_1 - послать самолеты по разным маршрутам;

A_2 - послать самолеты по одному маршруту;

B_1 - поместить зенитки на разных маршрутах;

B_2 - поместить зенитки на одном маршруте.

Если страна A выберет стратегию A_1 , а страна B - стратегию B_1 , то самолеты будут сбиты, т.е. вероятность разрушения моста равна нулю.

Если страна A выберет стратегию A_2 , а страна B - стратегию B_1 , то хотя бы один самолет достигнет цели, т.е. вероятность разрушения моста равна единице.

Если страна A выберет стратегию A_1 , а страна B - стратегию B_2 , то хотя бы один самолет достигнет цели, т.е. вероятность разрушения моста равна единице.

Если страна A выберет стратегию A_2 , а страна B - стратегию B_2 , то вероятность достижения цели равна 0,5, т.е. вероятность разрушения моста равна 0,5. Найти решения этой игры.

Решение

Матрица игры имеет вид

$$C = \begin{pmatrix} 0 & 1,0 \\ 1 & 0,5 \end{pmatrix}.$$

Решим задачу графическим методом. Проведем анализ игры на наличие седловой точки. Нижняя цена игры равна 0,5, верхняя равна 1.

Построим график нижней огибающей (16), предварительно запишем уравнения прямых (17):

$$w_1 = 0 \cdot p + (1 - p) = -p + 1,$$

$$w_2 = p + 0,5(1 - p) = 0,5p + 0,5.$$

График этих прямых, построенных в системе координат pOw , представлены рис. 7. Нижняя огибающая выделена на этом рисунке полужирной ломаной линией. Точка максимума нижней огибающей лежит на пересечении

двух прямых. Решая уравнение $-p + 1 = 0,5 + 0,5p$, получим $p_{opt} = \frac{1}{3}$. Цены игры, являющиеся математическим ожиданием выигрыша игрока A , равна

$$v = E(A, \bar{p}_{opt}, \bar{q}_{opt}) = w_{1, opt} = -\frac{1}{3} + 1 = \frac{2}{3}.$$

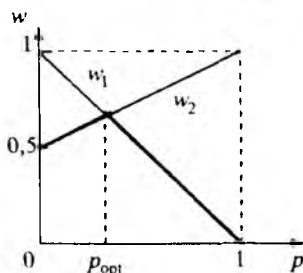


Рис. 7

Для определения оптимальной смешанной игры игрока B подставим в уравнение

$$a_{1k}q + a_{1l}(1-q) = a_{2k}q + a_{2l}(1-q).$$

Соответствующие значения из матрицы игры:

$$0 \cdot q + (1-q) = q + 0,5 \cdot (1-q) \text{ или } 1,5q = 0,5.$$

Отсюда находим $q = \frac{1}{3}$.

Таким образом, цена игры и оптимальные стратегии игроков A и B равны

$$v = \frac{2}{3}; \quad \bar{p}_{opt} = \left(\frac{1}{3}; \frac{2}{3}\right); \quad \bar{q}_{opt} = \left(\frac{1}{3}; \frac{2}{3}\right).$$

Это означает, что если страна A будет случайным образом посылать самолеты по разным маршрутам в течение десяти дней из тридцати, а по одному маршруту – в течение двадцати дней, то в среднем она будет иметь 66,7% удачных вылетов. Если страна B будет случайным образом помещать зенитки по разным маршрутам в течение десяти дней из тридцати, а по одно-

му маршруту – в течение двадцати дней, то в среднем она не позволит бомбить мост чаще, в 66,7% случаев.

6. Правило доминирования

Важным приемом, позволяющим уменьшить размеры платежной матрицы игры, является правило доминирования. Оно основано на отбрасывании тех чистых стратегий платежной матрицы, которые не вносят никакого вклада в искомые оптимальные смешанные стратегии. Отбрасывание подобных стратегий позволяет заменить первоначальную матрицу меньших размеров.

Одна из возможностей снижения размеров матрицы заключается в сравнении ее строк и столбцов.

Считают, что i -я строка называется *доминируемой*, а j -я строка – *доминирующей*. При этом говорят, что стратегия A_j игрока A доминирует стратегию A_i . Считают, что игрок A поступит разумно, если будет избегать стратегий, которым в матрице игры соответствуют доминируемые строки.

Считают также, что k -й столбец матрицы не меньше его l -го столбца, если одновременно выполняются неравенства

$$a_{1k} \geq a_{1l}, a_{2k} \geq a_{2l}, \dots, a_{nk} \geq a_{nl}.$$

В этом случае k -й столбец называется *доминируемым*, а l -й столбец – *доминирующим*. При этом говорят, что стратегия B_l игрока B доминирует стратегию B_k . Считают, что игрок B поступит разумно, если будет избегать стратегий, которым в матрице игры отвечают доминируемые столбцы.

Если в матрице игры одна из строк (один из столбцов) доминирует другую строку (другой столбец), то число строк (столбцов) в этой матрице можно уменьшить путем отбрасывания доминируемой строки (доминируемого столбца).

Оптимальные смешанные стратегии в игре с матрицей, полученной усечением исходной за счет доминируемых строк и столбцов, дадут опти-

мальной решение в исходной игре. Вероятности, соответствующие доминируемым чистым стратегиям, следует взять равными нулю.

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.

Пример 7. Заменить исходную матрицу

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

на матрицу выигрышей меньших размеров и решить игру.

Решение

Первая строка совпадает с последней, т.е. они дублируют друг друга. Поэтому одну из этих строк можно вычеркнуть. В результате получим

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

Первая строка доминирует вторую, поэтому вторую строку можно отбросить. Тогда матрица примет вид

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

В этой матрице четвертый столбец доминирует третий. В результате получим

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

Решим игру 2×3 графическим способом. Проведем анализ игры на наличие седловой и точки. Нижняя цена игры равна -1 , верхняя равна 0 . Седловой точки нет. Решение надо искать в смешанных стратегиях.

Построим график нижней огибающей (16). Предварительно запишем уравнения прямых (17):

$$w_1 = -p + 2(1 - p) = -3 + 2;$$

$$w_2 = 0 \cdot p + (1 - p) = -p + 1;$$

$$w_3 = p - 2(1 - p) = 3p - 2.$$

Графики этих прямых, построенных в системе координат p/w , представлены на рис. 8.

Решая уравнение $-3p + 2 = 3p - 2$, получим $p_{opt} = \frac{2}{3}$. Цена игры

$$v = -3p_{opt} + 2 = -3 \cdot \frac{2}{3} - 2 = 0.$$

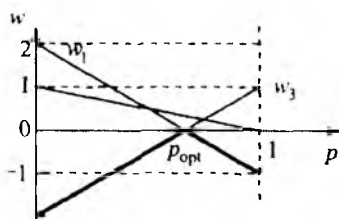


Рис. 8

В наивысшей точке нижней огибающей пересекаются прямые w_3 и w_1 (рис. 8), при этом прямая w_3 имеет положительный наклон, а прямая w_1 — отрицательный. Составим уравнение:

$$q - (1 - q) = -2q + 2(1 - q) \text{ или } 6q = 3.$$

Отсюда находим $q_{opt} = \frac{1}{2}$.

Таким образом, цена игры 2×3 и оптимальные стратегии игроков равны

$$v = 0; \quad \bar{p}_{opt} = \left(\frac{2}{3}; \frac{1}{3}\right); \quad \bar{q}_{opt} = \left(\frac{1}{2}; 0; \frac{1}{2}\right).$$

Возвращаясь к исходной игре 4×4 , получим окончательный ответ:

$$v = 0; \quad \bar{p}_{opt} = \left(\frac{2}{3}; 0; \frac{1}{3}; 0\right); \quad \bar{q}_{opt} = \left(\frac{1}{2}; 0; 0; \frac{1}{2}\right).$$

Таким образом, при использовании правила доминирования используется следующий алгоритм решения матричной игры:

1. Проводится проверка наличия или отсутствия равновесия в чистых стратегиях. При наличии равновесной ситуации указывается отсутствующие оптимальные стратегии игроков и цена игры.

2. Проводится поиск доминирующих стратегий и усечения матрицы.

3. Ищется цена игры и оптимальные смешанные стратегии.

Упражнение

Используя правило доминирования, найдите решение матричной игры:

$$C = \begin{pmatrix} 4 & 2 & 0 & 1 & 0 & 2 \\ 4 & 2 & 0 & 2 & 1 & 1 \\ 4 & 3 & 1 & 3 & 1 & 2 \\ 4 & 3 & 4 & -1 & 2 & 2 \\ 4 & 3 & 3 & -2 & 2 & 2 \end{pmatrix}$$

Вопросы для самопроверки

1. Определите предмет и задачи теории игр.
2. Дайте понятие матричных игр.
3. Когда достигается равновесная ситуация матричной игры?
4. Что такое седловая точка в матричной игре?
5. Какие стратегии матричной игры называются смешанными?
6. Назовите теоремы теории матричных игр.
7. Приведите алгоритм графического решения матричных игр.
8. Рассмотрите на примере процедуру снижения размеров матрицы при помощи правила доминирования.

§26. Сведение матричных игр к задачам линейного программирования

Ключевые слова: матричная игра, линейное программирование.

1. Схема сведения матричных игр к задачам линейного программирования

Рассмотрим один из методов нахождения решения игр путем сведения матричных игр к задачам линейного программирования.

При решении игр с помощью линейного программирования часто приходится использовать следующую теорему:

Теорема. Пусть $\overline{p_{opt}}$ и $\overline{q_{opt}}$ - оптимальные смешанные стратегии игроков A и B в игре $m \times n$ с матрицей (a_{jk}) и ценой игры v . Тогда $\overline{p_{opt}}$ и $\overline{q_{opt}}$ будут оптимальными в игре с матрицей $(ba_{jk} + c)$ и ценой игры $v' = bv + c$, где $b > 0$.

Рассмотрим случай игры $m \times n$, когда $\alpha \neq \beta$, а все элементы платежной матрицы $a_{jk} \geq 0$. Последнего условия всегда можно добиться, используя теорему 3 из предыдущей лекции. В этом случае и цена игры $v > 0$. Найдем сначала оптимальную смешанную стратегию $\overline{q_{opt}}$ игрока B . Применяя ее, игрок B проиграет не более v при любой частой стратегии A_i игрока A , т.е. равенство теоремы 2 из предыдущей лекции можно записать в виде неравенства:

$$\sum_{k=1}^n a_{ik} q_k \leq v, \quad i = 1, 2, \dots, m. \quad (1)$$

Разделив обе части неравенства (21) из предыдущей лекции на v , получим:

$$\sum_{k=1}^n a_{ik} \frac{q_k}{v} \leq 1, \quad i = 1, 2, \dots, m. \quad (2)$$

Введя обозначения

$$y_k = \frac{q_k}{v}, \quad k = 1, 2, \dots, n, \quad (3)$$

получим

$$\sum_{k=1}^n a_{ik} y_k \leq 1, \quad i = 1, 2, \dots, m. \quad (4)$$

$$y_k \geq 0, \quad k = 1, 2, \dots, n. \quad (5)$$

Кроме того, y_k удовлетворяют условию

$$\sum_{k=1}^n y_k = \sum_{k=1}^n \frac{q_k}{v} = \frac{1}{v} \sum_{k=1}^n q_k = \frac{1}{v}. \quad (6)$$

Игрок B стремится сделать свой проигрыш v как можно меньше, т.е. как можно больше величину

$$\varphi = \sum_{k=1}^n y_k = \frac{1}{v}. \quad (7)$$

Учитывая сказанное, приходим к следующей задаче линейного программирования:

$$\varphi = \sum_{k=1}^n y_k \rightarrow \max \quad (8)$$

при условиях

$$\sum_{k=1}^n a_{ik} y_k \leq 1, \quad i = 1, 2, \dots, m, \quad (9)$$

$$y_k \geq 0, \quad k = 1, 2, \dots, n. \quad (10)$$

Решая эту задачу, находим оптимальный вектор

$$\overline{y_{opt}} = (y_{1,opt}, y_{2,opt}, \dots, y_{n,opt})$$

и максимальное значение целевой функции $\varphi_{opt} = \varphi_{\max}$, а затем, используя формулы (7) и (3), цену игры и компоненты оптимальной смешанной стратегии:

$$v = \frac{1}{\varphi_{\max}}; \quad q_{k,opt} = v y_{k,opt}, \quad k = 1, 2, \dots, n. \quad (11)$$

Аналогично приходим к задаче линейного программирования для нахождения оптимальной смешанной стратегии $\overline{p_{opt}}$ игрока A :

$$\psi = \sum_{i=1}^m x_i \rightarrow \min \quad (12)$$

при условиях

$$\sum_{i=1}^n a_{ik} x_i \geq 1, k = 1, 2, \dots, n, \quad (13)$$

$$x_i \geq 0, i = 1, 2, \dots, m. \quad (14)$$

Решая эту задачу, находим оптимальный вектор $\overline{x_{opt}} = (x_{1,opt}, x_{2,opt}, \dots, x_{n,opt})$ и минимальное значение целевой функции $\psi_{opt} = \psi_{max}$, а затем цену игры и компоненты оптимальной смешанной стратегии:

$$v = \frac{1}{\psi_{max}}; p_{i,opt} = vx_{i,opt}, i = 1, 2, \dots, m^2. \quad (15)$$

2. Сведение одной матричной игры к задаче линейного программирования

Пример. Фирмы *A* и *B* производят однородный сезонный товар, пользующийся спросом *n* единиц времени. Доход от продажи товара в единицу времени составляют *C* ден. ед. Фирма *B*, будучи более состоятельной, в ходе конкурентной борьбы стремится вытеснить фирму *A* с рынка сбыта, способствуя своими действиями минимизации ее дохода, не считаясь при этом с временными потерями своего дохода в надежде наверстать упущенное в будущем. Действующее законодательство не позволяет прибегать к демпинговым ценам. Единственным допустимым способом достижения своих целей для фирм *A* и *B* остается повышение качества товара и выбор момента времени поставки его на рынок. Уровень спроса на товар и выбор момента времени поставки его на рынок. Уровень спроса на товар зависит от его качества, и в данный момент реализуется тот товар, качество которого выше. Повышение же качества требует дополнительных затрат времени на совершенствование технологии его изготовления и переналадки оборудования. В связи в этом будем предполагать, что качество товара тем выше, чем позже он поступает на рынок.

Придать описанной ситуации игровую схему и дать рекомендации фирмам *A* и *B* по оптимальным срокам поставки товара на рынок, обеспечи-

являющимся фирме A наибольший средний доход, а фирме B - наименьшие потери

Решение

Фирма A выбирает некий момент времени поставки товара на рынок (обозначим его через i) с целью максимизировать свой доход. Фирма B выбирает момент времени k поставки товара на рынок с целью минимизировать доход фирмы A .

Функцию выигрышей игрока A можно записать в виде

$$a_{ik} = \begin{cases} C(k-i), & i < k; \\ 0.5C(n-i+1), & i = k; \\ C(n-i+1), & i > k. \end{cases}$$

Действительно, если $i < k$, то фирма A , не имея конкурентов в течение $k-i$ единиц времени, получит за этот период доход $C(k-i)$ ден. ед. В момент k на рынке появляется товар фирмы B более высокого качества, и фирма A теряет рынок (первая строка в (16)).

Если $i = k$, то товар фирм A и B имеет одинаковое качество и реализуется с одинаковым спросом. Поэтому доходы фирм A и B за период времени $n-i+1$ равны друг другу и вычисляются по формуле $0.5C(n-i+1)$ (вторая строка в (16)).

Если $i > k$, то фирма A , предлагая товар более высокого качества, в течение $n-i+1$ единиц времени единолично получит за этот период доход, равный $C(n-i+1)$ ден. ед. (третья строка (16))

По формулам (16) можно построить платежную матрицу. Например, для количества единиц времени $n=5$ платежная матрица имеет вид, представленный в табл. 1.

Проведем анализ игры на наличие седловой точки. Нижняя цена игры равна C , верхняя равна $2C$. Седловой точки нет. Решение надо искать в смешанных стратегиях.

Таблица 1

	B_1	B_2	B_3	B_4	B_5	p_i
A_1	2,5C	C	2C	3C	4C	p_1
A_2	4C	2C	C	2C	3C	p_2
A_3	3C	3C	1,5C	C	2C	p_3
A_4	2C	2C	2C	C	C	p_4
A_5	C	C	C	C	0,5C	p_5
q_k	q_1	q_2	q_3	q_4	q_5	

Упростим платежную матрицу, умножим все элементы на $\frac{1}{C}$. В результате получим матрицу, представленную в табл. 2.

Таблица 2

	B_1	B_2	B_3	B_4	B_5	p_i
A_1	2,5	1	2	3	4	p_1
A_2	4	2	1	2	3	p_2
A_3	3	3	1,5	1	2	p_3
A_4	2	2	2	1	1	p_4
A_5	1	1	1	1	0,5	p_5
q_k	q_1	q_2	q_3	q_4	q_5	

В этой матрице первая строка доминирует пятую. Отпуская доминируемую строку, получим матрицу, представленную в табл. 3.

Таблица 3

	B_1	B_2	B_3	B_4	B_5	p_i
A_1	2,5	1	2	3	4	p_1
A_2	4	2	1	2	3	p_2
A_3	3	3	1,5	1	2	p_3
A_4	2	2	2	1	1	p_4
q_k	q_1	q_2	q_3	q_4	q_5	

В полученной матрице элементы второго столбца доминируют первый, а четвертый столбец доминирует пятый. Поэтому первый и пятый столбцы можно отпустить. Тогда получим матрицу (табл. 4).

	B_2	B_3	B_4	p_i
A_1	1	2	3	p_1
A_2	2	1	2	p_2
A_3	3	1,5	1	p_3
A_4	2	2	1	p_4
q_k	q_2	q_3	q_4	

В результате выполненных преобразований установлено, что в оптимальных смешанных стратегиях $p_{3,opt} = 0$, $q_{1,opt} = 0$ и $q_{3,opt} = 0$, а цена $v = C \cdot v'$,

где v' - цена игры с упрощенной матрицей табл. 4.

По табл. 4 составляем задачу линейного программирования (12) – (14):

$$\psi = x_1 + x_2 + x_3 + x_4 \rightarrow \min$$

при условиях

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 2x_4 \geq 1, \\ 2x_1 + x_2 + 1,5x_3 + x_4 \geq 1, \\ 3x_1 + 2x_2 + x_3 + x_4 \geq 1, \end{cases} \quad (17)$$

$$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0.$$

Приведем задачу к канонической форме:

$$\psi = x_1 + x_2 + x_3 + x_4 \rightarrow \min$$

при условиях

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 2x_4 - x_6 = 1, \\ 2x_1 + x_2 + 1,5x_3 + 2x_4 - x_7 = 1, \\ 3x_1 + 2x_2 + x_3 + x_4 - x_8 = 1, \end{cases} \quad (18)$$

$$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0; x_4 \geq 0; x_6 \geq 0; x_7 \geq 0; x_8 \geq 0.$$

В качестве базисных неизвестных принимаем x_6, x_7, x_8 . Тогда, производя в (18) перенос свободных неизвестных в правую часть, получим систему уравнений:

$$\begin{cases} x_6 = x_1 + 2x_2 + 3x_3 + 2x_4 - 1, \\ x_7 = 2x_1 + x_2 + 1,5x_3 + 2x_4 - 1, \\ x_8 = 3x_1 + 2x_2 + x_3 + x_4 - 1. \end{cases} \quad (19)$$

Составим первую симплекс-таблицу (табл. 5).

Таблица 5

Базисные неизвестные	Свободные члены	x_1	x_2	x_3	x_4	x_6	x_7	x_8
x_6	-1	-1	-2	-3	-2	1	0	0
x_7	-1	-2	-1	-1,5	-2	0	1	0
x_8	-1	-3	-2	-1	-1	0	0	1
ψ	0	-1	-1	-1	-1	0	0	0

Свободные члены являются отрицательными, поэтому первое базисное решение недопустимо. Выбираем первое уравнение (отмечено горизонтальной стрелкой) и заменяем основную переменную x_6 на переменную x_1 . Перестроим табл. 5. Для этого умножаем выделенную строку на $-\frac{1}{1}$ и записываем результат вместо этой строки табл. 6.

Умножаем первую строку табл. 6 на -2, складываем с соответствующими значениями второй строки табл. 5 и результат вписываем во вторую строку табл. 6. Заметим, что в столбце x_1 при этой операции появился ноль.

Таблица 6

Базисные неизвестные	Свободные члены	x_1	x_2	x_3	x_4	x_6	x_7	x_8
x_1	1	1	2	3	2	-1	0	0
x_7	1	0	3	4,5	2	-2	1	0
x_8	2	0	4	8	5	-3	0	1
ψ	1	0	1	2	1	-1	0	0

Умножаем первую строку табл. 6 на -3, складываем с соответствующими значениями третьей строки табл. 5 и результат вписываем в третью строку табл. 6.

Умножаем первую строку табл. 6 на 1, складываем с соответствующими значениями четвертой строки табл. 5 и результат вписываем в последнюю строку табл. 6.

В последней строке табл. 6 имеются три положительных коэффициента. Принимаем в качестве разрешающего столбец x_1 , отметим его вертикальной стрелкой. Положительными являются также числа во всех строках базисных неизвестных. Находим отношения:

$$-\frac{\alpha}{\alpha_3} = \frac{1}{3}, \quad -\frac{\beta}{\beta_3} = \frac{1}{4,5} = \frac{2}{9}, \quad -\frac{\gamma}{\gamma_3} = \frac{2}{8} = \frac{1}{4}.$$

Наименьшим из них является

$$-\frac{\beta}{\beta_3} = \frac{2}{9}.$$

Тот факт, что первым из базисных неизвестных обращается в ноль

x_6 , служит для нас сигналом к замене базиса $\{x_1, x_7, x_8\}$ на базис x_1, x_3, x_8 . Выделяем горизонтальной стрелкой строку при базисном неизвестном x_1 . Разрешающим элементом здесь является 4,5. Перестроим табл. 6. Для этого умножаем выделенную строку на дробь $\frac{1}{4,5} = \frac{2}{9}$ и записываем результат вместо этой строки в новую таблицу (табл. 7).

Таблица 7

Базисные неизвестные	Свободные члены	x_1	x_2	x_3	x_4	x_6	x_7	x_8
x_1	$\frac{1}{3}$	1	0	0	$\frac{2}{3}$	$\frac{1}{3}$	$-\frac{2}{3}$	0
x_2	$\frac{2}{9}$	0	$\frac{2}{3}$	1	$\frac{4}{9}$	$-\frac{4}{9}$	$\frac{2}{9}$	0
x_8	$\frac{2}{9}$	0	$-\frac{4}{3}$	0	$\frac{13}{9}$	$\frac{5}{9}$	$-\frac{16}{9}$	1
ψ	$\frac{5}{9}$	0	$-\frac{1}{3}$	0	$\frac{1}{9}$	$-\frac{1}{9}$	$-\frac{4}{9}$	0

Умножаем вторую строку табл. 7 на -3, складываем с соответствующими значениями первой строки табл. 6 и результат вписываем в первую строку табл. 7.

Умножаем вторую строку табл. 7 на -8, складываем с соответствующими значениями третьей строки табл. 6 и результат вписываем в третью строку табл. 7.

Умножаем вторую строку табл. 7 на -2, складываем с соответствующими значениями четвертой строки табл. 6 и результат вписываем в последнюю строку табл. 7.

В последней строке имеется положительный коэффициент в столбце x_2 . Выделяем этот столбец стрелкой. Выделяем стрелкой также строку x_3 .

Перестроим табл. 7. Для этого умножаем выделенную строку на дробь $\frac{9}{13}$ и записываем результат вместо этой строки в табл. 8.

Таблица 8.

Базисные неизвестные	Свободные члены	x_1	x_2	x_3	x_4	x_5	x_7	x_8
x_1	$\frac{3}{13}$	1	$\frac{8}{13}$	0	0	$\frac{1}{13}$	$\frac{2}{13}$	$-\frac{6}{13}$
x_3	$\frac{2}{13}$	0	$\frac{14}{13}$	0	0	$-\frac{8}{13}$	$\frac{22}{39}$	$-\frac{4}{9}$
x_4	$\frac{2}{13}$	0	$-\frac{12}{13}$	1	1	$\frac{5}{13}$	$-\frac{16}{13}$	$\frac{9}{13}$
ψ	$\frac{7}{13}$	0	$-\frac{3}{9}$	0	0	$-\frac{2}{13}$	$-\frac{4}{13}$	$-\frac{1}{13}$

Укажем вторую строку табл. 8 на $-\frac{2}{3}$, складываем с соответствующими значениями первой строки табл. 7.

Умножаем вторую строку табл. 8 на $-\frac{4}{9}$, складываем с соответствующими значениями третьей строки табл. 7.

Умножаем вторую строку табл. 8 на $-\frac{1}{9}$, складываем с соответствующими значениями четвертой строки табл. 6.

В табл. 8 последняя строка не имеет положительных чисел в последних семи столбцах. Значит, достигнуто оптимальное решение. Таким образом, оптимальное решение имеет вид

$$x_{1,opt} = \frac{3}{14}, x_{2,opt} = 0, x_{3,opt} = \frac{2}{13}, x_{4,opt} = \frac{2}{13}.$$

Минимальным значением целевой функции является свободный член в последней строке, т.е. $\psi_{opt} = \frac{7}{13}$.

Цена игры и компоненты оптимальной смешанной стратегии фирмы А определяются соотношениями (15):

$$v' = \frac{1}{\psi_{min}} = \frac{13}{7}; p_{1,opt} = v x_{1,opt} = \frac{13}{7} \cdot \frac{3}{17} = \frac{3}{7}; p_{2,opt} = 0;$$

$$p_{3,opt} = \frac{13}{7} \cdot \frac{2}{13} = \frac{2}{7}; p_{4,opt} = \frac{13}{7} \cdot \frac{2}{13} = \frac{2}{7}; p_{5,opt} = 0.$$

Теперь по табл. 4 составляем задачу линейного программирования (12)-(14):

$$\varphi = y_1 + y_2 + y_3 \rightarrow \max \quad (20)$$

при условиях

$$\begin{cases} y_2 + 2y_3 + 3y_4 \leq 1, \\ 2y_2 + y_3 + 2y_4 \leq 1, \\ 3y_2 + 1,5y_3 + y_4 \leq 1, \\ 2y_2 + 2y_3 + y_4 \leq 1, \\ y_1 \geq 0; y_2 \geq 0; y_3 \geq 0. \end{cases} \quad (21)$$

Задачу на максимум сведем к задаче на минимум, для чего целевую функцию умножим на -1. Приведем задачу к канонической форме:

$$\varphi = -y_1 - y_2 - y_3 \rightarrow \min \quad (22)$$

$$\begin{cases} y_2 + 2y_3 + 3y_4 + y_6 = 1, \\ 2y_2 + y_3 + 2y_4 + y_7 = 1, \\ 3y_2 + 1,5y_3 + y_4 + y_8 = 1, \\ 2y_2 + 2y_3 + y_4 + y_9 = 1, \\ y_2 \geq 0; y_3 \geq 0; y_4 \geq 0; y_6 \geq 0; y_7 \geq 0; y_8 \geq 0; y_9 \geq 0. \end{cases} \quad (23)$$

$$y_2 \geq 0; y_3 \geq 0; y_4 \geq 0; y_6 \geq 0; y_7 \geq 0; y_8 \geq 0; y_9 \geq 0.$$

В качестве базисных неизвестных принимаем y_6, y_7, y_8, y_9 . Тогда, произведя в (23) перенос свободных неизвестных в правую часть, получим систему уравнений:

$$\begin{cases} y_6 = 1 - y_2 - 2y_3 - 3y_4, \\ y_7 = 1 - 2y_2 - y_3 - 2y_4, \\ y_8 = 1 - 3y_2 - 1,5y_3 - y_4, \\ y_9 = 1 - 2y_2 - 2y_3 - y_4. \end{cases} \quad (24)$$

Сопоставим (24) с (19), получим следующие соответствия:

$$\begin{array}{cccccccc} x_1 & x_2 & x_3 & x_4 & x_6 & x_7 & x_8 & \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ y_6 & y_7 & y_8 & y_9 & y_2 & y_3 & y_4 & \end{array}$$

Из последней строки табл. 8 и приведенных соответствий находим решение двойственной задачи:

$$y_2 = \frac{2}{13}; y_3 = \frac{4}{13}; y_4 = \frac{1}{13}.$$

Определим компоненты оптимальной смешанной стратегий игрока B по формуле с учетом полученных выше результатов

$$q_{1, \text{opt}} = v \cdot y_{k, \text{opt}}, \quad k = 1, 2, \dots, n.$$

$$q_{1, \text{opt}} = 0; \quad q_{2, \text{opt}} = \frac{13}{7} \cdot \frac{2}{13} = \frac{2}{7}; \quad q_{3, \text{opt}} = \frac{1}{7}; \quad q_{4, \text{opt}} = \frac{1}{7}; \quad q_{5, \text{opt}} = 0.$$

Таким образом, фирма должна поставлять свой товар на рынок в первую, третью и четвертую единицы времени с вероятностями $\frac{3}{7}$; $\frac{2}{7}$ и $\frac{2}{7}$, соответственно, и совсем не поставлять во вторую и пятую единицы времени. В этом случае ожидаемый доход фирмы A будет равен $\frac{13}{7} C$. Фирме B следует поставлять свою продукцию на рынок во вторую, третью и четвертую единицы времени с вероятностями $\frac{2}{7}$; $\frac{4}{7}$; и $\frac{1}{7}$, соответственно, и совсем не поставлять в первую и пятую единицы времени. В этом случае ее ожидаемые потери не превысят $\frac{13}{7} C$.

Упражнение

Методом сведения игры к задаче линейного программирования найдите решение. Матрица игры имеет вид;

$$C = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 5 & 0 \\ 0 & 2 & 5 \end{pmatrix}.$$

Упражнение

Борьба за рынки. Фирмы A и B производят однородный товар, который могут сбыть на одном из двух рынков. Фирма B является более состоятельной. Фирмы выбирают для борьбы один из рынков независимо друг от друга. Если фирмы выберут один и тот же рынок, то в конкурентной борьбе фирма A потерпит поражение. Если фирмы выберут разные рынки, то фирма A , не встречая противодействия, захватит выбранный рынок. Пусть платежные матрицы игроков имеют вид:

$$A = \begin{pmatrix} -10 & 2 \\ 1 & -1 \end{pmatrix}; \quad B = \begin{pmatrix} 5 & -2 \\ -1 & 1 \end{pmatrix}.$$

Из этих матриц следует, что если фирмы выберут один и тот же рынок, то победа достанется фирме B . При выборе стратегий (A_1, B_1) выигрыш игрока B равен 5, а при выборе стратегий (A_2, B_2) его выигрыш равен 1. Это говорит о том, что первый рынок более выгоден, например, удобно расположен, хорошо посещаем и т. д. Проигрыш игрока A при выборе (A_1, B_1) равен -10, а при выборе стратегий (A_2, B_2) - -1. Если фирмы выберут разные рынки, то при выборе стратегий (A_1, B_2) выигрыш игрока A равен 2, а при выборе стратегий (A_2, B_1) его выигрыш равен 1. Здесь игрок A выигрывает больше на более выгодном рынке. Потери, которые несет при этом игрок B , оказываются прямо противоположными. Определить параметры равновесной ситуации.

Вопрос для самопроверки

1. Приведите алгоритм решения игр с помощью линейного программирования.

§ 27. Принятие решения в условиях риска и неопределенности

Ключевые слова: неопределенность – характеристика внешней среды, ситуация риска, многокритериальная задача, процедура (метод) свертывания критериев, аддитивная оптимизация, нормализация критериев, метод последовательных уступок, критерий ожидаемого значения, критерий предельного уровня, критерий наиболее вероятного события (исхода), критерий Лапласа, критерий Вальда, критерий Сэвиджа, критерий Гурвица, конфликтная ситуация, игра конфликтующих сторон (игроков).

1. Основные понятия

Рассмотренные задачи линейного программирования формулировались и решались в предположении наличия полной информации. Их можно отнести к совокупности задач принятия решений в условиях определенности. Например, в транспортной задаче стоимость перевозки (издержки) c_{ij} связанные с доставкой груза от i -го поставщика к j -му потребителю, считались фиксированной величиной. Если x_{ij} - оптимальное значение переменной, определяющей объем перевозок груза от i -го поставщика к j -му потребителю, то общий вклад в издержки от транспортировки грузов равен произведению c_{ij} , которое также является фиксированной величиной при заданном значении x_{ij} . В реальных экономических условиях приходится решать отдельные задачи при ограниченности, неточности исходной информации о самом объекте и внешней среде, в которой он функционирует и развивается.

При принятии управленческих решений о функционировании и развитии экономического объекта необходимо учитывать важную характеристику внешней среды - неопределенность.

Под **неопределенностью** следует понимать отсутствие, неполноту, недостаточность информации об объекте, процессе, явлении или неуверенность в достоверности информации. В условиях рыночной экономики существует множество источников возникновения неопределенности для различных эко-

номических объектов. Например, к основным источникам возникновения неопределенности на транспорте можно отнести следующие:

1. Существенная зависимость транспортного процесса от погодных условий. Например, погодные условия могут вызвать непредвиденные последствия в перевозках сельскохозяйственной продукции.

2. Наличие, кроме транспортного предприятия, других участников транспортного процесса — поставщиков грузов, потребителей грузов, ГАИ и др. Результат их влияния на транспортный процесс носит неопределенный и неоднозначный характер.

3. Наличие в работе автотранспорта элементов вероятности и случайности (надежность подвижного состава, неравномерность спроса на транспортные услуги во времени и др.).

4. Недостаточность, неполнота информации об объекте, процессе, явлении, по отношению к которому принимается решение: ограниченность в сборе и обработке информации, постоянная ее изменчивость.

5. Наличие в общественной жизни страны противоборствующих тенденций, столкновение противоречивых интересов.

6. Невозможность однозначной оценки объекта при сложившихся в данных условиях уровне и методах научного познания.

7. Относительная ограниченность сознательной деятельности лица, принимающего решение, существующие различия в социально-психологических установках, идеалах, намерениях, оценках, стереотипах поведения.

Неопределенность обуславливает появление ситуаций, не имеющих однозначного исхода (решения). Среди различных видов ситуаций, с которыми в процессе производства сталкиваются предприятия, особое место занимают ситуации риска.

Под **ситуацией риска** следует понимать сочетание, совокупность различных обстоятельств и условий, создающих обстановку того или иного вида деятельности. Ей сопутствуют три условия. Это

- наличие неопределенности;
- необходимость выбора альтернативы (отказ от выбора таковых является разновидностью альтернативы);
- возможность оценить вероятность осуществления выбираемых альтернатив.

Таким образом, если существует возможность количественно и качественно определить степень вероятности того или иного варианта, то это и будет ситуация риска.

Для того чтобы снять ситуацию риска, руководители предприятий вынуждены принимать решения и стремиться реализовать их. Этот процесс находит свое выражение в понятии «риск». Несмотря на то что риск объективно присутствует во всех сферах общественной жизни и в большинстве видов управленческой деятельности, обнаруживается, что понятие «риск» до сих пор не получило универсальной трактовки.

Следует упомянуть об экономическом риске применительно к процессам принятия решений в условиях неопределенности и риска, иными словами, в условиях дефицита информации или неуверенности в достоверности информации. В этом случае риск предстает в виде совокупности вероятных экономических, политических, нравственных и других положительных и неблагоприятных последствий, которые могут наступить при реализации выбранных решений. Определим риск как целенаправленные действия, в ходе которых имеется возможность количественно и качественно оценить вероятность достижения желаемого результата, неудачи и отклонения от цели (положительного или отрицательного свойства).

Процесс установления рыночных отношений в нашей стране порождает различные виды рискованных ситуаций, более того, в работе предприятий риск становится необходимым и обязательным его компонентом.

Чтобы проиллюстрировать различие между ситуациями, когда приходится принимать решения в условиях риска или в условиях неопределенно-

сти, рассмотрим задачу оптимального выбора ассортимента выпускаемой продукции.

В условиях риска доход c_j - от реализации единицы продукции j не является фиксированной величиной. Напротив, это случайная величина, точное числовое значение которой не известно, но описывается с помощью функции распределения $f(c_j)$. Часть дохода, $c_j x_j$, определяемая продукцией j , также случайная величина, если даже значение переменной x_j , определяющей уровень выпуска продукции j , задано.

В условиях неопределенности функция распределения $f(c_j)$ неизвестна. В действительности неопределенность не означает полного отсутствия информации о задаче. Например, известно, что c_j может принимать пять значений, но неизвестны вероятности этих значений. Эта ситуация рассматривается как принятие решений в условиях неопределенности.

Таким образом, с точки зрения полноты исходных данных определенность и неопределенность представляют два крайних случая, а риск определяет промежуточную ситуацию, в которой приходится принимать решение.

Степень неинформированности данных определяет, каким образом задача формализуется и решается.

При решении задач в условиях неопределенности внешней среды наиболее часто возникают две ситуации. При первой ситуации сама система препятствует принятию решений, например, задача составления графика выпуска на работу подвижного состава, занимающегося перевозкой сельхозпродукции, в зависимости от того, будет дождь или нет. В этой задаче природа будет восприниматься как «доброжелательный» противник.

Во второй ситуации возможно наличие конкуренции, когда два (или более) участника находятся в конфликте, и каждый стремится как можно больше выиграть у другого (других). Эта ситуация отличается от обычных

процессов принятия решений в условиях неопределенности тем, что лицу, принимающему решение, противостоит мыслящий противник. Теория, в которой рассматриваются задачи принятия решений в условиях неопределенности при наличии противника («доброжелательного» или мыслящего), известна как теория игр.

Теория игр принадлежит к наиболее молодым математическим дисциплинам. Ее возникновение относится к 1944 г., когда вышла в свет монография Неймана и Моргенштерна «Теория игр и экономического поведения». В дальнейшем теория игр превратилась в самостоятельное математическое направление, имеющее практическое приложение.

Теория игр - это теория математических моделей, интересы участников которых различны, причем они достигают своей цели различными путями. Столкновение противоположных интересов участников приводит к возникновению конфликтных ситуаций. Необходимость анализировать такие ситуации, в свою очередь, привела к возникновению теории игр, задачей которой является выработка рекомендаций по рациональному образу действия участников конфликта.

Чтобы исключить трудности, возникающие при анализе практических конфликтных ситуаций в результате наличия многих несущественных факторов, строится упрощенная модель ситуации. Такая модель называется игрой. Конфликтная ситуация в игровой модели развивается по определенным правилам.

2. Принятие решений в условиях полной определенности

Математические модели исследуемых явлений или процессов могут быть заданы в виде таблиц, элементами которых являются значения частных критериев эффективности функционирования системы, вычисленные для каждой из сравниваемых стратегий при строго заданных внешних условиях. Для рассматриваемых условий принятие решений может производиться:

- по одному критерию;

- по нескольким критериям.

Пример 1. Одной из фирм требуется выбрать оптимальную стратегию по обеспечению нового производства оборудованием. С помощью экспериментальных наблюдений были определены значения частных критериев функционирования соответствующего оборудования (a_{ij}), выпускаемого заводами-изготовителями. Рассмотрим данные для выбора оптимальной стратегии в условиях полной определенности:

Варианты оборудования (стратегии, решения)	Частные критерии эффективности оборудования			
	производи- тельность, д.е	стоимость оборудов. д.е	энергоем- кость, д.е	надежность, у.е
Оборудование завода 1, x_1	$a_{11} = 5$	$a_{12} = 7$	$a_{13} = 5$	$a_{14} = 6$
Оборудование завода 2, x_2	$a_{21} = 3$	$a_{22} = 4$	$a_{23} = 7$	$a_{24} = 3$
Оборудование завода 3, x_3	$a_{31} = 4$	$a_{32} = 6$	$a_{33} = 2$	$a_{34} = 4$

На основе экспертных оценок были также определены веса частных критериев λ_j , $j = 1, 2, 3, 4$:

Очевидно, выбор оптимальной стратегии (варианта оборудования) по одному критерию в данной задаче не вызывает затруднений. Например, если оценивать оборудование по надежности, то лучшим является оборудование завода 1 (стратегия x_1).

Выбор оптимального решения по комплексу нескольких критериев (в нашем примере - по четырем критериям) является **задачей многокритериальной**.

Один из подходов к решению многокритериальных задач управления связан с процедурой образования обобщенной функции $F(a_1; a_2; a_3; \dots; a_n)$ монотонно зависящей от критериев $a_1; a_2; a_3; \dots; a_n$. Данная процедура называется **процедурой (методом) свертывания критериев**. Существует несколько методов свертывания, например:

- метод аддитивной оптимизации;
- метод многоцелевой оптимизации и др.

Рассмотрим подробнее метод **аддитивной оптимизации**.

Пусть

$$F(a_j) = \sum_{j=1}^n \lambda_j a_{ij} . \quad (1)$$

Здесь выражение (1) определяет аддитивный критерий оптимальности. Величины λ_j являются весовыми коэффициентами, которые определяют в количественной форме степень предпочтения j -го критерия по сравнению с другими критериями. Другими словами, коэффициенты λ_j определяют важность j -го критерия оптимальности. При этом более важному критерию приписывается больший вес, а общая важность всех критериев равна единице, т. е.

$$\sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = 1, 2, \dots, n \quad (2)$$

Обобщенная функция цели (1) может быть использована для свертывания частных критериев оптимальности, если:

- частные (локальные) критерии количественно соизмеримы по важности, т. е. каждому из них можно поставить в соответствие

некоторое число λ_j , которое численно характеризует его важность по отношению к другим критериям;

- частные критерии являются однородными (имеют одинаковую размерность; в нашем примере критерии «стоимость оборудования» и «производительность оборудования» в условных денежных единицах будут однородными).

В этом случае для решения задачи многокритериальной оптимизации оказывается справедливым применение аддитивного критерия оптимальности.

Допустим, в примере 2.1 необходимо выбрать оптимальный вариант оборудования по двум однородным локальным критериям:

- производительность (д. е.);
- стоимость оборудования (д. е.).

На основе экспертных оценок были определены весовые коэффициенты этих двух частных критериев: $\lambda_1 = 0,667$; $\lambda_2 = 0,333$. Вычислим аддитивный критерий оптимальности для трех вариантов:

$$F_1(a_{1j}) = \lambda_1 a_{11} + \lambda_2 a_{12} = 0,667 \cdot 5 + 0,333 \cdot 7 = 5,666;$$

$$F_2(a_{2j}) = \lambda_1 a_{21} + \lambda_2 a_{22} = 0,667 \cdot 3 + 0,333 \cdot 4 = 3,333;$$

$$F_3(a_{3j}) = \lambda_1 a_{31} + \lambda_2 a_{32} = 0,667 \cdot 4 + 0,333 \cdot 6 = 4,666.$$

Очевидно, первый вариант оборудования по двум частным стоимостным критериям будет оптимальным, так как $F_{\min} = F_1(a_{1j}) = 5,666$. В примере 1 четыре локальных критерия не однородны, т.е. имеют различные единицы измерения. В этом случае требуется нормализация критериев. Под **нормализацией критериев** понимается такая последовательность процедур, с помощью которой все критерии приводятся к единому, безразмерному масштабу измерения. К настоящему времени разработано большое количество схем нормализации. Рассмотрим некоторые из них.

Определим максимум и минимум каждого локального критерия, т. е.

$$a_i^- = \max a_{ij}, \quad i = 1, 2, \dots, m; \quad (3)$$

$$a_j^- = \min a_{ij}, \quad i = 1, 2, \dots, m. \quad (4)$$

Выделим группу критериев a_j , $j = 1, 2, \dots, l$, которые максимизируются при решении задачи, и группу критериев a_j , $j = l + 1, l + 2, \dots, n$, которые минимизируются при решении задачи.

Тогда в соответствии с принципом максимальной эффективности нормализованные критерии определяются из следующих соотношений:

$$\hat{a}_{ij} = \frac{a_{ij}}{a_j^-}, \quad j = 1, 2, \dots, l; \quad (5)$$

$$\hat{a}_{ij} = 1 - \frac{a_{ij}}{a_j^+}, \quad j = l + 1, \dots, n, \quad (6)$$

или

$$\hat{a}_{ij} = \frac{a_{ij} - a_j^-}{a_j^+ - a_j^-}, \quad j = 1, 2, \dots, l; \quad (7)$$

$$\hat{a}_{ij} = \frac{a_j^+ - a_{ij}}{a_j^+ - a_j^-}, \quad j = l + 1, \dots, n. \quad (8)$$

Оптимальным будет тот вариант (стратегия), который обеспечивает максимальное значение функции цели:

$$F_i = \sum_{j=1}^n \lambda_j \hat{a}_{ij}, \quad i = 1, 2, \dots, m. \quad (9)$$

В соответствии с принципом минимальной потери нормализованные критерии определяются из соотношений

$$\hat{a}_{ij} = 1 - \frac{a_{ij}}{a_j^+}, \quad j = 1, 2, \dots, l; \quad (10)$$

$$\hat{a}_{ij} = \frac{a_{ij}}{a_j^-}, \quad j = l + 1, \dots, n, \quad (11)$$

или

$$\hat{a}_{ij} = \frac{a_j^+ - a_{ij}}{a_j^+ - a_j^-}, \quad j = 1, 2, \dots, l; \quad (12)$$

$$\hat{a}_{ij} = \frac{a_{ij} - a_j^-}{a_j^+ - a_j^-}, \quad j = l + 1, \dots, n. \quad (13)$$

При этом оптимальным будет тот вариант (стратегия), который обеспечивает минимальное значение функции цели (9).

Пример 2. Используя данные примера 1, определите оптимальную стратегию выбора оборудования из трех возможных ($m = 3$) с учетом четырех локальных критериев ($n = 4$).

Решение. Определим max и min каждого локального критерия:

$$a_1^+ = 5; \quad a_2^+ = 7; \quad a_3^+ = 7; \quad a_4^+ = 6.$$

При решении задачи максимизируются первый (производительность) и четвертый (надежность) критерии, а минимизируются второй (стоимость оборудования) и третий (энергоёмкость) критерии.

Исходя из принципа максимизации эффективности, нормализуем критерии:

$$\hat{a}_{11} = \frac{a_{11}}{a_1^+}; \quad \hat{a}_{11} = \frac{a_{11}}{a_1^+} = \frac{5}{5} = 1; \quad \hat{a}_{21} = \frac{a_{21}}{a_1^+} = \frac{3}{5} = 0,6; \quad \hat{a}_{31} = \frac{a_{31}}{a_1^+} = \frac{4}{5} = 0,8.$$

$$\hat{a}_{14} = \frac{a_{14}}{a_4^+}; \quad \hat{a}_{14} = \frac{a_{14}}{a_4^+} = \frac{6}{6} = 1; \quad \hat{a}_{24} = \frac{a_{24}}{a_4^+} = \frac{3}{6} = 0,5; \quad \hat{a}_{34} = \frac{a_{34}}{a_4^+} = \frac{4}{6} = \frac{2}{3};$$

$$\hat{a}_{12} = 1 - \frac{a_{12}}{a_2^+}; \quad \hat{a}_{21} = \frac{a_{12}}{a_2^+} = 1 - \frac{7}{7} = 0; \quad \hat{a}_{22} = 1 - \frac{a_{22}}{a_2^+} = 1 - \frac{4}{7} = \frac{3}{7}; \quad \hat{a}_{32} = 1 - \frac{a_{32}}{a_2^+} = 1 - \frac{6}{7} = \frac{1}{7}.$$

$$\hat{a}_{13} = 1 - \frac{a_{13}}{a_3^+}; \quad \hat{a}_{13} = 1 - \frac{a_{13}}{a_3^+} = 1 - \frac{5}{7} = \frac{2}{7}; \quad \hat{a}_{23} = 1 - \frac{a_{23}}{a_3^+} = 1 - \frac{7}{7} = 0; \quad \hat{a}_{33} = \frac{a_{33}}{a_3^+} = 1 - \frac{2}{7} = \frac{5}{7}.$$

Определим обобщенную функцию цели по каждому варианту:

$$F_1 = \lambda_1 \hat{a}_{11} + \lambda_2 \hat{a}_{12} + \lambda_3 \hat{a}_{13} + \lambda_4 \hat{a}_{14} = 0,4 \cdot 1 + 0,2 \cdot 0 + 0,1 \cdot \frac{2}{7} + 0,3 \cdot 1 \approx 0,729;$$

$$F_2 = \lambda_1 \hat{a}_{21} + \lambda_2 \hat{a}_{22} + \lambda_3 \hat{a}_{23} + \lambda_4 \hat{a}_{24} = 0,4 \cdot 0,6 + 0,2 \cdot \frac{3}{7} + 0,1 \cdot 0 + 0,3 \cdot 0,5 \approx 0,476;$$

$$F_3 = \lambda_1 \hat{a}_{31} + \lambda_2 \hat{a}_{32} + \lambda_3 \hat{a}_{33} + \lambda_4 \hat{a}_{34} = 0,4 \cdot 0,8 + 0,2 \cdot \frac{1}{7} + 0,1 \cdot \frac{5}{7} + 0,3 \cdot \frac{2}{3} \approx 0,603.$$

Оптимальным является первый вариант оборудования, так как

$$F_{\max} = F_1 = 0,729.$$

Рассмотренный подход к решению многокритериальных задач зачастую применяется при решении экономических задач, связанных с оценкой качества промышленной продукции и оценкой уровня технического совершенства технических устройств и систем по нескольким показателям.

Другим возможным методом решения многокритериальных задач является **метод последовательных уступок**. В начале критерии ранжируются и нумеруются в порядке убывания важности. Абсолютное значение коэффициентов важности λ_i на этом этапе не играет никакой роли. Оптимизируется первый по важности критерий a_1 , и определяется его экстремальное значение a_1^* . Затем назначается величина допустимого отклонения критерия от оптимального значения (уступка) Δa_1 , и ищется экстремальное значение второго по важности критерия a_2 при условии, что отклонение первого от оптимального значения не превзойдет величины уступки. Затем назначается уступка для второго критерия, и задача оптимизируется по третьему критерию и т.д. Таким образом, многокритериальная задача оптимизации заменяется последовательностью однокритериальных задач. Решение каждой предыдущей задачи используется при решении последующих для формирования дополнительных условий, состоящих в ограничении на величину уступки.

3. Принятие решений в условиях риска

Основными критериями оценки принимаемых решений в условиях риска являются:

- ожидаемое значение результата;
- ожидаемое значение результата в сочетании с минимизацией его дисперсии;

- известный предельный уровень результата;
- наиболее вероятное событие (исход) в будущем.

Критерий ожидаемого значения используется в случаях, когда требуется определить экстремальное значение (max или min) результативного показателя (прибыль, расходы, экономические потери и т.д.). Применение этого критерия рассмотрим на конкретном примере, связанном с постановкой задачи проведения ремонтно-профилактических воздействий автомобилей. Оптимальное количество ремонтных воздействий, определенное минимизацией суммарных затрат на заданной наработке L_k с учетом рисков пропуска отказов и выполнения лишних ТО, приравнивается к количеству ТО на указанном пробеге. Модель данной задачи является моделью вероятностного спроса на ремонты с мгновенным восстановлением. Здесь минимизируются суммарные издержки за пробег L_k , которые определяются затратами на плановый ремонт S_p , профилактику $S_{ТО}$ и незапланированный аварийный ремонт $S_{ш}$, рассматриваемый как штраф за пропуск отказа:

$$S = S_p + S_{ТО} + S_{ш} \rightarrow \min, \quad (14)$$

Составляющие суммарных затрат формулы (9.14) зависят от количества ремонтно-профилактических операций за наработку L_k , определяемых по формуле

$$n = \frac{L_k}{L_{от}}, \quad (15)$$

где $L_{от}$ - наработка до отказа.

Наработка до отказа - величина случайная, определяемая плотностью распределения $f(L_{от})$, $L_{от} < L_k$. В силу случайности $L_{от}$ величина n также будет случайной с плотностью распределения

$$f(n) = \left| -\frac{L_k}{n^2} \right| \cdot f\left(\frac{L_k}{n}\right). \quad (16)$$

Используя $f(n)$ как весовую функцию и выражая составляющие суммарных затрат через соответствующие стоимости из (14), получим

$$S = C_p \cdot n_p + \int_0^{n_p} C_{TO}(n_p - n) f(n) dn + \int_{n_p}^{\infty} C_{ш}(n - n_p) f(n) dn \rightarrow \min, \quad (17)$$

где C_p - средняя стоимость предупредительного (планового) ремонта; C_{TO} - средняя стоимость профилактики (или убыток от недоиспользования ресурса замененных при ТО деталей); $C_{ш}$ - ущерб (штраф) от пропуска отказа (или стоимость устранения аварийного отказа). Очевидно, $C_{ш} > C_{TO}$.

Интеграл (16) в пределах $[0, n_p]$ соответствует риску выполнения лишних ТО (избыточность затрат на ТО), а интеграл в пределах $[n_p, \infty)$ - риску пропуска аварийных отказов (избыточность затрат на ТР по потребности). Из уравнения (17) находим оптимальное количество ремонтов n_p на пробеге L_k (обычно L_k - пробег до КР). Далее, заменяя необходимые ремонты обслуживаниями, при которых выполняется комплекс операций по предупреждению отказов, включая предупредительные замены деталей, получим

$$L_{TO} = \frac{L_k}{n_p}. \quad (18)$$

Пример 3. Определить оптимальную периодичность ТО (у.е.) при $L_k = 200$ тыс. км, $C_{ш} = 69$, $C_p = 24$, $C_{TO} = 15$, если наработки до отказа имеют нормальное распределение с параметрами $\bar{L}_{OT} = 20$ тыс. км и $\sigma_L = 5$ тыс. км.

$$f(L_{OT}) = \frac{1}{\sigma_L \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{L_{OT} - \bar{L}_{OT}}{\sigma_L} \right)^2}. \quad (19)$$

Решение. Выполнив преобразование распределения (19) по формуле (15), получим ($n \geq 1$):

$$f(n) = \frac{1}{n^2 \sigma_L \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{L_1 / \sigma_L - \bar{L}_0 \tau}{\sigma_L} \right)^2} \quad (20)$$

После подстановки выражения (20) в (17) получим задачу оптимизации, для решения которой воспользуемся математическим пакетом EUREKA.

Решая задачу, получим оптимальную периодичность $L_{TO} = 15,3$ тыс.км при $n_p = 13,08$, которая обеспечивает минимальные суммарные издержки S .

Критерий ожидаемого значения позволяет получить достоверные оценки в случае, когда одно и то же решение приходится принимать достаточно большое число раз, так как замена математического ожидания выборочными данными правомерна лишь при большом объеме выборки.

Если необходимость в принятии решения встречается редко, то выборочное значение может значительно отличаться от математического ожидания, а применение критерия ожидаемых значений может приводит к ошибочным результатам. В таких случаях рекомендуется применять следующий вид:

$$\begin{aligned} M(X) + K \cdot D(X) &\rightarrow \min. \\ M(X) + K \cdot D(X) &\rightarrow \max. \end{aligned} \quad (21)$$

где X - случайная величина (например, суммарные издержки); $D(X)$ - дисперсия этой случайной величины; K - заданная постоянная.

Постоянную K иногда интерпретируют как уровень несклонности к риску. Считается, что K определяет «степень важности» дисперсии $D(X)$ по отношению к $M(X)$. Например, предприниматель, особенно остро реагирующий на большие отрицательные отклонения прибыли вниз от $M(X)$, может выбрать K много больше единицы. Это придает больший вес дисперсии и приводит к решению, уменьшающему большие потери прибыли.

Критерий предельного уровня не позволяет получить оптимальное решение, найти максимум прибыли и минимум расходов. Этот критерий дает

возможность определить приемлемый (допустимый) способ действий. Например, транспортная фирма распродает автомобили, бывшие в эксплуатации. По каждой модели автомобиля определенного возраста определяется лимитная цена, т.е, минимально допустимая цена продажи автомобиля. Продажа автомобилей по цене ниже лимитной приведет к убыточной работе транспортной фирмы. Это и есть предельный уровень, позволяющий транспортной фирме согласиться на первое же превышающее этот уровень предложение цены. Такой критерий не определяет оптимальное решение, поскольку одно из последующих предложений может оказаться более выгодным, чем принятое.

Одно из преимуществ критерия предельного уровня заключается в том, что для него нет необходимости задавать в явном виде плотность распределения случайных величин. В нашем примере случайная величина - рыночная цена автомобиля. Транспортная фирма располагает информацией о распределении рыночных цен на подобные автомобили в неявном виде. Иначе при полном отсутствии информации о распределении рыночных цен фирма установила бы предельные цены на автомобили очень высокими или, наоборот, очень низкими.

Критерий наиболее вероятного события (исхода) основан на преобразовании случайной ситуации в детерминированную путем замены случайной величины единственным значением, имеющим наибольшую вероятность реализации.

4. Принятие решений в условиях неопределенности

Неопределенность является характеристикой внешней среды (природы), в которой принимается управленческое решение о развитии (или функционировании) экономического объекта. Здесь будем рассматривать неопределенность «природы», вызванную отсутствием или недостатком информации о действительных условиях (факторах), при которых развивается объект управления. Внешняя среда («природа») может находиться в одном из множества возможных состояний. Это множество может быть конечным и

Пескконечным. Будем считать, что множество состояний конечно или по крайней мере количество состояний можно пронумеровать.

Пусть S_i - состояние «природы», при этом $i = 1, 2, \dots, n$, где n - число возможных состояний. Все возможные состояния известны, не известно только, какое состояние будет иметь место в условиях, когда планируется реализация принимаемого управленческого решения. Будем считать, что множество управленческих решений (планов) R_j также конечно и равно m . Реализация R_j плана в условиях, когда «природа» находится в S_i состоянии, приводит к определенному результату, который можно оценить, введя количественную меру. В качестве этой меры могут служить выигрыши от принимаемого решения (плана); потери от принимаемого решения, а также полезность, риск и другие количественные критерии.

Данные, необходимые для принятия решения в условиях неопределенности, обычно задаются в форме матрицы, строки которой соответствуют возможным действиям (управленческим решениям) R_j , а столбцы - возможным состояниям «природы» S_i .

Допустим, каждому R_j -му действию и каждому возможному S_i -му состоянию «природы» соответствует результат (исход), определяющий результат (выигрыш, полезность) при выборе j -го действия и реализации i -го состояния, - V_{ji} .

	S_1	S_2	...	S_i	...	S_n
R_1	V_{11}	V_{12}	...	V_{1i}	...	V_{1n}
R_2	V_{21}	V_{22}	...	V_{2i}	...	V_{2n}
...	...	...	...	...	...	...
R_j	V_{j1}	V_{j2}	...	V_{ji}	...	V_{jn}
...	...	...	...	...	...	...
R_m	V_{m1}	V_{m2}	...	V_{mi}	...	V_{mn}

(22)

Следовательно, математическая модель задачи принятия решений определяется множеством состояний $\{S_i\}$, множеством планов (стратегий) $\{R_j\}$ и матрицей возможных результатов (V_{ji}) . В качестве результатов в отдельных задачах рассматривается матрица рисков (r_{ji}) .

Риск - мера несоответствия между разными возможными результатами принятия определенных стратегий (действий).

Элементы матрицы рисков (r_{ji}) связаны с элементами матрицы полезностей (выигрышей) следующим соотношением:

$$r_{ji} = V_i - V_{ji}, \quad (23)$$

где $V_i = \max_j V_{ji}$ - максимальный элемент в столбце i матрицы полезностей.

Если матрица возможных результатов (V_{ji}) представляет собой матрицу потерь (затрат), то элементы матрицы рисков (r_{ji}) следует определять по формуле

$$r_{ji} = V_{ji} - V_i, \quad (24)$$

где $V_i = \min_j V_{ji}$ - минимальный элемент в столбце i матрицы потерь (результатов).

Таким образом, риск – это разность между результатом, который можно получить, если знать действительное состояние «природы», и результатом, который будет получен при j -й стратегии.

Матрица рисков дает более наглядную картину неопределенной ситуации, чем матрица выигрышей (полезностей).

Непосредственный анализ матриц выигрышей (V_{ji}) или рисков (r_{ji}) не позволяет в общем случае принять решение по выбору оптимальной стратегии (плана), за исключением тривиального случая, когда выигрыши при одной стратегии выше, чем при любой другой для каждого состояния «приро-

ны» (элементы матрицы выигрышей в некоторой строке больше, чем в любой из других). Другими словами, имеется в наличии «доминирующая» стратегия.

Для принятия решения в условиях неопределенности используется ряд критериев. Рассмотрим некоторые из них. Это критерий Лапласа, критерий Виллиса, критерий Сэвиджа, критерий Гурвица.

4.1. Критерий Лапласа

Этот критерий опирается на «принцип недостаточного основания» Лапласа, согласно которому все состояния «природы» S_i , $i = 1, 2, \dots, n$ полагаются равновероятными. В соответствии с этим принципом каждому состоянию S_i ставится вероятность q_i , определяемая по формуле

$$q_i = \frac{1}{n}. \quad (25)$$

При этом исходной может рассматриваться принятие решения в условиях риска, когда выбирается действие R_j , дающее наибольший ожидаемый выигрыш. Для принятия решения для каждого действия R_j вычисляют среднее арифметическое значение выигрыша:

$$M_j(R) = \frac{1}{n} \sum_{i=1}^n V_{ji}. \quad (26)$$

Среди $M_j(R)$ выбирают максимальное значение, которое будет соответствовать оптимальной стратегии R_j .

Другими словами, находится действие R_j^* , соответствующее

$$\max_{R_j} \left\{ \frac{1}{n} \sum_{i=1}^n V_{ji} \right\} \quad (27)$$

Если в исходной задаче матрица возможных результатов представлена матрицей рисков (r_{ji}) , то критерий Лапласа принимает следующий вид:

$$\min_{R_j} \left\{ \frac{1}{n} \sum_{i=1}^n r_{ji} \right\}. \quad (28)$$

Пример 4. Одно из транспортных предприятий должно определить уровень своих провозных возможностей так, чтобы удовлетворить спрос клиентов на транспортные услуги на планируемый период. Спрос на транспортные услуги не известен, но ожидается (прогнозируется), что он может принять одно из четырех значений: 10, 15, 20 или 25 тыс. т. Для каждого уровня спроса существует наилучший уровень провозных возможностей транспортного предприятия (с точки зрения возможных затрат). Отклонения от этих уровней приводят к дополнительным затратам либо из-за превышения провозных возможностей над спросом (из-за простоя подвижного состава), либо из-за неполного удовлетворения спроса на транспортные услуги. Ниже приводится таблица, определяющая возможные прогнозируемые затраты на развитие провозных возможностей:

Варианты провозных возможностей транспортного предприятия	Варианты спроса на транспортные услуги			
	1	2	3	4
1	6	12	20	24
2	9	7	9	28
3	23	18	15	19
4	27	24	21	15

Необходимо выбрать оптимальную стратегию.

Решение

Согласно условию задачи, имеются четыре варианта спроса на транспортные услуги, что равнозначно наличию четырех состояний «природы»: S_1 , S_2 , S_3 , S_4 . Известны также четыре стратегии развития провозных возможностей транспортного предприятия: R_1 , R_2 , R_3 , R_4 . Затраты на развитие провозных возможностей при каждой паре S_i и R_j заданы следующей матрицей:

$$V = \begin{pmatrix} 6 & 12 & 20 & 24 \\ 9 & 7 & 9 & 28 \\ 23 & 18 & 15 & 19 \\ 27 & 24 & 21 & 15 \end{pmatrix}.$$

Принцип Лапласа предполагает, что S_1, S_2, S_3, S_4 равновероятны.

Следовательно, $p_i = P\{S = S_i\} = \frac{1}{n} = \frac{1}{4} = 0,25; \quad i = 1, 2, 3, 4$ и ожидаемые за-

раты при различных действиях R_1, R_2, R_3, R_4 составляют:

$$W\{R_1\} = 0,25 \cdot (6 + 12 + 20 + 24) = 15,5;$$

$$W\{R_2\} = 0,25 \cdot (9 + 7 + 9 + 28) = 13,25;$$

$$W\{R_3\} = 0,25 \cdot (23 + 18 + 15 + 19) = 18,75;$$

$$W\{R_4\} = 0,25 \cdot (27 + 24 + 21 + 15) = 21,75.$$

Таким образом, наилучшей стратегией развития провозных возможностей в соответствии с критерием Лапласа будет R_2 .

4.2. Критерий Вальда (минимаксный или максиминный критерий)

Применение данного критерия не требует знания вероятностей состояний S_i . Этот критерий опирается на принцип наибольшей осторожности, поскольку он основывается на выборе наилучшей из наихудших стратегий R_j .

Если в исходной матрице (по условию задачи) результат V_{ji} представляет потери лица, принимающего решение, то при выборе оптимальной стратегии используется минимаксный критерий. Для определения оптимальной стратегии R_j необходимо в каждой строке матрицы результатов найти наибольший элемент $\max_i \{V_{ji}\}$, а затем выбирается действие R_j (строка j), которому будет соответствовать наименьший элемент из этих наибольших элементов, т.е. действие, определяющее результат, равный

$$W = \min_j \max_i \{V_{ji}\}. \quad (29)$$

Если в исходной матрице по условию задачи результат V_{ji} представляет выигрыш (полезность) лица, принимающего решение, то при выборе оптимальной стратегии используется максиминный критерий.

Для определения оптимальной стратегии R_j в каждой строке матрицы результатов находят наименьший элемент $\min_i \{V_{ji}\}$, а затем выбирается действие R_j (строка j), которому будут соответствовать наибольшие элементы из этих наименьших элементов, т.е. действие, определяющее результат, равный

$$W = \max_j \min_i \{V_{ji}\}. \quad (30)$$

Пример 5. Рассмотрим пример 4. Так как V_{ji} в этом примере представляет потери (затраты) применяем минимаксный критерий. Необходимые результаты вычисления приведены в следующей таблице:

R_j	Затраты, д.е. (V_{ji})				$\max_i \{V_{ji}\}$	$W = \min_j \max_i \{V_{ji}\}$
	S_1	S_2	S_3	S_4		
R_1	6	12	20	24	24	-
R_2	9	7	9	28	28	-
R_3	23	18	15	19	19	23
R_4	27	24	21	15	15	-

Таким образом, наилучшей стратегией развития провозных возможностей в соответствии с минимаксным критерием «лучшим из худших» будет третья, т.е. R_3 .

Минимаксный критерий Вальда иногда приводит к нелогичным выводам из-за своей чрезмерной «пессимистичности». «Пессимистичность» этого критерия исправляет критерий Сэвиджа.

4.3. Критерий Сэвиджа использует матрицу рисков (r_{ji})

Элементы данной матрицы можно определить по формулам (23), (24), которые перепишем в следующем виде:

$$r_{ji} = \begin{cases} \max_j \{V_{ji}\} - V_{ji}, & \text{если } V - \text{выигрыш,} \\ V_{ji} - \min_j \{V_{ji}\}, & \text{если } \text{потери.} \end{cases} \quad (31)$$

Это означает, что r_{ji} есть разность между наилучшим значением в столбце i и значениями V_{ji} при том же i . Отметим, что независимо от того, является ли V_{ji} доходом (выигрышем) или потерями (затратами) r_{ji} в обоих случаях определяет величину потерь лица, принимающего решение. Следовательно, можно применять к r_{ji} только минимаксный критерий. Критерий Сэвиджа рекомендует в условиях неопределенности выбирать ту стратегию R_j , при которой величина риска принимает наименьшее значение в самой неблагоприятной ситуации (когда риск максимален).

Пример 6. Рассмотрим пример 4. Заданная матрица определяет потери (затраты). По формуле (31) вычислим элементы матрицы рисков (r_{ji}):

$$(r_{ji}) = \begin{pmatrix} 0 & 5 & 11 & 9 \\ 3 & 0 & 0 & 13 \\ 17 & 11 & 6 & 4 \\ 21 & 17 & 12 & 0 \end{pmatrix}.$$

Полученные результаты вычислений с использованием критерия минимального риска Сэвиджа оформим в следующей таблице:

R_j	S_i	Величина риска, д.е. (V_{ji})				$\max \{r_{ji}\}$	$W = \min \max \{r_{ji}\}$
		S_1	S_2	S_3	S_4		
R_1		0	5	11	9	11	11
R_2		3	0	0	13	13	-
R_3		17	11	6	4	17	-
R_4		21	17	12	0	21	-

Введение величины риска r_{ji} привело к выбору первой стратегии R_1 , обеспечивающей наименьшие потери (затраты) в самой неблагоприятной ситуации (когда риск максимален).

Применение критерия Сэвиджа позволяет любыми путями избежать большого риска при выборе стратегии, а значит, избежать большого проигрыша (потерь).

4.4. Критерий Гурвица основан на следующих двух предположениях: «природа» может находиться в самом невыгодном состоянии с вероятностью $(1-\alpha)$ и в самом выгодном состоянии с вероятностью α , где α - коэффициент доверия. Если результат V_{ji} - прибыль, полезность, доход и т.п., то критерий Гурвица записывается так:

$$W = \max_j \left[\alpha \max_i V_{ji} + (1-\alpha) \min_i V_{ji} \right]. \quad (32)$$

Когда V_{ji} представляет затраты (потери), то выбирают действие,

$$W_{\min}^* = \min_j \left[\alpha \min_i V_{ji} + (1-\alpha) \max_i V_{ji} \right]. \quad (33)$$

Если $\alpha = 0$, получим пессимистический критерий Вальда.

Если $\alpha = 1$, то приходим к решающему правилу вида

$$\max_j \min_i V_{ji}$$

или к так называемой стратегии «здорового оптимиста», т. е. критерий слишком оптимистичный.

Критерий Гурвица устанавливает баланс между случаями крайнего пессимизма и крайнего оптимизма путем взвешивания обоих способов поведения соответствующими весами $(1-\alpha)$ и α , где $0 \leq \alpha \leq 1$. Значение α от 0 до 1 может определяться в зависимости от склонности лица, принимающего решение, к пессимизму или к оптимизму. При отсутствии ярко выраженной склонности $\alpha = 0,5$ представляется наиболее разумной.

Пример 7. Критерий Гурвица используем в примере 4. Положим $\alpha = 0,5$. Результаты необходимых вычислений приведены ниже:

W_j	$\min_i V_{ji}$	$\max_i V_{ji}$	$\alpha \min_i V_{ji} + (1-\alpha) \max_i V_{ji}$	$\min_i W_j$
W_1	6	24	15	15
W_2	7	28	17,5	-
W_3	15	23	19	-
W_4	15	27	21	-

Оптимальное решение заключается в выборе W' .

Таким образом, в примере предстоит сделать выбор, какое из возможных решений предпочтительнее:

- по критерию Лапласа — выбор стратегии R_2 ;
- по критерию Вальда — выбор стратегии R_3 ;
- по критерию Сэвиджа — выбор стратегии R_1 ;
- по критерию Гурвица при $\alpha = 0.5$ — выбор стратегии R_1 , а если лицо, принимающее решение, — пессимист ($\alpha = 0$), то выбор стратегии R_3 .

Это определяется выбором соответствующего критерия (Лапласа, Вальда, Сэвиджа или Гурвица).

Выбор критерия принятия решений в условиях неопределенности является наиболее сложным и ответственным этапом в исследовании операций. При этом не существует каких-либо общих советов или рекомендаций. Выбор критерия должно производить лицо, принимающее решение, с учетом конкретной специфики решаемой задачи и в соответствии со своими целями, а также опираясь на прошлый опыт и собственную интуицию.

В частности, если даже минимальный риск недопустим, то следует применять критерий Вальда. Если, наоборот, определенный риск вполне приемлем и лицо, принимающее решение намерено вложить в некоторое предприятие столько средств, чтобы потом оно не сожалело, что вложено слишком мало, то выбирают критерий Сэвиджа.

Упражнения

1. Для пяти проектов технических систем определены относительные единичные показатели технического совершенства конструкции и коэффици-

енты весомости единичных показателей. Численные значения единичных показателей и коэффициентов весомости приведены в следующей таблице:

Варианты технических систем	Относительные единичные показатели					
	слож- ности	веса	времени подго- товки	автома- тизации	мощно- сти	унифика- ции
I	1,0	0.088	1.0	1,0	0.72	0.614
II	0.72	1,0	0.8	0.78	0.81	0.420
III	0.658	0.358	0,765	0.782	0.525	0.915
IV	0,425	0.97	0.755	0.70	0.98	0.31
V	0,467	0,555	0.865	0,705	0,865	0,650
Коэффици- енты веса	0.157	0.124	0,210	0.195	0.174	0.140

Проведите ранжирование проектов технических систем по комплексному критерию.

2. Одной из фирм требуется выбрать оптимальную стратегию по техническому обеспечению процесса управления производством. С помощью статистических данных и информации соответствующих заводов-изготовителей были определены локальные критерии функционирования необходимого оборудования. Исходные данные представлены в следующей таблице:

Варианты оборудования	Локальные критерии эффективности оборудования*			
	производи- тельность, д. е.	стоимость оборудования, д. е.	объем памяти, у. е.	Надежность, у. е.
I	100	5	5	8
II	150	6	8	5
III	120	4	6.5	6
IV	200	7	6	4
Коэффици- енты веса	0.25	0,20	0.32	0,23

* Значения локальных критериев даны в условных единицах.

3. Для шести проектов транспортных устройств определены относительные единичные показатели технического совершенства конструкций. Численные значения единичных показателей и соответствующие весовые коэффициенты приведены в следующей таблице:

Варианты транспортных устройств	Относительные единичные показатели					
	скорости, K_1	прочности, K_2	перегрузки, K_3	устойчивости, K_4	металлоемкости, K_5	мощности, K_6
I	1.0	0.798	0.92	1.0	1.0	0,77
II	1.0	1,0	0.65	0.92	0.94	0.92
III	1.0	0.93	0.924	1.0	0,98	0,95
IV	0.87	0.96	0,91	0.915	0.99	0.85
V	0.85	0.97	1.0	0.90	0.7	0,82
VI	0.88	0.78	0.75	0,967	0,8	ко
Коэффициенты веса	0.210	0,195	0,174	0,157	0.124	0.140

Проведите ранжировку проектов технических систем по комплексному критерию.

4. Рассмотрим следующую платежную матрицу (матрицу доходов):

	S_1	S_2	S_3	S_4	S_5	S_6
R_1	15	12	1	-3	18	20
R_2	2	15	9	7	1	3
R_3	0	6	15	21	-2	5
R_4	8	20	12	3	0	4

Вероятности состояний природы $\{S_i\}$ неизвестны. Определите оптимальную стратегию R_1 , используя критерии Лапласа и максимина. Сравните полученные решения.

5. Сравните решения в задаче 4 при использовании критериев Сэвиджа и Гурвица (положите $\alpha = 0,4$).

6. Намечается крупномасштабное производство легковых автомобилей. Имеются четыре варианта проекта автомобиля R_j , $j=1,2,3,4$. Определена экономическая эффективность V_j каждого проекта в зависимости от рентабельности производства. По истечении трех сроков S_i , $i=1,2,3$ рассматриваются как некоторые состояния среды (природы). Значения экономической эффективности для различных проектов и состояний природы приведены в следующей таблице (д. е.):

Проекты	Состояние природы		
	S_1	S_2	S_3
R_1	20	25	15
R_2	25	24	10
R_3	15	28	12
R_4	9	30	20

Требуется выбрать лучший проект легкового автомобиля для производства, используя критерии Лапласа, Вальда, Сэвиджа и Гурвица при $\alpha = 0,4$. Сравните решения и сделайте выводы.

7. Фирма рассматривает вопрос о строительстве станции технического обслуживания (СТО) автомобилей. Составлена смета расходов на строительство станции с различным количеством обслуживаемых автомобилей, а также рассчитан ожидаемый доход в зависимости от удовлетворения прогнозируемого спроса на предлагаемые услуги СТО (прогнозируемое количество обслуженных автомобилей в действительности). В зависимости от принятого решения - проектного количества обслуживаемых автомобилей в сутки (проект СТО) R_j и величины прогнозируемого спроса на услуги СТО - построена нижеследующая таблица ежегодных финансовых результатов (доход, д. е.):

Проекты СТО	Прогнозируемая величина удовлетворяемости спроса					
	0	10	20	30	40	50
20	-120	60	240	250	250	250
30	-160	15	190	380	390	390
40	-210	-30	150	330	500	500
50	-270	-80	100	280	470	680

Определите наилучший проект СТО с использованием критериев Лапласа, Вальда, Сэвиджа и Гурвица (при $\alpha = 0.5$).

Вопросы для самопроверки

1. Что следует понимать под неопределенностью?
2. Что следует понимать под ситуацией риска?
3. Что такое теория игр?
4. Что является многокритериальной задачей?
5. Что называется методом свертывания критериев? Перечислите методы свертывания.
6. Что понимается под нормализацией критериев?
7. Разъясните метод последовательных уступок.
8. Назовите основные критерии оценки принимаемых решений в условиях риска. Разъясните, в каких случаях используют каждое из этих критериев?
9. На что опирается критерий Лапласа?
10. На что опирается критерий Вальда?
11. Что использует критерий Сэвиджа?
12. На каких предположениях основан критерий Гурвица?

ЛИТЕРАТУРА

1. Robert M. Leekley. Applied Statistics for Business and Economics. USA. 2010.
2. Alpha C. Chiang, Kevin Wainwright. Fundamental Methods of Mathematical Economics. NY, 2005.
3. John E. Floyd. Statistics For Economists: a beginning. University of Toronto, 2010.
4. David G. Luenberger, Yinyu Ye. Linear and Nonlinear Programming. Springer, 2008
5. Таха Х. Введение в исследование операций. В 2-х кн. М.: Мир, 1991.
6. Walras L. Elements d'Economie Politique pure. Lausanne, 1874. (англ. перевод: Elements of Pure Economics. London, 1954.)
7. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. - М.: Радио и связь, 1983.
8. Данциг Дж. Линейное программирование, его обобщение и применение. М.: Прогресс. 1966.
9. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. М.: Наука, 1965.
10. Кузнецов Ю.Н., Кузубов В.И., Волощенко А.Б. Математическое программирование. М.: Высшая школа, 1980.
11. Габасов Р., Кириллова Ф.М. Методы оптимизации. Минск: БГУ, 1975.
12. Бережной В.И., Бережная Е.В. Математические методы моделирования экономических систем. М.: Финансы и статистика, 2001.
13. Бережной В.И., Бережная Е.В. Экономико - математические методы и модели в примерах и задачах. Ставрополь: Интеллект-сервис, 1996.
14. Бабаджанов Ш.Ш. Математическое программирование: Учебное пособие. Т.: Iqtisod – Moliya, 2006.

15. Бабаджанов Ш. Ш. Теория вероятностей и математическая статистика: Курс лекций. Т.: ТМИ, 2004.
16. Бабаджанов Ш.Ш. Материалы для самостоятельных работ по теории вероятностей и математической статистике: Учебное пособие. Т.: Iqtisod – Moliya. 2006.
17. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика: Учебник. М.: Инфра - М. 1997.
18. Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: ЮНИТИ, 2001.
19. Математическое программирование /Под. ред. Н.Ш. Кремера. М.: Финстатинформ, 1996.
20. Калихман И. Л., Войтенко М. А. Динамическое программирование в примерах и задачах. М.: Высшая школа, 1979.
21. К.С.Сафаева, Э.Адигамова, Ш.Ш.Бабаджанов, Э.Мамуров Математическое программирование: Сборник задач. Т.: Iqtisod – Moliya, 2006.
22. Солодовников А. С. Теория вероятностей: Учебное пособие. М.: Просвещение, 1983.
23. Чистяков В. П. Курс теории вероятностей. 2 изд. М.: Наука, 1982.
24. Гнеденко Б. В. Курс теории вероятностей. 5 изд. М.: Наука, 1971.
25. Коваленко И. Н., Филиппова Д. А. Теория вероятностей и математическая статистика. 2 изд. М.: Высшая школа. 1982.
26. Севастьянов Б. А., Чистяков В. П. Зубков А.М. Сборник задач по теории вероятностей. М.: Наука, 1980.
27. Емельянов Г. В., Скитович В. П. Задачник по теории вероятностей и математической статистики. М.: МГУ, 1967.
28. Э.А.Вуколов, А.В.Ефимов, В.Н.Земсков, А.Ф.Каракулин, А.С.Поспелов, А.М.Терещенко Сборник задач по математике для вузов : Специальные курсы. - М.: Наука, 1984.
29. Акулич И.Л. Математическое программирование в примерах и задачах. С-П. М.: Лань. 2011.

СОДЕРЖАНИЕ

Предисловие.....	3
-------------------------	----------

ГЛАВА 1. СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

§1. Пространство элементарных исходов (событий). Определения вероятности.....	4
§2. Теоремы сложения и умножения вероятностей. Формула полной вероятности и формулы Байеса.....	49
§3. Последовательность независимых испытаний. Предельные теоремы в схеме Бернулли.....	72

ГЛАВА 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§4. Случайные величины и законы их распределения.....	97
§5. Числовые характеристики случайных величин.....	126
§6. Наиболее известные дискретные и непрерывные законы распределения. Закон больших чисел и центральная предельная теорема.....	158

ГЛАВА 3. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

§7. Статистические оценки параметров и требования, предъявляемые к ним.....	190
§8. Статистические и корреляционные зависимости.....	217
§9. Критерий согласия Пирсона и Колмогорова-Смирнова.....	229

ПРИЛОЖЕНИЕ. Вероятностные таблицы.....	242
---	------------

ГЛАВА 4. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

§10. Построение математических моделей некоторых простейших экономических задач.....	246
§11. Общая задача линейного программирования.....	269
§12. Графический метод решения задачи линейного программирования.....	280
§13. Симплексный метод решения задачи линейного программирования.....	290
§14. Метод искусственного базиса. Вырожденные задачи.....	309
§15. Теория двойственности в линейном программировании.....	321

§16. Анализ оптимального решения экономической задачи.....	333
§17. Двойственный симплексный метод.....	350
§18. Общая постановка и математическая модель транспортной задачи.....	354
§19. Нахождение оптимального плана транспортной задачи.....	369
§20. Целочисленное линейное программирование.....	390

ГЛАВА 5. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

§21. Постановка задачи нелинейного программирования. Задачи на безусловный минимум.....	407
§22. Задачи на условный минимум.....	426
§23. Выпуклое программирование.....	445
§24. Элементы динамического программирования.....	470

ГЛАВА 6. ЭЛЕМЕНТЫ ТЕОРИИ ИГР И ПРИНЯТИЯ РЕШЕНИЙ

§25. Элементы теории игр. Матричные игры.....	492
§26. Сведение матричных игр к задачам линейного программирования.....	518
§27. Принятие решения в условиях риска и неопределенности.....	530
ЛИТЕРАТУРА.....	558

Шопулат Шомашрабович БАБАДЖАНОВ

ЭКОНОМИЧЕСКАЯ МАТЕМАТИКА

Учебное пособие

Редактор Э. Хуснутдинова
Худ. редактор К. Бойхужаев
Компьютерная верстка З. Улугбекова

Лиц. изд. АІ 305. Подписано в печать 02.08.2018.
Формат 60x84 ^{1/16}. Усл.печ.л. 32,7. Уч.-изд.л. 34.0.
Тираж 100 экз. Заказ № 19.

Издательство «IQTISOD-MOLIYA».
100000, Ташкент, ул. Амира Темура, 60 ^a.

Отпечатано в типографии «HUMOYUNBEK-ISTIQLOL MO'JIZASI».
100000, Ташкент, ул. Амира Темура, 60 ^a.