МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

Норик Л. А., Шевченко А. К.

ВЫСШАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА

Учебное пособие для иностранных студентов отрасли знаний 0306 «Менеджмент и администрирование»

Часть 2

Авторы Л. А. Норик

А. К. Шевченко

Ответственный за выпуск Малярец Л.М.

Харьков, Вид. ХНЕУ, 2013

УДК
ББК
Л

Рецензенты:
доктор физ.-мат. наук, профессор, декан факультета прикладной математики и менеджмента Харьковского национального университета радиоэлектроники *Дорошенко В. А.,* доктор экон. наук, профессор, зав. кафедры статистики и экономического прогнозирования Харьковского национального экономического университета *Раевнева Е. В.*

Рекомендовано к изданию решением ученого совета Харьковского национального экономического университета.
Протокол № 7 от 25.03.2013 г.

Норик Л.А.
Высшая и прикладная математика. Учебное пособие для иностранных студентов отрасли знаний 0306 «Менеджмент и администрирование». Ч. 2 / Л. А. Норик, А. К. Шевченко. – Х. : Изд. ХНЭУ, 2013. – 404 с.

Представлен теоретический и практический материал в соответствии с рабочей программой учебной дисциплины «Высшая и прикладная математика» для формирования у студентов целостной системы знаний математического аппарата теории вероятностей, математической статистики, математического программирования и исследования операций, умений и навыков его применения к решению экономических задач.
Рекомендовано для иностранных студентов (отрасль знаний «Менеджмент и администрирование») всех форм обучения, а также для преподавателей и аспирантов.

ISBN
УДК
ББК

НАВЧАЛЬНЕ ВИДАННЯ

Норік Лариса Олексіївна
Шевченко Олександра Кирилівна

ВИЩА ТА ПРИКЛАДНА МАТЕМАТИКА

Навчальний посібник для іноземних студентів галузі знань 0306 «Менеджмент та адміністрування»

Частина 2 (рос. мовою)

Відповідальний за випуск **Л. М. Малярець**

Відповідальний редактор **Л. М. Сєдова**

Редактор **Семенова І. М.**

Коректор

Подано теоретичний і практичний матеріал відповідно до робочої програми навчальної дисципліни «Вища і прикладна математика» для формування у студентів цілісної системи знань математичного апарату теорії ймовірності, математичної статистики, математичного програмування і дослідження операцій, умінь і навиків його застосування до вирішення економічних завдань.

Рекомендовано для іноземних студентів (галузь знань «Менеджмент і адміністрування») всіх форм навчання, а також для викладачів і аспірантів.

План 2013 р. Поз. № 24-П

Підп. до друку Формат 60×90 1/16.
Папір MULTICOPY Друк Rizo.
Ум.- друк. арк. 25 Обл.-вид. арк. Тираж прим. 300
Зам. №

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи ***Дк №481 від 13.06.2001 р.***

Видавець і виготівник – видавництво ХНЕУ, 61001, м. Харків, просп. Леніна, 9а

Введение

Управление любой социально-экономической системой реализуется как процесс, подчиняющийся определенным закономерностям. Их знание помогает определить условия, необходимые и достаточные для осуществления данного процесса. Для этого все параметры, характеризующие процесс и внешние условия, должны быть изучены, количественно определены, измерены.

Математическая наука, которая изучает специфические закономерности случайных явлений, называется *теорией вероятностей*. Найденные закономерности широко применяются в планировании, управлении и прогнозировании. *Математическая статистика*, опираясь на теорию вероятностей, изучает методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления статистических закономерностей*. Математическое программирование* и *исследование операций* помогают разработать и обосновать практическое применение методов наиболее эффективного управления различными организационными системами.

Разделы «Теория вероятностей и математическая статистика» и «Математическое программирование. Исследование операций» являются прикладными математическими составляющими учебной дисциплины «Высшая и прикладная математика» с использованием математических методов функционального анализа, алгебры, аналитической геометрии, математического анализа, теории интегрального и дифференциального исчисления и др.

Правильность исходных предпосылок теории вероятностей и математической статистики, как и всякой другой прикладной теории, проверяется практикой. Сегодня теория вероятностей и математическая статистика успешно применяются в физике, химии, инженерии, астрономии, медицине, генетике, социологии, лингвистике и др. На базе теории вероятностей и математической статистики сформировались и развиваются такие разделы, как теория случайных процессов, теория массового обслуживания, математическая теория надежности и т. д.

Цель математического программирования и исследования операций – количественное обоснование принимаемых решений по организации управления.

В предлагаемом пособии рассматриваются основные понятия теории вероятностей, математической статистики, математического программирования и исследования операций, используются простые математические конструкции, описываются основные методы решения задач.

Основными **задачами** изучения учебной дисциплины «Высшая и прикладная математика» являются предоставление студентам знаний по основным разделам высшей математики, повышение уровня фундаментальной математической подготовки студентов с усилением ее прикладной направленности, а также получение необходимой математической базы для изучения других дисциплин математического цикла.

В процессе изучения данной дисциплины студент получает аналитические исследовательские **компетентности**, которые необходимы современному экономисту в любых сферах его деятельности, а именно:

уметь проводить основные математические вычисления, самостоятельно применять полученные знания для решения соответствующих задач и ситуационных упражнений;

уметь анализировать, обрабатывать полученные результаты с учетом полученных данных и делать выводы на достаточно высоком профессиональном уровне;

уметь отслеживать основные тенденции и направления развития математической науки, самостоятельно работать с научно-методической литературой;

уметь использовать полученные знания для дальнейшего формирования соответствующих экономико-математических моделей и их решения (определение балансовых отношений, вычисление коэффициентов расходов, определение зависимости спроса и предложения, сравнение эффективности финансовых операций и др.);

уметь, в случае сложного задания, использовать метод разложения от сложного к простому, то есть приводить отдельные сложные части к более простым с последующим их решением.

Раздел 3. Теория вероятностей и математическая статистика

24. Основные понятия теории вероятностей

24.1. Стохастический эксперимент, его роль и место при моделировании социально-экономических и естественных процессов

В природе, технике и экономике нет явлений, в которых не присутствовали бы элементы случайности. Существует два подхода к изучению этих явлений. Один из них – *классический*. Он состоит в том, что выделяются основные факторы, определяющие данное явление, а влиянием множества остальных, приводящих к случайным отклонениям от результата, пренебрегают. Выявляется основная закономерность, свойственная данному явлению, позволяющая однозначно предсказать результат по заданным условиям. Однако при моделировании социально-экономических и естественных процессов необходимо учитывать не только основные факторы, но и множество второстепенных, приводящих к случайным искажениям результата, то есть вносящих в него элемент неопределенности. Этот подход заложен в основе *стохастического* эксперимента, при котором элемент неопределенности изучается специальными методами. В практической деятельности часто приходится сталкиваться со случайными событиями, то есть с событиями, которые могут произойти или не произойти по причинам, неподдающимся непосредственному учету в данных условиях. Изучение количественных закономерностей, которым подчинены массовые случайные события, и есть ***предмет теории вероятностей***. Таким образом, теория вероятностей изучает свойства и закономерности массовых случайных явлений.

24.2. Алгебра случайных событий

Пусть происходит некоторое испытание (эксперимент, исследование) со случайным результатом.

Под ***испытанием*** понимается определенная совокупность условий и действий, которая может быть воспроизведена сколь угодно большое число раз. Результат реализации этих условий или действий является ***событием***. События разделяются на достоверные, невозможные и случайные.

Событие называется ***случайным***, если оно может произойти или не произойти в результате испытания.

Событие называется ***достоверным*** (Ω), если оно обязательно произойдет в результате испытания, и ***невозможным*** (∅), если оно не может произойти в результате данного испытания.

Случайные события обозначаются большими буквами латинского алфавита A, B, C и др.

События называются ***совместными***, если они могут появиться вместе в одном и том же испытании.

События A и B называются ***несовместными***, если они не могут произойти вместе в одном и том же испытании.

Два события называются ***независимыми***, если появление одного из них не зависит от появления другого. Два события называются ***зависимыми***, если появление одного из них зависит от появления другого.

События называются ***равновозможными***, если по условиям испытания ни одно из этих событий не является объективно более возможным, чем другое.

События называются ***единственно возможными***, если кроме них не могут произойти никакие другие события.

Противоположным относительно A событием (или *дополнением*) называется событие $\overline{A}$, которое состоит в том, что A не происходит. Два несовместных единственно возможных события являются противоположными.

Если события независимы, то независимы также и соответствующие им противоположные события.

События A_1, A_2, ... , A_n образуют ***полную группу несовместных событий***, если в некотором испытании обязательно происходит одно из них и никакое другое событие произойти не может. Полную группу событий составляет совокупность всех единственно возможных событий в данном испытании. ***Элементарными событиями*** некоторого испытания называются всевозможные результаты этого испытания.

Множество всех элементарных событий некоторого эксперимента (испытания) называется ***пространством элементарных событий***.

Элементы пространства элементарных событий будем обозначать ω. Само пространство элементарных событий представляет собой событие, состоящее из всех возможных исходов, то есть происходящее всегда, и носит название *достоверного события*.

В реальном испытании элементарным событиям соответствуют *взаимоисключающие события*.

Например, при бросании игральной кости естественно выбрать $\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, где ω_i означает результат испытания, который заключается в выпадении i очков.

Кроме взаимоисключающих событий, есть возможность указать и другие случайные события. В приведенном выше примере можно говорить о других случайных событиях, которые заключаются, например, в том, что при бросании игральной кости получим парное количество очков. Это случится в том случае, когда происходит одно из событий ω_2, ω_4 или ω_6.

Алгеброй событий $\mathfrak{R}$ будем называть непустую систему некоторых подмножеств, удовлетворяющую следующим ***аксиомам***:

если подмножество A принадлежит $\mathfrak{R}$ (является событием), то дополнение $\overline{A}$ также принадлежит $\mathfrak{R}$ (является событием);

если подмножества A и B принадлежат $\mathfrak{R}$ (являются событиями), то и объединение $A \cup B$ также принадлежит $\mathfrak{R}$ (является событием).

Алгебру событий можно определить как систему пространства элементарных событий, замкнутую относительно конечного числа теоретико-множественных операций.

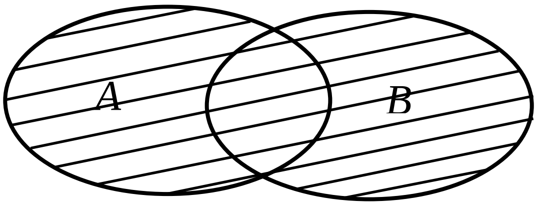
Аналогично операциям над множествами можно определить понятие *операций с элементарными событиями*.

Объединением (суммой) двух событий A и B называется событие $A \cup B$ (или $A + B$), которое заключается в появлении или события A, или события B, или обоих вместе (рис. 24.1а).

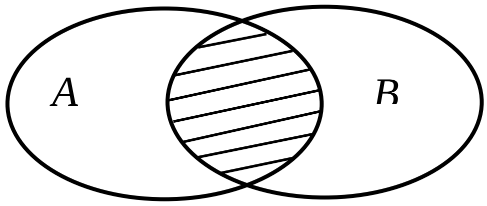
В реальном испытании событие, которое соответствует $A + B$, заключается в том, что произошло хотя бы одно из событий A или B.

Пересечением (произведением) двух событий A и B называется событие $A \cap B$ (или $A \cdot B$), которое заключается в одновременном появлении обоих событий (рис. 24.1б).

Событие $A \cdot B$ происходит тогда, когда происходит и A, и B.



а) $A+B$



б) $A \cdot B$

Рис. 24.1. **Операции над событиями**

Обозначим достоверное событие Ω, невозможное – Ø.

Для противоположного относительно A события $\overline{A}$, которое заключается в невыполнении события A, справедливо следующее:

$$A \cdot \overline{A} = \varnothing; \quad A + \overline{A} = \Omega.$$

Имеет место утверждение:

сумма противоположных событий – это событие, которое противоположно произведению событий, то есть $\overline{A} + \overline{B} = \overline{A \cdot B}$;

верно и обратное:

событие, противоположное произведению событий, равно сумме противоположных событий, то есть $\overline{A \cdot B} = \overline{A} + \overline{B}$.

Для несовместных событий A и B выполняется: $A \cdot B = \varnothing$.

24.3. Вероятности на дискретном пространстве элементарных событий

Пусть пространство элементарных событий конечно. И пусть каждому событию A, принадлежащему алгебре событий $\Re$, соответствует число $P(A)$.

Числовая функция $P(A)$ называется ***вероятностью***, если она удовлетворяет следующим аксиомам:

1) $P(A) \geq 0$ (аксиома неотрицательности);

2) $P(\Omega) = 1$, где Ω – достоверное событие (аксиома нормированности);

3) $P(A \cup B) = P(A) + P(B)$ (аксиома аддитивности), если события A и B принадлежат $\Re$, то событие $A \cup B$ также принадлежит $\Re$.

Таким образом, ***вероятностью события*** называется численная мера степени объективной возможности этого события.

Из аксиом вероятности выводится ряд свойств вероятности:

1) вероятность невозможного события равна нулю: $P(\varnothing)=0$;

2) вероятность события принадлежит интервалу $[0,1]$, то есть $0 \le P(A) \le 1$.

Доказательство.

Любое событие A можно представить в виде произведения этого события и достоверного события, то есть $A = A \cdot \Omega$.

Поскольку $A \cdot \Omega \subset \Omega$, то $P(A) = P(A \cdot \Omega) \le P(\Omega) = 1$, то есть $P(A) \le 1$. Из аксиомы неотрицательности вероятности события вытекает, что $P(A) \ge 0$.

Таким образом, $0 \le P(A) \le 1$.

Событие называется ***маловероятным***, если в данной системе испытаний вероятность его появления пренебрежительно мала. Уровень вероятности, которым можно пренебречь, называется ***уровнем значимости*** (α). Как правило, на практике выбирают уровень значимости, который равен $\alpha_1 = 0{,}01$ или $\alpha_2 = 0{,}05$. Но возможны и другие уровни значимости.

Известно, что события A_1, A_2, ... , A_n образуют полную группу несовместных событий, если в некотором испытании обязательно происходит одно из них и никакое другое событие произойти не может. То есть сумма событий, которые образуют полную группу, является достоверным событием, поэтому сумма их вероятностей равна единице:

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1. \qquad (24.1)$$

Два противоположных события A и $\overline{A}$ образуют полную группу событий, то есть $A + \overline{A} = \Omega$ – достоверное событие.

Поэтому:

$$P(A + \overline{A}) = 1, \qquad P(A) = 1 - P(\overline{A}).$$

Если обозначить $P(A) = p$ и $P(\overline{A}) = q$, то $p = 1 - q$.

24.4. Теоремы сложения вероятностей для несовместных и совместных событий

Теорема 1. Вероятность появления суммы двух несовместных событий равна сумме вероятностей этих событий:

$$P(A+B)=P(A)+P(B). \qquad (24.2)$$

Следствие. Вероятность появления суммы нескольких попарно несовместных событий равна сумме их вероятностей*:*

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n).$$

Теорема 2. Вероятность появления суммы двух совместных событий равна сумме вероятностей этих событий минус вероятность их совместного появления:

$$P(A+B)=P(A)+P(B)-P(A\cdot B). \qquad (24.3)$$

Пример 24.1. Пусть вероятность того, что стрелок при попадании в мишень выбьет 10 очков, равна 0,4; 9 очков – 0,2; 8 очков – 0,2; 7 очков – 0,1; 6 очков и меньше – 0,1. Найти вероятность того, что стрелок при одном выстреле выбьет не менее 9 очков.

Решение.

Искомое событие (обозначим его C) состоится, если стрелок выбьет или 9 (событие A), или 10 очков (событие B). События A и B несовместные. Поэтому $P(C)=P(A)+P(B)=0{,}2+0{,}4=0{,}6$.

Пример 24.2. Вероятность сдачи экзамена первым студентом равна 0,7, вторым – 0,6. Какова вероятность того, что кто-нибудь из них сдаст экзамен?

Решение.

Пусть событие A – экзамен сдаст первый студент, событие B – экзамен сдаст второй студент. События A и B совместные. Поэтому $P(A+B)=P(A)+P(B)-P(AB)=0{,}7+0{,}6-0{,}7\cdot 0{,}6=0{,}88$.

Вопросы для самодиагностики

1. Что является предметом теории вероятностей?

2. Что называется алгеброй событий? Привести аксиомы вероятности.

3. Что называют событиями и как их классифицируют?

4. Сформулировать теоремы сложения вероятностей. Привести примеры их реализации.

Упражнения

24.1. Стрелок стреляет по мишени, разделенной на 4 концентрические зоны. Вероятности попадания в эти области соответственно равны 0,4, 0,3, 0,2 и 0,1. Найти вероятность попадания либо в первую, либо во вторую зоны.

24.2. В ящике 10 белых, 15 красных, 20 синих и 25 черных шаров. Вынимается один шар. Какова вероятность того, что выбранный шар имеет цвет: а) белый или синий; б) синий или красный?

24.3. Вероятность сдачи экзамена студентом на «5» равна 0,3; на «4» – 0,45; на «3» – 0,1; на «2» – 0,05. Какова вероятность того, что студент получит положительную оценку?

24.4. На складе готовой продукции находятся изделия, среди которых 5 % нестандартных. Найти вероятность того, что при выдаче изделия со склада оно будет стандартным.

24.5. Для поломки двигателя достаточно выхода с работы одного из двух его основных элементов. Найти вероятность того, что при запуске двигатель будет поломан, если вероятность выхода с работы первого элемента двигателя равна 0,4, а второго – 0,7.

25. Классическое определение вероятности и элементы комбинаторного анализа. Статистическое и геометрическое определения вероятности

25.1. Классическое определение вероятности

В классической схеме ***вероятность*** *события* A определяется как отношение числа исходов m, которые благоприятствуют ему, к общему числу n равновозможных, единственно возможных и несовместных исходов испытания, то есть:

$$P(A) = \frac{m}{n}. \qquad (25.1)$$

Пример 25.1. В отдел технического контроля поступили 15 изделий первого сорта и 5 изделий – второго. Какова вероятность выбрать изделие первого сорта?

Решение.

По условию задачи $n = 15 + 5 = 20$, $m = 15$. Вероятность события A (выбора изделия первого сорта) равна: $P(A) = \dfrac{15}{20} = 0{,}75$.

Пример 25.2. Найти вероятность того, что наугад взятое двузначное число окажется кратным или 2, или 5, или тому и другому одновременно.

Решение.

Пусть A – событие, которое заключается в том, что наугад взятое двузначное число кратно 2, а B – событие, состоящее в том, что взято число кратное 5. Нужно найти $P(A+B)$. События A и B совместные. Двузначные числа – это 10, 11, ... , 98, 99. Всего их 90. Из них 45 кратны 2 (способствуют появлению A), 18 кратны 5 (способствуют появлению B) и, наконец, 9 чисел кратны и 2, и 5 одновременно (способствуют появлению $A \cdot B$).

Поэтому по теореме сложения для совместных событий (24.3) имеем:

$$P(A) = \frac{45}{90} = 0{,}5; \quad P(B) = \frac{18}{90} = 0{,}2; \quad P(A \cdot B) = \frac{9}{90} = 0{,}1.$$

Следовательно $P(A+B) = 0{,}5 + 0{,}2 - 0{,}1 = 0{,}6$.

25.2. Основные понятия комбинаторного анализа

Комбинаторика изучает количество комбинаций, которые подчинены определенным условиям и которые можно составить из элементов. Рассмотрим основные формулы комбинаторики, которые используются в теории вероятностей.

Перестановки – это комбинации, которые состоят из одних и тех же n разных элементов и отличаются только порядком их размещения:

$$P_n = n!, \tag{25.2}$$

где $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$, $0! = 1$.

Пример 25.3. Заданы цифры 1, 2, 3, 4, 5. Сколько пятизначных чисел можно составить из этих цифр, если каждое из них входит в число только один раз?

Решение.

Число пятизначных чисел равно: $P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

Размещениями называют комбинации, которые составлены из n разных элементов по m элементов, которые отличаются или составом элементов, или их порядком:

$$A_n^m = n \cdot (n-1)(n-2) \cdot \ldots \cdot (n-m+1). \qquad (25.3)$$

Пример 25.4. Сколько можно составить сигналов из 6 флажков разного цвета, если взять их по 2?

Решение.

Число сигналов равно: $A_6^2 = 6 \cdot 5 = 30$.

Сочетания – это комбинации, которые составлены из n разных элементов по m элементов, которые отличаются хотя бы одним элементом:

$$C_n^m = \frac{n!}{m!(n-m)!} = \frac{n(n-1) \cdot \ldots \cdot (n-m+1)}{m!}. \qquad (25.4)$$

Пример 25.5. Сколькими способами можно выбрать 3 детали из ящика, в котором 15 деталей?

Решение.

По условию задачи $n = 15, m = 3$.

Число способов равно: $C_{15}^3 = \frac{15!}{3!12!} = \frac{1 \cdot 2 \cdot 3 \cdot \ldots \cdot 15}{(1 \cdot 2 \cdot 3)(1 \cdot 2 \cdot 3 \cdot \ldots \cdot 12)} = 455$

или

$$C_{15}^3 = \frac{15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3} = 455.$$

Подчеркнем, что числа перемещений, размещений и соединений связаны равенством:

$$A_n^m = P_n \cdot C_n^m.$$

При решении задач в комбинаторике используют такие *правила*:

1) ***правило сумм***: если объект A может быть выбран из совокупности объектов m способами, а второй объект B – n способами, то выбрать или A, или B можно $m+n$ способами;

2) ***правило произведения***: если объект A можно выбрать из совокупности объектов m способами и после каждого такого выбора объект B можно выбрать n способами, то пара объектов (A, B) в таком порядке может быть выбрана $m \cdot n$ способами.

25.3. Геометрическое и статистическое определения вероятности

Статистическое определение вероятности

Относительной частотой $w(A)$ события A называют отношение числа m его появлений в n испытаниях к числу всех испытаний, то есть:

$$w(A) = \frac{m}{n}.$$

Если n достаточно большое, то относительная частота $w(A)$ колеблется вокруг некоторой постоянной величины $P(A)$, которую называют ***вероятностью события*** A.

Пример 25.6. В магазин поступили 100 телевизоров, среди них 5 с неявным дефектом. Какова вероятность приобрести телевизор с неявным дефектом?

Решение. $P(A) = w(A) = \frac{5}{100} = 0{,}05$, где A – событие, которое заключается в том, что телевизор имеет неявный дефект.

Классическое и статистическое определения вероятности имеют принципиальную разницу. Вероятность, согласно классическому определению, вычисляют до испытания (эксперимента), а относительную частоту, согласно статистическому определению, – после испытания.

Геометрическое определение вероятности.

Пусть отрезок l составляет часть отрезка L. На отрезке L случайным образом отмечена точка (то есть точка может быть в любом месте отрезка L). Тогда вероятность того, что точка попадет на отрезок l, пропорциональна длине этого отрезка и не зависит от его размещения на L. В данных предположениях вероятность того, что точка попадет на l, определяется равенством:

$$P = \frac{Длина\, l}{Длина\, L}.$$

Рассмотренный вопрос может быть обобщен и для плоских (пространственных) фигур. Если обозначить через g часть плоской (пространственной) фигуры G, то вероятность попадания точки в g пропорциональна ее площади (объему) и не зависит ни от ее размещения в G, ни от формы g, то есть:

$$P = \frac{S_g}{S_G} \ (P = \frac{V_g}{V_G}).$$

Таким образом, вероятность события A равна отношению меры множества, элементарные события которого способствуют событию A, к мере множества всех элементарных событий испытания.

Вопросы для самодиагностики

1. Привести классическое определение вероятности.
2. В чем заключается основное правило комбинаторики?
3. Что такое перестановки, размещения, сочетания?
4. Привести статистическое определение вероятности.
5. Описать геометрическое определение вероятности.

Упражнения

25.1. Игральная кость бросается два раза. Чему равна вероятность того, что сумма очков: будет делится на 3; будет больше 7?

25.2. В ящике находится 10 стандартных и 5 нестандартных деталей. Какова вероятность того, что среди наугад взятых 6 деталей будет 4 стандартных и 2 нестандартных?

25.3. В коробке лежит 10 шаров: 6 белых и 4 черных. Найти вероятность того, что из пяти взятых наугад шаров будет 4 белых.

25.4. Среди 17 студентов, из которых 8 девушек, распределяются 7 билетов в театр. Какова вероятность того, что билеты получат 4 девушки?

25.5. На пяти карточках написаны буквы «л», «д», «о», «а», «к». После тщательного перемешивания берут по одной карточке и кладут последовательно рядом. Какова вероятность того, что в результате получится слово «лодка»?

25.6. Монета брошена 2 раза. Найти вероятность того, что хотя бы один раз появится «герб».

25.7. Какова вероятность того, что при заполнении карточки спортивной лотереи «6» из «36» будет угадано 4 номера?

26. Условная вероятность, понятие независимости событий. Формула полной вероятности. Формула Байеса

26.1. Условная вероятность, теоремы умножения вероятностей

События называются ***зависимыми***, если появление одного из них зависит от появления другого. Вероятность события A, которая вычисляется при условии, что событие B уже произошло, называется ***условной вероятностью*** события A и обозначается $P_B(A)$. Условная вероятность обладает всеми свойствами безусловной вероятности.

Пример 26.1. В ящике лежит 11 деталей, 3 из них нестандартные. Из ящика дважды берут по одной детали, не возвращая их обратно. Найти вероятность того, что во второй раз из ящика будет извлечена стандартная деталь – событие A, если в первый раз была извлечена нестандартная деталь – событие B.

Решение.

После первого извлечения в ящике из 10 деталей осталось 8 стандартных, и, следовательно, искомая вероятность $P_B(A) = 0{,}8$.

Теоремы умножения вероятностей

Теорема 1. Вероятность совместного появления двух зависимых событий равна произведению вероятности одного из этих событий на условную вероятность второго при условии, что первое уже произошло:

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A). \tag{26.1}$$

Теорема 2. Вероятность совместного появления двух независимых событий равна произведению вероятностей этих событий:

$$P(A \cdot B) = P(A) \cdot P(B). \tag{26.2}$$

Пример 26.2. Среди 50 электрических лампочек 3 – нестандартные. Найти вероятность того, что 2 взятые одновременно электролампочки окажутся нестандартными.

Решение.

Вероятность события A (первая лампочка окажется нестандартной) равна 3/50. Вероятность того, что вторая лампочка будет нестандартной (событие B) при условии, что первая лампочка оказалась нестандартной, равна 2/49, потому что общее число лампочек и число нестандартных среди них уменьшились на единицу.

По теореме умножения вероятностей для двух зависимых событий (26.1) имеем:

$$P(A \cdot B) = P(A) \cdot P_A(B) = \frac{3}{50} \cdot \frac{2}{49} \approx 0{,}0024.$$

26.2. Независимость событий

Можно уточнить понятие независимости событий. События A и B ***независимы***, если условная вероятность события B при условии A совпадет с безусловной вероятностью события B, то есть $P_A(B) = P(B)$. Несколько событий называются ***независимыми в совокупности***, если вероятность каждого из них не изменяется в связи с наступлением или ненаступлением любого другого события или их комбинации.

Теорема. Вероятность появления *хотя бы одного* из независимых в совокупности событий $A_1, A_2, \ldots, A_n$ определяется формулой

$$P(A) = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n, \tag{26.3}$$

где $q_i = 1 - p_i$ – вероятности соответствующих противоположных событий $\overline{A_i}$ $(i = \overline{1,n})$.

Доказательство.

Пусть в результате испытания могут произойти независимые в совокупности события A_1, A_2, ... , A_n. Рассмотрим событие A, которое состоит в том, что произойдет хотя бы одно из этих событий. Тогда $\overline{A}$ – это событие, которое состоит в том, что не появится ни одно из этих событий, то есть

$$\overline{A} = \overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_n} .$$

События A и $\overline{A}$ образуют полную группу событий. Поэтому $P(A) = 1 - P(\overline{A})$. Следовательно, вероятность того, что произойдет хотя бы одно из этих событий, равна разности между единицей и произведением вероятностей противоположных событий:

$$P(A) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \ldots \cdot P(\overline{A_n}) = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n .$$

Следствие. Если все события $A_1, A_2, \ldots, A_n$ имеют одинаковую вероятность $P(A_1) = P(A_2) = \ldots = P(A_n) = p$, то $P(\overline{A_1}) = P(\overline{A_2}) = \ldots = P(\overline{A_n}) = q$.

Отсюда имеем:

$$P(A) = 1 - q^n .$$

Пример 26.3. Три студента сдают экзамен. Вероятность того, что первый студент сдаст экзамен, равна 0,6; для второго – 0,7; для третьего – 0,75.

Найти вероятность того, что хотя бы один студент сдаст экзамен.

Решение.

Пусть A_1, A_2, A_3 – события, которые состоят в том, что экзамен будет сдан соответственно первым, вторым, третьим студентами, а событие A – в том, что экзамен сдаст хотя бы один студент.

По условию задачи известно, что $p_1 = 0{,}6$; $p_2 = 0{,}7$ и $p_3 = 0{,}75$. Тогда $q_1 = 0{,}4$, $q_2 = 0{,}3$, $q_3 = 0{,}25$.

Следовательно, вероятность того, что экзамен сдаст хотя бы один студент, равна:

$$P(A)=1-q_1q_2q_3=1-0{,}4\cdot 0{,}3\cdot 0{,}25=0{,}97.$$

Определим *необходимое количество испытаний*.

Пусть произведено n испытаний. Вероятность желаемого результата (успеха) для каждого из них равна p.

Нужно определить количество испытаний, которое необходимо для получения желаемого результата, с надежностью не менее чем P.

Обозначим события следующим образом:

пусть A – это событие, которое заключается в том, что желаемый результат достигнут, то есть успешным было хотя бы одно испытание;

тогда $\overline{A}$ – событие, которое заключается в том, что все испытания проведены без достижения желаемого результата.

По условию $P(A)\ge \mathrm{P}$, то есть $P(A)=1-P(\overline{A})=1-q^n\ge \mathrm{P}$, где $q=1-p$.

Тогда

$$(1-p)^n\le 1-\mathrm{P}.$$

Прологарифмируем полученное неравенство:

$$n\ln(1-p)\le \ln(1-\mathrm{P}).$$

Так как $\ln(1-p)<0$, то

$$n\ge \frac{\ln(1-\mathrm{P})}{\ln(1-p)}. \tag{26.4}$$

Например, при $\mathrm{P}=0{,}98$ и $p=0{,}004$ $n\ge 976$.

26.3. Формула полной вероятности

Пусть некоторое событие A может произойти с одним из событий (их называют ***гипотезами***) $B_1, B_2, \ldots, B_n$, образующих полную группу несовместных событий.

Необходимо найти вероятность события A.

По условию событие A можно записать в виде

$$A = A \cdot B_1 + A \cdot B_2 + \ldots + A \cdot B_n .$$

События $A \cdot B_1, A \cdot B_2, \ldots, A \cdot B_n$ несовместные. Поэтому:

$$P(A) = P(A \cdot B_1) + P(A \cdot B_2) + \ldots + P(A \cdot B_n).$$

По теореме произведения вероятностей для зависимых событий (26.1) имеем:

$$P(A) = P(B_1) P_{B_1}(A) + P(B_2) P_{B_2}(A) + \ldots + P(B_n) P_{B_n}(A),$$

или

$$P(A) = \sum_{i=1}^{n} P(B_i) P_{B_i}(A). \qquad (26.5)$$

Формула (26.5) называется ***формулой полной вероятности***.

Пример 26.4. С первого станка на сборку поступает 20 %, со второго – 30 %, с третьего – 50 % деталей. Первый станок дает в среднем 0,2 % брака, второй – 0,3 %, третий – 0,1 %. Найти вероятность того, что на сборку поступила бракованная деталь.

Решение.

Пусть A – событие, которое состоит в том, что на сборку поступила бракованная деталь, B_1, B_2, B_3 – события, которые заключаются в том, что наугад выбранная деталь изготовлена соответственно на первом, втором и третьем станках.

Тогда $P(B_1) = 0{,}2$; $P(B_2) = 0{,}3$; $P(B_3) = 0{,}5$.

$P_{B_1}(A) = 0{,}002$, $P_{B_2}(A) = 0{,}003$, $P_{B_3}(A) = 0{,}001$ – вероятности того, что наугад взятая деталь бракованная, при условии, что она изготовлена соответственно на первом, втором и третьем станках.

По формуле полной вероятности (26.5) вероятность события A равна:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A) =$$

$$= 0{,}2 \cdot 0{,}002 + 0{,}3 \cdot 0{,}003 + 0{,}5 \cdot 0{,}001 = 0{,}0018.$$

26.4. Формула Байеса

Пусть события $B_1, B_2, \dots, B_n$ (гипотезы) образуют полную группу несовместных событий. Событие A может произойти с одной из этих гипотез. В результате испытания событие A произошло. Требуется определить вероятность того, что оно произошло с гипотезой B_i $\left(i = \overline{1,n}\right)$.

Если событие A произошло, то по условию произошло и некоторое событие $A \cdot B_i$. Вычислим вероятность события $A \cdot B_i$ по теореме произведения вероятностей для зависимых событий:

$$P(AB_i) = P(A)P_A(B_i) = P(B_i) \cdot P_{B_i}(A).$$

Откуда имеем ***формулу Байеса***:

$$P_A(B_i) = \frac{P(B_i) \cdot P_{B_i}(A)}{P(A)}, \tag{26.6}$$

где $P(A)$ – это полная вероятность события A (26.5).

Таким образом, формула Байеса позволяет оценить относительный вклад каждого элемента формулы полной вероятности. При этом недостатком формулы Байеса и формулы полной вероятности является то, что надо знать *априорные* (до испытания) вероятности гипотез, которые не всегда известны. Вероятности $P_A(B_i)$ – это *апостериорные* (после испытания) вероятности гипотез.

Пример 26.5. В условиях примера 26.4 найти вероятность того, что обнаруженная бракованная деталь изготовлена на первом станке.

Решение.

Вычислим условную вероятность по формуле Байеса (26.6) для первого станка:

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A)} =$$

$$= \frac{0{,}2 \cdot 0{,}002}{0{,}2 \cdot 0{,}002 + 0{,}3 \cdot 0{,}003 + 0{,}5 \cdot 0.001} = 0{,}222.$$

Вопросы для самодиагностики

1. Что называется условной вероятностью?
2. Сформулировать теоремы умножения вероятностей.
3. Что называется априорной и апостериорной вероятностями случайных событий?
4. Привести формулу полной вероятности.
5. Привести формулу Байеса.

Упражнения

26.1. Найти вероятность поражения цели при совместной стрельбе тремя орудиями, если вероятности поражения цели орудиями соответственно равны 0,9, 0,8 и 0,7 (события A, B и C).

26.2. На перевозку груза направлены 4 автомобиля. Вероятность того, что каждая машина в исправном состоянии, равна 0,8. Найти вероятность того, что в исправном состоянии, то есть перевозит груз, хотя бы один из выделенных для этого автомобилей.

26.3. Вероятность обслуживания клиента одним служащим банка равна 0,6. Какое минимальное число служащих должно работать в банке, чтобы вероятность обслуживания клиента была не менее 0,95?

26.4. В двух урнах находятся белые и красные шары: в первой – 4 белых и 5 красных, во второй – 7 белых и 3 красных. Из второй урны наудачу взяли шар и переложили его в первую урну. Найти вероятность того, что наудачу взятый после этого из первой урны шар будет белым.

26.5. В двух ящиках находятся детали: в первом – 10 штук и из них 3 нестандартных, а во втором – 20 штук и из них 8 нестандартных. Из каждого ящика наудачу вынуто по одной детали, а потом из этих двух деталей наудачу взята одна. Найти вероятность того, что эта деталь окажется стандартной.

26.6. Вероятность изготовления изделия с браком равна 0,08. После изготовления все изделия подвергаются проверке, в результате которой изделия без брака признаются годными с вероятностью 0,95, а изделия с браком – с вероятностью 0,06. Найти долю изделий, выпущенных после проверки, а также вероятность того, что выпущенное после проверки изделие окажется без брака.

26.7. В среднем из каждых 100 клиентов отделения банка 60 обслуживаются первым служащим и 40 – вторым. Вероятность того, что клиент будет обслужен без помощи заведующего отделением, только самим служащим, составляет 0,9 и 0,75 соответственно для первого и второго служащих банка. Найти вероятность полного обслуживания клиента первым служащим.

26.8. Вероятность попадания в мишень каждым из двух стрелков равна 0,3. Стрелки стреляют по очереди, причем каждый может сделать два выстрела. Попавший в мишень первым получает приз. Найти вероятность того, что стрелки получат приз.

27. Модель повторных испытаний схемы Бернулли. Теоремы Муавра – Лапласа и Пуассона

27.1. Повторные независимые испытания. Схема Бернулли

Испытания называются *однородными независимыми,* если они происходят независимо друг от друга, в одинаковых условиях и так, что вероятность появления события во всех испытаниях одинакова.

Пусть происходят n однородных независимых испытаний, в каждом из которых может произойти или не произойти определенное событие A (такую серию повторных независимых испытаний называют ***схемой Бернулли***). Вероятность появления события A в каждом испытании равна p $(q=1-p)$.

Тогда вероятность того, что в результате n независимых испытаний событие A произойдет ровно m раз, вычисляется по ***формуле Бернулли***:

$$P_n(m)=C_n^m p^m q^{n-m}=\frac{n!}{m!\,(n-m)!}p^m q^{n-m}. \qquad (27.1)$$

Доказательство.

Пусть A – событие, которое происходит в одном испытании, B – событие, состоящее в том, что событие A произойдет m раз в n испытаниях. Тогда:

$$B=\underbrace{AA\ldots A}_{m}\underbrace{\overline{A}\,\overline{A}\ldots\overline{A}}_{n-m}+\underbrace{AA\ldots A\overline{A}A}_{m+1}\underbrace{\overline{A}\,\overline{A}\ldots\overline{A}}_{n-m-1}+\ldots+\underbrace{\overline{A}\,\overline{A}\ldots\overline{A}}_{n-m}\underbrace{AA\ldots A}_{m}.$$

Каждое слагаемое отражает факт того, что событие A происходит m раз и не происходит $n-m$ раз, поэтому вероятность его $p^m q^{n-m}$. Число слагаемых равно C_n^m – числу сочетаний из n элементов по m.

Таким образом:

$$P(B) = P_n(m) = C_n^m p^m q^{n-m}.$$

Пример 27.1. Стрелок производит 5 выстрелов в тире; вероятность попадания в цель при каждом выстреле равна 0,9. Найти вероятность того, что из пяти выстрелов было не менее двух успешных.

Решение.

Событие A – из пяти выстрелов было не менее двух успешных – является объединением четырех событий: «2 попадания», «3 попадания», «4 попадания», «5 попаданий». Но проще найти вероятность противоположного события $\overline{A}$ – из пяти выстрелов одно попадание или ни одного:

$$P(\overline{A}) = P_5(0) + P_5(1) = C_5^0 \cdot (0{,}9)^0 \cdot (0{,}1)^5 + C_5^1 \cdot (0{,}9)^1 \cdot (0{,}1)^4 = 0{,}00046.$$

Следовательно, $P(A) = 1 - 0{,}00046 = 0{,}99954$.

Число m_0 появления события A в n независимых испытаниях называется ***наивероятнейшим***, если вероятность появления события m_0 раз наибольшая. Наивероятнейшее число m_0 появления события A в n испытаниях, в каждом из которых оно может произойти с вероятностью p (и не произойти с вероятностью $q = 1 - p$), определяется неравенством:

$$np - q \le m_0 \le np + p, \qquad (27.2)$$

где m_0 – целое число.

Доказательство.

Из условия $m = m_0$ – наибольшее, тогда:

$$P_n(m) \ge P_n(m-1) \quad \text{и} \quad P_n(m) \ge P_n(m+1).$$

Из первого условия $P_n(m) \geq P_n(m-1)$ следует, что $\dfrac{P_n(m)}{P_n(m-1)} \geq 1$.

По формуле Бернулли имеем:

$$\frac{C_n^m p^m q^{n-m}}{C_n^{m-1} p^{m-1} q^{n-m+1}} \geq 1, \quad \frac{n(n-1)\ldots(n-m+2)(n-m+1)(m-1)!\,p}{m!\,n(n-1)\ldots(n-m+2)\,q} \geq 1,$$

$$(n-m+1)p \geq mq, \quad np - mp + p \geq mq, \quad np + p \geq m(p+q).$$

Поскольку $p+q=1$, то $np+p \geq m$.

Из второго условия $P_n(m) \geq P_n(m+1)$ с помощью аналогичных преобразований имеем:

$(n-m)p \leq (m+1)q, \quad np - q \leq m(p+q)$, откуда $m \geq np - q$.

То есть доказано, что $np - q \leq m_0 \leq np + p$.

Если при вычислении значения $np+p$ получим целое число, то имеем два значения наивероятнейшего числа m_0, если $np+p$ является дробным, то наивероятнейшее число одно.

Пример 27.2. Доля изделий высшего сорта на данном предприятии составляет 31 %. Найти наивероятнейшее число изделий высшего сорта в случайно отобранной партии из 75 изделий.

Решение.

Неравенство (27.2) при $n=75$, $p=0{,}31$ и $q=1-0{,}31=0{,}69$ имеет вид: $75\cdot 0{,}31 - 0{,}69 \leq m_0 \leq 75\cdot 0{,}31 + 0{,}31$, то есть $22{,}56 \leq m_0 \leq 23{,}56$, откуда следует, что $m_0 = 23$, потому что это единственное целое число, которое находится между числами 22,56 и 23,56.

Пример 27.3. Контролер проверяет 24 изделия. Вероятность того, что изделие стандартное, равна 0,6. Найти наивероятнейшее число стандартных изделий.

Решение.

По условию задачи $n=24$, $p=0{,}6$ и $q=1-0{,}6=0{,}4$.

Тогда получаем: $24\cdot 0{,}6 - 0{,}4 \leq m_0 \leq 24\cdot 0{,}6 + 0{,}6$, $14 \leq m_0 \leq 15$, откуда следует, что $m_0 = 14$ и $m_0 = 15$.

27.2. Локальная теорема Муавра – Лапласа

Если вероятность p появления события A в каждом из n независимых испытаний постоянна, а число испытаний достаточно большое ($npq \geq 20$), то вероятность того, что в этих испытаниях событие A произойдет m раз, вычисляется по формуле Муавра – Лапласа:

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}, \qquad (27.3)$$

где $x = \dfrac{m - np}{\sqrt{npq}}$; (27.4)

$\varphi(x)$ – *дифференциальная функция Лапласа*;

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Таблицу значений функции $\varphi(x)$ можно найти в приложении А. Исследование функции $\varphi(x)$ приведено ниже.

Отметим лишь, что в таблице приведены значения $\varphi(x)$ для положительных значений x, потому что $\varphi(x)$ – четная функция, то есть $\varphi(-x) = \varphi(x)$.

Для значений $|x| > 4$ следует считать, что $\varphi(x) \approx 0$.

Пример 27.4. Вероятность рождения мальчика равна 0,515. Найти вероятность того, что из 200 рожденных детей количество мальчиков и девочек будет одинаковым.

Решение.

В данном случае $n = 200$; $m = 100$; $p = 0{,}515$; $q = 1 - p = 0{,}485$; $\sqrt{npq} = \sqrt{200 \cdot 0{,}515 \cdot 0{,}485} \approx \sqrt{49{,}955} \approx 7{,}068$.

Значение x, которое соответствует $m = 100$ согласно формуле (27.4), равно:

$$x = \frac{100 - 200 \cdot 0{,}515}{7{,}068} \approx -0{,}424.$$

Так как $\varphi(-0{,}424)=\varphi(0{,}424)=0{,}3647$, то по формуле (27.3) получим:

$$P_{200}(100)=\frac{\varphi(-0{,}424)}{\sqrt{49{,}955}}\approx\frac{0{,}3647}{7{,}068}\approx 0{,}052.$$

27.3. Формула Пуассона

Если в каждом испытании вероятность p появления события A постоянна и достаточно мала, а число испытаний n достаточно большое, то вероятность того, что событие A произойдет m раз, приблизительно равна:

$$P_n(m)\approx\frac{\lambda^m}{m!}e^{-\lambda}, \qquad (27.5)$$

где $\lambda=np$, $np\le 10$.

Доказательство.

По условию $np=\lambda$, то есть $p=\dfrac{\lambda}{n}$, $q=1-\dfrac{\lambda}{n}$.

По формуле Бернулли $P_n(m)=C_n^m p^m q^{n-m}$ имеем:

$$P_n(m)=\frac{n(n-1)\cdot\ldots\cdot(n-m+1)}{m!}\left(\frac{\lambda}{n}\right)^m\left(1-\frac{\lambda}{n}\right)^{n-m}=$$

$$=\frac{n(n-1)\cdot\ldots\cdot(n-m+1)}{n^m}\frac{\lambda^m}{m!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-m}.$$

Если $n\to\infty$, то $\lim\limits_{n\to\infty}\dfrac{n(n-1)\cdot\ldots\cdot(n-m+1)}{n^m}=1$, $\lim\limits_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^n=e^{-\lambda}$,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-m}=1.$$

То есть $P_n(m)\approx\dfrac{\lambda^m}{m!}e^{-\lambda}$.

Для упрощения расчетов по формуле (27.5) можно использовать таблицу значений функции Пуассона, которая приведена в приложении Б.

27.4. Интегральная теорема Муавра – Лапласа

Если вероятность p появления события в каждом испытании постоянна, а число испытаний n достаточно велико, то вероятность того, что событие A произойдет не менее m_1 и не более m_2 раз ($m_1 < m_2$), приближенно равна:

$$P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1), \tag{27.6}$$

где

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{m_2 - np}{\sqrt{npq}}. \tag{27.7}$$

В формуле (27.6) функция $\Phi(x)$ – это *интегральная функция Лапласа*, которая определяется равенством:

$$\Phi(x) = \int_0^x \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Значение функции $\Phi(x)$ приведено в приложении В, где можно найти значение этой функции лишь для $0 \le x \le 4$. Для $x < 0$ используют ту же таблицу, так как функция $\Phi(x)$ является нечетной, то есть $\Phi(-x) = -\Phi(x)$. Для $x > 4$ можно принять $\Phi(x) = 0{,}5$.

Пример 27.5. Вероятность того, что деталь изготовлена с нарушением стандартов, $p = 0{,}2$. Найти вероятность того, что среди 400 случайно отобранных деталей нестандартных окажется от 70 до 100.

Решение.

По формулам (27.7) при $m_1 = 70$, $m_2 = 100$, $n = 400$, $p = 0{,}2$, $q = 0{,}8$ вычисляем x_1 и x_2:

$$x_1 = \frac{70 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = -1{,}25, \quad x_2 = \frac{100 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 2{,}5.$$

По формуле (27.6) вычислим вероятность искомого события:

$$P_{400}(70 \le m \le 100) = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

По таблице значений функции $\Phi(x)$ находим $\Phi(2{,}5)=0{,}4938$, $\Phi(1{,}25)=0{,}3944$.

Тогда $P_{400}(70\le m\le 100)=0{,}4938+0{,}3944=0{,}8882$.

27.5. Исследование дифференциальной функции Лапласа

Проведем исследование дифференциальной функции $\varphi(x)$:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \qquad (27.8)$$

1. Область определения: $(-\infty,\infty)$.

2. Четность: $\varphi(-x)=\varphi(x)$ – функция четная, то есть график функции симметричен относительно оси OY.

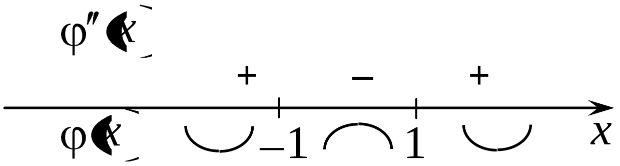
3. Исследование на экстремум:

$\varphi'(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}(-x)$, $\varphi'(x)=0$, если $x=0$;

$\varphi''(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}(x^2-1)$, $\varphi''(0)=-\frac{1}{\sqrt{2\pi}}<0$, то есть $x=0$ является точкой максимума. Если $x_{\max}=0$, то $\varphi_{\max}=\frac{1}{\sqrt{2\pi}}$.

4. Определение точек перегиба: $\varphi''(x)=0$, если $x^2-1=0$ $\quad x=\pm 1$.

Тогда:



,

то есть $x=\pm 1$ являются точками перегиба.

Если $x=1$, то $\varphi(1)=0{,}24$, если $x=-1$, то $\varphi(-1)=0{,}24$.

5. Асимптоты: вертикальных асимптот нет. Найдем наклонные асимптоты: $y=kx+b$,

$$k=\lim_{x\to\infty}\frac{\varphi(x)}{x}=\lim_{x\to\infty}\frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{x^2}{2}}}{x}=0, \quad b=\lim_{x\to\infty}(\varphi(x)-kx)=\lim_{x\to\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=0.$$

То есть график функции имеет горизонтальную асимптоту $y=0$.

График дифференциальной функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ изображен на рис. 27.1.

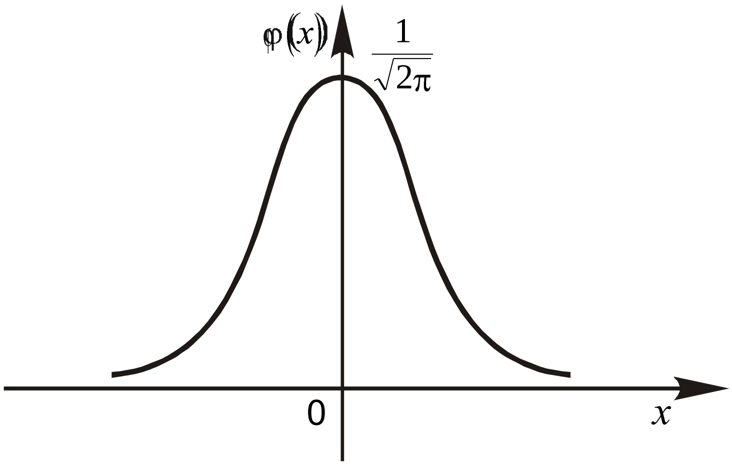


Рис. 27.1. **График дифференциальной функции** $\varphi(x)$

В дальнейшем функцию $\varphi(x)$ будем использовать при изучении нормального закона распределения непрерывной случайной величины.

27.6. Исследование интегральной функции Лапласа

Проведем исследование интегральной функции $\Phi(x)$:

$$\Phi(x)=\int_0^x \varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt. \quad (27.9)$$

1. Область определения: $(-\infty,\infty)$.
2. Четность:

$$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_0^{-x} e^{-\frac{t^2}{2}}dt=\left|\begin{array}{l}-t=u\\ dt=-du\\ 0<t<-x\\ 0<u<x\end{array}\right|=-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{u^2}{2}}du=-\Phi(x),$$

то есть функция $\Phi(x)$ нечетная, ее график симметричен относительно начала координат.

3. Исследование на экстремум:

$\Phi'(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – производная интеграла по верхнему пределу.

$\Phi'(x) \neq 0$, поэтому точек экстремума нет.

4. Определение точек перегиба: $\Phi''(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (-x)$, $\Phi''(x) = 0$, если $x = 0$.

$\Phi''(x)$	$+$		$-$	
$\Phi(x)$	$\cup$	0	$\cap$	x

, то есть $x = 0$ является точкой перегиба.

Если $x = 0$, то $\Phi(0) = 0$.

5. Асимптоты: вертикальных асимптот нет, найдем наклонные асимптоты $y = kx + b$.

$$k = \lim_{x\to\infty} \frac{\Phi(x)}{x} = \lim_{x\to\infty} \frac{\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx}{x} = 0,$$

$$b = \lim_{x\to\infty} \Phi(x) = \lim_{x\to\infty} \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \underbrace{\int_0^\infty e^{-\frac{x^2}{2}} dx}_{\text{интеграл Пуассона}} = \frac{1}{2}.$$

То есть, учитывая симметрию, $y = \pm\dfrac{1}{2}$ – горизонтальные асимптоты. График интегральной функции $\Phi(x)$ представлен на рис. 27.2.

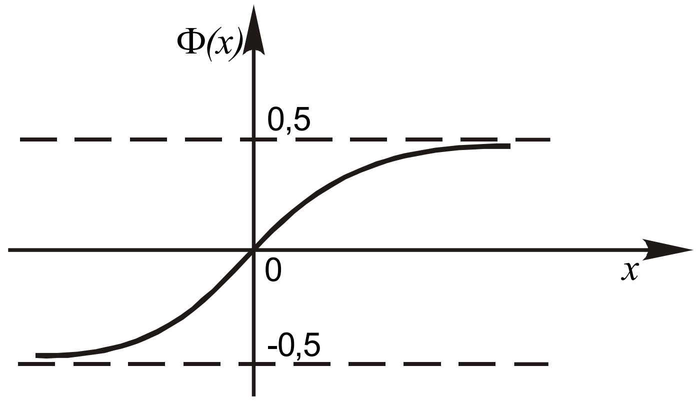


Рис. 27.2. **График интегральной функции Лапласа**

Функция $\Phi(x)$ применяется также при изучении непрерывных случайных величин.

27.7. Вероятность отклонения относительной частоты от постоянной вероятности в независимых испытаниях

Известно, что отношение числа испытаний, в которых событие A появилось, к общему числу фактически проведенных испытаний называют *относительной частотой события* $w(A)=\dfrac{m}{n}$, где m – число появлений события A, n – общее число испытаний.

Различие между вероятностью и относительной частотой состоит в том, что первая вычисляется до испытания, а вторая – после него.

Одной из важных характеристик независимых испытаний с постоянной вероятностью появления события A в каждом испытании $(0\le p\le 1)$ является ***отклонение относительной частоты от вероятности*** p.

Теорема. Пусть в n независимых испытаниях вероятность события A постоянна и равна p $(0\le p\le 1)$.

Тогда вероятность того, что абсолютная величина отклонения относительной частоты от своей вероятности меньше чем на ε, равна $2\Phi(x)$, где x определяется формулой $\varepsilon = x\sqrt{\dfrac{pq}{n}}$, $\Phi(x)$ – интегральная функция Лапласа:

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=2\Phi(x). \qquad (27.10)$$

Доказательство.

По интегральной теореме Муавра – Лапласа (27.6) имеем:

$$P(m_1<m<m_2)=\Phi(x_2)-\Phi(x_1),$$

где $x_2=\dfrac{m_2-np}{\sqrt{npq}}$, $x_1=\dfrac{m_1-np}{\sqrt{npq}}$.

Отсюда $m_2=np+x_2\sqrt{npq}$, $m_1=np+x_1\sqrt{npq}$. (27.11)

Подставим (27.11) в (27.6) и преобразуем неравенство. Тогда:

$$P\left(np + x_1\sqrt{npq} < m < np + x_2\sqrt{npq}\right) = P\left(x_1\sqrt{npq} \le m - np \le x_2\sqrt{npq}\right) =$$

$$= P\left(x_1\sqrt{\frac{pq}{n}} \le \frac{m}{n} - p \le x_2\sqrt{\frac{pq}{n}}\right).$$

Пусть $x_2 = x$ и $x_1 = -x$.

Тогда:

$$P\left(-x\sqrt{\frac{pq}{n}} \le \frac{m}{n} - p \le x\sqrt{\frac{pq}{n}}\right) = P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = \Phi(x) - \Phi(-x) = 2\Phi(x),$$

то есть

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi(x), \text{ где } \varepsilon = x\sqrt{\frac{pq}{n}}.$$

Теорема доказана.

Пример 27.6. Для определения качества выпускаемой продукции отобраны 100 изделий. Вероятность того, что изделие высокого качества, равна 0,1.

Найти:

а) вероятность P того, что относительная частота отклонится от вероятности на величину $\le \varepsilon = 0{,}01$;

б) точность ε, с которой вероятность отклонения относительной частоты от вероятности составляет $P = 0{,}95$;

в) сколько надо взять изделий, чтобы с точностью $\varepsilon = 0{,}02$ вероятность отклонения относительной частоты от вероятности была $P = 0{,}9$.

Решение:

а) $P\left(\left|\frac{m}{n} - 0{,}1\right| \le 0{,}01\right) = 2\Phi(x)$, где $0{,}01 = x\sqrt{\frac{pq}{n}}$, откуда

$$x = \frac{0{,}01 \cdot \sqrt{100}}{\sqrt{0{,}1 \cdot 0{,}9}} = \frac{0{,}1}{\sqrt{0{,}09}} = \frac{0{,}1}{0{,}3} = 0{,}33.$$

По таблице значений интегральной функции $2\Phi(x) = 2 \cdot 0{,}1293$.

Следовательно $P = 0{,}2586$;

б) по условию $P\left(\left|\dfrac{m}{n}-p\right|\le\varepsilon\right)=0{,}95$.

Тогда $0{,}95=2\Phi(x)$, откуда $x=1{,}96$, $\varepsilon=1{,}96\sqrt{\dfrac{0{,}1\cdot 0{,}9}{100}}$.

Следовательно, искомая точность $\varepsilon=1{,}96\dfrac{0{,}3}{10}=0{,}0588\approx 0{,}06$;

в) по условию задачи $P=0{,}9$, $\varepsilon=0{,}02$.

Тогда $P\left(\left|\dfrac{m}{n}-p\right|\le 0{,}02\right)=0{,}9=2\Phi(x)$, откуда $x=1{,}65$.

По формуле $\varepsilon=x\sqrt{\dfrac{pq}{n}}$, тогда: $n=\dfrac{x^2pq}{\varepsilon^2}$.

Следовательно, $n=\dfrac{1{,}65^2\cdot 0{,}1\cdot 0{,}9}{0{,}02^2}=612{,}1$, то есть для контроля необходимо взять не менее 613 изделий.

Вопросы для самодиагностики

1. Какие испытания называются однородными независимыми?
2. Привести формулу Бернулли.
3. Привести формулу Пуассона.
4. Сформулировать локальную теорему Муавра – Лапласа.
5. Сформулировать интегральную теорему Муавра – Лапласа.
6. Как используется интегральная теорема Муавра – Лапласа для вычисления вероятности попадания случайного события в заданный интервал?
7. Как вычисляется вероятность отклонения относительной частоты от вероятности случайного события?

Упражнения

27.1. Вероятность покупки бракованного комплекта посуды равна 0,1. Найти вероятность того, что из 7 купленных комплектов 5 будет без брака.

27.2. Определить, что вероятнее для соперников равной силы при игре в шахматы: выиграть одну партию из двух или две партии из четырех.

27.3. Вероятность выпуска бракованного изделия равна 0,3. Найти вероятность того, что среди 100 выпущенных изделий будет 60 изделий без брака.

27.4. Вероятность выпуска бракованных деталей равна 0,3. Найти вероятность того, что среди 100 выпущенных деталей будет не менее 75 стандартных.

27.5. Вероятность появления события равна 0,7 в каждом из 2 100 независимых испытаний. Найти вероятность появления события: а) не менее 1 470 раз; б) не менее 1 470 и не более 1 500 раз; в) не более 1 469 раз.

27.6. В страховой компании 10 тыс. клиентов, застраховавших свою недвижимость. Страховой взнос составляет 2 000 грн, вероятность несчастного случая p = 0,005, страховая выплата клиенту при несчастном случае составляет 200 тыс. грн.

Определить размер прибыли страховой компании с вероятностью P: 1) 0,9; 2) 0,995.

27.7. Вероятность получения нестандартной детали p = 0,1. Найти вероятность того, что среди случайно взятых 200 деталей относительная частота появления нестандартной детали отклонится от вероятности p по абсолютной величине не более чем на 0,03.

27.8. Вероятность появления события в каждом из 900 независимых испытаний равна 0,5. Найти вероятность отклонения относительной частоты появления события от его вероятности по абсолютной величине не более чем на 0,02.

28. Случайные величины, их законы распределения и числовые характеристики

28.1. Определение случайных величин и их классификация

Случайной величиной называется функция, ставящая в соответствие каждому элементарному исходу число.

При этом множество элементарных исходов принадлежит алгебре событий.

Случайная величина – это переменная величина, значения которой зависят от ряда случайных факторов, причем в результате испытаний она может принимать случайные, заранее неизвестные значения.

Тот факт, что случайная величина принимает определенное значение, называется ***случайным событием***. Случайные величины обозначают большими буквами латинского алфавита X, Y, Z и др., а их возможные значения – соответствующими маленькими буквами. Например, X – случайная величина, ее возможные значения – x_1, x_2, ... , x_n.

Различают дискретные и непрерывные случайные величины. Случайная величина называется ***дискретной***, если в результате испытания она может принимать конкретные, вполне определенные изолированные значения, их может быть конечное или бесконечное число. Например, размер обуви является дискретной случайной величиной. Случайная величина называется ***непрерывной***, если в результате испытаний она может принимать любые значения, принадлежащие некоторому интервалу $X \in [a, b]$. Например, рост людей является непрерывной случайной величиной.

Понятие *независимости* случайных величин отражает отсутствие связи между ними, то есть, зная значения, которые принимает случайная величина X, мы никакой новой информации о значениях случайной величины Y не получим.

28.2. Закон распределения дискретной случайной величины

Для того чтобы полностью охарактеризовать случайную величину, необходимо знать ее значения и вероятности появления этих значений. Если известны возможные значения дискретных случайных величин и вероятности их появления, то говорят, что задан ***закон распределения*** этих случайных величин, или ***ряд распределения***.

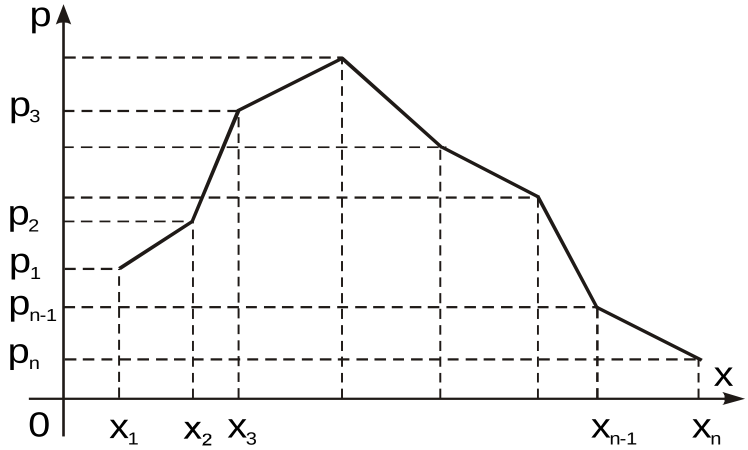
Ряд распределения дискретной случайной величины записывают в виде *таблицы*:

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

, причем $\sum_{i=1}^{n} p_i = 1$.

Графическое изображение дискретной случайной величины

По оси абсцисс откладывают возможные значения переменной x_i, а по оси ординат – соответствующие вероятности p_i и соединяют для наглядности полученные точки отрезками (рис. 28.1).



В итоге получают *полигон распределения*, или *многоугольник распределения.*

Рис. 28.1. **Полигон распределения**

28.3. Числовые характеристики дискретных случайных величин и их свойства

28.3.1. Математическое ожидание дискретной случайной величины и его свойства

Характеристикой среднего значения случайной величины X является ***математическое ожидание***. Математическое ожидание дискретной случайной величины вычисляется как сумма произведений возможных значений случайной величины на их вероятности p_i.

Пусть задан ряд распределения случайной величины X:

x_i	x_1	x_2	$\dots$	x_i	...	x_n
p_i	p_1	p_2	$\dots$	p_i	...	p_n

Обозначим математическое ожидание случайной величины $M(X)$, тогда получим:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i . \tag{28.1}$$

Математическое ожидание часто называют центром распределения, так как оно характеризует среднее значение случайной величины.

Свойства математического ожидания

1. Математическое ожидание постоянной величины равно самой постоянной:

$$M(C) = C, \text{ где } C = const.$$

Доказательство.

Пусть распределение вероятностей случайной величины X задано в виде таблицы:

x_i	C	C	$\dots$	C
p_i	p_1	p_2	$\dots$	p_n

Тогда $M(C) = \sum_{i=1}^{n} C \cdot p_i = C \sum_{i=1}^{n} p_i = C$, поскольку $\sum_{i=1}^{n} p_i = 1$.

2. Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX) = C \cdot M(X).$$

Доказательство.

Пусть распределение вероятностей случайной величины CX задано в таблице:

Cx_i	Cx_1	Cx_2	$\dots$	Cx_n
p_i	p_1	p_2	$\dots$	p_n

Тогда $M(CX) = \sum_{i=1}^{n} Cx_i p_i = C \sum_{i=1}^{n} x_i p_i = C \cdot M(X)$.

3. Математическое ожидание суммы случайных величин равно сумме их математических ожиданий:

$$M(X + Y) = M(X) + M(Y).$$

Доказательство.

Пусть X и Y независимые случайные величины с законами распределения:

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

y_j	y_1	y_2	$\dots$	y_m
q_j	q_1	q_2	$\dots$	q_m

Тогда распределение случайной величины $X+Y$ будет:

x_i+y_j	x_1+y_1	x_1+y_2	$\dots$	x_1+y_m	x_2+y_1	x_2+y_2	$\cdots$	x_n+y_m
p_{ij}	p_1q_1	p_1q_2	$\dots$	p_1q_m	p_2q_1	p_2q_2	$\cdots$	p_nq_m

Действительно, обозначим события: $A\colon X=x_i$; $B\colon Y=y_j$. Для того чтобы произошло событие $C\colon X+Y=x_i+y_j$, необходимо, чтобы произошло и событие A, и событие B, то есть $C=AB$.

Тогда

$$P(C)=P(AB)=P(A)\cdot P(B)=p_iq_j.$$

$$M(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)p_iq_j=\sum_{i=1}^{n}x_ip_i\sum_{j=1}^{m}q_j+\sum_{i=1}^{n}p_i\sum_{j=1}^{m}y_jq_j=$$

$$=\sum_{i=1}^{n}x_ip_i+\sum_{j=1}^{m}y_jq_j=M(X)+M(Y).$$

$\sum_{i=1}^{n}p_i=1$ и $\sum_{j=1}^{m}q_j=1$, так как значения x_i и значения y_j образуют полную группу событий X и Y соответственно.

Следствие 1. Математическое ожидание разности случайных величин равно разности их математических ожиданий:

$$M(X-Y)=M(X)-M(Y).$$

Доказательство.

$$M(X-Y)=M(X+(-Y))=M(X)+M(-Y)=M(X)-M(Y).$$

Разность между случайной величиной и ее математическим ожиданием называется *отклонением.*

Следствие 2. Математическое ожидание отклонения случайной величины от ее математического ожидания равно нулю: $M\left(X-M(X)\right)=0$.

Доказательство.

$$M\left(X-M(X)\right)=M(X)-M\left(M(X)\right)=M(X)-M(X)=0.$$

4. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий:

$$M(XY)=M(X)\cdot M(Y).$$

Доказательство.

Пусть X и Y независимые случайные величины с законами распределения:

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

y_j	y_1	y_2	$\dots$	y_m
q_j	q_1	q_2	$\dots$	q_m

Найдем закон распределения случайной величины XY:

$x_i y_j$	$x_1 y_1$	$x_1 y_2$	$\dots$	$x_1 y_m$	$x_2 y_1$	$x_2 y_2$	$\cdots$	$x_n y_m$
p_{ij}	$p_1 q_1$	$p_1 q_2$	$\dots$	$p_1 q_m$	$p_2 q_1$	$p_2 q_2$	$\cdots$	$p_n q_m$

Действительно, если события A: $X=x_i$; B: $Y=y_j$, то событие AB: $XY=x_i y_j$, то есть

$$P(AB)=P(A)\cdot P(B)=p_i q_j.$$

$$M(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_i q_j=\sum_{i=1}^{n}x_i p_i\sum_{j=1}^{m}y_j q_j=M(X)\cdot M(Y).$$

Пример 28.1. В лотерее 100 билетов. На 20 билетов нет выигрыша, на 25 билетов можно выиграть 1 гривну, на 20 билетов выпадает выигрыш по 2 гривны, на 15 – по 3 гривны, на 10 – по 5 гривен и на 10 – по 10 гривен. Необходимо найти математическое ожидание выигрыша.

Решение.

Ряд распределения случайной величины X – размер выигрыша в лотерее – имеет вид:

x_i	0	1	2	3	5	10
p_i	0,2	0,25	0,2	0,15	0,1	0,1

$$M(X)=0\cdot 0{,}2+1\cdot 0{,}25+2\cdot 0{,}2+3\cdot 0{,}15+5\cdot 0{,}1+10\cdot 0{,}1=2{,}6.$$

28.3.2. Дисперсия дискретной случайной величины и ее свойства

Характеристикой степени рассеяния значений случайной величины около ее математического ожидания является ***дисперсия***.

Дисперсией $D(X)$ дискретной случайной величины есть математическое ожидание квадрата отклонения случайной величины от своего математического ожидания, то есть:

$$D(X)=M\left((X-M(X))^2\right)=\sum_{i=1}^{n}(x_i-M(X))^2\cdot p_i\,. \qquad (28.2)$$

Преобразуем формулу (28.2) для вычисления дисперсии.

$$D(X)=M\left((X-M(X))^2\right)=M\left(X^2-2XM(X)+M^2(X)\right)=$$

$$=M(X^2)-2M(X)M(X)+M(M^2(X))=M(X^2)-2M^2(X)+M^2(X)=$$

$$=M(X^2)-M^2(X),$$

то есть:

$$D(X)=M(X^2)-M^2(X)=\sum_{i=1}^{n}x_i^2p_i-(M(X))^2. \qquad (28.3)$$

Дисперсия случайной величины равна математическому ожиданию ее квадрата минус квадрат ее математического ожидания.

Свойства дисперсии

1. Дисперсия постоянной величины равна нулю:

$$D(C)=0, \text{ где } C=const.$$

Доказательство.

Пусть $X = C$, тогда:

$$D(X) = D(C) = M\left[(C - M(C))^2\right] = M\left[(C - C)^2\right] = M(0) = 0.$$

2. Постоянный множитель можно выносить за знак дисперсии в квадрате:

$$D(C \cdot X) = C^2 \cdot D(X).$$

Доказательство.

Действительно,

$$D(CX) = M\left[(CX - M(CX))^2\right] = M\left[(CX - CM(X))^2\right] =$$

$$= M\left[C^2 (X - M(X))^2\right] = C^2 M\left[(X - M(X))^2\right] = C^2 D(X).$$

3. Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$D(X + Y) = D(X) + D(Y).$$

Доказательство.

По определению дисперсии и на основании свойств имеем:

$$D(X + Y) = M\left[(X + Y - M(X + Y))^2\right] = M\left[(X + Y - M(X) - M(Y))^2\right] =$$

$$= M\left[((X - M(X)) + (Y - M(Y)))^2\right] = M\left[(X - M(X))^2 + (Y - M(Y))^2 +\right.$$

$$\left. + 2(X - M(X))(Y - M(Y))\right] = M\left[(X - M(X))^2\right] + M\left[(Y - M(Y))^2\right] +$$

$$+ 2M(X - M(X))M(Y - M(Y)) = M\left[(X - M(X))^2\right] + M\left[(Y - M(Y))^2\right] =$$

$$= D(X) + D(Y), \quad (M(X - M(X)) = 0 \text{ и } M(Y - M(Y)) = 0).$$

4. Дисперсия разности двух случайных величин равна сумме их дисперсий:

$$D(X - Y) = D(X) + D(Y).$$

Доказательство.

$$D(X - Y) = D(X + (-Y)) = D(X) + D(-Y) = D(X) + D(Y).$$

28.3.3. Среднее квадратическое отклонение. Коэффициент вариации. Мода, медиана. Начальный и центральный моменты

Среднее квадратическое отклонение является мерой рассеяния значений случайной величины около ее среднего значения.

Среднее квадратическое отклонение $\sigma(X)$ (σ_x) случайной величины X – это квадратный корень из дисперсии, то есть:

$$\sigma(X)=\sqrt{D(X)}. \tag{28.4}$$

Коэффициент вариации – это отношение среднего квадратического отклонения к математическому ожиданию, выраженное в процентах:

$$v(X)=\frac{\sigma(X)}{M(X)}\cdot 100\%. \tag{28.5}$$

Коэффициент вариации дает возможность сравнить степень рассеяния значений разных по природе случайных величин.

Мода M_o – это значение x_i случайной величины X с максимальной вероятностью.

Медиана m_e – это значение x_i случайной величины X, которое разделяет ряд распределения пополам.

Начальный момент k-го порядка ν_k – это математическое ожидание k-й степени случайной величины X:

$$\nu_k=M(X^k). \tag{28.6}$$

Так, начальный момент первого порядка $\nu_1=M(X)$ – это математическое ожидание случайной величины X, $\nu_2=M(X^2)$ – математическое ожидание квадрата случайной величины X и т. д.

Центральный момент k-го порядка μ_k – это математическое ожидание k-й степени отклонения случайной величины X от своего математического ожидания:

$$\mu_k=M\left((X-M(X))^k\right). \tag{28.7}$$

Так, $\mu_1=0$; $\mu_2=D(X)$.

Пример 28.2. Задан ряд распределения дискретной случайной величины X:

x_i	1	2	3
p_i	0,4	0,3	0,3

Найти числовые характеристики случайной величины X.

Решение.

1. Математическое ожидание:

$$M(X)=x_1p_1+x_2p_2+x_3p_3=1\cdot 0{,}4+2\cdot 0{,}3+3\cdot 0{,}3=1{,}9.$$

2. Дисперсия:

$$D(X)=(x_1-M(X))^2p_1+(x_2-M(X))^2p_2+(x_3-M(X))^2p_3=$$

$$=(1-1{,}9)^2\cdot 0{,}4+(2-1{,}9)^2\cdot 0{,}3+(3-1{,}9)^2\cdot 0{,}3\approx 0{,}69,$$

или

$$D(X)=x_1^2p_1+x_2^2p_2+x_3^2p_3-(M(X))^2=1\cdot 0{,}4+4\cdot 0{,}3+9\cdot 0{,}3=0{,}69.$$

3. Среднее квадратическое отклонение:

$$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}69}\approx 0{,}84.$$

4. Мода: $M_o=1$ $(max\ p_i=0{,}4)$.

5. Медиана: $m_e=2$.

6. Коэффициент вариации:

$$\nu(X)=\frac{0{,}84}{1{,}9}\cdot 100\,\%=44\,\%.$$

7. Начальный момент: $\nu_1=1{,}9$; $\nu_2=4{,}3$.

8. Центральный момент: $\mu_1=0$; $\mu_2=0{,}69$.

Моменты высшего порядка можно использовать для того, чтобы отдифференцировать влияние больших по величине, но маловероятных значений случайной величины.

Пример 28.3. Задан ряд распределения дискретной случайной величины X:

x_i	1	2	10
p_i	0,5	0,48	0,02

Определить начальные моменты случайной величины X.

Решение.

$$\nu_1 = M(X) = 1\cdot 0{,}5 + 2\cdot 0{,}48 + 10\cdot 0{,}02 = 1{,}66.$$

$$\nu_2 = M(X^2) = 1\cdot 0{,}5 + 4\cdot 0{,}48 + 100\cdot 0{,}02 = 4{,}42.$$

$$\nu_3 = M(X^3) = 1\cdot 0{,}5 + 2^3\cdot 0{,}48 + 10^3\cdot 0{,}02 = 24{,}34.$$

$$\nu_4 = M(X^4) = 1\cdot 0{,}5 + 2^4\cdot 0{,}48 + 10^4\cdot 0{,}02 = 208{,}18.$$

Момент ν_4 полностью зависит от значения $x_3 = 10$.

Таким образом, отдифференцировано влияние большого, но маловероятного значения случайной величины.

28.4. Математическое ожидание и дисперсия среднего арифметического n независимых случайных величин

Пусть имеем n независимых случайных величин: $X_1, X_2, \ldots, X_n$ с математическими ожиданиями $M(X_1), M(X_2), \ldots, M(X_n)$ соответственно.

Пусть X – случайная величина, которая равна:

$$X = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Согласно свойствам математического ожидания имеем:

$$M(X) = \frac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n}, \qquad (28.8)$$

то есть математическое ожидание среднего арифметического n случайных величин равно среднему арифметическому их математических ожиданий.

Пусть $D(X_1), D(X_2), \ldots, D(X_n)$ – дисперсии этих случайных величин и $max(D(X_1), D(X_2), \ldots, D(X_n)) = D$.

Согласно условию $D(X_1) \le D, D(X_2) \le D, \ldots, D(X_n) \le D$ получим:

$$D(X) = \frac{D(X_1) + D(X_2) + \ldots + D(X_n)}{n^2} \le \frac{nD}{n^2} = \frac{D}{n}, \tag{28.9}$$

то есть дисперсия среднего арифметического n независимых случайных величин, имеющих ограниченные дисперсии, в n раз меньше наибольшей дисперсии.

Если случайные величины $X_1, X_2, \ldots, X_n$ одинаково распределены, то есть:

$$M(X_1) = M(X_2) = \ldots = M(X_n) = a$$

и

$$D(X_1) = D(X_2) = \ldots = D(X_n) = D.$$

Тогда $$M(X) = \frac{na}{n} = a, \qquad D(X) = \frac{D}{n}, \tag{28.10}$$

то есть математическое ожидание n одинаково распределенных случайных величин равно их общему математическому ожиданию, а дисперсия в n раз меньше их общей дисперсии.

Отсюда имеем: $\sigma(X) = \frac{\sigma}{\sqrt{n}}$, то есть среднее квадратическое отклонение среднего арифметического n независимых одинаково распределенных случайных величин равно $\frac{\sigma}{\sqrt{n}}$ $(\sigma = \sqrt{D})$.

28.5. Основные законы распределения дискретных случайных величин

Задать *закон распределения* случайной величины значит задать ее ряд распределения, то есть указать возможные значения случайной величины и их вероятности. В зависимости от способа вычисления вероятностей различают законы распределения.

28.5.1. Биномиальное распределение

Пусть происходит n независимых испытаний, в каждом из которых вероятность появления события равна p. Обозначим через X число испытаний, в которых событие A появилось. Случайная величина X может принимать значения $0, 1, 2, \ldots, m, \ldots, n$. Вероятность того, что случайная величина примет значение $X = m$, можно вычислить по формуле Бернулли (27.1), то есть:

$$P_n(X = m) = C_n^m p^m q^{n-m}, \tag{28.11}$$

где $q = 1 - p$.

Закон распределения, в котором вероятность случайной величины вычисляется по формуле Бернулли, называется ***биномиальным законом распределения***.

Данный закон распределения имеет вид:

x_i	0	1	2	$\ldots$	$n-1$	n
p_i	q^n	$C_n^1 pq^{n-1}$	$C_n^2 p^2 q^{n-2}$	$\ldots$	$C_n^{n-1} p^{n-1} q$	p^n

Если n – большое число, то вероятности p_i вычисляются по формуле Муавра – Лапласа (27.3):

$$P_n(X = m) = \frac{\varphi(x)}{\sqrt{npq}}, \text{ где } x = \frac{m - np}{\sqrt{npq}}.$$

Найдем *числовые характеристики биномиального закона распределения*. Случайную величину X можно рассматривать как сумму n независимых одинаково распределенных случайных величин X_i $(i = \overline{1, n})$ с рядом распределения:

x_i	0	1
p_i	q	p

где x_i – появление события A в i-м опыте, то есть $x_i = 1$, если событие A появилось, и $x_i = 0$, если событие A не появилось.

Тогда $M(X_i) = 0 \cdot q + 1 \cdot p = p$.

Поскольку математическое ожидание суммы случайных величин равно сумме математических ожиданий, то:

$$\boldsymbol{M}(\boldsymbol{X})=\sum_{i=1}^{n} M(X_i)=\sum_{i=1}^{n} p=\boldsymbol{np}. \qquad (28.12)$$

Дисперсия случайной величины X_i равна:

$$D(X_i)=(0-p)^2\cdot q+(1-p)^2\cdot p=p^2q+pq^2=pq(p+q)=pq.$$

Поскольку дисперсия суммы случайных величин равна сумме их дисперсий, то:

$$\boldsymbol{D}(\boldsymbol{X})=\sum_{i=1}^{n} D(X_i)=\sum_{i=1}^{n} pq=\boldsymbol{npq}. \qquad (28.13)$$

Среднее квадратическое отклонение биномиального распределения определяется формулой:

$$\boldsymbol{\sigma}(X)=\sqrt{\boldsymbol{npq}}. \qquad (28.14)$$

Пример 28.4. Вероятность сдачи экзамена на «5» для каждого из трех студентов равна 0,4. Составить закон распределения количества отличных оценок, полученных студентами на экзамене. Найти математическое ожидание и дисперсию.

Решение.

Дискретная случайная величина X (количество студентов, которые получили «5») имеет такие возможные значения:

$x_1=0$ (ни один студент не сдал экзамен на «5»);

$x_2=1$ (один студент сдал экзамен на «5»);

$x_3=2$ (два студента сдали экзамен на «5»);

$x_4=3$ (три студента сдали экзамен на «5»).

Сдача экзамена на «5» студентами – события независимые, вероятности сдать экзамен каждым студентом одинаковые, поэтому используем формулу Бернулли.

По условию задачи $n=3$, $p=0{,}4$, $q=0{,}6$.

Тогда:

если $x_1 = 0$, то $P_3(0) = q^3 = 0{,}6^3 = 0{,}216;$

если $x_2 = 1$, то $P_3(1) = C_3^1 pq^2 = 3 \cdot 0{,}4 \cdot 0{,}6^2 = 0{,}432;$

если $x_3 = 2$, то $P_3(2) = C_3^2 p^2 q = 3 \cdot 0{,}4^2 \cdot 0{,}6 = 0{,}288;$

если $x_4 = 3$, то $P_3(3) = p^3 = 0{,}4^3 = 0{,}064.$

Закон распределения случайной величины X имеет вид:

x_i	0	1	2	3
p_i	0,216	0,432	0,288	0,064

, $\sum\limits_i p_i = 1.$

По формулам (28.12), (28.13) получим: $M(X) = n \cdot p = 3 \cdot 0{,}4 = 1{,}2;$ $D(X) = npq = 3 \cdot 0{,}4 \cdot 0{,}6 = 0{,}72;$ $\sigma(X) = \sqrt{0{,}72} \approx 0{,}85.$

28.5.2. Закон распределения Пуассона

Если n достаточно велико, а вероятность p очень мала, то вероятность того, что случайная величина примет значение $X = m$, вычисляют по формуле Пуассона (27.5):

$$P_n(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}. \tag{28.15}$$

Закон распределения в этом случае называют ***законом распределения Пуассона***.

Данный закон распределения имеет вид:

x_i	0	1	2	…	n
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2}$	…	$\frac{\lambda^n e^{-\lambda}}{n!}$

, где $\lambda = np$.

Определим *числовые характеристики случайной величины, распределенной по закону Пуассона*.

Поскольку закон Пуассона является предельным для биномиального закона при достаточно больших n $(n \to \infty)$ и достаточно малых p

$(q \to 1)$, то математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, определяются по формулам:

$$\boldsymbol{M}(X) = np = \lambda, \quad \boldsymbol{D}(X) = npq \approx np = \lambda. \qquad (28.16)$$

Таким образом, $M(X) \approx D(X) = \lambda$.

Пример 28.5. Составить закон распределения случайной величины X, если в $1\,000$ независимых испытаниях событие появляется с вероятностью $0{,}001$. Найти $M(X)$ и $D(X)$.

Решение.

Имеем независимые испытания с одинаковой малой вероятностью $(p = 0{,}001)$ и большим числом испытаний $(n = 1\,000)$. Поэтому вероятности появления каждого отдельного значения вычислим по формуле (28.15) при $\lambda = np = 1$.

Возможные значения случайной величины X: $0, 1, 2, 3, \ldots$

Запишем закон распределения величины X:

x_i	0	1	2	3	4	5	6
p_i	0,36788	0,36788	0,18394	0,06131	0,01533	0,00306	0,00051

$\sum_i p_i \approx 0{,}9999 \approx 1.$

Для вычисления математического ожидания и дисперсии $M(X), D(X)$ используем формулу (28.16):

$$M(X) = D(X) = np = 1\,000 \cdot 0{,}001 = 1.$$

Вопросы для самодиагностики

1. Что называется случайной величиной и как их классифицируют?

2. Охарактеризовать основные числовые характеристики дискретной случайной величины.

3. Что называется начальным и центральным теоретическими моментами?

4. Чему равны математическое ожидание и дисперсия среднего арифметического n случайных величин?

5. Описать биномиальное распределение случайной величины.
6. Описать закон распределения Пуассона.

Упражнения

28.1. В денежной лотерее на 100 билетов разыгрывается один выигрыш в 20 грн, два выигрыша по 10 грн и 10 выигрышей по 1 грн. Найти закон распределения возможного выигрыша на один билет.

28.2. Партия из 8 изделий содержит 5 стандартных. Наудачу отбираются 3 изделия. Составить закон распределения числа стандартных изделий среди отобранных.

28.3. Банк выдает 5 кредитов. Вероятность невозврата кредита равна 0,2 для каждого из заемщиков. Составить закон распределения количества заемщиков, не вернувших кредит по окончании срока кредитования.

28.4. На базу отправлено 10 000 изделий. Вероятность того, что изделие в пути получит повреждение, равна 0,0003. Найти вероятность того, что на базу прибудут 4 поврежденных изделия.

28.5. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение количества очков, выпадающих при бросании игральной кости.

28.6. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение числа выигрышных лотерейных билетов, если вероятность выигрыша по одному билету равна 0,015, причем куплено 200 билетов.

29. Непрерывные и абсолютно непрерывные случайные величины. Функция и плотность распределения вероятностей. Числовые характеристики

29.1. Определение непрерывных случайных величин

Случайную величину называют ***непрерывной***, если ее возможные значения заполняют некоторый числовой интервал. Число возможных значений непрерывной случайной величины бесконечно.

Непрерывная случайная величина характеризуется двумя функциями:

1) функцией распределения вероятностей $F(x)$ (интегральной функцией распределения);

2) плотностью распределения вероятностей $f(x)$ (дифференциальной функцией распределения).

Случайная величина называется также *непрерывной,* если ее функция распределения непрерывна и дифференцируема.

29.2. Функция распределения вероятностей и ее свойства

Функция распределения вероятностей $F(x)$ – это вероятность того, что случайная величина X в результате испытания примет значение меньше x:

$$F(x) = P(X < x). \tag{29.1}$$

Пример. $P(X < 2) = F(2)$.

Свойства функции распределения:

1. Значение функции $F(x)$ принадлежат отрезку $[0,1]$ (по определению):

$$0 \le F(x) \le 1.$$

2. Интегральная функция является неубывающей функцией:

$$F(x + \Delta x) \ge F(x), \text{ если } \Delta x > 0.$$

Действительно, если $\Delta x > 0$, то:

$$F(x+\Delta x) = P(X < x+\Delta x) = P(X < x) + P(x \le X < x+\Delta x) =$$

$$= F(x) + P(x \le X < x+\Delta x) \ge F(x).$$

3. Вероятность попадания непрерывной случайной величины в заданный интервал (α, β) равна разности функции $F(x)$ на концах интервала:

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha). \tag{29.2}$$

Действительно,

$$P(\alpha < X < \beta) = P(X < \beta) - P(X < \alpha) = F(\beta) - F(\alpha).$$

4. Вероятность попадания случайной величины в точку равна нулю:

$$P(X = C) = 0.$$

Действительно, $P(X = C) = P(C < X < C) = F(C) - F(C) = 0$.

Следствие.
$$P(\alpha < X < \beta) = P(\alpha \le X < \beta) = P(\alpha < X \le \beta) = P(\alpha \le X \le \beta).$$

5. Если значения случайной величины X принадлежат интервалу (a, b), то при $x \le a$ $F(x) = 0$, при $x > b$ $F(x) = 1$.

Пример 29.1. Случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0, & x < 2, \\ \dfrac{x}{2} - 1, & 2 < x \le 4, \\ 1, & x > 4. \end{cases}$$

Найти вероятность того, что в результате испытания случайная величина X примет значение: а) в интервале (2; 3); б) меньше 0,2; в) меньшее 3; г) не меньше 3; д) не меньше 5.

Решение:

а) $P(2 < X < 3) = F(3) - F(2) = \left(\dfrac{3}{2} - 1\right) - 0 = 0{,}5$;

б) $X < 0{,}2$, $P(X < 0{,}2) = F(0{,}2) = 0$;

в) $X < 3$, $P(X < 3) = F(3) = \dfrac{3}{2} - 1 = 0{,}5$;

г) $X \ge 3$, так как $P(X \ge 3) + P(X < 3) = 1$, то

$$P(X \ge 3) = 1 - P(X < 3) = 0{,}5;$$

д) $X \ge 5$, $P(X \ge 5) = 1 - P(X < 5) = 1 - F(5) = 1 - 1 = 0$.

Для дискретной случайной величины аналогом интегральной функции распределения является эмпирическая функция распределения (кумулята), графиком которой является ступенчатая линия.

Пример 29.2. Задан ряд распределения случайной величины X :

x_i	2	4	7
p_i	0,5	0,2	0,3

Найти функцию распределения $F(x)$ и построить ее график.

Решение.

$$F(x)=\begin{cases}0, & x\le 2,\\ 0{,}5, & 2<x\le 4,\\ 0{,}5+0{,}2=0{,}7, & 4<x\le 7,\\ 0{,}7+0{,}3=1, & x>7.\end{cases}$$

Построим график (рис. 29.1) полученной функции.

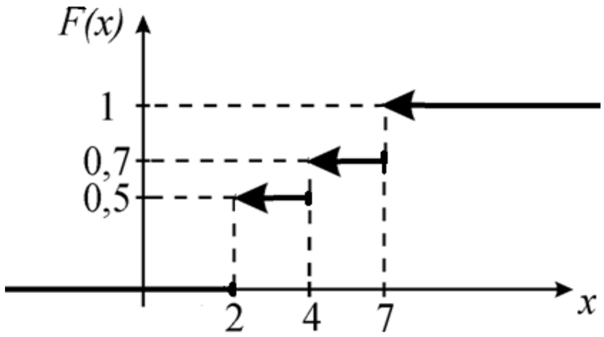


Рис. 29.1. **График функции распределения вероятностей**

Точками разрыва графика является значения x, в которых $F(x)$ изменяет свое значение. Если случайная величина задана интервалами, то эмпирическую функцию можно построить ломаной линией.

Пример 29.3. Заданы возможные интервалы значений случайной величины и их вероятности:

X	[1; 3)	[3; 5)	[5; 8)
p	0,5	0,2	0,3

Решение.

Найдем функцию распределения вероятностей $F(x)$ и построим ее график (рис. 29.2).

$$F(x)=\begin{cases} 0, & x \le 1, \\ 0{,}5, & 1 < x \le 3, \\ 0{,}5+0{,}2=0{,}7, & 3 < x \le 5, \\ 0{,}7+0{,}3=1, & 5 < x \le 8, \\ 1, & x > 8. \end{cases}$$

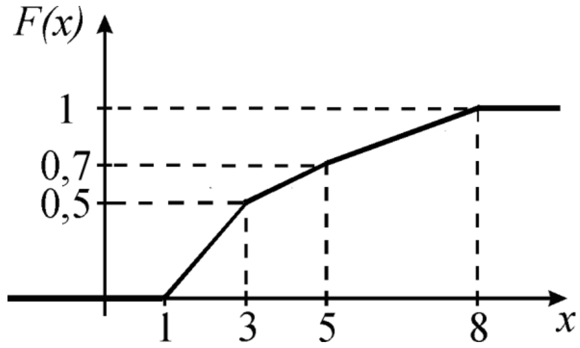


Рис. 29.2. **Эмпирическая интегральная функция распределения**

29.3. Плотность распределения вероятностей и ее свойства

Дифференциальная функция распределения вероятностей (плотность распределения вероятностей) $f(x)$ есть производная от интегральной функции распределения $F(x)$:

$$f(x)=F'(x). \tag{29.3}$$

Случайная величина X называется ***абсолютно непрерывной***, если существует такая функция $f(x)$ (плотность распределения вероятностей случайной величины X), для которой выполняется равенство:

$$F(x)=\int_{-\infty}^{x} f(x)\,dx.$$

Таким образом, поиск *интегральной функции*, если задана дифференциальная, связан с решением обратной задачи:

$$F(x)=\int_{-\infty}^{x} f(x)dx. \qquad (29.4)$$

Действительно $\int_{-\infty}^{x} f(x)dx=F(x)\Big|_{-\infty}^{x}=F(x)-F(-\infty)=F(x)$, поскольку $F(-\infty)=P(X<-\infty)=0$ ($X<-\infty$ – событие невозможное).

Вероятность того, что случайная величина примет значение, принадлежащее интервалу (a,b):

$$P(a<X<b)=F(b)-F(a)=\int_{-\infty}^{b} f(x)dx-\int_{-\infty}^{a} f(x)dx=\int_{a}^{b} f(x)dx,$$

то есть

$$P(a<X<b)=\int_{a}^{b} f(x)dx. \qquad (29.5)$$

Исходя из геометрического смысла определенного интеграла, можем сделать следующее заключение: вероятность $P(a<X<b)$ численно равна площади фигуры, которая ограничена прямыми $x=a$, $x=b$, $y=0$ и кривой $y=f(x)$.

Свойства дифференциальной функции распределения $f(x)$:

1. Функция $f(x)$ неотрицательная:

$$f(x)\geq 0.$$

Это свойство следует из того, что производная от неубывающей функции $F(x)$ является функцией неотрицательной.

2. Если $X\in(-\infty,\infty)$, то:

$$\int_{-\infty}^{\infty} f(x)dx=1. \qquad (29.6)$$

Действительно $\int_{-\infty}^{\infty} f(x)dx=P(-\infty<X<\infty)=1$, так как $-\infty<X<\infty$ – достоверное событие.

3. Если $X \in (a,b)$, то $\int\limits_a^b f(x)dx = 1$.

Пример 29.4. Задана плотность распределения $f(x)$ случайной величины X:

$$f(x) = \begin{cases} 0, & x \le 0, \\ cx, & 0 < x \le 1, \\ 0, & x > 1. \end{cases}$$

Найти: а) параметр c; б) функцию $F(x)$; в) $P(0{,}5 < X < 1)$. Изобразить графики функций $F(x)$ и $f(x)$.

Решение:

а) поскольку $\int\limits_a^b f(x)dx = 1$, то $\int\limits_0^1 f(x)dx = 1$.

Откуда имеем $\int\limits_0^1 cx\,dx = c\dfrac{x^2}{2}\Big|_0^1 = \dfrac{c}{2} = 1$.

Следовательно, $c = 2$ и $f(x) = \begin{cases} 0, & x < 0, \\ 2x, & 0 \le x \le 1, \\ 0, & x > 1; \end{cases}$

б) так как $F(x) = \int\limits_{-\infty}^{x} f(x)dx$, то:

если $x \le 0$, $F(x) = \int\limits_{-\infty}^{x} 0dx = 0$;

если $x \le 1$, $F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{0} 0dx + \int\limits_0^x 2xdx = x^2\Big|_0^x = x^2$;

если $x > 1$, $F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{0} 0dx + \int\limits_0^1 2xdx + \int\limits_0^x 0dx = x^2\Big|_0^1 = 1$.

Следовательно, $F(x) = \begin{cases} 0, & x \le 0, \\ x^2, & 0 < x \le 1, \\ 1, & x > 1; \end{cases}$

в) $P(0{,}5 < X < 1) = F(1) - F(0{,}5) = 1 - 0{,}25 = 0{,}75$,

или

$$P(0{,}5 < X < 1) = \int_{0,5}^{1} 2x\,dx = x^2\Big|_{0,5}^{1} = 0{,}75.$$

Графики интегральной и дифференциальной функций распределения $F(x)$ и $f(x)$ изображены на рис. 29.3.

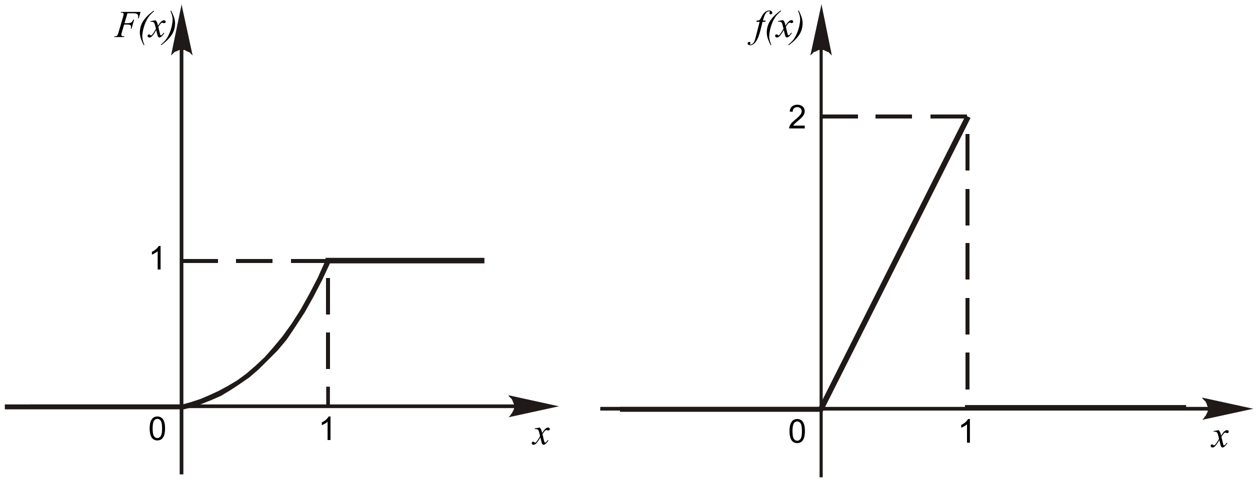


Рис. 29.3. **Графики функций** $F(x)$ **и** $f(x)$

Вероятностный смысл плотности распределения

Если $F(x)$ – функция распределения вероятностей непрерывной случайной величины X, то:

$$f(x) = F'(x), \text{ то есть } f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

Известно, что $F(x + \Delta x) - F(x) = P(x < X < x + \Delta x)$.

Учитывая, что $F(x + \Delta x) - F(x) = \Delta F(x)$, а $\Delta F(x) \approx dF(x) = f(x)\Delta x$, имеем:

$$P(x < X < x + \Delta x) \approx f(x)\Delta x.$$

Таким образом, вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу $(x, x + \Delta x)$ (при $\Delta x \to 0$), приближенно равна произведению дифференциальной функции на длину интервала Δx, то есть дифференциальная функция выступает в роли плотности распределения вероятностей.

29.4. Числовые характеристики непрерывной случайной величины

Пусть X – непрерывная случайная величина, которая может принимать всевозможные значения на отрезке $[a,b]$ и имеет плотность распределения вероятностей $f(x)$. Разобьем промежуток $[a,b]$ на n частей точками $a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$.

Получим отрезки $[x_0, x_1]$, $[x_1, x_2]$, … , $[x_{n-1}, x_n]$. Выберем на каждом из них произвольную точку x_i $(i = \overline{1,n})$. Согласно вероятностному смыслу плотности распределения $f(x_i)\Delta x_i$ равна вероятности попадания случайной величины X на интервал Δx_i.

Используя формулу для математического ожидания дискретной случайной величины X, получим:

$$M(X_n) = \sum_{i=1}^{n} x_i f(x_i)\Delta x_i .$$

Если $\max\limits_{i \to 0} \Delta x_i \to 0$ и $n \to \infty$, то дискретная величина X_n будет все меньше отличаться от непрерывной величины X.

Функция $xf(x)$ – непрерывна, тогда имеем:

$$\lim_{\substack{\max \Delta x_i \to 0 \\ n \to \infty}} \sum_{i=1}^{n} x_i f(x_i)\Delta x_i = \int_a^b xf(x)dx.$$

Следовательно, ***математическое ожидание непрерывной случайной величины*** X, возможные значения которой принадлежат отрезку $[a,b]$, вычисляется как определенный интеграл:

$$M(X) = \int_a^b xf(x)dx. \qquad (29.7)$$

Аналогично, если $X \in (-\infty, \infty)$, то $M(X) = \int_{-\infty}^{\infty} xf(x)dx.$

Дисперсия непрерывной случайной величины X вычисляется как математическое ожидание квадрата отклонения случайной величины от своего математического ожидания.

Если $X \in (-\infty, \infty)$, то:

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 f(x) dx, \tag{29.8}$$

или

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (M(X))^2. \tag{29.9}$$

Если $X \in [a, b]$, то:

$$D(X) = \int_{a}^{b} (x - M(X))^2 f(x) dx, \tag{29.10}$$

или

$$D(X) = \int_{a}^{b} x^2 f(x) dx - (M(X))^2. \tag{29.11}$$

Свойства математического ожидания и дисперсии дискретных величин сохраняются и для непрерывных величин.

Среднее квадратическое отклонение непрерывной случайной величины X:

$$\sigma(X) = \sqrt{D(X)}. \tag{29.12}$$

Модой M_o непрерывной случайной величины называют такое значение случайной величины, для которого дифференциальная функция максимальна.

Медианой m_e называют такое значение случайной величины, для которого выполняется равенство $P(X > m_e) = P(X < m_e)$. Геометрически медиану можно определить как точку, в которой ордината функции $f(x)$ разделяет пополам площадь под кривой распределения (дифференциальной).

Моменты непрерывной случайной величины:

а) ***начальный*** момент k-го порядка:

$$\nu_k = \int_{-\infty}^{\infty} x^k f(x) dx; \tag{29.13}$$

б) ***центральный*** момент k-го порядка:

$$\mu_k = \int_{-\infty}^{\infty} (x - M(x))^k f(x)\,dx. \qquad (29.14)$$

Пример 29.5. Задана дифференциальная функция распределения непрерывной случайной величины X:

$$f(x) = \begin{cases} 0, & x < 1, \\ x - \dfrac{1}{2}, & 1 \le x \le 2, \\ 0, & x > 2. \end{cases}$$

Найти интегральную функцию распределения $F(x)$. Изобразить графики функций $f(x)$ и $F(x)$. Найти $M(X)$ и $D(X)$.

Решение.

Функцию распределения $F(x)$ найдем по формуле (29.4):

$$F(x) = \int_{-\infty}^{x} f(x)\,dx.$$

Если $x < 1$, то $f(x) = 0$, откуда $F(x) = 0$;

если $1 \le x \le 2$, то:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{1} 0\,dx + \int_{1}^{x} \left(x - \frac{1}{2}\right) dx = \left(\frac{x^2}{2} - \frac{x}{2}\right)\Bigg|_1^x = \frac{x^2}{2} - \frac{x}{2};$$

если $x > 2$, то:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{1} 0\,dx + \int_{1}^{2} \left(x - \frac{1}{2}\right) dx + \int_{2}^{x} 0\,dx = \left(\frac{x^2}{2} - \frac{x}{2}\right)\Bigg|_1^2 = 1.$$

Окончательно имеем $F(x)$:

$$F(x) = \begin{cases} 0, & x < 1, \\ \dfrac{x^2}{2} - \dfrac{x}{2}, & 1 \le x \le 2, \\ 1, & x > 2. \end{cases}$$

Построим графики функций $f(x)$ и $F(x)$ (рис. 29.4):

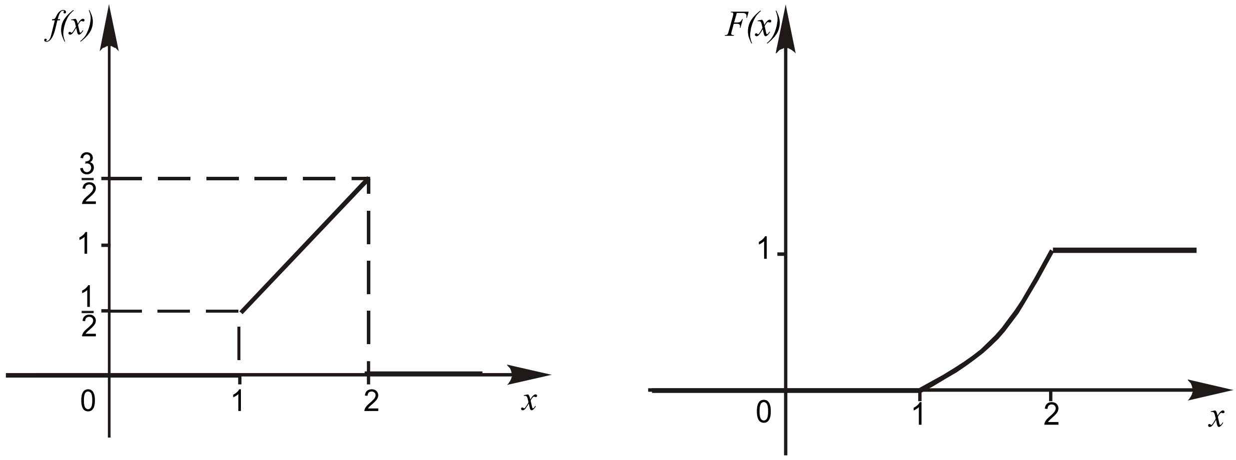


Рис. 29.4. **Графики плотности распределения $f(x)$**
и функции распределения $F(x)$

Найдем числовые характеристики:

$$M(X)=\int_1^2 x\cdot\left(x-\frac{1}{2}\right)dx=\int_1^2\left(x^2-\frac{x}{2}\right)dx=\left(\frac{x^3}{3}-\frac{x^2}{4}\right)\Bigg|_1^2=1{,}58;$$

$$D(X)=\int_1^2 x^2\cdot\left(x-\frac{1}{2}\right)dx-1{,}58^2=\int_1^2\left(x^3-\frac{x^2}{2}\right)dx-1{,}58^2=$$

$$=\left(\frac{x^4}{4}-\frac{x^3}{6}\right)\Bigg|_1^2-1{,}58^2=0{,}08;$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}08}=0{,}28.$$

Вопросы для самодиагностики

1. Привести определение непрерывной случайной величины. Что называется функцией распределения и каковы ее свойства?
2. Что называется плотностью распределения вероятностей и каковы ее свойства?

3. Каков вероятностный смысл плотности распределения вероятностей?

4. Охарактеризовать числовые характеристики непрерывной случайной величины.

Упражнения

29.1. Случайная величина X задана функцией распределения

$$F(x)=\begin{cases}0, & x\le 0,\\ \sin x, & 0<x\le \pi/2,\\ 1, & x>\pi/2.\end{cases}$$

Найти плотность распределения $f(x)$.

29.2. Непрерывная случайная величина X задана плотностью распределения на всей числовой оси:

$$f(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

Найти вероятность того, что X примет значение на интервале (–1, 1).

29.3. Задана плотность распределения случайной величины X:

$$f(x)=\begin{cases}0, & x\le 0\\ a(3x-x^2), & 0<x\le 3\\ 0, & x>3\end{cases}$$

Найти коэффициент a и функцию распределения $F(x)$.

29.4. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X, заданной плотностью распределения на отрезке [0; 1]: $f(x)=x$.

29.5. Найти основные числовые характеристики непрерывной случайной величины X, заданной функцией распределения вероятностей на положительной полуоси Ox: $F(x)=1-e^{-x}$.

29.6. Случайная величина X задана на интервале (0, 5) плотностью распределения $f(x)=2x/25$; вне этого интервала $f(x)=0$. Найти дисперсию X.

29.7. Непрерывная случайная величина X задана на всей оси Ox функцией распределения $F(x)=1/2+arctgx/\pi$. Найти вероятность того, что величина X примет значение, заключенное в интервале (0, 1).

30. Равномерный, показательный и нормальный законы распределения вероятностей. Преобразование последовательностей нормально распределенных случайных величин

30.1. Равномерный закон распределения вероятностей и его числовые характеристики

Равномерным распределением вероятностей непрерывной случайной величины называют такое распределение, при котором дифференциальная функция является постоянной величиной на интервале $[a,b]$, а вне этого интервала равна нулю:

$$f(x)=\begin{cases}0, & x<a,\\ c, & a\le x\le b,\\ 0, & x>b.\end{cases} \tag{30.1}$$

Определим параметр c.

Исходя из свойства функции $f(x)$, имеем:

$$\int_a^b f(x)\,dx=1,\quad \int_a^b c\,dx=cx\Big|_a^b=c(b-a),\quad c(b-a)=1\Rightarrow c=\frac{1}{b-a}.$$

Итак, для равномерного распределения вероятностей дифференциальная функция $f(x)$ имеет вид:

$$f(x)=\begin{cases}0, & x<a,\\ \dfrac{1}{b-a}, & a\le x\le b,\\ 0, & x>b.\end{cases} \tag{30.2}$$

Найдем интегральную функцию распределения $F(x)$ для $X \in [a,b]$:

$$F(x) = \int_a^x f(x)dx = \int_a^x \frac{1}{b-a}dx = \frac{x}{b-a}\bigg|_a^x = \frac{x-a}{b-a}.$$

Таким образом, $F(x)$ запишем в виде:

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b. \end{cases} \qquad (30.3)$$

Графики дифференциальной функции $f(x)$ и интегральной функции $F(x)$ равномерного распределения вероятностей приведены на рис. 30.1.

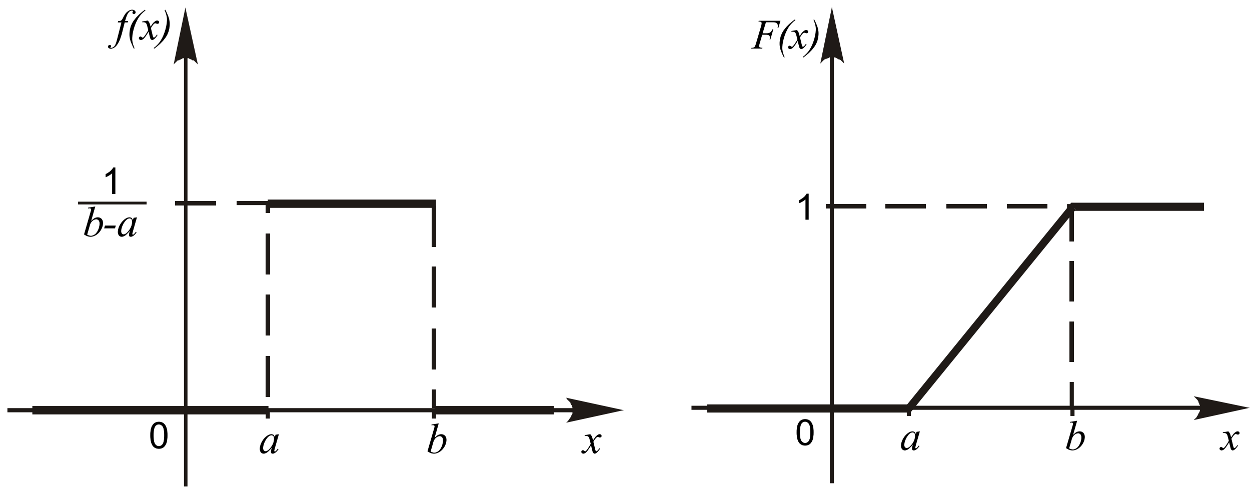


Рис. 30.1. **Графики дифференциальной и интегральной функций равномерного распределения вероятностей**

Определим вероятность попадания случайной величины X в интервал $(\alpha, \beta) \in [a,b]$:

$$\boldsymbol{P(\alpha < X < \beta)} = F(\beta) - F(\alpha) = \frac{\beta - a}{b-a} - \frac{\alpha - a}{b-a} = \boldsymbol{\frac{\beta-\alpha}{b-a}}. \qquad (30.4)$$

Вычислим числовые характеристики случайной величины, распределенной по равномерному закону.

Математическое ожидание $M(X)$ определим по формуле (29.7):

$$\boldsymbol{M(X)} = \int_a^b x \cdot f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\,\frac{x^2}{2}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \boldsymbol{\frac{a+b}{2}}. \quad (30.5)$$

Дисперсия $D(X)$ (формула (29.9)):

$$\boldsymbol{D(X)} = \int_a^b x^2 \cdot f(x)\,dx - (M(X))^2 = \frac{1}{b-a}\int_a^b x^2\,dx - \left(\frac{a+b}{2}\right)^2 =$$

$$= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \boldsymbol{\frac{(b-a)^2}{12}}. \quad (30.6)$$

Среднее квадратическое отклонение $\sigma(X)$ вычислим по формуле (29.12):

$$\boldsymbol{\sigma(X)} = \sqrt{D(X)} = \boldsymbol{\frac{b-a}{2\sqrt{3}}}. \quad (30.7)$$

Пример 30.1. Автобус прибывает на остановку с интервалом 5 минут. Найти вероятность того, что автобус появится в последние две минуты. Найти $M(X)$, $D(X)$, $\sigma(X)$.

Решение.

По условию задачи $b-a=5$, $\alpha=3$, $\beta=5$.

$$P(3<X<5) = F(5) - F(3) = \frac{x-0}{5-0}\bigg|_{x=5} - \frac{x-0}{5-0}\bigg|_{x=3} = \frac{5-3}{5} = \frac{2}{5} = 0{,}4;$$

$$M(X) = \frac{0+5}{2} = 2{,}5; \quad D(X) = \frac{25}{12} = 2{,}08; \quad \sigma(X) = 1{,}44.$$

Равномерный закон распределения вероятностей применяется при работе с округленными числами. Например, если число округлено до целого, то ошибка округления Δ распределена равномерно на отрезке $[-0{,}5; 0{,}5]$.

30.2. Показательный закон распределения вероятностей

Непрерывная случайная величина X распределена по ***показательному закону***, если плотность распределения вероятностей:

$$f(x)=\begin{cases}0, & x<0,\\ \lambda\cdot e^{-\lambda x}, & x\geq 0,\end{cases} \quad \text{с параметром } \lambda>0. \qquad (30.8)$$

Проверим, что функция, которая задана в таком виде, удовлетворяет свойствам дифференциальной функции распределения.

Действительно $f(x)\geq 0$ и

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{0}^{\infty}\lambda e^{-\lambda x}dx=\lambda\lim_{b\to\infty}\int_{0}^{b}e^{-\lambda x}dx=\lambda\lim_{b\to\infty}\left(-\frac{1}{\lambda}e^{-\lambda x}\right)\Bigg|_{0}^{b}=1.$$

Интегральная функция показательного распределения $F(x)$ имеет вид:

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{0}0dx+\int_{0}^{x}\lambda e^{-\lambda x}dx=-e^{-\lambda x}\Big|_{0}^{x}=1-e^{-\lambda x}.$$

Окончательно:

$$F(x)=\begin{cases}0, & x<0,\\ 1-e^{-\lambda x}, & x\geq 0.\end{cases} \qquad (30.9)$$

Графики функций $f(x)$ и $F(x)$ приведены на рис. 30.2.

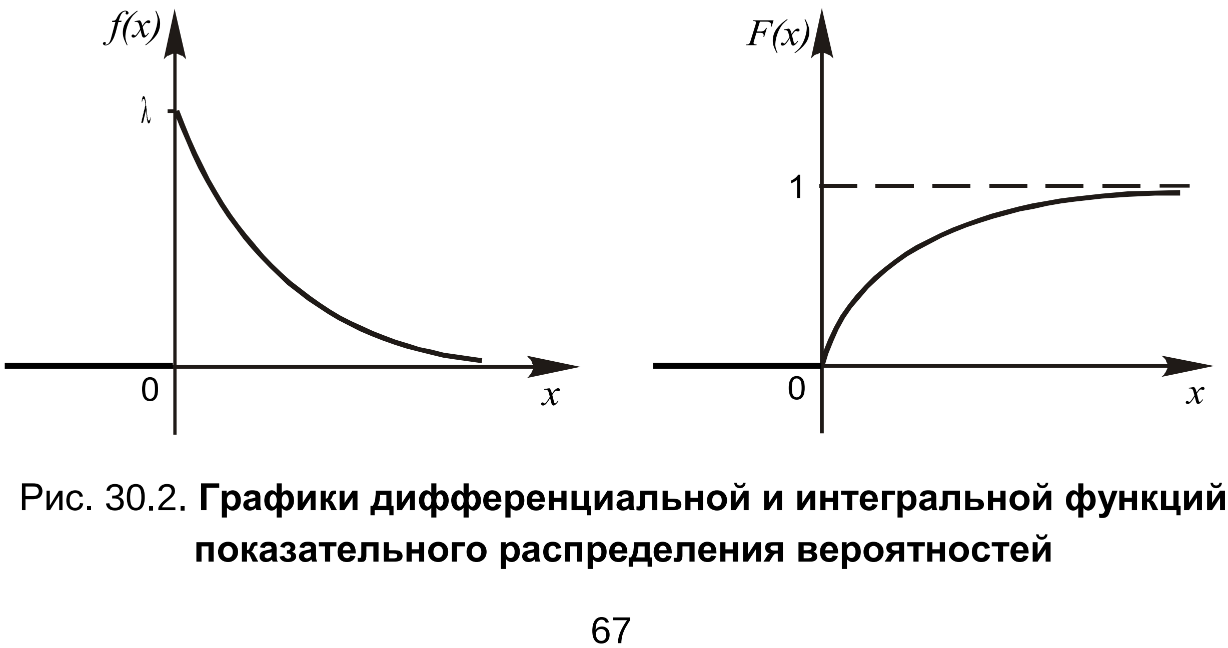


Рис. 30.2. **Графики дифференциальной и интегральной функций показательного распределения вероятностей**

Вероятность попадания случайной величины X в интервал (α, β) определяется формулой:

$$\boldsymbol{P(\alpha < X < \beta)} = F(\beta) - F(\alpha) = \left(1 - e^{-\lambda\beta}\right) - \left(1 - e^{-\lambda\alpha}\right) = \boldsymbol{e^{-\lambda\alpha} - e^{-\lambda\beta}}. \qquad (30.10)$$

Вычислим числовые характеристики показательного закона распределения вероятностей.

Математическое ожидание $M(X)$:

$$\boldsymbol{M(X)} = \int_a^{\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda \lim_{b\to\infty} \int_0^b x \cdot e^{-\lambda x} dx = \begin{vmatrix} u = x & du = dx \\ dv = e^{-\lambda x} dx & v = -\dfrac{1}{\lambda} e^{-\lambda x} \end{vmatrix} = \qquad (30.11)$$

$$= \lambda \lim_{b\to\infty} \left(-\frac{x}{\lambda} e^{-\lambda x} \Big|_0^b + \frac{1}{\lambda} \int_0^b e^{-\lambda x} dx \right) = \lim_{b\to\infty} \left(0 - e^{-\lambda b}\right) - \frac{1}{\lambda} \lim_{b\to\infty} e^{-\lambda x} \Big|_0^b = 0 + \frac{1}{\lambda} = \boldsymbol{\frac{1}{\lambda}}.$$

Дисперсия $D(X)$:

$$\boldsymbol{D(X)} = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx - \left(M(X)\right)^2 = \lambda \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} =$$

$$= \begin{vmatrix} u = x^2 & du = 2x dx \\ dv = e^{-\lambda x} dx & v = -\dfrac{1}{\lambda} e^{-\lambda x} \end{vmatrix} == \lambda \lim_{b\to\infty} \left(-\frac{x^2}{\lambda} e^{-\lambda x} \Big|_0^b + \frac{2}{\lambda} \int_0^b x \cdot e^{-\lambda x} dx \right) - \frac{1}{\lambda^2} = \qquad (30.12)$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \boldsymbol{\frac{1}{\lambda^2}}.$$

Среднее квадратическое отклонение $\sigma(X)$:

$$\boldsymbol{\sigma(X)} = \sqrt{D(X)} = \boldsymbol{\frac{1}{\lambda}}. \qquad (30.13)$$

Отметим, что при показательном распределении математическое ожидание равно среднему квадратическому отклонению:

$$M(X) = \sigma(X) = \frac{1}{\lambda}.$$

Показательный закон используют в теории массового обслуживания.

Пример 30.2. Среднее время обслуживания покупателя составляет 20 минут и распределено по показательному закону. Какова вероятность простоять в очереди от 20 до 40 минут?

Решение. По условию $M(X)=20$, тогда $1/\lambda=20$.

$$P(20<X<40)=F(40)-F(20)=e^{-20\cdot 1/20}-e^{-40\cdot 1/20}=e^{-1}-e^{-2}=$$
$$=0{,}368-0{,}135=0{,}233.$$

Функцией, которая определяет вероятность безотказной работы элемента за промежуток времени длиной t, является ***функция надежности*** $R(t)$: $R(t)=P(T\geq t)$. События $(T\geq t)$ и $(T<t)$ противоположные. Функция распределения $F(t)=P(T<t)$ определяет вероятность отказа элемента за время длиной t.

Таким образом, для показательного распределения:

$$\boldsymbol{R(t)}=P(T\geq t)=1-P(T<t)=1-F(t)=\boldsymbol{e^{-\lambda t}}, \qquad (30.14)$$

где λ – интенсивность отказов.

Пример 30.3. Случайная величина T – время работы лампы накаливания – имеет показательное распределение. Определить вероятность того, что время работы лампы будет составлять не меньше 600 часов, если среднее время работы лампы – 400 часов.

Решение. $M(X)=400$, тогда $\lambda=\dfrac{1}{M(X)}=\dfrac{1}{400}$, $t=600$.

$$R(t)=P(T\geq 600)=e^{-\frac{1}{400}\cdot 600}=e^{-1{,}5}=0{,}223.$$

30.3. Нормальный закон распределения вероятностей и его стандартное представление

Случайная величина X распределена по нормальному закону, если ее плотность распределения вероятностей при $x\in(-\infty;\infty)$ определяется формулой:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (30.15)$$

где σ, a – параметры распределения.

Проверим, что $f(x)$ удовлетворяет свойствам дифференциальной функции распределения. Действительно $f(x)>0$.

Вычислим $\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$

Сделаем замену $t=\frac{x-a}{\sigma}$, $dx=\sigma dt$. Пределы интегрирования при этом сохраняются.

Тогда имеем:

$$\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}} \sigma dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1,$$

где $\int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ – интеграл Пуассона.

Определим интегральную функцию распределения $F(x)$ для нормального закона:

$$F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Сделаем замену $t=\frac{x-a}{\sigma}$; $dx=\sigma dt$.

Новые пределы интегрирования: $x\to-\infty,\ t\to-\infty$;

$x\to x,\ t\to\frac{x-a}{\sigma}$.

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} \sigma dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt.$$

Здесь $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} e^{-\frac{t^2}{2}} dt = \frac{1}{2}$ – интеграл Пуассона,

$$\frac{1}{\sqrt{2\pi}}\int_0^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}}dt=\Phi\left(\frac{x-a}{\sigma}\right), \text{ поскольку } \frac{1}{\sqrt{2\pi}}\int_0^{x} e^{-\frac{t^2}{2}}dt=\Phi(x).$$

Таким образом:

$$\boldsymbol{F(x)}=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{2}\sqrt{2\pi}+\Phi\left(\frac{x-a}{\sigma}\right)=\boldsymbol{\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right)}. \qquad (30.16)$$

Если ввести центрированную и нормированную величину $t=\dfrac{x-a}{\sigma}$, такую, что $a=0$, $\sigma=1$, то

$$\boldsymbol{f(x)=\varphi(t)}, \quad \boldsymbol{F(x)=\frac{1}{2}+\Phi(t)}, \qquad (30.17)$$

где $\varphi(t)$, $\Phi(t)$ – дифференциальная и интегральная функции Лапласа:

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}, \quad \Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt.$$

Графики дифференциальной $f(x)$ и интегральной $F(x)$ функций нормального распределения приведены на рис. 30.3.

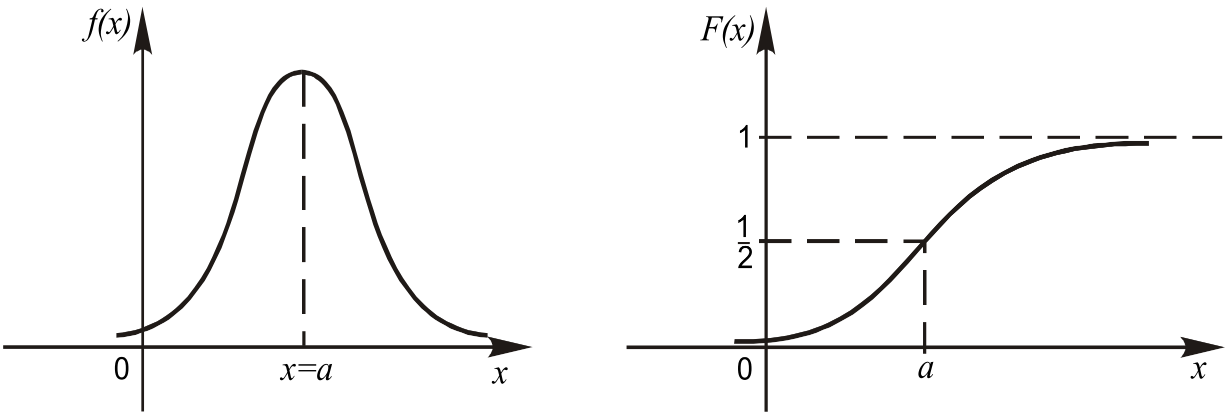


Рис. 30.3. **Графики дифференциальной и интегральной функций нормального распределения вероятностей**

Докажем, что параметр a – математическое ожидание $M(X)$, а параметр σ – среднее квадратическое отклонение $\sigma(X)$.

По формуле для $M(X)$ непрерывной случайной величины:

$$M(X)=\int_{-\infty}^{\infty} x\cdot f(x)dx,$$

для нормального распределения имеем:

$$M(X)=\int_{-\infty}^{\infty} x\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

Сделаем замену $t=\dfrac{x-a}{\sigma}$; $x=\sigma t+a$; $dx=\sigma dt$. Новые пределы интегрирования при этом равны исходным:

$$M(X)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma t+a)e^{-\frac{t^2}{2}}\sigma dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}te^{-\frac{t^2}{2}}dt+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt.$$

Первый интеграл равен нулю, как интеграл от нечетной функции по промежутку, симметричному относительно начала координат.

Второй интеграл является интегралом Пуассона:

$$\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}.$$

Таким образом $\boldsymbol{M(X)}=\dfrac{a}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\boldsymbol{a}$.

Дисперсия $D(X)$ непрерывной случайной величины X:

$$D(X)=\int_{-\infty}^{\infty}(x-M(X))^2 f(x)dx.$$

Тогда для нормального распределения:

$$D(X)=\int_{-\infty}^{\infty}(x-a)^2\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

Сделаем ту же замену, как и для $M(X)$, то есть $t=\dfrac{x-a}{\sigma}$, тогда:

$$D(X)=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t\cdot t\,e^{-\frac{t^2}{2}}dt=\left|\begin{array}{ll} u=t & du=dt \\ dv=te^{-\frac{t^2}{2}}dt & v=\int e^{-\frac{t^2}{2}}d\left(\frac{t^2}{2}\right)=-e^{-\frac{t^2}{2}}\end{array}\right|=$$

$$=-\underbrace{\frac{\sigma^2}{\sqrt{2\pi}}te^{-\frac{t^2}{2}}\Bigg|_{-\infty}^{\infty}}_{=0}+\frac{\sigma^2}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt}_{\text{интеграл Пуассона}}=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\boldsymbol{\sigma}^2.$$

То есть получим:

$$\boldsymbol{a}=\boldsymbol{M}(\boldsymbol{X}),\quad \boldsymbol{\sigma}=\sqrt{\boldsymbol{D}(\boldsymbol{X})}. \tag{30.18}$$

Для нормального распределения кривая распределения – функция $f(x)$ – достигает максимума при $x=a$ и симметрична относительно линии $x=a$.

Найдем вероятность попадания случайной величины X, распределенной по нормальному закону с параметрами a и σ, в промежуток (α,β):

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)=\frac{1}{2}+\Phi\left(\frac{\beta-a}{\sigma}\right)-\left(\frac{1}{2}+\Phi\left(\frac{\alpha-a}{\sigma}\right)\right)=$$

$$=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right).$$

Таким образом:

$$\boldsymbol{P}(\boldsymbol{\alpha}<\boldsymbol{X}<\boldsymbol{\beta})=\boldsymbol{\Phi}\left(\frac{\boldsymbol{\beta}-\boldsymbol{a}}{\boldsymbol{\sigma}}\right)-\boldsymbol{\Phi}\left(\frac{\boldsymbol{\alpha}-\boldsymbol{a}}{\boldsymbol{\sigma}}\right). \tag{30.19}$$

Пример 30.4. Случайная величина X распределена по нормальному закону и имеет плотность распределения:

$$f(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{(x-3)^2}{8}}.$$

Найти числовые характеристики величины X и вероятность попадания ее в интервал (1; 7).

Решение.

По определению функции $f(x)$ имеем: $a=3,\ \sigma=2$.

Таким образом, $M(X)=3,\ \sigma(X)=2,\ D(X)=4$.

Тогда вероятность попадания случайной величины X в интервал (1; 7) по формуле (30.19) равна:

$$P(1<X<7)=\Phi\left(\frac{7-3}{2}\right)-\Phi\left(\frac{1-3}{2}\right)=\Phi(2)-\Phi(-1)=0{,}477+0{,}341=0{,}818.$$

Пример 30.5. Автоматический станок штампует детали. Длина детали является случайной величиной, распределенной по нормальному закону с параметрами $a=10$ см, $\sigma^2=0{,}0004$. Найти вероятность брака, если допустимые размеры детали должны быть $10\pm0{,}05$см.

Решение.

$$P_{\text{брака}}+P(9{,}95<X<10{,}05)=1;$$

$$P_{\text{брака}}=1-P(9{,}95<X<10{,}05);\ \alpha=9{,}95;\ \beta=10{,}05;\ \sigma=0{,}02.$$

$$P(9{,}95<X<10{,}05)=\Phi\left(\frac{10{,}05-10}{0{,}02}\right)-\Phi\left(\frac{9{,}95-10}{0{,}02}\right)=$$

$$=\Phi\left(\frac{0{,}05}{0{,}02}\right)-\Phi\left(\frac{-0{,}05}{0{,}02}\right)=2\Phi(2{,}5)=2\cdot0{,}494=0{,}988.$$

Таким образом, вероятность брака равна:

$$P_{\text{брака}}=1-0{,}988=0{,}012.$$

Найдем *вероятность отклонения нормально распределенной случайной величины от ее математического ожидания по модулю на величину, не превышающую* ε ($\varepsilon>0$), то есть найдем $P(|X-a|<\varepsilon)$:

$$P(|X-a|<\varepsilon)=P(a-\varepsilon<X<a+\varepsilon)=\Phi\left(\frac{a+\varepsilon-a}{\sigma}\right)-\Phi\left(\frac{a-\varepsilon-a}{\sigma}\right)=$$

$$=\Phi\left(\frac{\varepsilon}{\sigma}\right)-\Phi\left(-\frac{\varepsilon}{\sigma}\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Таким образом, $$\boldsymbol{P}(|\boldsymbol{X}-\boldsymbol{a}|<\boldsymbol{\varepsilon})=\mathbf{2}\Phi\left(\frac{\boldsymbol{\varepsilon}}{\boldsymbol{\sigma}}\right). \quad (30.20)$$

Пример 30.6. Деталь, изготовленная на станке, считается стандартной, если отклонение ее размера от проектного не больше 10 мм. Случайные отклонения распределены по нормальному закону: $\sigma=5$, $a=0$. Найти, какой процент стандартных деталей изготовляется на станке.

Решение.

По условию $\varepsilon=10$, $\sigma=5$.

Имеем: $P(|X|<\varepsilon)=2\Phi\left(\frac{\varepsilon}{\sigma}\right)=2\Phi\left(\frac{10}{5}\right)=2\Phi(2)=2\cdot 0{,}477=0{,}954$.

Следовательно, на станке изготавливается приблизительно 95% стандартных деталей.

Правило трех сигм

Преобразуем формулу (30.20).

Пусть $\varepsilon=\sigma\cdot t$, тогда $P(|X-a|<\sigma t)=2\Phi(t)$.

Если $t=1$, то $\varepsilon=\sigma$, тогда: $P(|X-a|<\sigma)=2\Phi(1)=0{,}6826$.

Это значит, что 68 % значений случайной величины находятся на промежутке $(a\pm\sigma)$.

Если $t=2$, то $\varepsilon=2\sigma$, тогда: $P(|X-a|<2\sigma)=2\Phi(2)=0{,}9544$.

Это значит, что 95 % значений случайной величины находятся на промежутке $(a\pm 2\sigma)$.

И последнее: $t=3\Rightarrow\varepsilon=3\sigma$, имеем:

$$P(|X-a|<3\sigma)=2\Phi(3)=0{,}9973.$$

Отсюда *правило трех сигм*: нормально распределенная случайная величина X принимает все свои значения на промежутке $(a\pm 3\sigma)$ с достоверностью приблизительно равной 100%.

То есть из 10 000 значений нормально распределенной случайной величины лишь 27 выйдут за пределы интервала $(a\pm 3\sigma)$.

Пример 30.7. На станке изготавливают шары, диаметр которых является случайной величиной X, распределенной по нормальному закону, имеющей среднее значение $a = 2$ мм и $\sigma = 0{,}1$ мм. Какие размеры диаметра шаров можно гарантировать с надежностью 99,73 %?

Решение.

По условию задачи $P = 0{,}9973$, $3\sigma = 0{,}3$.

То есть $a - 3\sigma < X < a + 3\sigma$, $2 - 0{,}3 < X < 2 + 0{,}3$, $1{,}7 < X < 2{,}3$.

При изучении распределений, которые отличаются от нормального, возникает необходимость оценить это отличие. С этой целью вводят такие характеристики, как *асимметрия* и *эксцесс*. Для нормального распределения асимметрия и эксцесс равны нулю. Большие значения асимметрии и эксцесса указывают на значительное отклонение от нормального распределения, при малых значениях асимметрии и эксцесса можно допустить близость этого распределения к нормальному.

Асимметрией распределения называют величину:

$$A_S = \frac{\mu_3}{\sigma^3}, \qquad (30.21)$$

где μ_3 – центральный момент третьего порядка;

σ – среднее квадратическое отклонение.

Асимметрия характеризует отклонение кривой распределения $f(x)$ от центра симметрии нормального распределения $x = a$, то есть моды. Если $A_S > 0$, то максимум функции $f(x)$ отходит влево; если $A_S < 0$ – вправо, при этом значение максимума сохраняется (рис. 30.4).

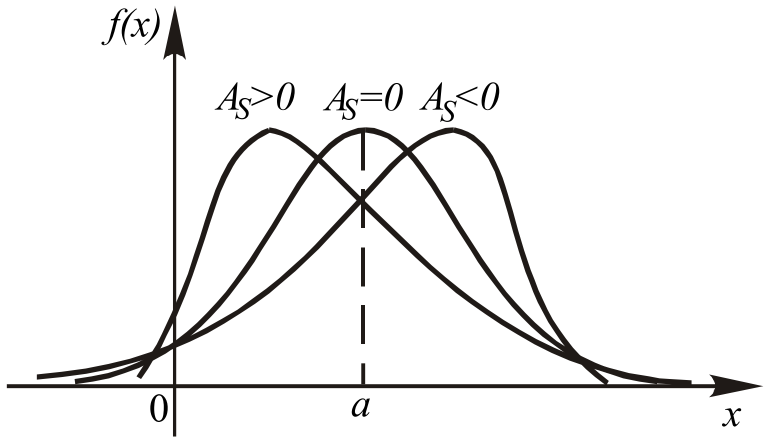


Рис. 30.4. **Асимметрия распределения**

Эксцессом распределения называют величину:

$$E_S = \frac{\mu_4}{\sigma^4} - 3, \qquad (30.22)$$

где μ_4 – центральный момент четвертого порядка.

Эксцесс распределения характеризует смещение максимума кривой распределения вдоль оси симметрии $x = a$ (рис. 30.5).

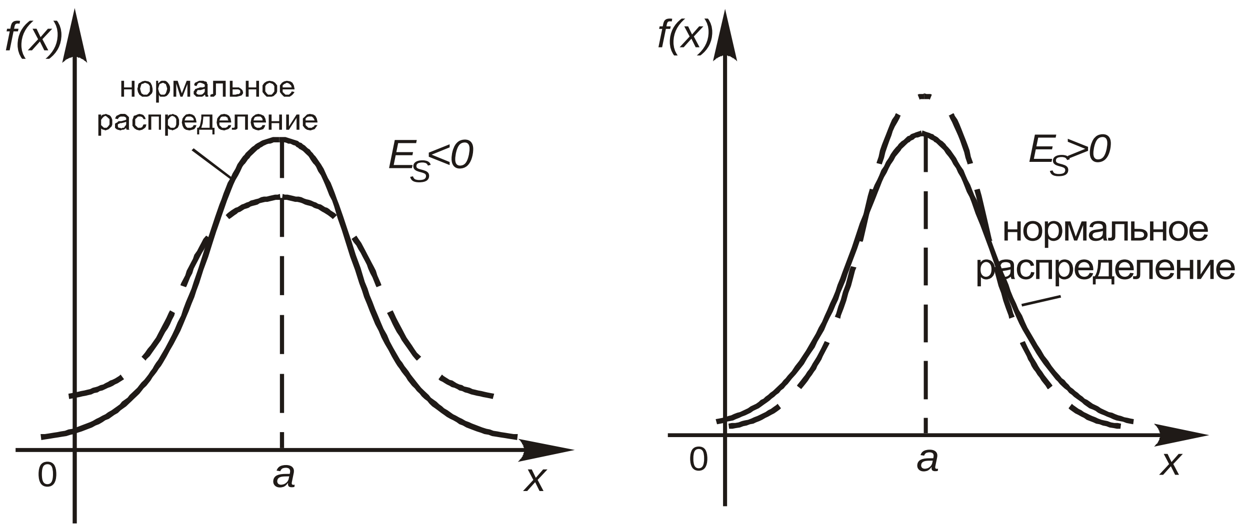


Рис. 30.5. **Эксцесс распределения**

Пример 30.8. Дано $\mu_3 = 0{,}4$, $\mu_4 = 6{,}17$, $\sigma^2 = 1{,}66$. Найти A_S, E_S.

Решение.

Асимметрия: $A_S = \dfrac{0{,}4}{1{,}66\sqrt{1{,}66}} \approx 0{,}19.$

Эксцесс: $E_S = \dfrac{6{,}17}{1{,}66^2} - 3 = -0{,}76.$

Можно сказать, что кривая распределения будет отходить влево ($A_S > 0$) относительно $x = a$ и максимум будет меньше, чем у кривой нормального распределения ($E_S < 0$).

Нормально распределенные случайные величины широко распространены на практике.

Если $X = \dfrac{X_1 + X_2 + ... + X_n}{n}$, где X_i – независимые случайные величины, $M(X_i) = a$, $D(X_i) = \dfrac{\sigma^2}{n}$, $\sigma(X_i) = \dfrac{\sigma}{\sqrt{n}}$, то

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right), \text{ то есть}$$

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right). \tag{30.23}$$

Содержание формулы (30.23) таково: *с вероятностью P можно утверждать, что доверительный интервал $(a-\varepsilon; a+\varepsilon)$ покрывает неизвестный параметр a с надежностью $P = 2\Phi(t)$ и точностью оценки $\varepsilon = \dfrac{t\sigma}{\sqrt{n}}$*. Оценку $|X - a| < t\dfrac{\sigma}{\sqrt{n}}$ называют ***классической***.

Из формулы $\varepsilon = \dfrac{t\sigma}{\sqrt{n}}$, которая определяет точность классической оценки, можно сделать выводы:

1) с ростом n число ε убывает, то есть точность оценки увеличивается;

2) увеличение вероятности $P = 2\Phi(t)$ приводит к росту t ($\Phi(t)$ – возрастающая функция) и тем самым к росту ε.

То есть увеличение вероятности классической оценки влечет за собой уменьшение ее точности.

30.4. Распределения Стьюдента и Фишера

Пусть X_0 и $X_1, X_2, ..., X_n$ попарно независимые случайные величины, каждая из которых распределена по нормальному закону с параметрами $a = 0, \sigma = 1$.

Последовательность случайных величин $X_1, X_2, ..., X_n$ преобразована в случайную величину $\chi^2(n) = X_1^2 + X_2^2 + ... + X_n^2$. $\chi^2(n)$ распределена по так называемому закону «хи-квадрат» (распределение Пирсона) с n степенями свободы, $M\left(\chi^2(n)\right) = n, D\left(\chi^2(n)\right) = 2n$. Тогда закон распределения, по которому распределена случайная величина $t = \dfrac{X_0}{\sqrt{\chi^2(n)/n}}$, называется ***распределением Стьюдента.*** Параметром распределения Стьюдента является n – число степеней свободы случайной величины t.

Основные числовые характеристики распределения Стьюдента:

$$M(t) = 0,\ D(t) = \frac{n}{n-2} \text{ при } n > 2.$$

Значения случайной величины t определяются по табл. Е.1 приложения Е.

Пусть есть две случайные величины, имеющие распределение Пирсона ($X_1^2 + X_2^2 + ... + X_n^2 = \chi^2(n)$ и $X_1^2 + X_2^2 + ... + X_m^2 = \chi^2(m)$, где $X_1, X_2, ..., X_n$ попарно независимые случайные величины, распределенные по нормальному закону): $\chi^2 \sim \chi^2(n)$ и $\chi^2 \sim \chi^2(m)$ с разным числом степеней свободы. Тогда распределение случайной величины $F = \dfrac{\chi^2(n)/n}{\chi^2(m)/m}$ называется ***распределением Фишера – Снедекора*** с двумя параметрами n и m.

Основные числовые характеристики распределения Фишера:

$$M(F) = \frac{m}{m-2} \text{ при } m > 2,\ D(F) = \frac{2m^2(n+m-2)}{n \cdot (m-2)^2(m-4)} \text{ при } m > 4.$$

Значения случайной величины F определяются по таблицам приложения Ж и З. При $n > 30$ и $m > 30$ распределение случайной величины $F(n; m)$ приближается к нормальному.

Вопросы для самодиагностики

1. Описать равномерный закон распределения вероятностей непрерывной случайной величины и его числовые характеристики.

2. Описать показательный закон распределения вероятностей непрерывной случайной величины и его числовые характеристики.

3. Описать нормальный закон распределения вероятностей непрерывной случайной величины.

4. Как вычисляется вероятность попадания случайной величины в заданный интервал?

5. В чем заключается правило трех сигм?

Упражнения

30.1. Найти дисперсию и среднее квадратическое отклонение случайной величины X, распределенной равномерно в интервале (2, 8).

30.2. Найти математическое ожидание и функцию распределения случайной величины X, которая имеет равномерное распределение на промежутке (2; 9). Построить график функции распределения.

30.3. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 4$. Написать формулу и построить график ее функции плотности вероятности.

30.4. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 5$. Найти математическое ожидание и дисперсию случайной величины X.

30.5. Размер мужских рубашек является случайной величиной с нормальным законом распределения, математическим ожиданием 39 и дисперсией 9. Какой процент от общего объема заказа следует предусмотреть магазину для рубашек 40-го размера воротничка при условии, что этот размер находится в интервале (39,5; 40,5)?

30.6. Найти формулу плотности вероятности нормально распределенной случайной величины X, если математическое ожидание равно 3, а дисперсия равна 16.

30.7. Случайная величина распределена по нормальному закону $a = 10$; $\sigma = 5$. Сравнить вероятности того, что случайная величина принимает значения, которые принадлежат интервалам (7; 12) и (1;6).

30.8. Дано математическое ожидание $a = 13$ и среднее квадратическое отклонение $\sigma = 4$ нормально распределенной случайной величины X. Найти вероятность того, что $|x - a| < 8$.

31. Случайные векторы и законы их распределений: совместные, маргинальные и условные. Условные и маргинальные числовые характеристики

31.1. Случайный вектор и совместный закон распределения вероятностей его компонент. Функция распределения компонент двумерного вектора

При моделировании ситуации, когда объект характеризуется несколькими случайными параметрами, возникает необходимость ввести многомерные случайные величины (*случайные векторы*). Например, погода в данной местности может быть охарактеризована системой случайных величин: X_1 – давление, X_2 – температура, X_3 – влажность, X_4 – скорость ветра и т. д.

Пусть $X_1, X_2, \ldots, X_n$ – случайные величины, определенные на множестве элементарных событий Ω. Под ***n*-мерной случайной величиной**, или ***случайным вектором***, понимают упорядоченный набор n случайных величин: $\overline{X} = (X_1, X_2, \ldots, X_n)$.

На многомерные случайные величины распространяются основные определения, которые относятся к одномерным случайным величинам.

Закон распределения многомерной случайной величины $\overline{X}$ можно задать с помощью функции распределения $F_{\overline{X}}(x_1, x_2, \ldots, x_n)$.

Функция распределения $F_{\overline{X}}(x_1, x_2, \ldots, x_n)$ n-мерной случайной величины $\overline{X} = (X_1, X_2, \ldots, X_n)$ определяется формулой:

$$F_{\overline{X}}(x_1, x_2, \ldots, x_n) = P\{X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\}. \qquad (31.1)$$

При этом $F_{\overline{X}}(x_1, x_2, \ldots, x_n)$ – неубывающая функция каждого аргумента. Для *независимых случайных величин* $X_1, X_2, \ldots, X_n$ функция распределения n-мерной случайной величины $\overline{X} = (X_1, X_2, \ldots, X_n)$ равна произведению функций распределения случайных величин $X_1, X_2, \ldots, X_n$: $F_{\overline{X}}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdot \ldots \cdot F_{X_n}(x_n)$.

Рассмотрим двумерную случайную величину (вектор) $Z = (X,Y)$. Каждую из величин X и Y назовем компонентами или составляющими случайного вектора.

Функцией распределения двумерной случайной величины (X,Y) (дискретной или непрерывной) называют функцию $F(x,y)$, которая для каждой пары чисел x, y определяет вероятность того, что X примет значение меньшее, чем x, при этом Y меньше, чем y:

$$F(x,y) = P(X < x, Y < y). \tag{31.2}$$

Геометрически это равенство можно объяснить так:

$F(x,y)$ является вероятностью того, что случайная точка (X,Y) попадет в бесконечный квадрант с вершиной (x,y), который находится левее и ниже этой вершины (рис. 31.1).

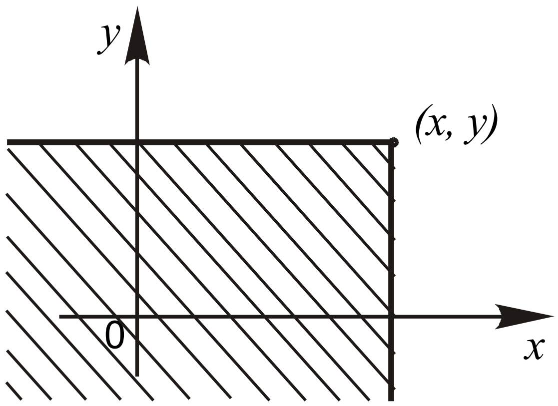


Рис. 31.1. **Геометрический смысл функции распределения $F(x,y)$**

Пример 31.1. Функция распределения $F(x,y)$ имеет вид $F(x,y) = \sin x \sin y$. Найти вероятность того, что компонента X примет значение меньше $\pi/6$, а компонента Y одновременно – меньше $\pi/3$.

Решение.

По определению функции распределения $F(x,y)$ имеем:

$F(x,y) = P(X < x, Y < y)$, то есть

$$P\left(X < \frac{\pi}{6}, Y < \frac{\pi}{3}\right) = F\left(\frac{\pi}{6}, \frac{\pi}{3}\right) = \sin\frac{\pi}{6} \cdot \sin\frac{\pi}{3} = \frac{1}{2} \cdot \frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{4} \approx 0{,}433.$$

Свойства функции распределения $F(x,y)$ аналогичны свойствам функции распределения одномерной случайной величины:

1. $0 \le F(x,y) \le 1$.

2. $F(x,y)$ – неубывающая функция по каждой из компонент:

$F(x_2,y) \ge F(x_1,y)$, если $x_2 > x_1$;

$F(x,y_2) \ge F(x,y_1)$, если $y_2 > y_1$.

3. Справедливы следующие предельные соотношения:

$$F(-\infty,y)=0,\quad F(x,-\infty)=0,\ F(-\infty,-\infty)=0,\ F(\infty,\infty)=1.$$

Доказательство.

События, условием осуществления которых является реализация соотношений $x<-\infty$ и $y<-\infty$, невозможны. Поэтому их вероятности равны нулю. Соответственно, функция распределения, которая содержит даже одно из подобных условий, равна нулю.

Событие, условием осуществления которого является реализация соотношений $x<+\infty$ и $y<+\infty$, достоверное. Поэтому его вероятность равна единице. Соответственно, функция распределения при таких условиях равна единице.

4. $F(x,\infty)=F_1(x)$, $F(\infty,y)=F_2(y)$.

Доказательство.

Поскольку событие $y=+\infty$ является достоверным, то значение функции распределения $F(x;\infty)$ определяется вероятностью события $X<x$, то есть эта функция является функцией распределения компоненты X двумерной случайной величины $(X;Y)$. Аналогично при $x=+\infty$.

5. Если компоненты X и Y независимы, то $F(x,y)=F_1(x)\cdot F_2(y)$.

31.2. Дискретные случайные векторы

Многомерная случайная величина $\overline{X}=(X_1,X_2,\ldots,X_n)$ является ***дискретной***, если каждая из случайных величин $X_1,X_2,\ldots,X_n$ является дискретной и существует конечное или бесконечное множество

n-мерных векторов $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ таких, что:

$$\sum_{i=1}^{\infty} P(\bar{X} = \bar{x}_i) = 1. \quad (31.3)$$

Закон распределения дискретной многомерной случайной величины полностью определяется набором векторов $\bar{x}_1, \bar{x}_2, \ldots$ и соответствующих им вероятностей $p_1 = P(\bar{X} = \bar{x}_1)$, $p_2 = P(\bar{X} = \bar{x}_2)$..., таких, что $p_1 + p_2 + \ldots = 1$.

Если $\bar{X} = (X_1, X_2, \ldots, X_n)$ – дискретная случайная величина, то для независимых дискретных случайных величин $X_1, X_2, \ldots, X_n$ имеем:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) P(X_2 = x_2) \ldots P(X_n = x_n).$$

Закон распределения дискретной двумерной случайной величины задается набором возможных значений этой случайной величины (x_i, y_j) и соответствующих им вероятностей $p(x_i, y_j)$ $(i = \overline{1, m};\ j = \overline{1, n})$, причем $\sum_i \sum_j p(x_i, y_j) = 1$. То есть закон распределения дискретной двумерной случайной величины задают таблицей, в которой находятся возможные значения (x_i, y_j) компонент (X, Y) и соответствующие им значения вероятностей $p(x_i, y_j)$ (табл. 31.1).

Таблица 31.1

Закон распределения двумерной случайной величины

Значения x_i	Значения y_j				$p_i = \sum_j p_{ij}$
	y_1	y_2	…	y_n	
x_1	p_{11}	p_{12}	…	p_{1n}	p_{x_1}
x_2	p_{21}	p_{22}	…	p_{2n}	p_{x_2}
…	…	…	…	…	…
x_m	p_{m1}	p_{m2}	…	p_{mn}	p_{x_m}
$p_j = \sum_i p_{ij}$	p_{y_1}	p_{y_2}	…	p_{y_n}	$\sum_i \sum_j p_{ij} = 1$

В каждой клетке, которая находится на пересечении строки x_i и столбца y_j, указана вероятность $p(x_i, y_j)$ того, что двумерная случайная величина примет значение (x_i, y_j), $p_{x_i} = \sum_i p_{ij}$, $p_{y_j} = \sum_j p_{ij}$.

Так, $p_{x_1} = p(x_1, y_1) + p(x_1, y_2) + ... + p(x_1, y_n)$,

$$p_{y_1} = p(x_1, y_1) + p(x_2, y_1) + ... + p(x_m, y_1).$$

Из закона распределения двумерной дискретной случайной величины можно найти законы распределения каждой из компонент (так называемые ***маргинальные законы распределения***).

То есть

x_i	x_1	x_2	$\ldots$	x_m
p_i	p_{x_1}	p_{x_2}	$\ldots$	p_{x_m}

,

y_j	y_1	y_2	$\cdots$	y_n
p_j	p_{y_1}	p_{y_2}	$\cdots$	p_{y_n}

.

Пример 31.2. Качество продукции характеризуется двумя случайными параметрами X и Y. Закон распределения двумерной случайной величины $Z = (X, Y)$ задан в табл. 31.2.

Таблица 31.2

Закон распределения двумерной случайной величины $Z = (X, Y)$

Значения x_i	Значения y_j			$p_i = \sum_j p_{ij}$
	1	2	3	
0	0,2	0	0	0,2
0,1	0,1	0,15	0	0,25
0,2	0,05	0,15	0,1	0,3
0,3	0,05	0,1	0,1	0,25
$p_j = \sum_i p_{ij}$	0,4	0,4	0,2	$\sum_i \sum_j p_{ij} = 1$

Первый столбец соответствует возможным значениям компоненты X $(x_i = 0;\ 0{,}1;\ 0{,}2;\ 0{,}3)$; первая строка – возможным значениям компоненты Y $(y_j = 1; 2; 3)$.

Найти законы распределения компонент дискретной двумерной случайной величины.

Решение.

x_i	0	0,1	0,2	0,3
p_i	0,2	0,25	0,3	0,25

, $\sum_i p_i = 1.$

y_j	1	2	3
p_j	0,4	0,4	0,2

, $\sum_j p_j = 1.$

31.3. Плотность совместного распределения и ее свойства

Многомерную случайную величину $\overline{X}$ называют ***абсолютно непрерывной***, если существует такая неотрицательная функция $f_{\overline{X}}(x_1, x_2, \ldots, x_n)$, что для любых $\overline{x} = (x_1, x_2, \ldots, x_n)$ функцию распределения $F_{\overline{X}}(\overline{x})$ можно представить в виде n-мерного интеграла:

$$F_{\overline{X}}(\overline{x}) = \int_{-\infty}^{x_1} dt_1 \int_{-\infty}^{x_2} dt_2 \ldots \int_{-\infty}^{x_n} f_{\overline{X}}(t_1, t_2, \ldots, t_n) dt_n .$$

Функция $f_{\overline{X}}(t_1, t_2, \ldots, t_n)$ называется ***плотностью совместного распределения*** многомерной непрерывной случайной величины $\overline{X}$. Как и функция распределения плотность распределения определяет *закон распределения n-мерной непрерывной случайной величины*.

Плотность распределения непрерывной многомерной случайной величины вычисляется по формуле:

$$f_{\overline{X}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1, \ldots, \partial x_n}. \qquad (31.4)$$

Если $\overline{X} = (X_1, X_2, \ldots, X_n)$ – непрерывная случайная величина, то для независимых непрерывных случайных величин $X_1, X_2, \ldots, X_n$ имеем:

$$f_{\overline{X}}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n).$$

Другие свойства многомерных непрерывных случайных величин рассмотрим на примере системы двумерных непрерывных случайных величин.

Плотность распределения двумерной непрерывной случайной величины определяется как вторая смешанная производная функции распределения $F(x, y)$:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}. \tag{31.5}$$

Геометрически эту функцию можно представить как поверхность, которую называют ***поверхностью распределения***.

Свойства плотности распределения:

1. $f(x,y) \geq 0$.

2. $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)\,dxdy = 1$.

3. $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)\,dxdy$. (31.6)

4. Если $f_1(x)$, $f_2(y)$ – плотности распределения каждой из компонент, то:

$$f_1(x) = \int\limits_{-\infty}^{\infty} f(x,y)\,dy, \quad f_2(y) = \int\limits_{-\infty}^{\infty} f(x,y)\,dx,$$

$$\int\limits_{-\infty}^{\infty} f_1(x)\,dx = 1, \quad \int\limits_{-\infty}^{\infty} f_2(y)\,dy = 1. \tag{31.7}$$

5. Если компоненты X и Y независимы, то $f(x,y) = f_1(x) \cdot f_2(y)$.

Пример 31.3. Система случайных величин (X,Y) распределена по закону $f(x,y) = \dfrac{C}{1 + x^2 + x^2y^2 + y^2}$.

Найти коэффициент C.

Выяснить, зависимы или нет величины X и Y.

Решение.

Согласно свойству 2 имеем: $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\frac{C}{1+x^2+x^2y^2+y^2}dxdy=1.$

Так как $1+x^2+x^2y^2+y^2=(1+x^2)(1+y^2)$, то получаем:

$$C\int\limits_{-\infty}^{\infty}\frac{dx}{1+x^2}\cdot\int\limits_{-\infty}^{\infty}\frac{dy}{1+y^2}=C\cdot\pi^2\Rightarrow C=\frac{1}{\pi^2}.$$

Каждый из приведенных выше интегралов вычислялся следующим образом:

$$\int\limits_{-\infty}^{\infty}\frac{dx}{1+x^2}=arctgx\Big|_{-\infty}^{\infty}=\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)=\pi.$$

Аналогично $\int\limits_{-\infty}^{\infty}\frac{dy}{1+y^2}=\pi.$

Следовательно, $f(x,y)=\frac{1}{\pi(1+x^2)}\cdot\frac{1}{\pi(1+y^2)}=f_1(x)\cdot f_2(y).$

То есть величины X и Y является независимыми.

Пример 31.4. Компоненты X и Y двумерной случайной величины независимы и распределены по показательному закону:

$$f_1(x)=\begin{cases}0, & x<0,\\ \lambda e^{-\lambda x}, & x\ge 0;\end{cases}\qquad f_2(y)=\begin{cases}0, & y<0,\\ \mu e^{-\mu x}, & y\ge 0.\end{cases}$$

Записать плотность распределения $f(x,y)$ системы (X,Y) и функцию распределения $F(x,y)$.

Решение.

Согласно свойству 5 можно записать:

$$f(x,y)=\begin{cases}0, & x<0, y<0,\\ \lambda\mu e^{-(\lambda x+\mu y)}, & x\ge 0,\ y\ge 0.\end{cases}$$

Для компонент X и Y, распределенных по показательному закону, имеем:

$$F_1(x)=\begin{cases}0, & x<0,\\ 1-e^{-\lambda x}, & x\ge 0;\end{cases} \quad F_2(y)=\begin{cases}0, & y<0,\\ 1-e^{-\mu x}, & y\ge 0.\end{cases}$$

Компоненты независимы, поэтому $F_1(x)\cdot F_2(y)=F(x,y)$:

$$F(x,y)=\begin{cases}0, & x<0 \text{ или } y<0,\\ (1-e^{-\lambda x})(1-e^{-\mu y}), & x\ge 0 \text{ и } y\ge 0.\end{cases}$$

Вероятность того, что случайная точка (X,Y) попадет в область D при известной плотности распределения $f(x,y)$, определяется по формуле:

$$P((X,Y)\in D)=\iint\limits_D f(x,y)\,dxdy. \qquad (31.8)$$

В частности, если область $D=(x_1<X<x_2,\ y_1<Y<y_2)$ – прямоугольник, то вероятность $P(X,Y)$ можно найти по формулам:

$$P(x_1<X<x_2,\ y_1<Y<y_2)=\int\limits_{x_1}^{x_2}\int\limits_{y_1}^{y_2} f(x,y)\,dxdy, \qquad (31.9)$$

или

$$P(x_1<X<x_2,\ y_1<Y<y_2)=(F(x_2,y_2)-F(x_1,y_2))-(F(x_2,y_1)-F(x_1,y_1)).$$

Вероятностный смысл плотности распределения двумерной случайной величины заключается в том, что функцию $f(x,y)$ можно определить как предел отношения вероятности попадания случайной точки (X,Y) в прямоугольник со сторонами Δx и Δy к площади этого прямоугольника, когда обе стороны прямоугольника приближаются к нулю (рис. 31.2).

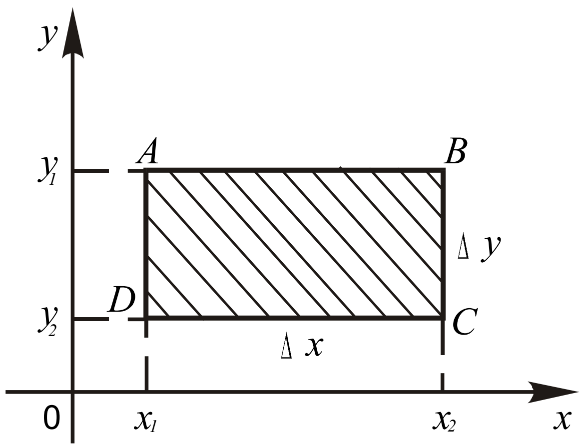


Рис. 31.2. **Иллюстрация к вероятностному смыслу функции $f(x,y)$**

Пример 31.5. Найти вероятность попадания случайной точки (X,Y) в прямоугольник, который образуют прямые: $x=\pi/6$, $x=\pi/2$, $y=\pi/4$, $y=\pi/3$, если известна функция распределения $F(x,y)=sinx\,siny$.

Решение.

Согласно формулам имеем:

$$P(\pi/6<X<\pi/2,\ \pi/4<Y<\pi/3)=(F(\pi/2,\pi/3)-F(\pi/6,\pi/3))-$$

$$-(F(\pi/2,\pi/4)-F(\pi/6,\pi/4))=\left(sin\frac{\pi}{2}\cdot sin\frac{\pi}{3}-sin\frac{\pi}{6}\cdot sin\frac{\pi}{3}\right)-$$

$$-\left(sin\frac{\pi}{2}\cdot sin\frac{\pi}{4}-sin\frac{\pi}{6}\cdot sin\frac{\pi}{4}\right)=\left(\frac{\sqrt{3}}{2}-\frac{\sqrt{3}}{4}\right)-\left(\frac{\sqrt{2}}{2}-\frac{\sqrt{2}}{4}\right)=\frac{1}{4}(\sqrt{3}-\sqrt{2})\approx 0{,}08,$$

или

$$f(x,y)=cosx\,cosy.$$

Тогда:

$$P(\pi/6<X<\pi/2,\ \pi/4<Y<\pi/3)=\int_{\pi/6}^{\pi/2}\int_{\pi/4}^{\pi/3}cosx\,cosy\,dxdy\approx 0{,}08.$$

31.4. Условные законы распределения случайного вектора

Рассмотрим дискретную двумерную случайную величину $Z=(X,Y)$. Возможные значения ее компонент: $x_1,x_2,\dots,x_m$ и $y_1,y_2,\ \dots,y_n$. Пусть в результате испытания случайная величина Y принимает значение y_1 ($Y=y_1$). При этом случайная величина X может принимать одно из возможных значений: $x_1,x_2,\ \dots,x_m$.

Обозначим условную вероятность того, что $X=x_i$, когда $Y=y_1$, через $p(x_i/y_1)$ $(i=\overline{1,m})$. В общем случае условные вероятности компоненты X при условии, что $Y=y_j$, обозначим через $p(x_i/y_j)$ $(i=\overline{1,m};$ $j=\overline{1,n})$, а условные вероятности компоненты Y при условии, что компонента $X=x_i$ – через $p(y_j/x_i)$.

Условной функцией $F(x, Y = y_j)$ распределения случайной величины X при условии, что случайная величина Y принимает значение y_j, называется условная вероятность $p(X < x / Y = y_j)$. Условная функция распределения обладает всеми свойствами, которые присущи обычной функции распределения.

Условным распределением компоненты $\boldsymbol{X}$ при $\boldsymbol{Y = y_j}$ называется совокупность условных вероятностей $p(x_i / y_j)$ в предположении, что случайная величина $Y = y_j$. Аналогично определяется условное распределение компоненты Y при $X = x_i$.

По закону распределения двумерной дискретной величины $Z = (X, Y)$ можно составить условные законы распределения компонент X и Y:

$$\text{для } X: \quad p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)},$$

$$\text{для } Y: \quad p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)}. \tag{31.10}$$

Отметим, что сумма вероятностей условного распределения для каждой из компонент равна единице.

Пример 31.6. По данным табл. 31.2 найти условные законы распределения компонент X, Y.

Решение.

Найдем условный закон распределения компоненты X при условии, что $Y = y_1 = 1$.

Для этого нужно найти: $p(x_1 / y_1)$, $p(x_2 / y_1)$, $p(x_3 / y_1)$, $P(x_4 / y_1)$.

По формулам (31.10) имеем:

$$p(x_1 / y_1) = \frac{p(x_1, y_1)}{p(y_1)} = \frac{0{,}2}{0{,}4} = 0{,}5; \quad p(x_2 / y_1) = \frac{p(x_2, y_1)}{p(y_1)} = \frac{0{,}1}{0{,}4} = 0{,}25;$$

$$p(x_3 / y_1) = \frac{p(x_3, y_1)}{p(y_1)} = \frac{0{,}05}{0{,}4} = 0{,}125; \quad p(x_4 / y_1) = \frac{p(x_4, y_1)}{p(y_1)} = \frac{0{,}05}{0{,}4} = 0{,}125,$$

то есть

$x_i / y=1$	0	0,1	0,2	0,3
$p_{x_i / y=1}$	0,5	0,25	0,125	0,125

Проверка: $\sum_{i=1}^{4} p(x_i / y_1) = 0{,}5 + 0{,}25 + 0{,}125 + 0{,}125 = 1.$

Аналогично найдем условный закон распределения компоненты Y при $X = x_3 = 0{,}2$:

$$p(y_1/x_3) = \frac{p(x_3, y_1)}{p(x_3)} = \frac{0{,}05}{0{,}3} = 0{,}17; \quad p(y_2/x_3) = \frac{p(x_3, y_2)}{p(x_3)} = \frac{0{,}15}{0{,}3} = 0{,}5;$$

$p(y_3/x_3) = \dfrac{p(x_3, y_3)}{p(x_3)} = \dfrac{0{,}1}{0{,}3} = 0{,}33$, то есть

$y_j / x = 0{,}2$	1	2	3
$p_{y_j / x=0{,}2}$	0,17	0,5	0,33

Проверка: $\sum_{j=1}^{3} p(y_j / x_3) = 0{,}17 + 0{,}5 + 0{,}33 = 1.$

В случае распределения непрерывной случайной величины $Z = (X, Y)$ возникают условные плотности распределения компоненты X, когда $Y = y_j$, и компоненты Y, когда $X = x_i$.

Условной плотностью $f_1(x/y)$ распределения компоненты X при значении $Y = y$ называется отношение плотности совместного распределения $f(x, y)$ системы (X, Y) к плотности распределения $f_2(y)$ компоненты Y:

$$f_1(x/y) = \frac{f(x, y)}{f_2(y)}. \qquad (31.11)$$

Аналогично условная плотность $f_2(y/x)$ компоненты Y при значении $X = x$ определяется по формуле:

$$f_2(y/x) = \frac{f(x,y)}{f_1(x)}. \tag{31.12}$$

Если известна плотность совместного распределения $f(x,y)$, то $f_1(x)$ и $f_2(y)$ можно найти по формулам (31.7).

Отметим такие свойства $f_1(x/y)$, $f_2(y/x)$:

$$f_1(x/y) \geq 0, \quad \int_{-\infty}^{\infty} f_1(x/y)\,dx = 1; \qquad f_2(y/x) \geq 0, \quad \int_{-\infty}^{\infty} f_2(y/x)\,dy = 1.$$

Пример 31.7. Случайная точка (X,Y) распределена с постоянной плотностью в квадрате R (рис. 31.3).

Записать выражение для условных плотностей распределения $f_1(x/y)$, $f_2(y/x)$.

Решение.

Площадь квадрата равна 2, поэтому

$$f(x,y) = \begin{cases} 1/2, & (x,y) \in R \\ 0, & (x,y) \notin R. \end{cases}$$

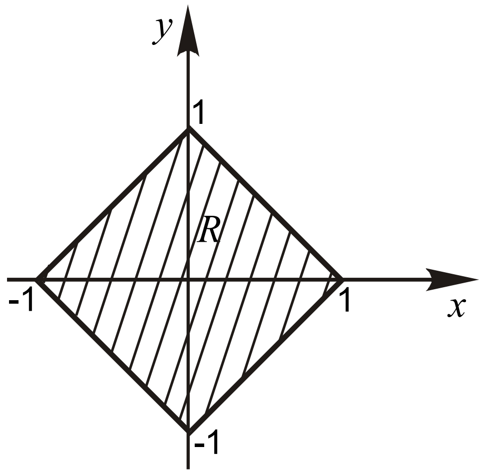


Рис. 31.3. **Чертеж к примеру**

По формулам (31.7) получим:

$$f_1(x)=\begin{cases}\frac{1}{2}\int\limits_{-(1-x)}^{1-x}dy=1-x, & x\in(0,1),\\ \frac{1}{2}\int\limits_{-(1+x)}^{1+x}dy=1+x, & x\in(-1,0),\\ 0, & x<-1,\ x>1\end{cases}\quad \text{или}\quad f_1(x)=\begin{cases}1-|x|, & |x|<1,\\ 0, & |x|>1.\end{cases}$$

Аналогично: $f_2(y)=\begin{cases}1-|y|, & |y|<1,\\ 0, & |y|>1.\end{cases}$

При $|y|<1$ имеем: $f_1(x/y)=\dfrac{f(x,y)}{f_2(y)}=\begin{cases}\dfrac{1}{2(1-|y|)}, & |x|<1-|y|,\\ 0, & |x|>1-|y|.\end{cases}$

Аналогично: $f_2(y/x)=\dfrac{f(x,y)}{f_1(x)}=\begin{cases}\dfrac{1}{2(1-|x|)}, & |y|<1-|x|,\\ 0, & |y|>1-|x|.\end{cases}$

Так как условная функция распределения обладает свойствами обычной функции распределения, то по ней можно определить ***условное математическое ожидание***.

Условное математическое ожидание дискретной случайной величины Y при $X=x_i$ вычисляют как произведение возможных значений компоненты Y на их условные вероятности:

$$M(Y/X=x_i)=\sum_{j=1}^{m}y_j\,p(y_j/x_i). \qquad (31.13)$$

Условное математическое ожидание $M(Y/x_i)$ является функцией от x – $M(Y/x_i)=f(x)$, которая называется ***функцией регрессии*** Y на X.

Аналогичные формулы для условного математического ожидания X при $Y=y_j$:

$$M(X/Y=y_j)=\sum_{i=1}^{m} x_i p(x_i/y_j), \quad M(X/y_j)=\varphi(y), \qquad (31.14)$$

где $\varphi(y)$ – функция регрессии X на Y.

Пример 31.8. Дискретная двумерная случайная величина $Z=(X,Y)$ задана в виде таблицы:

x_i \ y_j	0,1	0,2	0,3	p_i
1	0,1	0,3	0,2	0,6
2	0,06	0,18	0,16	0,4
p_j	0,16	0,48	0,36	1

Найти условное математическое ожидание компоненты X при $Y=y_1=0{,}1$.

Решение.

По формуле (31.14):

$$M(X/Y=y_1)=\sum_{i=1}^{2} x_i p(x_i/y_1).$$

Найдем условные вероятности:

$$p(x_1/y_1)=\frac{p(x_1,y_1)}{p(y_1)}=\frac{0{,}1}{0{,}16}=0{,}625,$$

$$p(x_2/y_1)=\frac{p(x_2,y_1)}{p(y_1)}=\frac{0{,}06}{0{,}16}=0{,}375.$$

Тогда получим:

$$M(X/Y=y_1)=1\cdot 0{,}625+2\cdot 0{,}375=1{,}375.$$

Условное математическое ожидание для непрерывных величин:

$$M(Y/X=x)=\int_{-\infty}^{\infty} y\, f_2(y/x)\,dy, \quad M(X/Y=y)=\int_{-\infty}^{\infty} x\, f_1(x/y)\,dx. \qquad (31.15)$$

31.5. Ковариация и коэффициент корреляции двумерного случайного вектора

Для двумерной случайной величины $Z=(X,Y)$ можно найти математическое ожидание и дисперсию каждой компоненты: $M(X)=m_x$, $M(Y)=m_y$, $D(X)=\sigma_x^2$, $D(Y)=\sigma_y^2$.

Однако эти характеристики недостаточно полно характеризуют величину Z, так как не воспроизводят степень зависимости между компонентами. Эту роль выполняют корреляционный момент (ковариация) μ_{xy} и коэффициент корреляции r_{xy}.

Корреляционным моментом (ковариацией) μ_{xy} двух случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий:

$$\mu_{xy}=M\left[(X-m_x)(Y-m_y)\right]. \quad (31.16)$$

Для вычисления корреляционного момента дискретных величин используют формулу:

$$\mu_{xy}=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_i-m_x)(y_j-m_j)p(x_i,y_j), \quad (31.17)$$

для непрерывных величин – формулу:

$$\mu_{xy}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(y-m_j)f(x,y)dxdy. \quad (31.18)$$

Ковариация обладает следующими *свойствами*:

1. $\mu_{xy}=\mu_{yx}$ – свойство симметрии.

2. $\mu_{xy}=0$, если X и Y – независимые величины.

3. $|\mu_{xy}|\le\sigma_x\sigma_y$.

4. $\mu_{xx}=D(X)$.

5. $\mu_{xy}=\pm\sqrt{D(X)\cdot D(Y)}$, тогда и только тогда, когда X и Y – зависимые величины.

6. Корреляционный момент имеет размерность, которая равна произведению размерностей величин X и Y.

При вычислении корреляционного момента вместо определения (31.16) используют преобразованную формулу:

$$\mu_{xy} = M(XY) - M(X) \cdot M(Y). \qquad (31.19)$$

Пример 31.9. Закон распределения дискретной двумерной случайной величины задан таблицей:

$Y \backslash X$	1	2	$\sum_{j=1}^{n} p_{ij}$
2	0,4	0,3	0,7
4	0,2	0,1	0,3
$\sum_{i=1}^{m} p_{ij}$	0,6	0,4	1=1

Найти μ_{xy} двух случайных величин X и Y.

Решение.

Вычислим математические ожидания случайных величин X и Y по их распределениям. Так, для случайной величины X по первой и четвертой строкам таблицы определяем:

$$M(X) = \sum_{j=1}^{n} x_j p_j = 1 \cdot 0{,}6 + 2 \cdot 0{,}4 = 1{,}4.$$

Для случайной величины Y по первому и последнему столбцам таблицы вычисляем:

$$M(Y) = \sum_{i=1}^{m} y_i p_i = 2 \cdot 0{,}7 + 4 \cdot 0{,}3 = 2{,}6.$$

Теперь для того, чтобы воспользоваться формулой (31.19), определяем математическое ожидание произведения случайных величин X и Y как математическое ожидание дискретной случайной величины, которая является произведением случайных величин X и Y:

$$M(XY)=\sum_{j=1}^{n}\sum_{i=1}^{m}x_j y_i p_{ij}=1\cdot 2\cdot 0{,}4+2\cdot 2\cdot 0{,}3+4\cdot 1\cdot 0{,}2+4\cdot 2\cdot 0{,}1=3{,}6.$$

По формуле (31.19) вычисляем корреляционный момент:

$$\mu_{xy}=3{,}6-1{,}4\cdot 2{,}6=-0{,}04.$$

Так как корреляционный момент не равен нулю, то компоненты X и Y двумерной случайной величини нельзя считать независимыми.

Коэффициентом корреляции r_{xy} двух случайных величин X и Y называется отношение корреляционного момента к произведению среднеквадратичных отклонений этих величин:

$$r_{xy}=\frac{\mu_{xy}}{\sigma_x\sigma_y}. \qquad (31.20)$$

Коэффициент корреляции является безразмерной величиной. Он равен нулю для независимых случайных величин.

Свойства коэффициента корреляции:

1. $r_{xy}=r_{yx}$.

2. $r_{xx}=1,\quad r_{yy}=1$.

3. $|r_{xy}|\le 1$.

4. $r_{xy}=0$ – величины независимы.

5. $r_{xy}=\pm 1$ – между X и Y существует линейная корреляционная зависимость.

Вопросы для самодиагностики

1. Привести основные определения многомерных случайных величин. Что называется системой двух случайных величин?
2. Как получить условные законы распределения компонент системы двух случайных величин?
3. Как вычислить условное математическое ожидание системы двух случайных величин?

4. Охарактеризовать числовые характеристики системы двух случайных величин.

5. Что называется корреляционным моментом и каковы его свойства?

6. Описать свойства коэффициента корреляции.

Упражнения

31.1. Дано распределение двумерной случайной величины (X, Y):

$Y \backslash X$	1	2	3
1	0,1	0,15	0,12
2	0,2	0,22	0,21

Найти распределения X, Y и $X + Y$.

31.2. Найти корреляционный момент и коэффициент корреляции двух случайных величин X и Y, распределения которых заданы в примере 31.1.

31.3. Дано распределение двумерной случайной величины (X, Y):

$Y \backslash X$	1	3	5
2	0,2	0,15	0
4	0,1	0,2	0,1
6	0	0,1	0,2

Найти ковариацию μ_{xy} и коэффициент корреляции r_{xy}.

31.4. Заданы дифференциальные функции независимых случайных величин X и Y:

$f_1(x)=\dfrac{1}{4}$ на интервале $(0,4)$, $f_1(x)=0$ вне интервала $(0,4)$;

$f_2(y)=\dfrac{1}{4}$ на интервале $(0,4)$, $f_2(y)=0$ вне интервала $(0,4)$.

Найти интегральную и дифференциальную функции распределения случайной величины $Z=X+Y$.

32. Закон больших чисел, центральная предельная теорема

32.1. Неравенство Чебышева, теорема Чебышева. Предельная теорема о вероятностной асимптотической сходимости (теорема Бернулли)

Вводимые в теории вероятностей понятия случайного события и случайной величины характеризуются *неопределенностью* факта возникновения случайного явления и *неточностью* его измерения. Тем не менее объективные законы, которые выражаются этими случайностями, гарантируют устойчивость этих показателей, закладываемых в функции распределения, и параметры вероятностных законов.

Известно, что математические законы теории вероятностей получены в результате формализации реальных статистических закономерностей, которые присущи массовым случайным событиям. При этом во время исследования массовых однородных случайных событий проявляются определенные закономерности типа стабильности.

Говорят, что последовательность случайных величин $X_1, X_2, \ldots, X_n, \ldots$ ***сходится по вероятности*** при $n \to \infty$ к случайной величине X, если для любого $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(|X_n - X| \le \varepsilon\right) = 0.$$

При большом количестве испытаний относительная частота события $w(A)$ обладает свойством устойчивости и по вероятности сходится к $P(A)$; среднее арифметическое для случайной величины сходится по вероятности к ее математическому ожиданию.

Все эти явления объединяют под общим названием ***закона больших чисел***, который можно в целом сформулировать так: в случае большого числа экспериментов, которые осуществляются для изучения определенного случайного события или случайной величины, средний их результат практически перестает быть случайным и может быть предсказан с большой надежностью.

Закон больших чисел состоит из нескольких теорем, в каждой из которых при определенных условиях утверждается факт сходимости средних характеристик во время проведения большого количества испытаний к определенным неслучайным, постоянным величинам.

Первые исследования по установлению предельной сходимости средних величин к их математическим ожиданиям для простейших классов последовательностей были сделаны Я. Бернулли и С. Пуассоном.

Полное решение проблемы предельной асимптотической сходимости было дано Чебышевым П. Л. в форме теоремы обобщенного вида, доказательство которой основано на неравенстве Чебышева. Неравенство Чебышева по существу формирует *достаточное условие* сходимости средних значений случайных величин к средним значениям их математических ожиданий.

Теорема (*неравенство Чебышева*). Если случайная величина имеет ограниченную дисперсию, то при любом $\varepsilon > 0$ имеет место неравенство:

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) \ge 1 - \frac{D(X)}{\varepsilon^2}. \tag{32.1}$$

Доказательство.

События $\left|X - M(X)\right| \le \varepsilon$ и $\left|X - M(X)\right| > \varepsilon$ образуют полную группу событий, то есть:

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) + P\left(\left|X - M(X)\right| > \varepsilon\right) = 1,$$

$$D(X) = M\left(X - M(X)\right)^2 = \sum_{i=1}^{n} \left(x_i - M(X)\right)^2 p_i .$$

Если отбросить слагаемые, в которых $\left|x_i - M(X)\right| \le \varepsilon$, то

$$D(X) \ge \sum_{j=1}^{k} \left(x_j - M(X)\right)^2 p_j \ge \varepsilon^2 \sum_{j=1}^{k} p_j = \varepsilon^2 P\left(\left|X - M(X)\right| > \varepsilon\right).$$

То есть $P\left(\left|X - M(X)\right| > \varepsilon\right) \le \dfrac{D(X)}{\varepsilon^2}$ и

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) = 1 - P\left(\left|X - M(X)\right| > \varepsilon\right) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

Теорема доказана.

Это неравенство дает возможность оценить ошибку, которую допускают, когда математическое ожидание заменяют средним значением ограниченной выборки.

Пример 32.1.

Электрическая сеть имеет 18 000 ламп, вероятность включения каждой из которых – 0,9.

Оценить вероятность того, что количество включенных ламп отклоняется от своего математического ожидания на величину не меньше, чем 200.

Решение.

По условию $n = 18\,000,\ p = 0{,}9.$

Тогда: $M(X) = np = 18\,000 \cdot 0{,}9 = 16\,200,$

$D(X) = npq = 18\,000 \cdot 0{,}9 \cdot 0{,}1 = 1\,620.$

По неравенству Чебышева:

$$P(|X - 16\,200| \le 200) \ge 1 - \frac{1\,620}{200^2} = 1 - 0{,}045 = 0{,}955,$$

$$P(|X - 16\,200| \ge 200) \ge 1 - 0{,}955 = 0{,}045.$$

Теорему Чебышева можно рассматривать как *необходимое и достаточное условие* предельной сходимости средних значений независимых случайных величин к средним значениям их математических ожиданий.

Суть теоремы заключается в том, что среднее арифметическое достаточно большого количества независимых случайных величин теряет характер случайной величины.

Теорема Чебышева. Пусть $X_1, X_2, \ldots, X_n$ – попарно независимые случайные величины с ограниченными дисперсиями, то есть существует такое D, что $D(X_1) \le D, D(X_2) \le D$, ... , $D(X_n) \le D$.

Тогда для любого $\varepsilon > 0$ вероятность отклонения среднего арифметического этих величин от среднего арифметического их математических ожиданий на величину не больше, чем ε, сколь угодно близка к единице, когда количество случайных величин достаточно велико:

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|\le\varepsilon\right)\xrightarrow[n\to\infty]{}1, \tag{32.2}$$

или

$$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|\le\varepsilon\right)=1.$$

Доказательство.

Пусть $X=\dfrac{X_1+X_2+\ldots+X_n}{n}$, тогда:

$$M(X)=\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n},$$

$$D(X)=\frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2}\le\frac{1}{n^2}\cdot nD=\frac{D}{n}.$$

По неравенству Чебышева имеем:

$$P(|X-M(X)|\le\varepsilon)\ge1-\frac{D(X)}{\varepsilon^2}\ge1-\frac{D}{n\varepsilon^2}.$$

Если $n\to\infty$, то $\lim\limits_{n\to\infty}P(|X-M(X)|\le\varepsilon)\ge1.$

Поскольку вероятность $P\le1$, то $\lim\limits_{n\to\infty}P(|X-M(X)|\le\varepsilon)=1,$

то есть:

$$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|\le\varepsilon\right)=1.$$

Теорема доказана.

Таким образом, на вероятностные величины распространяется *принцип детерминированности:* отдельная случайная величина может совершать хаотические блуждания, но если число этих величин будет достаточно большим, то движение совокупности станет почти определенным и образует средние показатели поведения совокупности.

Если в качестве случайных величин $X_1, X_2, \ldots, X_n$ взять конкретные значения случайной величины X, тогда их среднее арифметическое имеет свойство устойчивости, то есть при достаточно большом количестве испытаний среднее выборочное $\overline{X}_{\text{выб.}}$ приближается к математическому ожиданию a.

Следовательно,

$$\lim_{n\to\infty} P\left(\left|\overline{X}_{\text{выб.}} - a\right| < \varepsilon\right) = 1,\ \overline{X}_{\text{выб.}} \to a. \qquad (32.3)$$

Равенство (32.3) является следствием теоремы Чебышева и называется *законом больших чисел в форме* ***Хинчина*****.**

Теорема Бернулли является еще одной формой закона больших чисел. Это самая простая и исторически первая из теорем этой группы. Она устанавливает связь между относительной частотой случайного события и его вероятностью.

Теорема Бернулли (*предельная теорема о вероятностной асимптотической сходимости*).

Если вероятность события в n испытаниях постоянна и равна p, то вероятность отклонения относительной частоты от этой вероятности на величину не превосходящую ε стремится к единице при достаточно большом n:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \to 1 \text{ при } n \to \infty \text{или } \lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 1. \qquad (32.4)$$

Доказательство.

Пусть X_i – случайная величина, которая принимает значение 1, если при испытании наступает определенное событие, и значение 0 – если событие не наступает,

то есть $X_i = \begin{cases} 1, \textit{событиепроизошло} \\ 0, \textit{событие не произошло} \end{cases}$

Тогда имеем ряд распределения:

x_i	1	0
p_i	p	q

, где $q = 1 - p$.

$M(X_i) = 1 \cdot p + 0 \cdot q = p$.

Поскольку случайные величины X_i имеют ограниченную дисперсию, то можно использовать теорему Чебышева:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}\right|\le\varepsilon\right)=1.$$

Дробь $\dfrac{X_1+X_2+...+X_n}{n}$ равна относительной частоте.

Действительно, каждая из величин $X_1, X_2, \ldots, X_n$ с появлением события принимает значение 1.

Отсюда сумма $X_1+X_2+...+X_n$ равна m – количеству появлений события в n испытаниях.

Тогда

$$M(X_1)+M(X_2)+...+M(X_n)=np.$$

Таким образом,

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=1.$$

Теорема доказана.

32.2. Центральная предельная теорема. Теорема и неравенство Ляпунова

Установленные выше связи между последовательностями случайных величин и их математическими ожиданиями в первую очередь относятся к *дискретным законам* распределения.

Если рассмотреть случайные величины с *непрерывными законами* распределения, то последовательности независимых случайных величин при определенных условиях также обнаруживают тенденцию к образованию предельных зависимостей.

В классическом варианте рассматривают последовательность $X_1, X_2, ..., X_n$ непрерывных независимых и одинаково распределенных случайных величин с ограниченными математическими ожиданиями $M(X_1), M(X_2), ..., M(X_n)$ и дисперсиями $D(X_1), D(X_2), ..., D(X_n)$.

Для этих величин составляют *нормированные переменные*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} M(X_i)}{\sqrt{\sum_{i=1}^{n} D(X_i)}}$$

с функцией распределения $F_n(x) = P(Z_n < x)$.

Очевидно, что переменные Z_n имеют нулевые математические ожидания и единичные дисперсии. Функцию распределения $F_n(x)$ сравнивают со стандартной функцией нормального распределения $\Phi(x)$ случайной величины X:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz,$$

имеющей нулевое математическое ожидание и единичную дисперсию.

Центральная предельная теорема. При некоторых условиях для любого X обеспечивается сходимость по вероятности функции $F_n(x)$ к функции $\Phi(x)$:

$$F_n(x) \to \Phi(x) \text{ при } n \to \infty.$$

Формулировка центральной предельной теоремы основана на неравенстве Ляпунова.

Предельная непрерывная зависимость имеет более сложный характер, чем последовательность дискретных величин, поэтому требуется совершить переход от схемы последовательности величин к схеме подпоследовательностей величин, именуемых *сериями*.

Указанные серии задаются следующим образом: $X_{n1}, X_{n2}, \ldots, X_{nn}$, где $X_{nk} = \frac{X_k - M(X_k)}{\sqrt{\sum_{i=1}^{n} D(X_i)}}$, $k = \overline{1,n}$.

Тогда случайные величины внутри каждой серии *независимы* и

$$Z_n = X_{n1} + X_{n2} + \ldots + X_{nn}.$$

Условие предельной сходимости по Ляпунову задается в форме

$$\frac{1}{\sqrt{\left(\sum_{i=1}^{n} D(X_i)\right)^3}} \sum_{k=1}^{n} M\left[\left|X_k - M(X_k)\right|^3\right] < \varepsilon$$

для любого $\varepsilon > 0$ и $n \to \infty$.

Это означает, что приведенный момент третьего порядка должен быть пренебрежительно малым при достаточно больших n.

Из данного условия вытекает **неравенство Ляпунова**:

$$\max_{k=\overline{1,n}} P(|X_{nk}| > \varepsilon) = \max_{k=\overline{1,n}} P\left(|X_k - M(X_k)| > \varepsilon\sqrt{\sum_{i=1}^{n} D(X_i)}\right) \leq$$

$$\leq \max_{k=\overline{1,n}} \frac{1}{\varepsilon^3 \sqrt{\left(\sum_{i=1}^{n} D(X_i)\right)^3}} M\left[\left|X_k - M(X_k)\right|^3\right] \to 0$$

при любом $\varepsilon > 0$ и $n \to \infty$.

Стремление к нулю величины $\max_{k=\overline{1,n}} P(|X_{nk}| > \varepsilon)$ означает асимптотическую пренебрегаемость случайными величинами, образующими серии.

Теорема Ляпунова. Пусть дана произвольная схема серий $X_{n1}, X_{n2}, \ldots, X_{nn}$ асимптотически пренебрегаемых и независимых внутри каждой серии величин.

Тогда, если предельное распределение $Z_n = X_{n1} + X_{n2} + \ldots + X_{nn}$ существует и не является вырожденным (ранг матрицы ковариаций равен n), то оно будет нормальным тогда и только тогда, когда

$$\max_{k=\overline{1,n}} P(|X_{nk}| > \varepsilon) \to 0$$

для любого $\varepsilon > 0$ и $n \to \infty$.

Если в качестве случайных величин рассматривать результаты отдельных испытаний, то среднее значение этих результатов в серии случайных величин имеет распределение, близкое к нормальному.

Смысл теоремы Ляпунова для ее использования на практике заключается в следующем: если случайная величина равна сумме очень большого числа взаимно независимых случайных величин, влияние каждой из которых на сумму мало, то она имеет распределение, близкое к нормальному (*асимптотически нормальное*). В частности среднее значение имеет распределение, близкое к нормальному.

Таким образом, центральная предельная теорема определяет роль, которую играет нормальное распределение.

Основные выводы из закона больших чисел: при достаточно большом количестве наблюдений можно считать, что случайная величина имеет распределение, близкое к нормальному (вывод теоремы Ляпунова), при этом в качестве математического ожидания случайной величины можно принять среднее выборочное (теорема Чебышева), а в качестве вероятности события – относительную частоту (теорема Бернулли).

Вопросы для самодиагностики

1. Какие теоремы теории вероятностей называются предельными? По какому принципу разделяются предельные теоремы?
2. Сформулировать закон больших чисел.
3. Привести неравенство Чебышева.
4. Сформулировать теорему Чебышева и теорему Бернулли как следствие теоремы Чебышева.
5. Сформулировать теорему Ляпунова (центральную предельную теорему).

Упражнения

32.1. Сумма всех вкладов некоторого банка составляет 20 млн грн, а вероятность того, что случайно выбранный вклад будет меньше 10 000 грн, равна 0,8. Какие выводы можно сделать относительно количества вкладчиков данного банка?

32.2. Используя неравенство Чебышева, оценить вероятность того, что $|X - M(X)| < 0{,}3$, если $D(X) = 0{,}0025$.

Оценить вероятность этого же события, если известно, что X – нормально распределенная случайная величина.

32.3. Случайная величина X задана законом распределения:

X	1	3	5	6
$P(X=x_i)$	0,1	0,3	0,4	0,2

Используя неравенство Чебышева, оценить вероятность того, что:

абсолютная разница между случайной величиной X и ее математическим ожиданием меньше 5;

абсолютная разница между значением случайной величины X и ее математическим ожиданием не меньше 2.

32.4. Случайная величина X задана законом распределения:

X	3	5
$P(X=x_i)$	0,4	0,6

Используя неравенство Чебышева, оценить вероятность того, что абсолютная разница между значением случайной величины X и ее математическим ожиданием меньше 2.

32.5. Вероятность того, что при единичном обращении банкомат срабатывает правильно, равна 0,95.

Оценить вероятность того, что:

при 2 500 обращениях относительная частота случаев правильной работы банкомата отклонится (по абсолютной величине) от ее вероятности меньше, чем на 0,02;

при 2 000 обращениях количество случаев правильной работы банкомата находится в пределах от 1 860 до 1 940.

32.6. Определить количество деталей, которое необходимо для того, чтобы с вероятностью не меньше 0,98 можно было бы ожидать, что абсолютная величина отклонения относительной частоты стандартных деталей от вероятности, которая равна $p = 0{,}95$, была меньше 0,05.

33. Основные понятия математической статистики: выборочные наблюдения и выборочные оценки

33.1. Выборочный метод и его основные положения

Первой задачей математической статистики является обоснование методов сбора и группировки статистических данных, которые получены в результате экспериментов или наблюдений. *Вторая задача* – это разработка методов анализа статистических данных: оценки неизвестных вероятностей события, а также функции и параметров распределения; оценка зависимости случайной величины от других случайных величин; проверка статистических гипотез о виде и величинах параметров неизвестного распределения.

Множество объектов, из которых извлекается выборка, в статистике называется совокупностью. Вся совокупность, которая изучается, называется ***генеральной совокупностью***. Генеральную совокупность можно изучать путем сплошного изучения всех объектов или путем наблюдения за частью объектов.

Часть объектов, которую выбирают из генеральной совокупности, называется ***выборкой***, или ***выборочной совокупностью***.

Общее количество объектов генеральной совокупности или выборочной совокупности называется их *объемом*. Объем генеральной совокупности обозначают N, а объем выборочной совокупности – n. Результаты исследований любого признака генеральной совокупности будут достовернее, если выборку создать случайно.

Пример 33.1. Пусть из 2 000 изделий отобрано для исследования 100 изделий. Тогда объем генеральной совокупности N = 2 000, а объем выборки n = 100.

Выборка называется ***случайной***, если из генеральной совокупности элементы берутся наугад и каждый из них может попасть в нее с одинаковой вероятностью. На практике случайную выборку можно получить так: все элементы нумеруют от 1 до N, после чего выбирается последовательность n случайных чисел ($1 \le n \le N$). Выборку также можно получить с использованием датчика случайных чисел.

Отбор объектов может быть повторным или бесповторным.

Повторным называют отбор, когда отобранный объект возвращают в генеральную совокупность (до отбора следующего объекта). *Бесповторным* называют отбор, когда отобранный объект не возвращают в генеральную совокупность. Отбор объектов происходит случайным образом.

Если случайная выборка достаточно полно характеризует генеральную совокупность, то выборка называется ***репрезентативной***.

По выборке вычисляют ее числовые характеристики: математическое ожидание $M(X)$ и дисперсию $D(X)$. Если одна из двух выборок имеет меньшую дисперсию, то она полнее отображает генеральную совокупность.

33.2. Выборочное распределение

Признак любой совокупности принимает ряд значений, количество которых может быть бесконечным или ограниченным. Обозначим признак совокупности X, а его возможные значения $x_1, x_2, \ldots, x_l$.

Пусть имеем выборку из генеральной совокупности $x_1, x_2, \ldots, x_l$. Например: 10, 5, 15, 7, 20. Если записать эту выборку в виде возрастающей или убывающей последовательностей, то получим ***дискретный вариационный ряд***. В нашем примере это 5, 7, 10, 15, 20.

Некоторые значения дискретной случайной величины могут повторяться, тогда вариационный ряд записывают в виде таблицы, где указывают ***варианты*** – возможные значения (x_i) случайной величины X и их частоты (m_i). ***Частотой*** называют число появления отдельных значений случайной величины (вариант).

x_i	x_1	x_2	…	x_l
m_i	m_1	m_2	…	m_l

$$\sum_{i=1}^{l} m_i = n.$$

Например, выборку 10, 5, 5, 7, 15, 15, 15, 20 можно представить дискретным вариационным рядом:

x_i	5	7	10	15	20
m_i	2	1	1	3	1

$$n = 2+1+1+3+1 = 8.$$

Вместо частот можно указывать относительные частоты.

Относительной частотой w_i называют отношение частоты появления признака к общему объему выборки:

$$w_i = \frac{m_i}{n}, \qquad \sum_{i=1}^{l} w_i = 1. \tag{33.1}$$

Тогда имеем:

x_i	x_1	x_2	.	x_l
w_i	w_1	w_2	.	w_l

Относительная частота характеризует, какую часть совокупности составляют члены с одинаковыми значениями признака.

При большом объеме выборки $(n \geq 30)$ пользоваться дискретным вариационным рядом неудобно. В этом случае, а также в случае, когда данные получены в результате измерения непрерывной случайной величины, по выборке составляют ***интервальный вариационный ряд***. Для его построения весь интервал распределения признака разделяют на k равных частей длиной h. Затем подсчитывают частоту m_i или относительную частоту w_i попаданий в каждый интервал.

Под значением случайной величины считают середину интервала, то есть непрерывную величину приводят к дискретной.

$(x_{i-1}; x_i)$	$(x_0; x_1)$	$(x_1; x_2)$	...	$(x_{i-1}; x_i)$	...	$(x_{k-1}; x_k)$
m_i	m_1	m_2	...	m_i	...	m_k
x'_i	x'_1	x'_2	...	x'_i	...	x'_k

,

где $x'_i = \frac{x_{i-1} + x_i}{2}$, $h = x_i - x_{i-1}$.

Если вариационный ряд задается в виде приведенных таблиц, то его называют статистическим рядом распределения.

То есть ***статистическим распределением выборки*** называют перечень вариант случайной величины X $(x_1, x_2, \dots, x_l)$ и соответствующих им частот m_i или относительных частот w_i $(i = \overline{1,l})$.

Для интервальных рядов статистическое распределение задается в виде последовательности интервалов $(x_{i-1};\ x_i)$ и соответствующих им частот (за частоту принимают сумму частот признака, которые вошли в этот интервал).

Числовые характеристики выборки

Пусть задан вариационный ряд:

x_i	x_1	x_2	…	x_i	…	x_k
m_i	m_1	m_2	…	m_i	…	m_k

Для характеристики вариационного ряда используют следующие величины:

1. *Выборочное среднее* $\bar{x}$: $\bar{x} = \dfrac{\sum_{i=1}^{k} x_i m_i}{n}$, $n = \sum_{i=1}^{k} m_i$. (33.2)

Выборочное среднее характеризует среднее значение признака X.

2. *Выборочная дисперсия* D:

$$D = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 m_i$$

или (33.3)

$$D = \frac{1}{n}\sum_{i=1}^{k} x_i^2 m_i - (\bar{x})^2 = \overline{x^2} - (\bar{x})^2.$$

Дисперсия в статистике характеризует меру рассеяния признака X.

3. *Выборочное среднеквадратическое отклонение* σ:

$$\sigma = \sqrt{D}. \quad (33.4)$$

Выборочное среднеквадратическое отклонение является не только мерой вариации, но и показателем однородности статистической совокупности.

Поскольку $M(\bar{x}) = \bar{x}$, то выборочное среднее $\bar{x}$ является несмещенной оценкой.

Выборочная дисперсия D является смещенной оценкой.

Действительно, если $X_1, X_2, \ldots, X_n$ – одинаково распределенные случайные величины, такие, что $D(X_1) = D(X_2) = \ldots = D(X_n) = \sigma^2$, то

$$D\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}, \quad \text{поэтому } \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}.$$

По данным выборки нужно оценить генеральную дисперсию $D_{\text{ген.}}$.

$$M(D) = \frac{n-1}{n} D_{\text{ген.}}. \qquad (33.5)$$

Достаточно легко исправить выборочную дисперсию:

$$S^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{l} (x_i - \bar{x})^2 m_i}{n-1}, \qquad (33.6)$$

где S^2 – исправленная выборочная дисперсия.

Исправленная дисперсия является оценкой несмещенной, то есть

$$M(S^2) = M\left(\frac{n}{n-1}\sigma^2\right) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\text{ген.}} = D_{\text{ген.}}.$$

4. *Коэффициент вариации*: $$\nu = \frac{\sigma}{\bar{x}} \cdot 100\%. \qquad (33.7)$$

Коэффициент вариации является относительной мерой изменения признака. Его используют для сравнения случайных событий, разных по природе.

5. *Начальные моменты s-го порядка*: $$\nu_s = \overline{X^s} = \frac{\sum_{i=1}^{k} x_i^s m_i}{n}. \qquad (33.8)$$

6. *Центральные моменты s-го порядка*:

$$\mu_k = \overline{(X - \overline{X})^s} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^s m_i}{n}. \qquad (33.9)$$

7. *Асимметрия*: $A_S = \frac{\mu_3}{\sigma^3}$. (33.10)

A_S характеризует отклонение случайной величины от своего центрального положения влево или вправо.

8. *Эксцесс*: $E_S = \frac{\mu_4}{\sigma^4} - 3$. (33.11)

E_S характеризует отклонение случайной величины от своего центрального положения вверх или вниз.

9. *Мода* M_o – это варианта с максимальной частотой, то есть:

$$M_o = x_i \left(\underset{1 \le i \le k}{m_i} = \max \right). \qquad (33.12)$$

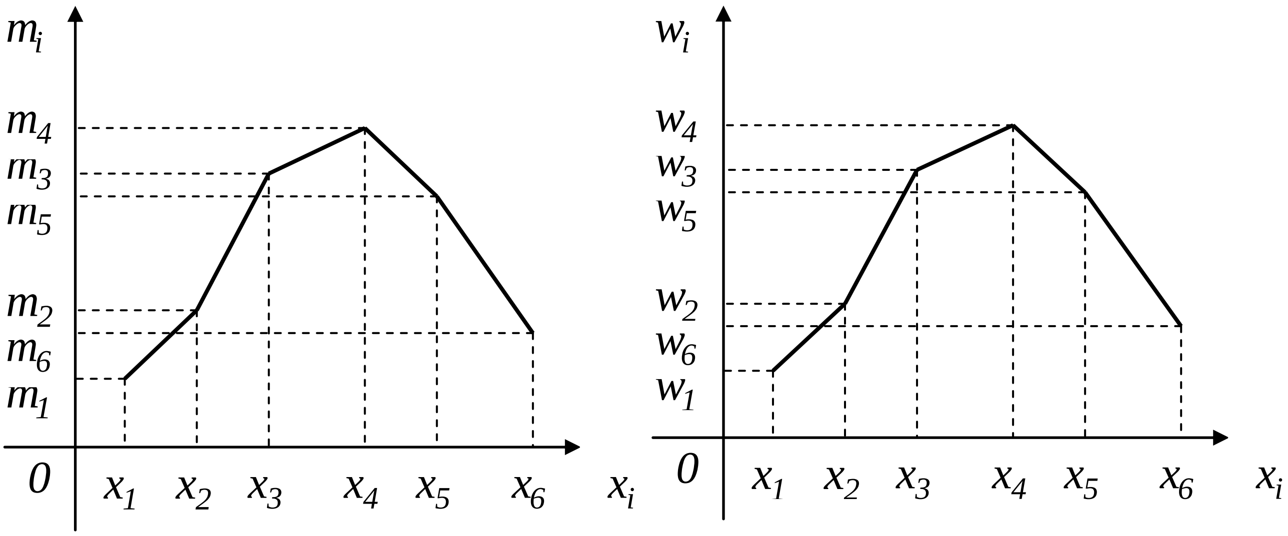
10. *Медиана* M_e – это варианта, которая разделяет вариационный ряд на две равные по количеству элементов части.

33.3. Эмпирическая функция распределения и гистограмма

Статистический ряд распределения случайной величины X, полученный по эмпирическим данным, называют также ***эмпирическим законом распределения***. Такой ряд можно изобразить графически. Для этого на оси абсцисс наносят варианты x_i, а на оси ординат – частоты m_i (рис. 33.1) (или относительные частоты w_i), то есть имеем точки $M_i(x_i, m_i)$ ($N_i(x_i, w_i)$). Соединив точки M_i (или N_i) соответствующих вариант и частот, получим ломаную линию, которую называют ***полигоном*** распределения.

Эмпирической функцией распределения выборки называется функция $F^*(x)$, которая для любого значения x определяет относительную частоту события, которое удовлетворяет условию $X < x$, то есть случайная величина примет значение меньше, чем x: $F^*(x) = \frac{m_x}{n}$,

где m_x – сумма частот вариант для значений аргумента, меньших, чем x; n – объем выборки.



а) **полигон частот** б) **полигон относительных частот**

Рис. 33.1. **Полигоны распределения**

Таким образом, эмпирическая функция распределения определяется путем последовательного сложения относительных частот всех вариант, меньших, чем x.

Графиком эмпирической функции распределения является ***кумулята***, или график накопленных относительных частот. Этот график – ступенчатая фигура, которая имеет точки разрыва при значениях абсцисс, равных числовым значениям, которые принимает случайная величина.

Для генеральной совокупности функцию распределения обозначим $F(x)$. Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция $F(x)$ определяет вероятность события $X < x$, а эмпирическая функция $F^*(x)$ – относительную частоту этого события.

Согласно теореме Бернулли можно сказать, что при больших n числа $F^*(x)$ и $F(x)$ мало отличаются друг от друга в том смысле, что:

$$\lim_{n\to\infty} P\left(\left|F(x) - F^*(x)\right| < \varepsilon\right) = 1 \quad (\varepsilon > 0).$$

Следовательно эмпирическая функция распределения выборки может быть использована для оценки теоретической функции распределения в генеральной совокупности.

Функция $F^*(x)$ имеет такие же свойства, что и функция $F(x)$:

1) значение эмпирической функции принадлежит отрезку $[0;1]$;

2) $F^*(x)$ – неубывающая функция: $F^*(x_2) \geq F^*(x_1)$, если $x_2 > x_1$;

3) если x_1 – наименьшая варианта, то $F^*(x) = 0$ при $x \leq x_1$, если x_k – наибольшая варианта, то $F^*(x) = 1$ при $x > x_k$.

Пример 33.2. Задана выборка рядом распределения:

x_i	1	6	11	16
m_i	20	10	40	30

Найти и построить эмпирическую функцию распределения $F^*(x)$.

Решение.

Найдем объем выборки: $\sum_{i=1}^{4} m_i = 20 + 10 + 40 + 30 = 100$.

Наименьшее значение величины X: $x_1 = 1$, поэтому $F^*(x) = 0$, если $X < 1$. Значение $X < 6$ наблюдалось 20 раз, следовательно $F^*(x) = \frac{20}{100} = 0{,}2$, если $1 \leq x < 6$. Значение $X < 11$, то есть 1 и 6 наблюдались $20 + 10 = 30$ раз, следовательно $F^*(x) = \frac{30}{100} = 0{,}3$, если $6 \leq x < 11$. Значение $X < 16$, то есть 1, 6 и 11 наблюдались $20 + 10 + 40 = 70$ раз, следовательно $F^*(x) = \frac{70}{100} = 0{,}7$, если $11 \leq x < 16$. Наибольшим значением величины X является $x_4 = 16$, следовательно $F^*(x) = 1$, если $X \geq 16$.

Таким образом, можно записать эмпирическую функцию распределения выборки в виде:

$$F^*(x) = \begin{cases} 0, & \text{если } x < 1; \\ 0{,}2, & \text{если } 1 \leq x < 6; \\ 0{,}3, & \text{если } 6 \leq x < 11; \\ 0{,}7, & \text{если } 11 \leq x < 16; \\ 1, & \text{если } x \geq 16. \end{cases}$$

График этой функции изображен на рис. 33.2.

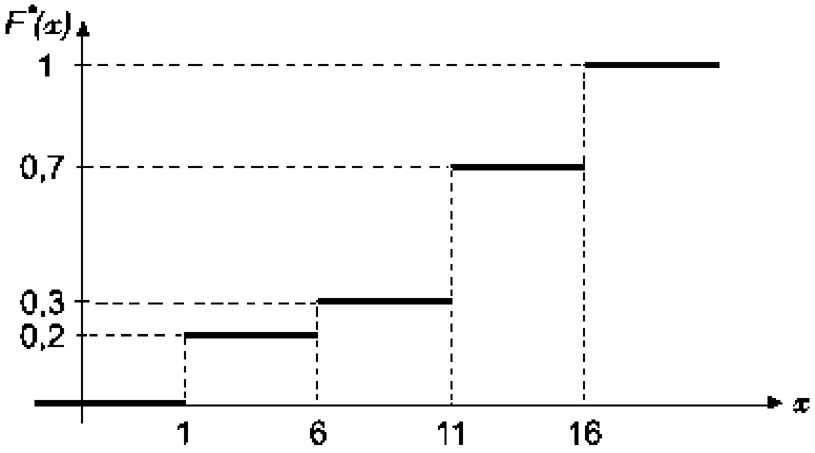


Рис. 33.2. **Эмпирическая функция распределения**

В случае интервальных вариационных рядов, полученных для непрерывной случайной величины, графическим изображением является ***гистограмма***. Различают гистограмму частот и гистограмму относительных частот.

Гистограммой частот называют ступенчатую фигуру, которая состоит из прямоугольников, площадь каждого из которых равна частоте попадания случайной величины в соответствующий интервал. Таким образом, основаниями прямоугольников являются интервалы, длиной h, а высоты равны плотности частоты $\frac{m_i}{h}$. Для построения гистограммы на оси абсцисс откладывают интервалы, а над ними проводят отрезки, которые параллельны оси абсцисс на расстоянии $\frac{m_i}{h}$. Площадь i-го прямоугольника равна $\frac{m_i}{h} \cdot h = m_i$ – сумме частот вариант i-го интервала. Таким образом, площадь гистограммы равна сумме всех частот, то есть объему выборки.

Гистограммой относительных частот называют ступенчатую фигуру, которая состоит из прямоугольников, основаниями которых являются интервалы длиной h, а высоты равны отношению $\frac{w_i}{h}$ (плотность относительной частоты).

Для построения гистограмм относительных частот на оси абсцисс откладывают интервалы, а над ними проводят отрезки, которые параллельны оси абсцисс на расстоянии $\frac{w_i}{h}$ от нее. Площадь гистограммы относительных частот равна сумме всех относительных частот, то есть единице. Пример гистограммы относительных частот приведен на рис. 33.3.

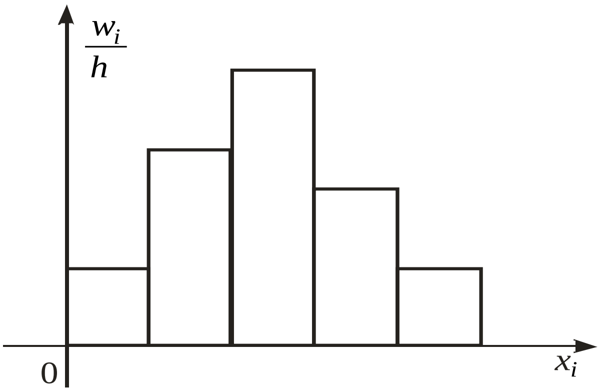


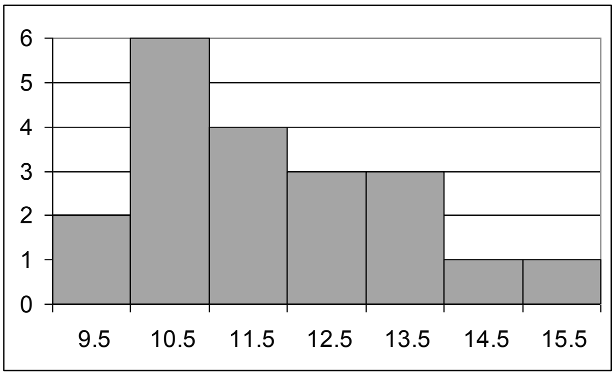
Рис. 33.3. **Гистограмма относительных частот** $\frac{w_i}{h}$

Можно строить также гистограммы частот m_i и относительных частот w_i.

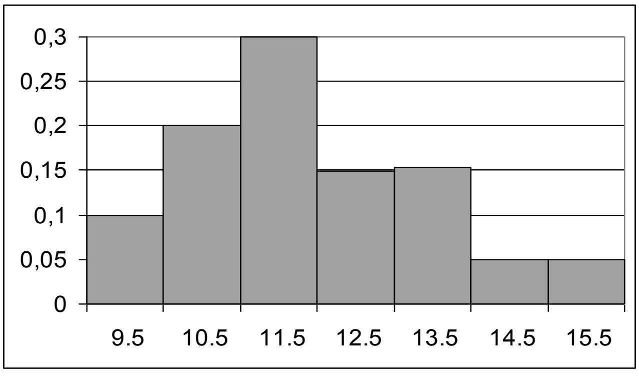
Пример 33.3. Для интервального вариационного ряда с объемом выборки $n = 20$ построить гистограмму частот и гистограмму относительных частот:

$x_{i-1} - x_i$	9 – 10	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16
m_i	2	6	4	3	3	1	1
w_i	0,1	0,3	0,2	0,15	0,15	0,05	0,05

Решение.



Гистограмма частот m_i



Гистограмма относительных частот w_i

Для построения гистограммы частот на оси абсцисс отложим промежутки и на них построим прямоугольники, высоты которых равны $\frac{m_i}{h}$ ($h=1$– длины интервалов). Площадь построенных прямоугольников равна объему выборки $n=20$. Для построения гистограммы относительных частот на оси абсцисс отложим промежутки и на них построим прямоугольники, высоты которых равны $\frac{w_i}{h}$. Общая площадь построенных прямоугольников равна единице.

33.4. Связь между характеристиками генеральной и выборочной совокупностей

Групповым средним называют среднее арифметическое значений признака (см. формулу (33.2)), которые принадлежат группе:

$$\bar{x}=\frac{\sum_{i=1}^{k} x_i m_i}{n}.$$

При наличии нескольких выборок в генеральной совокупности с групповыми средними $\bar{x}_1$, $\bar{x}_2$, ... , $\bar{x}_k$ и объемами n_1, n_2, ... , n_k введем термин общего среднего $\bar{x}_{\text{ген.}}$ для всей совокупности. Общее среднее равно среднему арифметическому групповых средних, взвешенных по объемам групп:

$$\bar{x}_{\text{ген.}}=\frac{\sum_{j=1}^{k} \bar{x}_j \cdot n_j}{N}, \qquad (33.13)$$

где j – номер группы; N – объем всей совокупности $\left(N=\sum_{j=1}^{k} n_j\right)$.

Для того чтобы охарактеризовать рассеяние значений признака X генеральной совокупности вокруг своего среднего значения, вводят генеральную дисперсию и генеральное среднеквадратическое отклонение по формулам:

$$S^2_{\text{ген.}}=\frac{\sum_{i=1}^{k} \left(x_i-\bar{x}_{\text{ген.}}\right)^2 \cdot N_i}{N}; \qquad \sigma_{\text{ген.}}=\sqrt{S^2_{\text{ген.}}}, \qquad (33.14)$$

где возможные значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $N_1, N_2, \ldots, N_k$ $\left(\sum_{i=1}^{k} N_i = N\right)$.

Пример 33.4. Найти общую среднюю заработную плату по всему предприятию, если величина часовой заработной платы рабочих по цехам предприятия задана таблицами:

I цех

x_i	m_i
1	4
2	3
4	1
5	2

II цех

x_i	m_i
1	6
2	2
3	8
4	4

III цех

x_i	m_i
1	5
2	8
3	3
5	4

,

где x_i – заработная плата (грн); m_i – количество рабочих: $n_1 = 10$, $n_2 = 20$, $n_3 = 20$, $N = 50$.

Решение.

Вначале найдем групповые средние по цехам:

$$\bar{x}_1 = \frac{1 \cdot 4 + 2 \cdot 3 + 4 \cdot 1 + 5 \cdot 2}{10} = 2{,}4;$$

$$\bar{x}_2 = \frac{1 \cdot 6 + 2 \cdot 2 + 3 \cdot 8 + 4 \cdot 4}{20} = 2{,}5; \quad \bar{x}_3 = \frac{1 \cdot 5 + 2 \cdot 8 + 3 \cdot 3 + 5 \cdot 4}{20} = 2{,}5.$$

Таким образом, средняя заработная плата за час по предприятию равна:

$$\bar{x}_{\text{ген.}} = \frac{2{,}4 \cdot 10 + 2{,}5 \cdot 20 + 2{,}5 \cdot 20}{50} = 2{,}48.$$

33.5. Статистические оценки и их свойства

Допустим, что экспериментальным путем найдена характеристика θ^*, которая служит оценкой неизвестного параметра θ закона распределения случайной величины X.

Приближенное значение параметра генеральной совокупности, вычисленное по выборке, называется ***оценкой параметра***.

Как правило, основными оценками параметров генеральной совокупности являются среднее выборочное $\overline{x}$, которое является аналогом математического ожидания $M(X)$, и выборочная дисперсия σ^2 – аналог дисперсии $D(X)$.

Пусть имеем параметр θ, а θ^* – его выборочная оценка.

Для того чтобы оценка θ^* достаточно полно характеризовала параметр генеральной совокупности, необходимо, чтобы она имела следующие свойства: несмещенность, состоятельность, эффективность.

Оценка θ^* параметра θ называется ***несмещенной***, если ее математическое ожидание равно оцениваемому параметру, то есть:

$$M(\theta^*) = \theta \text{ (например } \overline{x}_{\text{ген.}} \approx M(x)\text{).}$$

Оценка называется ***состоятельной***, если она при $n \to \infty$ сходится по вероятности к оцениваемому параметру, то есть:

$$\lim_{n\to\infty} P(\theta^* \to \theta) = 1.$$

Оценка называется ***эффективной***, если при заданном объеме наблюдений она имеет наименьшую дисперсию.

Вопросы для самодиагностики

1. Каковы основные задачи и цели математической статистики?
2. Как определяются генеральная и выборочная совокупности?
3. Каковы требования, которым должна отвечать выборочная совокупность?
4. Перечислить основные методы формирования выборочной совокупности.
5. Описать основные числовые характеристики закона распределения. Привести их свойства и геометрическое толкование.
6. Перечислить дополнительные числовые характеристики.
7. Что называется эмпирической функцией распределения? Каковы ее основные свойства?

8. Сформулировать основные принципы группирования эмпирических данных.

9. Как геометрически изобразить эмпирические распределения?

Упражнения

33.1. Выборка задана в виде распределения частот:

x_i	4	7	8	12	17
m_i	2	4	5	6	3

Найти распределение относительных частот и основные характеристики вариационного ряда.

33.2. Построить эмпирическую функцию по заданному распределению выборки:

x_i	2	4	6
m_i	10	15	25

33.3. Выборка задана таблицей распределения:

x_i	1	2	3	5
m_i	15	20	10	5

Найти выборочные характеристики: среднее, дисперсию и среднее квадратическое отклонение.

33.4. Выборка задана таблицей распределения

x_i	4	8	12	16	20	24	28
m_i	15	20	10	5	20	10	20

Найти выборочные характеристики: среднее, дисперсию и среднее квадратическое отклонение.

34. Методы параметрического и непараметрического оценивания параметров

34.1. Точечные оценки параметрической совокупности распределений

Оценка называется ***точечной***, если она задается одним числом. К ней относятся характеристики выборочной совокупности, которые являются точечными оценками соответствующих характеристик генеральной совокупности.

Пусть распределение признака X генеральной совокупности задано таблицей:

x_i	x_1	x_2	$\cdots$	x_k
M_i	M_1	M_2	$\cdots$	M_k

,

где x_i – значение признака X; M_i – частота этого значения ($i=\overline{1,k}$). Объем генеральной совокупности равен N ($N = M_1 + M_2 + ... + M_k$).

Среднее арифметическое (аналог математического ожидания) распределения признака X в генеральной совокупности, найденное по формуле (33.13), называется ***генеральным средним***. Если применяется выборочный метод, то ставится задача оценить его по данным выборки.

Пусть распределение признака X выборочной совокупности задано таблицей:

x_i	x_1	x_2	$\cdots$	x_l
m_i	m_1	m_2	$\cdots$	m_l

где m_i – частота значения x_i ($i=\overline{1,l}$), объем выборки равен n ($n = m_1 + m_2 + ... + m_l$).

Среднее арифметическое распределения выборки найдем по формуле (33.2). Выборочное среднее $\bar{x}$ является несмещенной оценкой среднего генерального $\bar{x}_{\text{ген.}}$, то есть: $M(\bar{x}) = \bar{x}_{\text{ген.}}$

Докажем, что $\bar{x}$ – несмещенная оценка $\bar{x}_{\text{ген.}}$.

Рассмотрим $\bar{x}$ как случайную величину, а значения признака X ($x_1, x_2, \ldots, x_n$) как независимые одинаково распределенные случайные величины $X_1, X_2, \ldots, X_n$. Поскольку эти величины одинаково распределены, то они имеют одинаковое математическое ожидание $M(X_i) = a$. Известно, что математическое ожидание среднего арифметического одинаково распределенных случайных величин равно их общему математическому ожиданию, то есть:

$$M(\bar{x}) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = a.$$

Математическое ожидание a каждой из величин равно математическому ожиданию признака X генеральной совокупности:

$$M(X) = \bar{x}_{\text{ген.}} = a.$$

Заменив a на $\bar{x}_{\text{ген.}}$, получим: $M(\bar{x}) = \bar{x}_{\text{ген.}}$.

То есть выборочное среднее является несмещенной оценкой генерального среднего.

Докажем, что выборочное среднее является также состоятельной оценкой генерального среднего.

Пусть случайные величины $X_1, X_2, \ldots, X_n$ имеют ограниченные дисперсии. Тогда можно использовать для этих величин теорему Чебышева:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\right| < \varepsilon\right) \underset{n\to\infty}{\to} 1,$$

согласно которой при росте n среднее арифметическое величин, которые рассматриваются ($\bar{x}$), стремится по вероятности к общему математическому ожиданию a, то есть к генеральному среднему $\bar{x}_{\text{ген.}}$ ($\bar{x}_{\text{ген.}} = a$).

Это означает, что выборочное среднее является состоятельной оценкой генерального среднего.

Близость выборочных средних к генеральному среднему не зависит от отношения объема выборки к объему генеральной совокупности, а зависит от объема выборки: чем объем выборки больше, тем меньше выборочное среднее отличается от генерального среднего.

Выборочная дисперсия является смещенной оценкой генеральной дисперсии $D_{\text{ген.}}$, то есть математическое ожидание выборочной дисперсии не равно генеральной дисперсии (33.5):

$$M(D) = \frac{n-1}{n} D_{\text{ген.}} .$$

Для доказательства обозначим $D_{\text{ген.}} = \sigma_0^2$, $D = \sigma^2$, $\bar{x}_0$ – генеральное среднее, $\bar{x}$ – выборочное среднее.

Запишем дисперсию выборки в виде: $\sigma^2 = \overline{x^2} - (\bar{x})^2$.

Тогда $M(\sigma^2) = M(\overline{x^2} - (\bar{x})^2) = M(\overline{x^2}) - M((\bar{x})^2)$,

$M(\overline{x^2}) = M\left(\dfrac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}\right) = \dfrac{1}{n} \cdot n \cdot \overline{x_0^2} = \overline{x_0^2}$ (для одинаково распределенных величин),

$$M((\bar{x})^2) = M((\bar{x} - \bar{x}_0 + \bar{x}_0)^2) = M(((\bar{x} - \bar{x}_0) + \bar{x}_0)^2) =$$

$$= M((\bar{x} - \bar{x}_0)^2 + 2(\bar{x} - \bar{x}_0)\bar{x}_0 + \bar{x}_0^{\,2}) =$$

$$= M((\bar{x} - \bar{x}_0)^2) + 2\bar{x}_0 M(\bar{x} - \bar{x}_0) + M(\bar{x}_0^{\,2}) =$$

$$= M((\bar{x} - \bar{x}_0)^2) + M(\bar{x}_0^{\,2}) = D(\bar{x}) + \bar{x}_0^{\,2}$$

(учитываем $M(\bar{x} - \bar{x}_0) = 0$ – центральное свойство математического ожидания).

Получаем:

$$M(\sigma^2) = \overline{x_0^2} - (D(\bar{x}) + (\bar{x}_0)^2) = \overline{x_0^2} - (\bar{x}_0)^2 - D(\bar{x}) = \sigma_0^2 - \frac{\sigma_0^2}{n} = \frac{n-1}{n}\sigma_0^2 .$$

Вернемся к обозначениям выборки и генеральной совокупности

$$M(D) = \frac{n-1}{n} D_{\text{ген.}}.$$

Чтобы сделать оценку несмещенной, умножим на $\frac{n}{n-1}$ последнее равенство:

$$\frac{n}{n-1} M(D) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\text{ген.}}$$

Откуда

$$M\left(\frac{n}{n-1} \cdot D\right) = D_{\text{ген.}}.$$

Величину $S_x^2 = \frac{n}{n-1} D$ называют ***исправленной статистической дисперсией выборки***. Она является несмещенной оценкой генеральной дисперсии:

$$M(S_x^2) = D_{\text{ген.}}.$$

Пример 34.1. Выборочная совокупность задана таблицей распределения:

x_i	3	3,5	4	4,5	5
m_i	5	4	9	4	8

Найти исправленную выборочную дисперсию.

Решение.

По определению исправленная выборочная дисперсия:

$$S_x^2 = \frac{n}{n-1} D.$$

Объем выборки: $n = 5 + 4 + 9 + 4 + 8 = 30$.

Выборочная дисперсия (33.3): $D = \frac{1}{n} \sum_{i=1}^{5} (x_i - \bar{x})^2 m_i$.

Выборочное среднее (33.2): $\bar{x} = \dfrac{\sum\limits_{i=1}^{5} x_i m_i}{n}$.

Находим выборочное среднее:

$$\bar{x} = \frac{3 \cdot 5 + 3{,}5 \cdot 4 + 4 \cdot 9 + 4{,}5 \cdot 4 + 5 \cdot 8}{30} = 4{,}1,$$

и выборочную дисперсию:

$$D = \frac{(3-4{,}1)^2 \cdot 5 + (3{,}5-4{,}1)^2 \cdot 4 + (4-4{,}1)^2 \cdot 9 + (4{,}5-4{,}1)^2 \cdot 4 + (5-4{,}1)^2 \cdot 8}{30},$$

$$D = 0{,}49.$$

Исправленная выборочная дисперсия: $S_x^2 = \dfrac{30}{29} \cdot 0{,}49 \approx 0{,}505.$

34.2. Метод максимального правдоподобия. Метод моментов

Метод максимального правдоподобия и метод моментов позволяют найти точку, около которой изменяется неизвестный параметр, который необходимо оценить.

Рассмотрим ***метод максимального правдоподобия***.

Пусть случайная величина X соответствует закону распределения с вероятностью $p(x,\theta)$ для дискретных случайных величин и плотностью распределения $f(x,\theta)$ – для непрерывных, где параметр θ оценивается по результатам эксперимента.

Ставится задача: при заданном наборе $X = (x_1, x_2, \ldots, x_n)$ найти такой параметр θ, чтобы вероятность $p(x,\theta)$ (в дискретном случае) или $f(x,\theta)$ (в непрерывном случае) была максимальной.

Пусть X – дискретная случайная величина. По результатам испытаний величина X принимает значение x_i $(i = \overline{1,n})$ с вероятностью $p(x_i,\theta)$.

Функцией правдоподобия дискретной случайной величины X называют функцию аргумента θ:

$$L(x_1, x_2, \ldots, x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta),$$

где x_i $(i = \overline{1,n})$ – фиксированные числа.

Оценкой параметра θ по методу максимального правдоподобия называют такое значение θ^*, при котором функция $L(x_i, \theta)$ принимает максимальное значение.

Так как функции L и $\ln L$ достигают максимума при одном значении θ, то вместо функции L применяют $\ln L$:

$$\ln L = \sum_{i=1}^{n} \ln p(x_i, \theta).$$

Чтобы найти точку максимума этой функции, необходимо первую производную функции по θ приравнять нулю и найти параметр θ^*:

$$\frac{\partial \ln L}{\partial \theta} = 0.$$

Это уравнение называют ***уравнением правдоподобия***.

Если найти вторую производную и определить точки, в которых она меньше нуля:

$$\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta=\theta^*} < 0,$$

то θ^* – точка максимума, она и является оценкой параметра θ.

Метод максимального правдоподобия наиболее полно использует данные выборки относительно параметра, который оценивается. Он эффективен также в случаях малых выборок.

В случае непрерывной случайной величины X ***функцией правдоподобия*** называют функцию:

$$L(X, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta),$$

где x_i $(i = \overline{1,n})$ – фиксированные числа.

Оценку максимального правдоподобия неизвестного параметра θ находят так же, как и для дискретной случайной величины.

Если требуется оценить несколько параметров $\theta_1, \theta_2, \ldots$, заменяют в приведенных выше формулах параметр θ вектором $\theta = (\theta_1, \theta_2 \ldots)$ и решают систему уравнений правдоподобия: $\dfrac{\partial \ln L}{\partial \theta_j} = 0$, $j = \overline{1,k}$.

Пример 34.2. Найти методом максимального правдоподобия оценку параметра p биномиального распределения:

$$P_n(m) = c_n^m p^m (1-p)^{n-m},$$

если в n_1 независимых испытаниях событие A появилось $x_1 = m_1$ раз и в n_2 независимых испытаниях событие A появилось $x_2 = m_2$ раз.

Решение.

Запишем функцию правдоподобия, учитывая, что $\theta = p$:

$$L = P_{n_1}(m_1) \cdot P_{n_2}(m_2) = c_{n_1}^{m_1} c_{n_2}^{m_2} p^{m_1+m_2} (1-p)^{(n_1+n_2)-(m_1+m_2)}.$$

Найдем логарифмическую функцию правдоподобия:

$$\ln L = \ln\left(c_{n_1}^{m_1} c_{n_2}^{m_2}\right) + (m_1+m_2)\ln p + \left((n_1+n_2)-(m_1+m_2)\right)\ln(1-p).$$

Определим первую производную этой функции и приравняем ее нулю:

$$\frac{d \ln L}{dp} = 0.$$

Получим:

$$\frac{m_1+m_2}{p} - \frac{(n_1+n_2)-(m_1+m_2)}{1-p} = 0.$$

Решим последнее уравнение относительно p.

$$p = \frac{m_1+m_2}{n_1+n_2}. \qquad (34.1)$$

Найдем вторую производную $\dfrac{d^2 \ln L}{dp^2}$:

$$\frac{d^2 \ln L}{dp^2} = -\frac{m_1 + m_2}{p^2} - \frac{(n_1 + n_2) - (m_1 + m_2)}{(1 - p)^2}.$$

Подставим p из (34.1) и убедимся, что вторая производная отрицательна. Таким образом, величина $p = \frac{m_1 + m_2}{n_1 + n_2}$ является точкой максимума функции $\ln L$ и ее можно принять за оценку максимального правдоподобия неизвестной вероятности p биномиального распределения:

$$\theta^* = p^* = \frac{m_1 + m_2}{n_1 + n_2}.$$

Пример 34.3. Найти по методу максимального правдоподобия оценку параметра λ для показательного закона распределения с плотностью распределения $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.

Решение.

Функция правдоподобия имеет вид:

$$L = \lambda e^{-\lambda x}\text{, где } \theta = \lambda .\ f(x) = \lambda e^{-\lambda x}.$$

Соответственно, логарифмическая функция правдоподобия:

$$\ln L = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i .$$

Уравнение правдоподобия: $\frac{d \ln L}{d\lambda} = 0$ или $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$.

Откуда имеем: $\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$.

Можно убедиться, что $\frac{d^2 \ln L}{d\lambda^2} < 0$ при определенном значении λ.

Таким образом, $\theta^* = \lambda^* = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}_в}$.

Рассмотрим ***метод моментов***.

Эмпирическая (выборочная) функция распределения с ростом числа наблюдений сколь угодно мало отличается от теоретической (настоящей) функции распределения с вероятностью, сколь угодно близкой к единице. Поэтому при большом числе наблюдений выборочные и теоретические моменты близки друг к другу.

Если теоретическая функция зависит от некоторых параметров, то для определения параметров можно использовать выборочные моменты. Используют начальные и центральные моменты:

$$\nu_k = M\left(X^k\right), \ \mu_k = M\left(X - M(X)\right)^k.$$

Метод моментов точечной оценки неизвестных параметров заключается в том, что приравнивают теоретические моменты распределения к соответствующим эмпирическим моментам того же порядка.

Пусть закон распределения случайной величины X содержит необходимые для оценки параметры θ_1, θ_2 (например, плотностью распределения непрерывной случайной величины X является $f(x, \theta_1, \theta_2)$).

Тогда моменты распределения величины X:

$$\nu_1 = M(X) = \int_{-\infty}^{\infty} x f(x, \theta_1, \theta_2)\,dx,$$

$$\mu_2 = D(X) = \int_{-\infty}^{\infty} (x - \nu_1)^2 f(x, \theta_1, \theta_2)\,dx.$$

Они представляют собой некоторые функции от параметров θ_1, θ_2.

Если рассмотреть записанные формулы, как уравнения относительно θ_1, θ_2, а затем заменить теоретические моменты ν_1, μ_2 соответствующими выборочными моментами – выборочным средним $\bar{x}$ и выборочной дисперсией D, то это и даст оценки θ_1 и θ_2 по методу моментов.

Пример 34.4. Для показательного закона распределения с плотностью распределения $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$ найти методом моментов начальную оценку неизвестного параметра λ.

Решение.

Сравним начальный теоретический момент первого порядка с начальным эмпирическим моментом первого порядка:

$$\nu_1 = M(X);\ M(X) = \bar{x}.$$

Для показательного закона распределения $\bar{x} = \frac{1}{\lambda}$. Тогда $\lambda = \frac{1}{\bar{x}}$.

Таким образом, точечная оценка параметра $\lambda^* = \frac{1}{\bar{x}}$.

34.3. Интервальные оценки. Доверительный интервал

При малом объеме выборки точечная оценка может значительно отличаться от оцениваемого параметра, то есть приводить к ошибкам. Поэтому следует использовать интервальную оценку, при которой исследуемый параметр покрывают доверительным интервалом.

Доверительным интервалом называется интервал, который покрывает все значения случайной величины с заданной надежностью или с заданным уровнем значимости.

Определим доверительный интервал для среднего генеральной совокупности. Поскольку по теореме Ляпунова при достаточно большом количестве испытаний все законы распределения приближаются к нормальному, то при построении доверительного интервала можно допустить, что выборка принадлежит нормально распределенной генеральной совокупности. Используя теорему Лапласа для отклонений нормально распределенной случайной величины от своего математического ожидания, получим:

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Пусть $\varepsilon = \sigma t$, тогда $P(|X - a| < \sigma t) = 2\Phi(t)$.

Если взять несколько выборок из генеральной совокупности и вычислить $\bar{x}_1$, $\bar{x}_2$, … , $\bar{x}_k$, то $\bar{x}$ можно рассматривать как случайную величину, $a = \bar{x}_{\text{ген.}}$, $\sigma = \frac{\sigma_{\text{в}}}{\sqrt{n}}$, $P\left(\left|\bar{x} - a\right| < \frac{\sigma_{\text{в}} t}{\sqrt{n}}\right) = 2\Phi(t)$, где n – объем выборочной совокупности.

То есть с надежностью $P=2\Phi(t)$ можно построить доверительный интервал для среднего генеральной совокупности:

$$\bar{x}-\frac{\sigma_{\text{в}}t}{\sqrt{n}}\le a\le\bar{x}+\frac{\sigma_{\text{в}}t}{\sqrt{n}}. \qquad (34.2)$$

При $n>30$ параметр t находят в таблице функции $\Phi(t)$ (см. приложение В), при $n\le 30$ параметр $t=t(\gamma,n)$ – табл. Е.1 из приложения Е (при этом вместо $\sigma_{\text{в}}$ надо поставить исправленное среднеквадратическое отклонение $\sqrt{S_x^2}$).

С надежностью $P=2\Phi(t)$ и точностью ε можно найти необходимый объем выборки.

Из формулы $\varepsilon=\frac{\sigma_{\text{в}}t}{\sqrt{n}}$ получим: $n=\frac{\sigma^2{}_{\text{в}}t^2}{\varepsilon^2}$.

Если генеральная совокупность ограничена, тогда надо сделать поправку на дисперсию и принять следующее:

$$\sigma=\frac{\sigma_{\text{в}}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}, \qquad \varepsilon=\frac{\sigma_{\text{в}}t}{\sqrt{n}}\sqrt{1-\frac{n}{N}},$$

где N – объем генеральной совокупности.

Соответственно, найдем n.

$$\varepsilon^2=\sigma^2t^2\left(\frac{1}{n}-\frac{1}{N}\right) \quad\Rightarrow\quad \frac{1}{n}-\frac{1}{N}=\frac{\varepsilon^2}{\sigma^2t^2}.$$

Окончательно:

$$n=\frac{1}{\frac{\varepsilon^2}{\sigma^2t^2}+\frac{1}{N}}=\frac{\sigma^2t^2N}{\varepsilon^2N+\sigma^2t^2}=\frac{\sigma^2t^2}{\varepsilon^2+\frac{\sigma^2t^2}{N}}.$$

Пример 34.5. Случайная величина имеет $\bar{x}=12{,}57$, $\sigma_{\text{в}}=5$, $n=100$. Построить доверительный интервал для среднего генеральной совокупности с надежностью 95%.

Решение.

$P=2\Phi(t)=0{,}95$, тогда $\Phi(t)=0{,}475$.

По таблице значений функции $\Phi(t)$ получим $t=1{,}96$.

Тогда:

$$12{,}57-\frac{5\cdot 1{,}96}{\sqrt{100}}\le a\le 12{,}57+\frac{5\cdot 1{,}96}{\sqrt{100}},\quad 11{,}59\le a\le 13{,}56.$$

Таким образом, с надежностью 95% любое число из этого интервала можно рассматривать как среднее генеральной совокупности.

Пример 34.6. Пусть $\overline{x}=2$, $\varepsilon=0{,}2$, $\sigma_в=1$. Определить, какого объема необходимо взять выборку, чтобы построить доверительный интервал для среднего генеральной совокупности длиной не более 2ε с надежностью 99%.

Решение.

$P=2\Phi(t)=0{,}99$, тогда $\Phi(t)=0{,}495$. По таблице значений функции $\Phi(t)$ получим $t=2{,}58$.

По формуле $\varepsilon=\dfrac{\sigma_в t}{\sqrt{n}}$ получаем $n=\dfrac{\sigma_в^2 t^2}{\varepsilon^2}$, откуда

$$n=\frac{1\cdot(2{,}58)^2}{(0{,}2)^2}=166{,}41.$$

Следовательно, $n=167$.

Пример 34.7. Пусть $\overline{x}=12$, $\sigma_в=4$, $n=16$. Построить доверительный интервал для среднего генеральной совокупности с надежностью 95%.

Решение.

$P=0{,}95$. По табл. Е.1 приложения Е получим значение параметра $t(0{,}95;16)=2{,}15$. $\sqrt{S_x^2}=\sqrt{\dfrac{n}{n-1}D}=\sqrt{\dfrac{n}{n-1}\sigma_в^2}=\sqrt{\dfrac{16}{15}\cdot 4^2}=4{,}13.$

Тогда:

$$12-\frac{4{,}13\cdot 2{,}15}{\sqrt{16}}\le a\le 12+\frac{4{,}13\cdot 2{,}15}{\sqrt{16}},\quad 9{,}78\le a\le 14{,}22.$$

Найдем ***доверительный интервал для оценки генерального среднеквадратического отклонения*** случайной величины. Если взять несколько выборок из генеральной совокупности и вычислить $\sigma_{1\text{в}}$, $\sigma_{2\text{в}}$, … , $\sigma_{k\text{в}}$, то $\sigma_{\text{в}}$ можно рассматривать как случайную величину.

Возьмем в качестве случайной величины среднеквадратическое отклонение s (s^2 – исправленная выборочная дисперсия).

Потребуем, чтобы выполнялось условие

$$P\left(\left|s-\sigma_{\text{ген.}}\right|<\varepsilon\right)=\gamma,$$

то есть получим доверительный интервал для генерального среднеквадратического отклонения с надежностью $P=\gamma$:

$$s-\varepsilon\le\sigma_{\text{ген.}}\le s+\varepsilon,$$

$$s\left(1-\frac{\varepsilon}{s}\right)\le\sigma_{\text{ген.}}\le s\left(1+\frac{\varepsilon}{s}\right),$$

$$s(1-q)\le\sigma_{\text{ген.}}\le s(1+q),\text{ где } q=\frac{\varepsilon}{s}. \qquad (34.3)$$

Значение величины $q=q(\gamma,n)$, которая зависит от заданной надежности и объема выборки, находят в специальной таблице Е.2 (приложение Е).

Пример 34.8. Пусть $n=50$, $s=2{,}82$. С надежностью 95% построить доверительный интервал для $\sigma_{\text{ген.}}$.

Решение.

По табл. Е.2 (приложение Е) находим $q(0{,}95;50)=0{,}21$.

Тогда:

$$2{,}82(1-0{,}21)\le\sigma_{\text{ген.}}\le2{,}82(1+0{,}21),$$

$$2{,}228\le\sigma_{\text{ген.}}\le3{,}412.$$

Таким образом, с надежностью 95 % любое число из этого интервала можно взять как генеральное среднеквадратическое отклонение.

Вопросы для самодиагностики

1. Что называют статистическими оценками? Каковы требования к статистическим оценкам?

2. Что такое смещенные и несмещенные статистические оценки?

3. Что называется исправленной дисперсией?

4. Чем отличаются точечные и интервальные статистические оценки?

5. В чем особенности метода максимального правдоподобия и метода моментов?

6. Как найти доверительный интервал для генерального среднего и генерального среднеквадратичного отклонения случайной величины, которая распределена по нормальному закону?

Упражнения

34.1. Случайная нормально распределенная величина имеет $\bar{x} = 15{,}7$, $\sigma = 6$, $n = 90$. Построить доверительный интервал для генерального среднего с надежностью 95 %.

34.2. Случайная величина распределена по нормальному закону (данные в таблице).

x_i	1	3	5	7	9
m_i	2	5	4	6	3

Вычислить с надежностью 0,95 доверительный интервал для оценки генерального среднего.

34.3. Количественная характеристика X генеральной совокупности распределена нормально. Из выборки объемом $n = 25$ вычислено среднеквадратичное отклонение $\sigma = 0{,}8$. Найти доверительный интервал, которому принадлежит генеральное среднеквадратичное отклонение σ с надежностью 0,95.

34.4. Количественная характеристика X генеральной совокупности распределена нормально. Из выборки объемом $n = 10$ вычислено исправленное среднеквадратичное отклонение $S = 0{,}15$. Найти доверительный интервал, которому принадлежит генеральное среднеквадратичное отклонение σ с надежностью 0,999.

35. Методы проверки статистических гипотез

35.1. Общее понятие о проверке гипотез. Основная гипотеза и альтернативная

При использовании методов математической статистики можно допустить определенный процент ошибочных решений, поскольку они основываются на случайных величинах и принимаются в условиях неопределенности. Доля ошибочных решений, которой можно пренебречь, называется ***уровнем значимости*** (α). Чаще всего она составляет 5% или 1%, то есть $\alpha = 0{,}05$ или $\alpha = 0{,}01$. Другими словами, надо выбрать такую вероятность, которую исследователь признает достаточной при установлении границ случайного колебания явления. Такую вероятность называют надежностью и берут, как правило, равной $0{,}95$ или $0{,}99$.

При сравнении нескольких статистических характеристик, которые вычислены по результатам выборок, возникает потребность установить, существенная между ними разница или нет. *Существенной* называют разницу, которая по величине превышает ту, которую можно было бы объяснить случайными колебаниями.

Пусть X – непрерывная или дискретная случайная величина.

Утверждение об отсутствии существенной разницы между ее эмпирической и теоретической характеристиками называется ***нулевой гипотезой*** (H_0). Вводят ***альтернативную гипотезу*** (H_1) – противоположную нулевой гипотезе. Правило, по которому проверяют нулевую гипотезу (принять ее или отклонить), называется ***критерием согласия***. Множество значений X, для которых нулевая гипотеза отклоняется, называется ***критической областью***.

35.2. Ошибки первого и второго рода. Мощность критерия

Правило, по которому принимается решение принять или отклонить нулевую гипотезу H_0, называется ***статистическим критерием***. Проверку выполнения условий статистических критериев выполняют на основе результатов выборки. То есть статистический критерий устанавливает, при каких результатах случайной выборки выдвинутая гипотеза принимается, при каких – отклоняется.

При проверке гипотез возможны ошибки двух родов. ***Ошибка первого рода*** заключается в том, что нулевая гипотеза H_0 отклоняется, в то время как она в действительности правильная. При этом уровень значимости обозначают $\alpha = 0{,}01$ или надежность 99 %. ***Ошибка второго рода*** заключается в том, что нулевая гипотеза H_0 принимается, а в действительности она неправильная. При этом уровень значимости обозначают $\beta = 0{,}05$ или надежность 95 %.

Ошибку первого рода, которую можно допустить, задают предварительно.

Значения параметра, при которых гипотеза H_0 отклоняется, образуют критическую область.

Задача проверки гипотезы сводится к нахождению критической области при данном уровне значимости.

Вероятность недопущения ошибки второго рода называют ***мощностью критерия***.

Проверка статистической гипотезы предусматривает последовательное выполнение таких этапов:

1. Оценка исходной информации и описание статистической модели выборочной совокупности.
2. Формулирование нулевой и альтернативной гипотез.
3. Определение уровня значимости, с помощью которого контролируется ошибка первого рода.
4. Выбор наиболее мощного критерия для проверки нулевой гипотезы.
5. Расчет фактического (эмпирического) значения статистики критерия.
6. Определение критической области и области согласия с нулевой гипотезой, то есть определение табличного (критического) значения критерия.
7. Сопоставление фактического и табличного значений критерия и формулирование выводов по результатам проверки нулевой гипотезы.

Выбор конкретного метода проверки гипотезы зависит от характера исследуемой гипотезы, свойств исходной информации и других условий.

35.3. Критерий Стьюдента относительно проверки гипотезы о существенности различия двух выборочных средних. Критерий согласия относительно частоты

Критерий Стьюдента для проверки гипотезы о существенности различия двух выборочных средних

Пусть имеем две выборки из генеральной совокупности. Требуется при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что различие средних незначимо, то есть нулевая гипотеза $H_0 : \overline{X}_1 = \overline{X}_2$ – выборки принадлежат одной генеральной совокупности.

Обозначим случайную величину по первой выборке X_1, объем выборки – n_1, по второй выборке – X_2, ее объем – n_2.

Рассмотрим $\overline{X}_1 - \overline{X}_2$. Если выборки из одной генеральной совокупности, то их средние $\overline{X}_1$ и $\overline{X}_2$ должны быть равными, то есть:

$$M(\overline{X}_1 - \overline{X}_2) = M(\overline{X}_1) - M(\overline{X}_2) = \overline{X}_1 - \overline{X}_2 = 0$$

$$\text{и} \quad D(\overline{X}_1 - \overline{X}_2) = D(\overline{X}_1) + D(\overline{X}_2).$$

Дисперсия среднего арифметического n одинаково распределенных случайных величин в n раз меньше их общей дисперсии, то есть:

$$D(\overline{X}_1) = \frac{S_1^2}{n_1}, \quad D(\overline{X}_2) = \frac{S_2^2}{n_2},$$

где S_1^2, S_2^2 – исправленные выборочные дисперсии случайных величин X_1 и X_2 соответственно.

По теореме Лапласа имеем:

$$P(|\overline{X}_1 - \overline{X}_2 - 0| \le \sigma t) = 2\Phi(t),$$

где $\sigma^2 = D(\overline{X}_1 - \overline{X}_2) = \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$.

Следовательно, $P\left(|\overline{X}_1 - \overline{X}_2| < t\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}\right) = 2\Phi(t)$,

$$t = \frac{\left| \overline{X_1} - \overline{X_2} \right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \qquad (35.1)$$

Если $t < t_{кр\alpha}$, то с уровнем значимости α гипотеза H_0 принимается, если $t > t_{кр\alpha}$, то гипотеза H_0 отклоняется с надежностью $P = 1 - \alpha$.

На практике приняты такие значения $t_{кр\alpha}$:

если $t < 1{,}96$, то с уровнем значимости $0{,}05$ нулевая гипотеза H_0 принимается, то есть различие средних незначимо, выборки принадлежат одной генеральной совокупности;

если $t > 2{,}58$, то с надежностью 99% гипотеза H_0 отклоняется, то есть выборки принадлежат разным генеральным совокупностям;

область $1{,}96 < t < 2{,}58$ считается областью неопределенности.

Если n_1 и n_2 малы, то $t_{кр\alpha}$ находят по таблице критических точек распределения Стьюдента (см. приложение Ж).

Пример 35.1. Произведены две серии испытаний по 15 в каждой, X – температура режима испытания в эксперименте.

x_{1i}	-3	-2	-1	0	1	2	3
m_{1i}	1	3	2	2	2	3	2

x_{2i}	-2	-1	0	1	2
m_{2i}	2	5	3	4	1

Проверить, значимо ли различие средних значений с надежностью 95 %.

Решение.

Формулируем нулевую гипотезу – $H_0 : \overline{X}_1 = \overline{X}_2$.

Вычисляем $\overline{X}_1 = 0{,}2$, $\overline{X}_2 = -0{,}2$, $D(X_1) = 3{,}63$, $D(X_2) = 1{,}36$; $S_1^2 = 3{,}89$, $S_2^2 = 1{,}46$.

Объемы выборок малы, поэтому по таблице критических точек распределения Стьюдента находим $t(k;\alpha) = t(15;0{,}05) = 2{,}13$.

Рассчитываем статистику критерия Стьюдента:

$$t=\frac{|0,2+0,2|}{\sqrt{\frac{3,89}{15}+\frac{1,46}{15}}}=0,67<2,13,$$

то есть гипотеза H_0 не будет отклонена.

Отсюда следует, что различие средних незначимо.

Критерий согласия относительно частоты

Пусть p – вероятность события, n – объем выборки из генеральной совокупности, m – частота события в этой выборке.

Гипотеза H_0: частота m близка к теоретической.

По теореме Лапласа строим критическую область частот:

$$M(X)=np,\ P(|m-np|<\sigma t)=2\Phi(t),$$

то есть с надежностью $P=2\Phi(t)$:

$$np-\sigma t<m<np+\sigma t. \tag{35.2}$$

Если эмпирическая частота не принадлежит этому интервалу, то с уровнем значимости $\alpha=1-p$ нулевая гипотеза H_0 отклоняется.

Пример 35.2. Объем выборки $n=280$, событие произошло $m=151$ раз. Вероятность события $p=1/2$. Можно ли считать выборочную частоту достаточно близкой к теоретической частоте (гипотеза H_0).

Решение.

Теоретическая частота $np=280\cdot 1/2=140$,

$$\sigma(X)=\sqrt{npq}=\sqrt{280\cdot\frac{1}{2}\cdot\frac{1}{2}}=\sqrt{70}\approx 8,37.$$

Выбираем надежность 95%, то есть $P=0,95$, $2\Phi(t)=0,95$.

По таблице $t=1,96$, тогда

$$140-8,37\cdot 1,96\le m\le 140+8,37\cdot 1,96,\quad 123,6\le m\le 156,4.$$

Частота $m = 151$ принадлежит полученному интервалу. Это значит, что гипотеза H_0 принимается – выборочная частота с надежностью 95% близка к теоретической.

Если $m = 157$, то строим доверительный интервал с надежностью $P = 99\,\%$, то есть $P = 0{,}99$, $2\Phi(t) = 0{,}99$, по таблице находим $t = 2{,}58$.

Тогда $140 - 8{,}37 \cdot 2{,}58 \le m \le 140 + 8{,}37 \cdot 2{,}58$, $118{,}4 \le m \le 161{,}6$, то есть m входит в область неопределенности.

Если же $m = 165$, то m вне интервала с надежностью 99 %, то есть гипотеза H_0 отвергается, выборочная частота не соответствует теоретической.

35.4. Критерий согласия χ^2 («хи-квадрат») относительно закона распределения

Часто закон распределения случайной величины в генеральной совокупности является неизвестным, но определенные предположения относительно его характера можно сделать, судя по гистограмме в выборочной совокупности. В этом случае проверяют нулевую гипотезу H_0: генеральная совокупность распределена по предполагаемому закону.

Критерием согласия называют критерий проверки гипотезы относительно предполагаемого неизвестного закона распределения.

35.4.1. Критерий согласия Пирсона

Согласно этому критерию наблюдаемое эмпирическое распределение выборочной совокупности, которое выражено эмпирическими частотами m_i сгруппированного ряда, сравнивается с предполагаемым теоретическим распределением генеральной совокупности, которое выражено теоретическими частотами $\tilde{m}_i$.

Если число наблюдений очень большое $(n \to \infty)$, то закон распределения случайной величины независимо от того, какому закону распределения подчинена генеральная совокупность, приближается к распределению χ^2 с k степенями свободы, а сам критерий называют критерием согласия «*хи-квадрат*», или *критерием Пирсона*.

Для проверки нулевой гипотезы H_0 требуется вычислить величину

$$\chi^2 = \sum_{i=1}^{s} \frac{(m_i - \tilde{m}_i)^2}{\tilde{m}_i}, \tag{35.3}$$

где s – количество интервалов сгруппированного ряда распределения;

m_i – эмпирические частоты;

$\tilde{m}_i$ – теоретические частоты.

Принято считать, что наблюдений m_i в каждом интервале должно быть не менее пяти процентов от общего числа наблюдений: $m_i \geq 0{,}05n$. Если их будет меньше, то необходимо укрупнить интервалы.

Найденная по формуле (35.3) величина сравнивается с критическими значениями $\chi^2_\alpha(k)$, которые находят в специальных справочных таблицах (см. приложение Д).

Число степеней свободы k определяется по формуле:

$$k = s - r - 1,$$

где s – количество укрупненных интервалов;

r – число параметров теоретического закона распределения (для нормального и равномерного закона $r = 2$, для показательного – $r = 1$).

Величина α определяет уровень значимости. Для критерия Пирсона будем рассматривать два уровня значимости: $\alpha = 0{,}05$ и $\alpha = 0{,}01$.

Если $\chi^2 < \chi^2_{0{,}05}(k)$, то нулевая гипотеза H_0 принимается, то есть предполагаемый закон распределения соответствует эмпирическим данным, при этом мы ошибаемся в пяти случаях из ста, принимая возможно ошибочную гипотезу (ошибка второго рода).

Если $\chi^2 > \chi^2_{0{,}01}(k)$, то нулевую гипотезу следует отвергнуть, то есть предполагаемый закон распределения не соответствует эмпирическим данным, при этом мы ошибаемся в одном случае из ста, отвергая возможно правильную гипотезу (ошибка первого рода).

Если $\chi^2_{0{,}05}(k) < \chi^2 < \chi^2_{0{,}01}(k)$, то это область неопределенности (гипотезу можно как принять так и отвергнуть) и необходимо использовать другие критерии.

Вычисление теоретических частот $\tilde{m}_i$.

Если имеем выборку случайной величины X, то по ее значениям находят $x_{\max}$ и $x_{\min}$. Тогда размах варьирования $R = x_{\max} - x_{\min}$, округлим $R \approx R_0$. Найдем шаг варьирования $h = \dfrac{R_0}{S}$, где S – число интервалов, тогда таблица вычисления $\tilde{m}_i$ выглядит следующим образом:

Интервалы	x_i	m_i	$h_i = \Delta x_i$	$t_i = \dfrac{x_i - \bar{x}}{\sigma}$	$\varphi(t_i)$	$f_i = \dfrac{\varphi(t_i)}{\sigma}$	$\tilde{p}_i = hf_i$	$\tilde{m}_i = n\tilde{p}_i$
$x_0^* - x_1^*$	x_1	m_1	h_1	t_1	$\varphi(t_1)$	f_1	$\tilde{p}_1$	$\tilde{m}_1$
$x_1^* - x_2^*$	x_2	m_2	h_2	t_2	$\varphi(t_2)$	f_2	$\tilde{p}_2$	$\tilde{m}_2$
…	…	…	…	…	…	…	…	…
$x_{S-1}^* - x_s^*$	x_s	m_s	h_s	t_s	$\varphi(t_s)$	f_s	$\tilde{p}_s$	$\tilde{m}_s$

В таблице $x_i = \dfrac{x_{i-1}^* + x_i^*}{2}$ – середина интервала; m_i – эмпирическая частота; $t_i = \dfrac{x_i - \bar{x}}{\sigma}$ – стандартизированная величина; $\varphi(t_i)$ – дифференциальная функция; $\tilde{p}_i = h_i f_i$ – теоретическая вероятность; $\tilde{m}_i$ – теоретическая частота.

Пример 35.3. Задан интервальный статистический ряд случайной величины X:

Интервалы X	10 – 14	14 – 18	18 – 22	22 – 26	26 – 30
Частоты	3	8	15	10	4

Проверить гипотезу о нормальном законе распределения.

Решение.

Выдвигаем гипотезу H_0: случайная величина X распределена по нормальному закону.

Для определения статистики χ^2 составим таблицу:

x_i	12	16	20	24	28	$\sum$
m_i	3	8	15	10	4	40
$x_i m_i$	36	128	300	240	112	816
$x_i^2 m_i$	432	2048	6000	5760	3136	17376
$t_i = \dfrac{x_i - \bar{x}}{\sigma}$	-1,97	-1,03	-0,09	0,84	1,78	–
$\varphi(t_i)$	0,0573	0,2347	0,3973	0,2803	0,0818	–
$f_i = \dfrac{\varphi(t_i)}{\sigma}$	0,013	0,055	0,093	0,066	0,019	–
$\tilde{m}_i = h_i n f_i$	2,15	8,79	14,88	10,50	3,06	39,4
χ_i^2	0,339	0,071	0,001	0,024	0,286	0,721

Для вычисления $\bar{x}$ и σ используем формулы (33.2) и (33.4):

$$\bar{x} = \frac{816}{40} = 20{,}4; \ \sigma = \sqrt{\frac{17376}{40} - (20{,}4)^2} = 4{,}27.$$

Получили, что $\chi^2 = 0{,}721$.

Определяем число степеней свободы $k = s - r - 1 = 5 - 2 - 1 = 2.$

По таблице приложения Д имеем $\chi^2_{0,05} = 6{,}0$ и $\chi^2_{0,01} = 9{,}2$.

Так как $\chi^2 = 0{,}721 < \chi^2_{0,05} = 6{,}0$, то гипотеза H_0 о теоретическом нормальном законе принимается.

35.4.2. Критерий согласия Пирсона по Романовскому

При достаточно большом числе испытаний ($n > 300$) величина χ^2 имеет закон распределения, близкий к нормальному.

Ее числовые характеристики: $M(\chi^2) = k$, $D(\chi^2) = 2k$, где k – число степеней свободы ($k = s - r - 1$).

Вычислим величину:

$$t = \left| \frac{\chi^2 - k}{\sqrt{2k}} \right|. \tag{35.4}$$

Если $t \le 1{,}96$, то с надежностью 95% гипотеза H_0 принимается, то есть закон распределения соответствует теоретическому;

если $t \ge 2{,}58$, то с уровнем значимости $\alpha = 0{,}01$ гипотеза H_0 отклоняется;

если $t \in (1{,}96; 2{,}58)$ – это область неопределенности.

35.5. Малые выборки. Критерий Стьюдента

Выборки, которые мы рассматривали выше, считались большими. Если требуется оценить генеральную совокупность по результатам малого числа наблюдений ($n < 30$), то формулы для больших выборок, которые основаны на нормальном законе распределения вероятностей, дают значительные неточности. Результаты малой выборки оценивают исправленной выборочной дисперсией:

$$S_x^2 = \frac{n}{n-1} D_в = \frac{\sum_i \left(x_i - \bar{x}\right)^2 m_i}{n-1}. \tag{35.5}$$

Чтобы по результатам малой выборки проверить гипотезу о соответствии среднего генеральной совокупности некоторому заданному номинальному значению, необходимо построить доверительный интервал для генерального среднего ($a = \bar{x}_{ген.}$) по формуле:

$$\bar{x}_в - \frac{S_x t_\gamma}{\sqrt{n}} < a < \bar{x}_в + \frac{S_x t_\gamma}{\sqrt{n}} \quad \text{и} \quad P\left(\left| \frac{\bar{x}_в - a}{S_x / \sqrt{n}} \right| < t_\gamma \right) = \gamma, \tag{35.6}$$

где γ – надежность;

t_γ – величина статистики t-критерия Стьюдента, которая зависит от уровня значимости $\alpha = 1 - \gamma$ и числа степеней свободы $k = n - 1$. Значения величины t_γ приведены в специальной таблице (см. приложение Ж).

Пример 35.4. Проведено 16 испытаний. При этом $S_x = 0{,}8$, $\overline{x}_в = 20{,}2$. С надежностью $\gamma = 0{,}95$ проверить гипотезу о том, что $a = \overline{x}_{\text{ген.}} = 20{,}2$.

Решение.

Выдвигаем нулевую гипотезу H_0: $a = \overline{x}_{\text{ген.}} = 20{,}2$.

Число степеней свободы $k = 16 - 1$, уровень значимости $\alpha = 1 - \gamma = 0{,}05$, тогда по таблице критических точек распределения Стьюдента (см. приложение Ж) $t(15; 0{,}05) = 2{,}13$.

Строим доверительный интервал для генерального среднего:

$$20{,}2 - \frac{2{,}13 \cdot 0{,}8}{\sqrt{16}} < a < 20{,}2 + \frac{2{,}13 \cdot 0{,}8}{\sqrt{16}},$$

$$19{,}774 < a < 20{,}626.$$

Так как $a = \overline{x}_{\text{ген.}} = 20{,}2$ принадлежит полученному интервалу, то с надежностью 95 % гипотеза H_0 принимается.

Заметим, что при $n \to \infty$ распределение Стьюдента совпадает с нормальным.

35.5. Критерий Фишера о равенстве двух дисперсий для нормальной статистической модели

Пусть имеем две выборки из нормальной генеральной совокупности, выборочные несмещенные дисперсии которых S_1^2, S_2^2. Необходимо проверить гипотезу H_0, которая заключается в том, что различие дисперсий незначимо. Для этого вычисляют значение статистики критерия Фишера – Снедекора:

$$F = \frac{S_{max}^2}{S_{min}^2}. \qquad (35.7)$$

Из таблиц критических точек распределения F Фишера – Снедекора (см. приложение З и приложение И) найдем $F_{\text{крит.}}(\alpha, k_1, k_2)$, где $k_1 = n_1 - 1$, $k_2 = n_2 - 1$ – числа степеней свободы большей и меньшей дисперсий, α – уровень значимости.

Если $F > F_{\text{крит.}}(\alpha, k_1, k_2)$, то с уровнем значимости α гипотеза H_0 отклоняется, то есть различие дисперсий значимо;

если $F < F_{\text{крит.}}(\alpha, k_1, k_2)$, то с надежностью $1-\alpha$ гипотеза H_0 принимается, то есть различие дисперсий незначимо.

Пример 35.5. Имеем две выборки случайных величин X и Y. Для случайной величины X: $n_1 = 10$, $S_x^2 = 1{,}23$; для случайной величины Y: $n_2 = 18$, $S_y^2 = 0{,}41$. Проверить гипотезу о незначимости различия дисперсий при $\alpha = 0{,}05$.

Решение.

Выдвигаем гипотезу H_0: различие дисперсий незначимо.

По формуле (35.7) вычисляем: $F = \dfrac{S_{\max}^2}{S_{\min}^2} = \dfrac{1{,}23}{0{,}41} = 3$.

$k_1 = 10 - 1 = 9$, $k_2 = 18 - 1 = 17$, $\alpha = 0{,}05$, тогда $F_{\text{крит.}}(0{,}05; 9; 17) = 2{,}5$.

Так как $F > F_{\text{крит.}}$, то различие дисперсий значимо, то есть гипотеза H_0 отклоняется.

Реализацию этого критерия удобно представить с помощью таблицы дисперсионного анализа.

В этом случае общую сумму квадратов отклонений значений случайной величины от ее общего выборочного среднего (*sum of squares total*)

$$SST = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 m_{ij}$$

($\bar{x}$ − общее среднее; p – количество выборок (групп); n_i – объем i-ой выборки; m_{ij} – частота значения x_{ij} в i-ой выборке в случае сгруппированных данных) представляют в виде суммы:

$$SST = SSR + SSE,$$

где SSR – сумма квадратов отклонений групповых средних относительно общего среднего, которые обьясняются влиянием исследуемого фактора (*sum of squares by regression*),

$$SSR = \sum_{i=1}^{p} n_i \cdot \left(\overline{x_i} - \overline{x}\right)^2,$$

($\overline{x_i}$ – выборочное среднее случайной величины X_i ($i = \overline{1, p}$));

SSE – сумма квадратов отклонений значений случайной величины от группового выборочного среднего, которые обьясняются неучтенными факторами (*sum of squares by errors*),

$$SSE = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x_i}\right)^2 m_{ij}.$$

Для проверки статистической гипотезы H_0 об однородности всей выборочной совокупности в целом рассматривают случайную величину F:

$$F = \frac{SSR(N - p)}{SSE(p - 1)},$$

где $\dfrac{SSR}{p-1}$ – удельная сумма квадратов отклонений, которые обьясняются влиянием исследуемого фактора;

$\dfrac{SSE}{N-p}$ – удельная сумма квадратов отклонений, которые обьясняются случайными причинами.

При этом следует учитывать, что число степеней свободы для SSR равно $df_R = p - 1$. Общее число степеней свободы для выборочной совокупности объемом N составляет $df_T = N - 1$. Для SSE число степеней свободы равно $df_E = df_T - df_R = N - p$.

По основной гипотезе H_0 считается, что для всех групп факторов, которые исследуются, математическое ожидание признака является постоянным, то есть: $M(X_1) = M(X_2) = \ldots = M(X_p)$. При этом предусматривается, что дисперсия признака под воздействием исследуемого фактора равна дисперсии признака, которая объясняется случайными ошибками.

В соответствии с альтернативной гипотезой H_1 дисперсия признака, которая обусловлена влиянием исследуемого фактора, существенно превышает дисперсию признака под воздействием случайных ошибок.

Случайная величина F распределена по статистике Фишера – Снедекора со степенями свободы $k_1 = df_R = p-1$ (число степеней свободы числителя – большей из дисперсий) и $k_2 = N-p$ (число степеней свободы знаменателя – меньшей из дисперсий). В соответствии с принятым предварительно уровнем значимости α и числом степеней свободы для обеих дисперсий находим критическую точку распределения Фишера – Снедекора F_{α, k_1, k_2}.

Если эмпирическое значение критерия не превышает критическое значение F_{α, k_1, k_2}, то гипотезу H_0 нет причин отклонить, следовательно, влияние фактора на признак X следует считать статистически несущественным. Если же эмпирическое значение критерия превышает критическое, то тем самым подтверждается наличие влияния определенного фактора на признак X.

Отчет о результатах дисперсионного анализа можно представить в форме таблицы (табл. 35.1).

Таблица 35.1

Форма таблицы однофакторного дисперсионного анализа

Изменчивость признака	Сумма квадратов отклонений, SS	Число степеней свободы, df	Статистическая оценка дисперсии, MS	Значение статистики критерия Фишера – Снедекора	
				эмпирическое	критическое
Межгрупповая	$\sum_{i=1}^{p} n_i \cdot (\bar{x}_i - \bar{x})^2$	$p-1$	$\frac{SSR}{p-1}$	$F_{эмп.}$	F_{α, k_1, k_2}
Внутригрупповая	$\sum_{i=1}^{p}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 m_{ij}$	$N-p$	$\frac{SSE}{N-p}$		
Общая	$\sum_{i=1}^{p}\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 m_{ij}$	$N-1$	$\frac{SST}{N-1}$		

Кроме решения основной проблемы, дисперсионный анализ также позволяет проверять гипотезу об однородности нескольких выборочных совокупностей.

Пример 35.6. Имеются данные о часовой заработной плате в условных единицах для рабочих по цехам предприятия:

I цех

x_i	m_i
1	4
2	3
4	1
5	2

II цех

x_i	m_i
1	6
2	2
3	8
4	4

III цех

x_i	m_i
1	5
2	8
3	3
5	4

,

где x_i – заработная плата; m_i – количество рабочих: $n_1 = 10$, $n_2 = 20$, $n_3 = 20$, $N = 50$.

Проверить на уровне значимости $\alpha = 0{,}05$, существенно ли различается зарплата рабочих по цехам.

Решение.

Имеем три группы наблюдений. Зарплата труда в каждом цехе предприятия является случайной величиной X_i, где $i = \overline{1{,}3}$.

Проверяем гипотезу H_0: $M(X_1) = M(X_2) = M(X_3)$, то есть качественный фактор (работа в определенном цехе) не влияет на зарплату рабочих.

Определим групповые средние, которые соответствуют фиксированному значению фактора, и общее среднее:

$$\overline{x_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j} m_{1j} = \frac{1\cdot 4 + 2\cdot 3 + 4\cdot 1 + 5\cdot 2}{10} = \frac{24}{10} = 2{,}4;$$

$$\overline{x_2} = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j} m_{2j} = \frac{1\cdot 6 + 2\cdot 2 + 3\cdot 8 + 4\cdot 4}{20} = \frac{50}{20} = 2{,}5;$$

$$\overline{x_3} = \frac{1}{n_3}\sum_{j=1}^{n_3} x_{3j} m_{3j} = \frac{1\cdot 5 + 2\cdot 8 + 3\cdot 3 + 5\cdot 4}{20} = \frac{50}{20} = 2{,}5;$$

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{3}\overline{x_j}n_j = \frac{2{,}4\cdot 10 + 2{,}5\cdot 20 + 2{,}5\cdot 20}{50} = \frac{124}{50} = 2{,}48.$$

Результаты расчетов квадратов отклонений значений случайных величин x_i от их групповых средних $\bar{x}_i$ и от общего выборочного среднего $\bar{x}$ представлены в табл. 35.2 и 35.3

Таблица 35.2

Результаты расчетов квадратов отклонений значений случайных величин x_i от их групповых средних $\bar{x}_i$

m_{1j}	$(x_{1j}-\overline{x_1})^2 m_{1j}$	m_{2j}	$(x_{2j}-\overline{x_2})^2 m_{2j}$	m_{3j}	$(x_{3j}-\overline{x_3})^2 m_{3j}$
4	7,84	6	13,5	5	11,25
3	0,48	2	0,5	8	2
1	2,56	8	2	3	0,75
2	13,52	4	9	4	25
$\sum$	24,4		25		39

Таблица 35.3

Результаты расчетов квадратов отклонений значений случайных величин x_i от общего выборочного среднего $\bar{x}$

m_{1j}	$(x_{1j}-\bar{x})^2 m_{1j}$	m_{2j}	$(x_{2j}-\bar{x})^2 m_{2j}$	m_{3j}	$(x_{3j}-\bar{x})^2 m_{3j}$
4	8,76	6	13,14	5	10,95
3	0,69	2	0,46	8	1,84
1	2,31	8	2,16	3	0,81
2	12,70	4	9,24	4	25,40
$\sum$	24,46		25,01		39,01

По данным табл. 35.2 находим остаточную вариацию как сумму квадратов отклонений внутри групп:

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\overline{x_i})^2 m_{ij} = 24{,}4 + 25 + 39 = 88{,}4.$$

По данным табл. 35.3 находим сумму квадратов отклонений значений случайной величины от общего среднего, то есть вычисляем полную вариацию:

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x}\right)^2 m_{ij} = 24{,}46 + 25{,}01 + 39{,}01 = 88{,}48.$$

Далее вычисляем сумму квадратов отклонений между группами (то есть вариацию, которая обусловлена влиянием исследуемого фактора):

$$SSR = \sum_{i=1}^{p} n_i \cdot (\overline{x}_i - \overline{x})^2 = 10 \cdot (2{,}4 - 2{,}48)^2 + 20 \cdot (2{,}5 - 2{,}48)^2 +$$
$$+ 20 \cdot (2{,}5 - 2{,}48)^2 = 0{,}08.$$

Проверяем выполнение соотношения:

$$SSR + SSE = SST.$$

Действительно,

$$0{,}08 + 88{,}4 = 88{,}48.$$

Следовательно, вычисления не содержат ошибок.

Теперь строим таблицу результатов однофакторного дисперсионного анализа (табл. 35.4).

Таблиця 35.4

Таблица результатов дисперсионного анализа

Компонента вариации	Сумма квадратов	Число степеней свободы	Удельная сумма квадратов отклонений
Между цехами	SSR= 0,08	df_R =2	MSR=0,04
Внутри цехов	SSE= 88,4	df_E =47	MSE=1,88
Полная	SST= 88,48	df_T =49	MST=1,81

По данным табл. 35.4 находим эмпирическое значение статистики F-критерия:

$$F = \frac{SSR(N-p)}{SSE(p-1)} = \frac{MSR}{MSE} = \frac{0{,}04}{1{,}88} = 0{,}021.$$

По статистике Фишера – Снедекора (см. приложение 3) при уровне значимости $\alpha = 0{,}05$ и числах степеней свободы $k_1 = p - 1$ и $k_2 = N - p$ находим предельное значение критерия:

$$F_{\alpha}(p-1;\, N-p) = F_{0,05}(2;\, 47) = 3{,}18.$$

Так как $F_{эмп.} = 0{,}021 < 3{,}18 = F_{0,05}(2;\, 47)$, то основную гипотезу нет причин отклонить.

Отсюда вывод: нельзя считать, что для данной совокупности цехов предприятия существует влияние уровня организации производства на зарплату рабочих.

Вопросы для самодиагностики

1. Охарактеризовать нулевую и альтернативную статистические гипотезы.
2. Что называется статистическим критерием проверки гипотезы? Как определяется мощность статистического критерия и уровень значимости? В чем заключаются статистические ошибки 1-го и 2-го рода?
3. Сформулировать критерий согласия Пирсона. Каковы условия его применения?
4. Описать критерий Стьюдента. Как сравниваются вариационные ряды с разными дисперсиями, но одинаковыми математическими ожиданиями?
5. В чем заключается критерий согласия Фишера – Снедекора?

Упражнения

35.1. По результатам исследования двух независимых выборочных совокупностей объемами $n_1 = 5$ и $n_2 = 6$, сформированных из генеральных совокупностей, соответствующих случайным величинам X и Y, вычислены выборочные средние этих величин $\bar{x} = 15{,}9$ и $\bar{y} = 14{,}1$, их исправленные дисперсии $S_x^2 = 14{,}76$ и $S_y^2 = 4{,}92$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу о равенстве математических ожиданий этих случайных величин $H_0 : M(X) = M(Y)$.

35.2. По результатам двух выборок объемами $n_1 = 10$ и $n_2 = 15$, которые сформированы из нормальной генеральной совокупности X и Y, вычислены исправленные выборочные дисперсии $S_x^2 = 21{,}2$ и $S_y^2 = 19{,}36$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу о равенстве дисперсий $H_0 : D(X) = D(Y)$.

35.3. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном законе распределения случайной величины в генеральной совокупности, если известны эмпирические частоты m_i и теоретические частоты $\tilde{m}_i$:

m_i	6	13	38	74	106	85	30	14
$\tilde{m}_i$	3	14	42	82	99	76	37	13

35.4. В некоторой местности был проведен анализ 100 проб воздуха на содержание в нем углеводорода. Результаты анализа приведены в таблице:

Интервалы, мг/м3	[1; 3)	[3; 5)	[5; 7)	[7; 9)	[9; 11)	[11; 13)	[13; 15)	[15; 17)	[17; 19)	[19; 21)	[21; 23)
m_i	2	4	6	10	18	20	16	11	7	5	1

Проверить гипотезу о нормальном распределении исследуемой случайной величины.

36. Элементы теории корреляции

36.1. Функциональная и статистическая зависимости

Изучение зависимостей между переменными начнем с двух переменных, которые обозначим X и Y. Существуют два основных вида зависимостей: функциональная и стохастическая (статистическая).

Функциональная зависимость между двумя переменными x и y – это, когда каждому значению одной переменной x соответствует определенное значение второй переменной – y, и имеет место уравнение $y = f(x)$.

При функциональной зависимости графики уравнений $y = f(x)$ и $x = \varphi(y)$ одинаковы, то есть не имеет значения, какую переменную считать независимой переменной, а какую функцией. Таким образом, функциональная зависимость не реагирует на направленность причинно-следственных связей.

Пример 36.1. Имеем уравнение $y = 2x - 4$. Отсюда $x = \frac{y}{2} + 2$. Графики этих линий совпадают. Обе прямые проходят через точки $x = 0, y = -4$ и $x = 2, y = 0$.

Для случайных величин X и Y не всегда можно установить функциональную зависимость. Между такими величинами существует зависимость, при которой с изменением одной величины изменяется распределение другой величины. Такая зависимость называется ***стохастической (статистической)***.

При статистической зависимости различают две компоненты:

1) стохастическая – связанная непосредственно с зависимостью между функциональным признаком и фактором – аргументом;

2) случайная – связанная с влиянием случайных факторов на зависимость между функциональным признаком и фактором – аргументом.

Отсутствие второй компоненты приводит к функциональной зависимости.

При статистических связях важны причинно-следственные связи, то есть важно, какую из переменных считать функциональным признаком, а какую – независимым фактором.

Частным случаем статистической зависимости является *корреляционная зависимость* между двумя случайными величинами, при которой с изменением одной из них изменяется среднее значение другой.

Введем понятие «условное среднее» $\overline{y}_{x_i}$. Пусть имеем случайную величину Y, связанную с величиной X, и каждому значению X соответствует несколько значений Y.

Например, при $x = 2$ величина Y принимает значения: $y_1 = 2$, $y_2 = 8$, $y_3 = 10$. Тогда среднее арифметическое значение Y, которое соответствует $x = 2$: $\overline{y}_{x=2} = \frac{6+8+10}{3} = 8$.

Условным средним $\overline{y}_{x_i}$ называют среднее арифметическое значений Y, которые соответствуют $x = x_i$.

Корреляционной зависимостью Y от X называют функциональную зависимость условного среднего $\overline{y}_x$ от x:

$$\overline{y}_x = f(x).$$

По результатам одного и того же испытания могут быть получены зависимости: $\overline{y}_x$ и $\overline{x}_y$, то есть необходимо указать, какую из переменных считать фактором-аргументом, а какую – функциональным признаком (где причина, а где следствие).

Уравнение $\overline{y}_x = f(x)$ называют ***уравнением регрессии*** Y ***на*** X, функцию $f(x)$ – ***регрессией*** Y ***на*** X, а график – ***линией регрессии*** Y ***на*** X.

Можно ввести и другую зависимость: $\overline{x}_y = \varphi(y)$ – уравнение регрессии X на Y.

Понятие регрессии и корреляции непосредственно связаны между собой. В то время как в корреляционном анализе оценивается теснота стохастической связи, в регрессионном анализе исследуется ее форма. Оба вида анализа предназначены для выяснения причинных соотношений между явлениями и для определения наличия или отсутствия связи.

Различают такие *виды корреляции и регрессии*:

1) *относительно характера* корреляции и регрессии имеем положительную или отрицательную корреляцию и регрессию. *Положительная* – когда с ростом (уменьшением) аргумента x растет (уменьшается) функциональный фактор y.

Например: между производительностью труда и заработной платой; между выполнением производственного плана и затратами рабочего времени; между техническим уровнем производства и производительностью труда. *Отрицательная* – когда с ростом (уменьшением) одной переменной значения второй уменьшаются (растут). Например: между производительностью труда и стоимостью изделия; между объемом продукции и расходами времени на единицу изделия;

2) *относительно числа переменных*: *парная* или *множественная* корреляция и регрессия.

Например: зависимость прибыли предприятия (y) от производительности труда (x): $\overline{y} = f(x)$ – парная регрессия; зависимость производительности труда (y) от уровня механизации производства (x_1), фонда рабочего времени (x_2), материалоемкости (x_3), квалификации рабочих (x_4): $\overline{y} = f(x_1, x_2, x_3, x_4)$ – множественная регрессия;

3) *относительно формы связи*: *линейная* или *нелинейная* корреляция и регрессия.

Например: зависимость суммы расходов производства (y) от количества выпускаемой продукции (x): $y = ax + b$ – линейная зависимость; зависимость себестоимости единицы продукции (y) от количества выпускаемой продукции (x) за какой-то период: $y = \frac{a}{x} + b$ – нелинейная зависимость (a – сумма расходов на прямые производственные потребности; b – постоянные расходы на единицу продукции).

36.2. Основные задачи теории корреляции. Корреляционная таблица, корреляционное поле, эмпирические линии регрессии

В теории корреляции решают *две основные задачи*:

1. Определение наличия корреляционной связи (если значения $\overline{y}_x$ одинаковы для всех значений X, то корреляционная зависимость отсутствует), установление формы зависимости (вида функции регрессии) и определение параметров уравнения регрессии.

2. Определение тесноты корреляционной связи. Теснота связи оценивается по величине рассеяния значений величины Y около условного среднего $\overline{y}_x$ и характеризует степень влияния изменчивости величины X на изменчивость величины Y.

Для определения наличия корреляционной связи эмпирические данные заносят в корреляционную таблицу. *Корреляционная таблица* формируется для сгруппированных наблюдений и содержит интервалы изменения величин X и Y и частоты совместного появления данной пары значений x и y (табл. 36.1).

Таблица 36.1

Корреляционная таблица

Интервалы Y \ Интервалы X	y_k \ x_i	$x_0^* - x_1^*$	$x_1^* - x_2^*$	…	$x_{i-1}^* - x_i^*$	…	$x_{l-1}^* - x_l^*$	m_y
		x_1	x_2	…	x_i	…	x_l	
$y_0^* - y_1^*$	y_1	m_{11}	m_{12}	…	m_{1i}	…	m_{1l}	m_{y_1}
$y_1^* - y_2^*$	y_2	m_{21}	m_{22}	…	m_{2i}	…	m_{2l}	m_{y_2}
…	…	…	…	…	…	…	…	…
$y_{k-1}^* - y_k^*$	y_k	m_{k1}	m_{k2}	…	m_{ki}	…	m_{kl}	m_{y_k}
…	…	…	…	…	…	…	…	…
$y_{s-1}^* - y_s^*$	y_s	m_{s1}	m_{s2}	…	m_{si}	…	m_{sl}	m_{y_s}
m_{x_i}		m_{x_1}	m_{x_2}	…	m_{x_i}	…	m_{x_l}	n

Обозначим n – общее количество наблюдений; m_{ki} – частоту совместного появления двух случайных величин x_i и y_k, причем $\sum_k \sum_i m_{ki} = n$.

По горизонтали корреляционной таблицы разместим интервалы фактора X, по вертикали – интервалы функционального признака Y. На пересечении столбца x_i и строки y_k заносим частоту m_{ki}. Через x_i и y_k обозначены середины интервалов.

В крайние столбец и строку запишем сумму по строкам и столбцам $m_{x_i} = \sum_{k=1}^{s} m_{ki}$ и $m_{y_k} = \sum_{i=1}^{l} m_{ki}$, причем $\sum_{i=1}^{l} m_{x_i} = \sum_{k=1}^{s} m_{y_k} = n$.

Из таблицы видим, что каждому значению одной из величин X или Y соответствует ряд распределения другой величины.

Например, для $X = x_1$ имеем ряд:

y	y_1	y_2	...	y_k	...	y_s
m	m_{11}	m_{21}	...	m_{k1}	...	m_{s1}

Для $X = x_2$ соответственно имеем:

y	y_1	y_2	...	y_k	...	y_s
m	m_{12}	m_{22}	...	m_{k2}	...	m_{s2}

Вычислим условное среднее:

$$\bar{y}_{x=x_1} = \frac{y_1 m_{11} + y_2 m_{21} + \ldots + y_k m_{k1} + \ldots + y_s m_{s1}}{m_{11} + m_{21} + \ldots + m_{k1} + \ldots + m_{s1}} = \frac{\sum_{k=1}^{s} y_k m_{k1}}{m_{x_1}},$$

$$\bar{y}_{x=x_2} = \frac{y_1 m_{12} + y_2 m_{22} + \ldots + y_k m_{k2} + \ldots + y_s m_{s2}}{m_{11} + m_{22} + \ldots + m_{k2} + \ldots + m_{s2}} = \frac{\sum_{k=1}^{s} y_k m_{k2}}{m_{x_2}}.$$

Обобщим:

$$\bar{y}_{x=x_i} = \frac{\sum_{k=1}^{s} y_k m_{ki}}{m_{x_i}}. \qquad (36.1)$$

То есть имеем зависимость:

x	x_1	x_2	...	x_i	...	x_l
$\bar{y}_x$	$\bar{y}_{x_1}$	$\bar{y}_{x_2}$	...	$\bar{y}_{x_i}$	...	$\bar{y}_{x_l}$

Аналогично для $Y = y_k$ условное среднее $\bar{x}_{y=y_k}$ вычисляется по формуле:

$$\bar{x}_{y=y_k} = \frac{\sum_{i=1}^{l} x_i m_{ki}}{m_{y_k}}. \qquad (36.2)$$

Зависимость имеет вид:

y	y_1	y_2	...	y_k	...	y_s
$\bar{x}_y$	$\bar{x}_{y_1}$	$\bar{x}_{y_2}$	...	$\bar{x}_{y_k}$	...	$\bar{x}_{y_s}$

Корреляционные таблицы дают возможность судить о наличии корреляционной зависимости: если $\bar{y}_x$ $(\bar{x}_y)$ изменяются от столбца к столбцу (от строки к строке), то между величинами X и Y есть корреляционная связь.

Корреляционное поле – это графическое изображение корреляционной таблицы. Корреляционным полем называется точечный график, в котором каждая точка является результатом отдельного наблюдения за двумя переменными величинами (рис. 36.1).

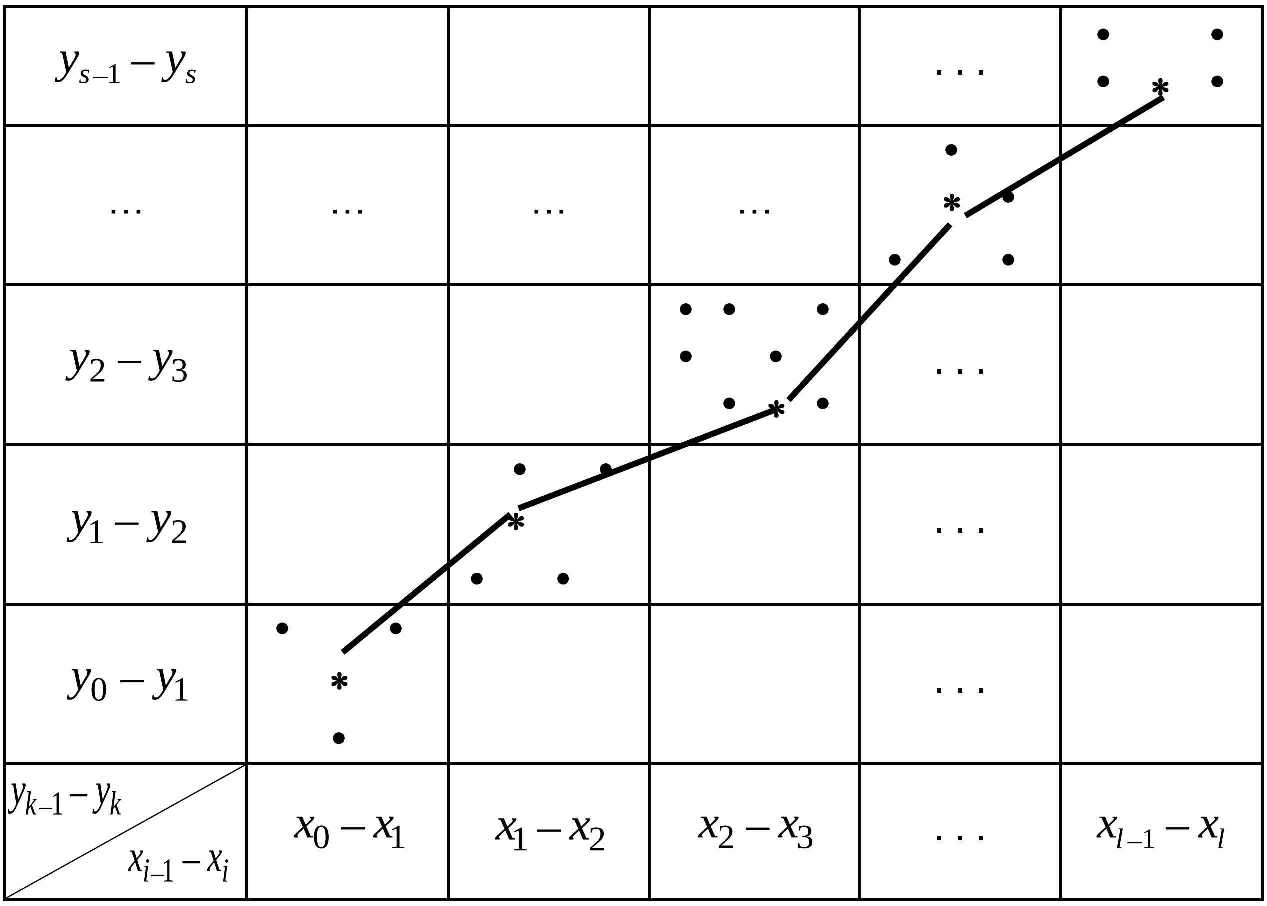


Рис. 36.1. **Корреляционное поле и эмпирическая линия регрессии**

Корреляционное поле можно построить по корреляционной таблице. Записываем интервалы величин X и Y, в каждую клетку на пересечении столбца и строки заносим точки, количество которых отвечают величинам m_{ki}. Если на корреляционном поле сгустки (центры группировки) смещаются, то можно судить о наличии корреляционной связи между величинами X и Y. На каждом интервале найдем $\bar{x}_{y_k}$ и $\bar{y}_{x_i}$ по формулам (36.1) и (36.2). Нанесем точки $(x_i, \bar{y}_{x_i})$ или $(\bar{x}_{y_k}, y_k)$ и соединим их ломаными линиями.

Ломаная линия $(\bar{y}_x)$, которая соединяет точки с координатами $(x_i, \bar{y}_{x_i})$, называется ***эмпирической линией регрессии*** Y ***на*** X.

Ломаная линия $(\bar{x}_y)$, которая соединяет точки с координатами $(\bar{x}_{y_k}, y_k)$, называется ***эмпирической линией регрессии*** X ***на*** Y.

36.3. Линейная регрессия. Определение параметров линейного уравнения парной регрессии

36.3.1. Выборочное уравнение прямой линии регрессии

Рассмотрим первую основную задачу теории корреляции. Эмпирическая линия регрессии строится на фоне корреляционного поля и показывает, как в среднем изменяется величина Y с увеличением величины X. Если число наблюдений увеличивать, то отдельные точки будут приближаться к линии регрессии.

Предельное положение эмпирической линии регрессии, к которой она приближается при неограниченном числе наблюдений, называется предельной ***теоретической линией регрессии***.

Вид уравнения теоретической линии $\bar{y}_x = f(x)$ или $\bar{x}_y = \varphi(y)$ можно установить по внешнему виду эмпирической линии регрессии, если ее можно сгладить прямой, то – линейная зависимость, если параболой, то – квадратичная зависимость и т. п.

При определении формы корреляционной связи между двумя переменными преимущество отдается линейной зависимости $\hat{y}_x = ax + b$ или $\hat{x}_y = a_1 y + b_1$ по нескольким причинам:

во-первых, простота, а в результате – менее сложные расчеты;

во-вторых, для малых промежутков изменения аргумента криволинейные зависимости можно с достаточной точностью приблизить прямолинейными, то есть вместо одной параболы можно рассматривать две прямые;

в-третьих, если величины X и Y нормально распределены, то форма корреляционной зависимости является линейной.

36.3.2. Определение параметров линейного парного уравнения регрессии

Пусть имеем эмпирические данные:

x	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
$y_{эмп.}$	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n

Построим теоретическое уравнение регрессии в виде:

$$\hat{y}_{теор.} = ax + b. \qquad (36.3)$$

Параметры a и b определим по методу наименьших квадратов так, чтобы сумма квадратов отклонений $\hat{y}_{i\ теор.}$ от $y_{i\ эмп.}$ была наименьшей:

$$\sum_{i=1}^{n} (\hat{y}_{i\ теор.} - y_{i\ эмп.})^2 = \sum_{i=1}^{n} \varepsilon_i^2 \quad (min).$$

Обозначим: $\sum_{i=1}^{n} \varepsilon_i^2 = F(a,b) = \sum_{i=1}^{n} (ax_i + b - y_i)^2 \quad \to min.$

Согласно необходимому условию экстремума функции двух переменных имеем:

$$\frac{\partial F(a,b)}{\partial a} = 0, \quad \frac{\partial F(a,b)}{\partial b} = 0.$$

Получаем систему уравнений:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 2\sum_{i=1}^{n} (ax_i + b - y_i) \cdot x_i = 0, \\ \dfrac{\partial F}{\partial b} = 2\sum_{i=1}^{n} (ax_i + b - y_i) \cdot 1 = 0. \end{cases}$$

Запишем эту систему в виде:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i. \end{cases} \Rightarrow \begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i, \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

Разделим на n оба уравнения системы:

$$\begin{cases} a\dfrac{\sum_{i=1}^{n} x_i^2}{n} + b\dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{n} x_i\, y_i}{n}, \\ a\dfrac{\sum_{i=1}^{n} x_i}{n} + b = \dfrac{\sum_{i=1}^{n} y_i}{n}. \end{cases}$$

Поскольку:

$$\frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}; \quad \frac{\sum_{i=1}^{n} x_i^2}{n} = \overline{x^2}; \quad \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}; \quad \frac{\sum_{i=1}^{n} x_i\, y_i}{n} = \overline{xy}. \qquad (36.4)$$

Получим:

$$\begin{cases} a\overline{x^2} + b\overline{x} = \overline{xy} \\ a\overline{x} + b = \overline{y} \end{cases}.$$

Умножим второе уравнение на $\overline{x}$ и вычтем из первого уравнения:

$$\begin{array}{r} a\overline{x^2} + b\overline{x} = \overline{xy} \\ -\quad a\left(\overline{x}\right)^2 + b\overline{x} = \overline{y}\cdot\overline{x} \\ \hline a\left(\overline{x^2} - \left(\overline{x}\right)^2\right) = \overline{xy} - \overline{x}\cdot\overline{y}. \end{array}$$

Тогда:

$$a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}. \qquad (36.5)$$

Из второго уравнения системы $b=\overline{y}-a\overline{x}$.

Параметр a называется ***коэффициентом регрессии*** Y ***на*** X и обозначается $\rho_{y/x}$. Величина $(\overline{xy}-\overline{x}\cdot\overline{y})$ называется ***ковариацией***, или корреляционным моментом и обозначается μ_{xy}; $\overline{x^2}-(\overline{x})^2=\sigma_x^2$ – дисперсия случайной величины X.

Таким образом:

$$\rho_{y/x}=\frac{\mu_{xy}}{\sigma_x^2}, \qquad b=\overline{y}-\rho_{y/x}\overline{x}. \tag{36.6}$$

Теоретическое уравнение регрессии Y на X имеет вид:

$$\hat{y}=\rho_{y/x}x+b. \tag{36.7}$$

Аналогично можно получить теоретическое уравнение регрессии X на Y:

$$\hat{x}=\rho_{x/y}y+b_1, \tag{36.8}$$

где $\rho_{x/y}=\dfrac{\mu_{xy}}{\sigma_y^2}$, $\sigma_y^2=\overline{y^2}-(\overline{y})^2$, $b_1=\bar{x}-\rho_{x/y}\bar{y}$.

Эмпирическая и теоретическая линии регрессии изображаются на корреляционном поле. Если теоретическое уравнение регрессии вычислено правильно, то ошибки ε_i незначительные. Точка $M(\bar{x},\bar{y})$ – центр регрессии, через который проходят обе теоретические линии регрессии.

Если данные сгруппированы, то по корреляционной таблице вычислим: $\bar{x}=\dfrac{\sum_i x_i m_{x_i}}{n}$; $\bar{y}=\dfrac{\sum_k y_k m_{y_k}}{n}$; $\overline{x^2}=\dfrac{\sum_i x_i^2 m_{x_i}}{n}$; $\overline{xy}=\dfrac{\sum_i\sum_k x_i y_k m_{ik}}{n}$.

Пример 36.2. Составить линейное парное уравнение регрессии $\hat{y}_x=\rho_{y/x}x+b$ по заданной корреляционной таблице, в которой приведены результаты экзаменов по физике (y) и математике (x) в одной из академических групп I курса:

y \ x	2	3	4	5	l_k
2	1	2			3
3	1	6	3		10
4		5	5		10
5			1	1	2
h_i	2	13	9	1	25

Решение.

Вычислим:

$$\bar{x} = \frac{2\cdot 2 + 3\cdot 13 + 4\cdot 9 + 5\cdot 1}{25} = \frac{84}{25} = 3{,}36;$$

$$\bar{y} = \frac{2\cdot 3 + 3\cdot 10 + 4\cdot 10 + 5\cdot 2}{25} = \frac{86}{25} = 3{,}44;$$

$$\overline{x^2} = \frac{4\cdot 2 + 9\cdot 13 + 16\cdot 9 + 25\cdot 1}{25} = \frac{294}{25} = 11{,}76;$$

$$\overline{y^2} = \frac{4\cdot 3 + 9\cdot 10 + 16\cdot 10 + 25\cdot 2}{25} = \frac{312}{25} = 12{,}48;$$

$$\overline{xy} = \frac{\sum_i \sum_k m_{ki} x_i y_k}{n} = \frac{1}{25}\Big(\big(1\cdot(2\cdot 2) + 2\cdot(3\cdot 2)\big) + \big(1\cdot(2\cdot 3) + 6\cdot(3\cdot 3) + 3\cdot(4\cdot 3)\big) +$$

$$+ \big(5\cdot(3\cdot 4) + 5\cdot(4\cdot 4)\big) + \big(1\cdot(4\cdot 5) + 1\cdot(5\cdot 5)\big)\Big) = \frac{1}{25}(16 + 96 + 140 + 45) = \frac{297}{25} = 11{,}88;$$

$$\rho_{y/x} = \frac{11{,}88 - 3{,}36\cdot 3{,}44}{11{,}76 - 3{,}36^2} = \frac{11{,}88 - 11{,}56}{11{,}76 - 11{,}29} = \frac{0{,}32}{0{,}47} \approx 0{,}68;$$

$$\bar{b} = 3{,}44 - 0{,}68\cdot 3{,}36 = 3{,}44 - 2{,}28 = 1{,}16;$$

$$\bar{y}_x = 0{,}68x + 1{,}16.$$

Проверка:

$x=3$;

$$\overline{y}_x=0{,}68\cdot 3+1{,}16=3{,}2;\ \overline{y}_{\text{эмп.}}=\frac{2\cdot 2+3\cdot 6+4\cdot 5}{13}=3{,}23;\ \varepsilon_1=0{,}03.$$

$x=4$;

$$\overline{y}_x=0{,}68\cdot 4+1{,}16=3{,}88;\ \overline{y}_{\text{эмп.}}=\frac{3\cdot 3+4\cdot 5+5\cdot 1}{9}=3{,}78;\ \varepsilon_1=0{,}1.$$

Пример 36.3. Зависимость между переменными величинами X и Y получена на основе эксперимента и представлена следующей таблицей:

x_i	1	1,5	2	2,5	3
y_i	2,15	2,3	2,6	2,8	2,5

В предположении линейной зависимости $\hat{y}_x=\rho_{y/x}x+b$ между X и Y найти коэффициенты $\rho_{y/x}$ и b.

Решение.

Исходные данные и результаты вычислений поместим в таблицу:

n	x_i	y_i	x_i^2	$x_i y_i$
1	1	2,15	1	2,15
2	1,5	2,3	2,25	3,45
3	2	2,6	4	5,2
4	2,5	2,8	6,25	7
5	3	2,5	9	7,5
$\sum$	10	12,35	22,5	25,3

Выполняя вычисления, получим:

$$\overline{x}=\frac{10}{5}=2;\ \overline{y}=\frac{12{,}35}{5}=2{,}47;\ \overline{x^2}=\frac{22{,}5}{5}=4{,}5;\ \overline{xy}=\frac{25{,}3}{5}=5{,}06.$$

$$\rho_{y/x}=0{,}24,\quad b=1{,}99.$$

Искомое уравнение имеет вид: $\overline{y}_x = 0{,}24x + 1{,}99$.

На рис. 36.2 изображены эмпирические точки и найденная теоретическая линия регрессии $\hat{y}_x$.

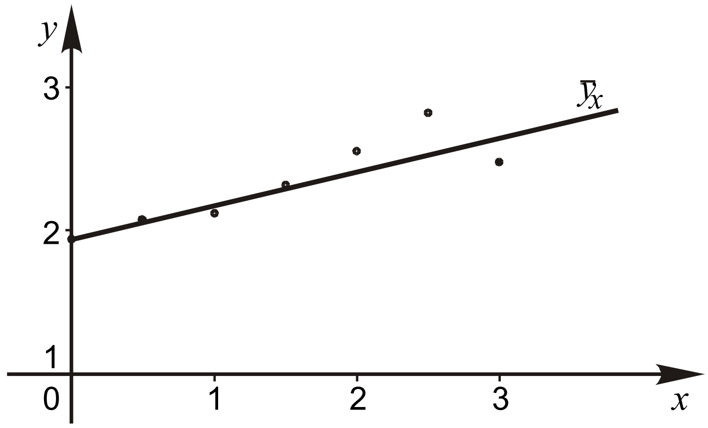


Рис. 36.2. **Теоретическая линия регрессии $\hat{y}_x$ и эмпирические точки**

Проверим уравнение:

$$\overline{y}_{x=1,5} = 0{,}24 \cdot 1{,}5 + 1{,}99 = 2{,}35; \qquad \overline{y}_{x=2,5} = 0{,}24 \cdot 2{,}5 + 1{,}99 = 2{,}6.$$

Отклонение: $\varepsilon_1 = |2{,}35 - 2{,}3| = 0{,}05$; $\varepsilon_2 = |2{,}8 - 2{,}6| = 0{,}2$.

36.3.3. Коэффициент регрессии, его толкование

Коэффициент регрессии по своему знаку совпадает со знаком ковариации: если знак $\rho_{y/x}$ «+», то регрессия положительная, если знак «–», то регрессия отрицательна.

Коэффициент регрессии $\rho_{y/x}$ – величина размерная. Он показывает, на сколько единиц своего измерения увеличится («+») (уменьшится («–»)) среднее значение функционального признака $\hat{y}_x$, если фактор X увеличить на единицу своего измерения.

Действительно, пусть x получило приращение Δx, тогда $\hat{y}_x$ – приращение $\Delta\hat{y}_x$. Новые данные x и $\hat{y}_x$ подставим в уравнение регрессии (36.7) $\hat{y} = \rho_{y/x}x + b$ и получим:

$$\hat{y} + \Delta\hat{y} = \rho_{y/x}(x + \Delta x) + b.$$

Отнимем из нового уравнения (36.7), получим:

$$\Delta\hat{y} = \rho_{y/x}\Delta x, \quad \text{или} \quad \Delta\hat{y}_x = \rho_{y/x}\Delta x.$$

Если $\Delta x = 1$, то $\Delta\hat{y}_x = \rho_{y/x}$.

Пример 36.4. $\hat{y}_x = -4{,}8x + 7$, единицы измерения: X – гривны, а Y – килограммы. Коэффициент $\rho_{y/x} = -4{,}8$ показывает, что с увеличением фактора X на 1 грн признак $\hat{y}_x$ уменьшится на $4{,}8$ кг.

36.4. Теснота связи в парной линейной корреляционной зависимости

36.4.1. Коэффициент корреляции, его свойства и толкование

Поскольку корреляционный момент (ковариация) μ_{xy} – характеристика размерная, то вводят дополнительную безразмерную характеристику – ***коэффициент корреляции*** – показатель тесноты линейной связи:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}. \tag{36.9}$$

Знаки коэффициентов r_{xy}, μ_{xy}, $\rho_{y/x}$, $\rho_{x/y}$ совпадают.

Свойства коэффициента корреляции r_{xy}:

1. Если случайные величины X и Y линейно независимы, то $r_{xy} = 0$ (так как $\mu_{xy} = 0$).

2. Если случайные величины линейно зависимы $y = ax + b$, то $r_{xy} = 1$.

Действительно,

$$xy = ax^2 + bx;\ \overline{xy} = a\overline{x^2} + b\bar{x};\ \bar{y} = a\bar{x} + b;$$

$$\sigma_y^2 = \sigma^2(ax + b) = a^2\sigma_x^2;\ \ \sigma_y = a\sigma_x;$$

$$\mu_{xy} = \overline{xy} - \bar{x}\cdot\bar{y} = a\overline{x^2} + b\bar{x} - \bar{x}(a\bar{x} + b) = a\left(\overline{x^2} - (\bar{x})^2\right) = a\sigma_x^2.$$

То есть $r = \dfrac{a\sigma_x^2}{\sigma_x a \sigma_x} = 1$.

3. $\left|r_{xy}\right| \le 1, \quad -1 \le r \le 1$.

Величина модуля коэффициента корреляции характеризует тесноту линейной корреляционной связи: с ростом $\left|r_{xy}\right|$ линейная корреляционная связь становится теснее.

Чем ближе $\left|r_{xy}\right|$ к единице, тем сильнее линейная корреляционная зависимость между фактором X и признаком Y; чем ближе $\left|r_{xy}\right|$ к нулю, тем линейная корреляционная зависимость слабее.

4. $r_{xy} = r_{yx}$ (по определению).

Приведем формулы для вычисления коэффициента корреляции:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}.$$

Так как $\rho_{y/x} = \dfrac{\mu_{xy}}{\sigma_x^2}$, $\rho_{x/y} = \dfrac{\mu_{xy}}{\sigma_y^2}$, то $\rho_{y/x} \cdot \rho_{x/y} = \dfrac{\mu_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = r^2$.

Следовательно,

$$r_{xy} = \pm\sqrt{\rho_{y/x} \cdot \rho_{x/y}},$$

знак «+», если $\rho_{y/x}$, $\rho_{x/y}$ положительные; знак «–», если $\rho_{y/x}$, $\rho_{x/y}$ отрицательные.

Пример 36.5. Для признаков x_i и y_k при объеме выборки $n = 100$ получено: $\sum x_i h_i = 2296$, $\sum y_k l_k = 269$, $\sum x_i y_k m_{ik} = 6272$, $s_x^2 = 7{,}45$, $s_y^2 = 10{,}21$, где h_i, l_k, m_{ik} – соответствующие частоты. Оценить тесноту связи.

Решение.

Из условия получим: $\bar{x} = 22{,}96$; $\bar{y} = 2{,}69$; $\overline{xy} = 62{,}72$;

$$\mu_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 62{,}72 - 22{,}96 \cdot 2{,}69 = 0{,}9576;$$

Коэффициент корреляции: $r = \dfrac{\mu_{xy}}{s_x \cdot s_y} = \dfrac{0{,}9576}{\sqrt{7{,}45 \cdot 10{,}21}} = 0{,}110.$

Так как число r близко к нулю, значит линейная корреляционная связь очень слабая.

Получим уравнение регрессии в стандартизированных переменных. Запишем уравнение регрессии (36.7) $\hat{y} = \rho_{y/x} x + b$.

Линия регрессии проходит через центр регрессии – точку $M(\overline{x}, \overline{y})$.

Подставим координаты точки M в уравнение регрессии:

$$\overline{y} = \rho_{y/x} \overline{x} + b. \tag{36.10}$$

Вычтем из уравнения (36.7) уравнение (36.10), тогда получим:

$$\hat{y} - \overline{y} = \rho_{y/x} (x - \overline{x}).$$

$$\rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2} = \frac{\mu_{xy}\sigma_y}{\sigma_x^2 \sigma_y} = r \frac{\sigma_y}{\sigma_x},$$

то есть
$$\hat{y} - \overline{y} = r \frac{\sigma_y}{\sigma_x} (x - \overline{x}),$$

или

$$\frac{\hat{y} - \overline{y}}{\sigma_y} = r \frac{(x - \overline{x})}{\sigma_x}. \tag{36.11}$$

Обозначим $\dfrac{\hat{y} - \overline{y}}{\sigma_y} = t_{\hat{y}}$; $\dfrac{x - \overline{x}}{\sigma_x} = t_x$, где $t_{\hat{y}}, t_x$ – стандартизированные переменные.

Тогда уравнение регрессии будет иметь вид: $t_{\hat{y}} = r \cdot t_x$.

Аналогично получим сопряженное уравнение, когда X – функциональный фактор, Y – фактор-аргумент:

$$t_{\hat{x}} = r \cdot t_y.$$

Если в уравнении (36.11) обозначить $\hat{y}-\overline{y}=\Delta\hat{y}; \quad x-\overline{x}=\Delta x$, то получим:

$$\frac{\Delta\hat{y}}{\sigma_y}=r\frac{\Delta x}{\sigma_x}.$$

Если взять в последнем равенстве $\Delta x=\sigma_x$, то получим: $\dfrac{\Delta\hat{y}}{\sigma_y}=r$, откуда $\Delta\hat{y}=r\sigma_y$.

Таким образом, коэффициент корреляции показывает, на какую часть своего среднеквадратического отклонения σ_y изменится среднее значение функционального признака Y, если фактор X увеличится на свое среднеквадратическое отклонение σ_x.

36.4.2. Эмпирическое корреляционное отношение, его свойства

Так как коэффициент корреляции является полноценным показателем тесноты связи лишь в случае линейной зависимости между переменными, то возникает необходимость в достоверном показателе интенсивности связи при любой форме зависимости.

Как известно, рассеяние функционального признака Y характеризует дисперсия:

$$\sigma_y^2=\frac{\sum\limits_k \left(y_k-\overline{y}\right)^2\cdot m_{y_k}}{n}. \tag{36.12}$$

Существует два вида рассеяния:

а) рассеяние, которое обусловлено влиянием фактора X $\left(\sigma_{\overline{y}_x}^2\right)$;

б) рассеяние, которое обусловлено влиянием других факторов, кроме X $\left(\sigma_{y/x}^2\right)$.

Запишем основное дисперсионное уравнение:

$$\sigma_y^2=\sigma_{\overline{y}_x}^2+\sigma_{y/x}^2. \tag{36.13}$$

Если $\sigma_{\overline{y}_x}^2=0$, то рассеяние Y полностью обусловлено влиянием побочных факторов, то есть между фактором X и признаком Y нет корреляционной связи;

если $\sigma^2_{y/x} = 0$, то рассеяние Y обусловлено только фактором X, то есть между X и Y функциональная зависимость.

Разделим обе части дисперсионного уравнения (36.13) на σ^2_y:

$$1 = \frac{\sigma^2_{\bar{y}_x}}{\sigma^2_y} + \frac{\sigma^2_{y/x}}{\sigma^2_y}. \tag{36.14}$$

Слагаемое $\dfrac{\sigma^2_{y/x}}{\sigma^2_y}$ характеризует относительное влияние побочных факторов на функциональный фактор Y. Слагаемое $\dfrac{\sigma^2_{\bar{y}_x}}{\sigma^2_y}$ характеризует относительное влияние фактора X на функциональный фактор Y, то есть характеризует тесноту корреляционной связи между X и Y.

Показателем тесноты связи между фактором X и признаком Y является ***эмпирическое корреляционное отношение***:

$$\eta_{y/x} = \sqrt{\frac{\sigma^2_{\bar{y}_x}}{\sigma^2_y}}, \tag{36.15}$$

где $\sigma^2_{\bar{y}_x} = \dfrac{\sum\limits_i \left(\overline{y_{x_i}} - \bar{y}\right)^2 m_{x_i}}{n} = \dfrac{\sum\limits_i \overline{y^2_{x_i}} m_{x_i}}{n} - \left(\bar{y}\right)^2$.

Величина $\eta^2_{y/x}$ показывает, какая часть полной изменчивости функционального признака Y объясняется влиянием фактора-аргумента X. Например, величина $\eta^2_{y/x} = 0{,}78$ означает, что 78% изменчивости признака Y обусловлено влиянием фактора X и только 22% изменчивости признака Y обусловлено влиянием других неучтенных факторов.

Корреляционные отношения

$$\eta_{y/x} = \frac{\sigma_{\bar{y}_x}}{\sigma_y}, \qquad \eta_{x/y} = \frac{\sigma_{\bar{x}_y}}{\sigma_x} \tag{36.16}$$

являются значениями, полученными по эмпирическим данным.

Величина η характеризует тесноту связи в корреляционной зависимости и имеет такие *свойства*:

1) $\eta \geq 0$ (по определению);

2) величина η принимает значения между нулем и единицей: $0 \leq \eta \leq 1$;

3) если $\eta = 1$, то связь между X и Y функциональная (поскольку $\sigma^2_{y/x} = 0$), то есть отсутствует рассеяние внутри группы, для каждого x_i одно y_i);

4) если $\eta = 0$, то $\sigma_{\bar{y}_x} = 0$, то между X и Y нет корреляционной зависимости;

5) если величина η приближается к нулю, то связь между X и Y слабая;

если величина η приближается к единице, то связь между X и Y сильная.

При парной линейной корреляционной зависимости имеем:

$$\sigma^2_{\bar{y}_x} = D(\rho_{y/x} x + b) = \rho^2_{y/x} D(x) + D(b) = \rho^2_{y/x} \cdot \sigma^2_x.$$

То есть

$$\eta_{y/x} = \left| \frac{\rho_{y/x}\sigma_x}{\sigma_y} \right| = \left| \frac{\mu_{xy}}{\sigma_x \sigma_y} \right| = |r|, \qquad (36.17)$$

где r – коэффициент корреляции.

Так как по определению $r = \dfrac{\mu_{xy}}{\sigma_x \sigma_y}$, то, если $\eta_{y/x} \approx |r|$, то имеет место линейная корреляционная зависимость. Если $\eta >> r$, то зависимость между X и Y нелинейная.

Действительно, если зависимость вида $\hat{y} = ax^2 + bx + c$, тогда

$$D(\hat{y}) = \sigma^2_{\hat{y}} = a^2\sigma^2(x^2) + \sigma^2(bx + c) = a^2\sigma^2(x^2) + \sigma^2_{\bar{y}},$$

где $\sigma^2_{\bar{y}}$ – линейная составляющая;

$a^2\sigma^2(x^2)$ – положительное слагаемое.

Отсюда $\eta_{нелинейн.} = \sqrt{\dfrac{a^2\sigma^2(x^2) + \sigma_{\bar{y}}^2}{\sigma_y^2}}$, то есть $\eta_{нелинейн.} >> |r|$, окончательно $|r| \le \eta \le 1$.

Пример 36.6. Найти корреляционное отношение $\eta_{y/x}$ и пояснить полученый результат, если $\bar{y} = 52{,}8$ и $S_y^2 = 102{,}16$.

$\overline{y_{x_i}}$	30	36,25	50,2	61,25	59,52	70
h_i	3	8	49	16	21	3

Решение.

Для вычисления корреляционного отношения $\eta_{y/x}$ заполним следующую таблицу:

$\overline{y_{x_i}}$	30	36,25	50,2	61,25	59,52	70
h_i	3	8	49	16	21	3
$\overline{y_{x_i}}^2 \cdot h_i$	2700	10512,5	123481,96	60025,0	74395,238	14700,0

$$\sum \overline{(y_{xi})}^2 \cdot h_i = 285814{,}69, \ \sum h_i = 100, \ \overline{(y_{xi})^2} = 2858{,}1469.$$

Следовательно,

$$\eta_{y/x}^2 = \frac{2858{,}1469 - (52{,}8)^2}{102{,}16} = 0{,}688 \text{ и } \eta_{y/x} = \sqrt{0{,}688} = 0{,}829.$$

Так как $\eta_{y/x}^2 = 0{,}688$, то можно сказать, что 68,8 % изменчивости фактора Y зависит непосредственно от изменчивости X.

36.4.3. Взаимное положение сопряженных теоретических линий регрессии

Теоретические линии регрессии $\hat{y}_x$ и $\hat{x}_y$ пересекаются в точке $M(\bar{x}, \bar{y})$. Из уравнений регрессии имеем такие угловые коэффициенты теоретических прямых:

$$\bar{y}_x = \rho_{y/x}x + b \;\Rightarrow\; k_1 = \rho_{y/x}, \qquad \bar{x}_y = \rho_{x/y}y + b_1 \;\Rightarrow\; k_2 = \frac{1}{\rho_{x/y}}.$$

Коэффициент корреляции $r^2 \le 1$, поэтому:

$$\rho_{y/x} \cdot \rho_{x/y} \le 1 \Rightarrow \rho_{y/x} \le 1/\rho_{x/y} \Rightarrow k_1 \le k_2.$$

То есть теоретическая линия регрессии $\hat{x}_y$ проходит круче теоретической линии регрессии $\hat{y}_x$. Угол между этими линиями определяется по формуле: $tg\varphi = \dfrac{k_2 - k_1}{1 + k_1 k_2}$.

Учитывая значение k_1 и k_2, получим:

$$tg\varphi = \frac{(1/\rho_{x/y}) - \rho_{y/x}}{1 + (\rho_{y/x}/\rho_{x/y})} = \frac{1 - \rho_{y/x} \cdot \rho_{x/y}}{\rho_{x/y} + \rho_{y/x}} = \frac{1 - r_{xy}^2}{r_{xy}} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \qquad (36.18)$$

Можно сделать такие выводы (рис. 36.3):

1) если $|r| = 1$, то $tg\varphi = 0$, $\varphi = 0$;

2) если $r = 0$, то $tg\varphi = \infty$, $\varphi = \pi/2$;

3) чем больше значение коэффициента корреляции, тем ближе друг к другу размещены сопряженные теоретические линии регрессии;

4) чем более слабая корреляционная зависимость, тем больше угол между сопряженными линиями регрессии.

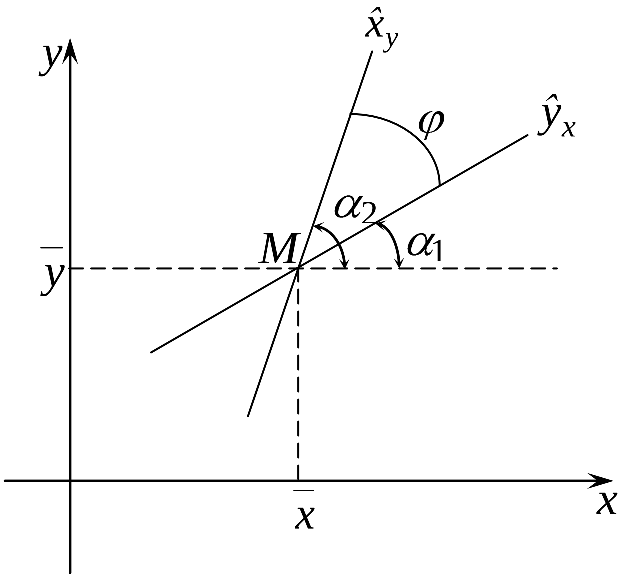


Рис. 36.3. **Взаимное положение теоретических линий регрессии $\hat{y}_x$ и $\hat{x}_y$**

36.4.4. Проверка значимости параметров корреляционной зависимости

1. *Проверка существенности корреляционной зависимости (значимости регрессии).*

Как известно, к эмпирическим показателям корреляции относятся коэффициент корреляции r и корреляционное отношение η. Коэффициент корреляции – величина случайная (если взять несколько выборок, коэффициент корреляции будет разным). Поэтому проверяется гипотеза о его значимости (существенности), то есть существенно ли r отличается от нуля или это отличие можно приписать влиянию случайных факторов.

Если рассматривать большой объем выборки, то можно допустить, что случайная величина r удовлетворяет нормальному закону распределения, тогда имеет место теорема Лапласа $P(|X-a|<t\sigma)=2\Phi(t)$.

Если $X=r$, $a=r_0$, то

$$P(|r-r_0|<t\sigma_r)=2\Phi(t),$$

где n – объем выборки;

$\sigma_r=\sqrt{\dfrac{1-r^2}{n-2}}$ для $n\le 50$;

$\sigma_r=\dfrac{1-r^2}{\sqrt{n}}$ для $n>50$.

Рассмотрим гипотезу H_0: $r_0=0$ – в генеральной совокупности нет корреляционной связи.

Тогда $P(|r-r_0|<t\sigma_r)=2\Phi(t)$, или $P\left(\dfrac{|r|}{\sigma_r}<t\right)=2\Phi(t)$.

Обозначим $\tau=\dfrac{|r|}{\sigma_r}=\dfrac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$ для $n\le 50$ и $\tau=\dfrac{|r|\sqrt{n}}{1-r^2}$ для $n>50$.

Если $\tau\le t$, то с надежностью $2\Phi(t)$ принимается гипотеза H_0 об отсутствии корреляционной зависимости (несущественности линейной связи).

Если $\tau>t$, то с надежностью $2\Phi(t)$ гипотеза H_0 отклоняется, то есть имеется линейная корреляционная связь и она существенна.

На практике для $n > 50$ приняты такие критерии:

если $\tau > 2{,}58$, то с уверенностью 99% можно утверждать, что линейная корреляционная зависимость существенна (значима). Корреляционная зависимость существует не только для данной выборки, а для всей генеральной совокупности;

если $\tau < 1{,}96$, то с уверенностью 95% можно утверждать, что корреляционная зависимость не является существенной (значимой). Корреляционная зависимость характерна только для данной выборки и может не существовать в генеральной совокупности;

если $1{,}96 < \tau < 2{,}58$, это область неопределенности, чаще говорят о несущественности корреляционной зависимости, вывод зависит от существа вопроса.

Если выборка небольшая ($n \leq 50$), то величина τ имеет распределение Стьюдента с $n-2$ степенями свободы. По таблице критических точек распределения Стьюдента (см. приложение Ж) находим $t_{\alpha}(n-2)$.

Если $\tau < t_{0,05}(n-2)$, то принимается гипотеза H_0 об отсутствии корреляционной зависимости.

Если $\tau > t_{0,01}(n-2)$, то гипотеза H_0 отклоняется, то есть имеется корреляционная связь и она существенна.

Область неопределенности: $t_{0,05}(n-2) \leq \tau \leq t_{0,01}(n-2)$.

2. *Проверка линейности корреляционной связи, ее значимости и адекватности модели. Доверительный чехол для регрессии.*

Известно, что если зависимость между нормально распределенными X и Y линейная (гипотеза H_0), то $|r| = \eta$.

Тогда по теореме Лапласа $P\left(\left|\left(\eta - |r|\right) - 0\right| < t\sigma_r\right) = 2\Phi(\,)$, то есть с надежностью $2\Phi(\,)$: $\left|\eta - |r|\right| \leq t\sigma_r$, где $\sigma_r = \sqrt{\dfrac{1-r^2}{n-2}}$ для $n \leq 50$; $\sigma_r = \dfrac{1-r^2}{\sqrt{n}}$ для $n > 50$.

Отсюда доверительный интервал для корреляционного отношения имеет вид:

$$|r| - t\sigma_r \leq \eta \leq |r| + t\sigma_r. \qquad (36.19)$$

Если η, вычисленное по выборке, принадлежит доверительному интервалу, то с надежностью $2\Phi(\ )$ принимается гипотеза H_0, то есть корреляционная связь линейная.

Если η не принадлежит доверительному интервалу, то с уровнем значимости $1-2\Phi(\ )$ гипотеза H_0 отклоняется, то есть корреляционная связь нелинейная.

На практике для $n > 50$ приняты такие критерии:

если η принадлежит интервалу с надежностью $\gamma = 95\%$, $t = 1{,}96$, то гипотеза H_0 принимается, то есть связь линейная;

если η не принадлежит интервалу с надежностью $\gamma = 99\%$, $t = 2{,}58$, то гипотеза H_0 отклоняется, то есть связь нелинейная;

если η не принадлежит интервалу с надежностью $\gamma = 95\%$, а принадлежит интервалу с надежностью $\gamma = 99\%$, то это область неопределенности.

Если выборка небольшая ($n \le 50$), то величину $t = t_\gamma(n-2)$ находят по таблице (см. табл. Е.1 приложения Е).

Пример 36.7. $n = 100$, $r = 0{,}764$, $\eta = 0{,}831$, $\alpha = 0{,}05$.

Исследовать существенность и линейность корреляционной связи.

Решение.

Проверим существенность корреляционной связи:

$$\tau = \frac{0{,}764 \cdot 10}{1-0{,}764^2} = \frac{7{,}64}{0{,}416} \approx 18{,}37.$$

Так как $\tau > 2{,}58$, то регрессия значима, корреляционная зависимость существенна с надежностью 99%.

Проверим линейность корреляционной связи с уровнем значимости $\alpha = 0{,}05$:

$$0{,}764 - 1{,}96\frac{0{,}416}{10} < \eta < 0{,}764 + 1{,}96\frac{0{,}416}{10};$$

$$0{,}682 < \eta < 0{,}846.$$

С надежностью 95% можно утверждать, что модель линейная.

Для проверки *значимости модели регрессии и ее адекватности (соответствия)* наблюдаемым данным можно использовать критерий Фишера, построенный на отношении двух несмещенных оценок дисперсий, который зависит от числа степеней свободы k_1 и k_2.

Значимость парной регрессии (гипотеза H_0) для сгруппированных данных по критерию Фишера определяется с помощью статистики:

$$F_{\eta_{y/x}} = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-k}{k-1}, \tag{36.20}$$

где n – число наблюдений;

k – число интервалов по фактору X;

$k_1 = k-1$, $k_2 = n-k$ – числа степеней свободы.

Расчетное значение статистики Фишера необходимо сравнить с табличными величинами с параметрами $k-1$, $n-k$ при уровне значимости $\alpha = 0{,}05$, $\alpha = 0{,}01$: $F_{0,05}(k-1; n-k)$; $F_{0,01}(k-1; n-k)$ (см. приложения З и И).

Если $F_{\eta_{y/x}} < F_{0,05}$, то регрессия не является значимой, корреляционная зависимости нет, гипотеза H_0 отклоняется;

если $F_{\eta_{y/x}} > F_{0,01}$, то корреляционная зависимость существенна (значима), гипотеза H_0 принимается;

если $F_{0,05} < F_{\eta_{y/x}} < F_{0,01}$, то более правильно считать регрессию незначимой.

Значимость регрессии (гипотеза H_0) можно проверить с помощью параметра F_R:

$$F_R = \frac{r^2}{1-r^2} \cdot \frac{n-2}{1}. \tag{36.21}$$

По специальным таблицам находим величины статистик критерия Фишера: $F_{0,01}(1; n-2)$ и $F_{0,05}(1; n-2)$.

Если $F_R > F_{0,01}$, то регрессия в целом значима, гипотеза H_0 принимается;

если $F_R < F_{0,05}$, то регрессия не является значимой, гипотеза H_0 отклоняется;

если $F_{0,05} < F_R < F_{0,01}$, то это область неопределенности, но более правильно считать регрессию незначимой.

Адекватность модели парной регрессии (гипотеза H_0) для сгруппированных данных по критерию Фишера проверяется с помощью величины F_{Ad}:

$$F_{Ad} = \frac{\eta^2 - r^2}{1-\eta^2} \cdot \frac{n-k}{k-2}. \qquad (36.22)$$

По специальным таблицам находим значения $F_{0,01}(k-2; n-k)$ и $F_{0,05}(k-2; n-k)$.

Если $F_{Ad} > F_{0,01}$, то модель неадекватна, то есть необходимо искать другую форму связи;

если $F_{Ad} < F_{0,05}$, то модель адекватна.

Если модель адекватна, то ее отклонения являются случайными ошибками, некоррелируемыми между собой и распределенными по нормальному закону. Наиболее надежные результаты (с наименьшей случайной ошибкой) определяются поблизости центра области рассеивания экспериментальных данных. По мере удаления от этого центра надежность расчетных значений уменьшается, что обуславливает необходимость определения границ применения корреляционной модели.

С надежностью 95% доверительные границы (границы ***доверительного чехла***) для расчетных значений регрессии вычисляются по формуле:

$$\Delta y = 2\sigma_y \cdot \sqrt{\frac{1-r^2}{n-2}} \cdot \sqrt{1 + \frac{(x-\bar{x})^2}{\sigma_x^2}}. \qquad (36.23)$$

Если в формуле (36.23) обозначить $X = x - \bar{x}$, $Y = \Delta y$, $a = \sigma_x$, $b = 2\sigma_y \cdot \sqrt{\dfrac{1-r^2}{n-2}}$, то получим уравнение сопряженной гиперболы:

$$Y = b\sqrt{1+\frac{X^2}{a^2}}, \text{ или } -\frac{X^2}{a^2}+\frac{Y^2}{b^2}=1.$$

Ширина гиперболической полосы (чехла) на интервале $(-a;a)$ приблизительно одинакова и равна $2b$. Далее границы полосы существенно расширяются, приближаясь к линейным асимптотам $\pm\dfrac{b}{a}X$ – продолжениям диагоналей прямоугольника с вершинами $(\pm a;\pm b)$.

В реальных переменных (x,y) самое узкое место доверительного чехла смещено вправо на $\bar{x}$ и доверительная полоса вытянута вдоль линии регрессии. На интервале $(\bar{x}-\sigma_x, \bar{x}+\sigma_x)$ величина доверительной ошибки $\pm\Delta y$ практически постоянна и равна $\pm 2\sigma_y \cdot \sqrt{\dfrac{1-r^2}{n-2}}$. Если задать эти постоянные границы на графике $\hat{y} = \rho_{y/x}x + b_0$, то получим параллелограмм с вершинами $\left(\bar{x}\pm\sigma_x, \bar{y}\pm 2\sigma_y \cdot \sqrt{\dfrac{1-r^2}{n-2}}\right)$. В этом параллелограмме проводим диагонали, которые продолжим за его пределами. Продолжения диагоналей являются границами 95%-го доверительного чехла для $|x-\bar{x}| > \sigma_x$.

Пример 36.8. Имеются следующие данные: $n = 60$, $k = 9$, $\sigma^2_{\bar{y}_x} = 1{,}598$, $\sigma^2_y = 2{,}383$, $r^2 = 0{,}448$, $\sigma_x = 0{,}254$, $\bar{x} = 1{,}125$, $\bar{y} = 8{,}25$.

Исследовать значимость и адекватность модели регрессии.

С надежностью 95% построить доверительный чехол вдоль теоретической линии регрессии $\hat{y} = 0{,}043x + 7{,}9$.

Решение.

Проверим гипотезу H_0 о значимости модели регрессии.

Используем формулу корреляционного отношения (36.15):

$$\eta^2_{y/x} = \frac{\sigma^2_{\bar{y}_x}}{\sigma^2_y} = \frac{1{,}598}{2{,}383} = 0{,}6706.$$

По формуле (36.20) получаем:

$$F_{y/x} = \frac{0{,}6706}{1-0{,}6706} \cdot \frac{60-9}{9-1} \approx 13{,}19.$$

По таблицам критерия Фишера:

$$F_{0,05}(8;51) = 2{,}10;\ F_{0,01}(8;51) = 2{,}82.$$

Имеем, что $F_{y/x} > F_{0,01}$, то есть корреляционная зависимость существенна (значима), гипотеза H_0 принимается.

Проверим гипотезу H_0 об адекватности модели регрессии.

По формуле (36.22) получаем:

$$F_{Ad} = \frac{0{,}6706-0{,}448}{1-0{,}6706} \cdot \frac{60-9}{9-2} = 4{,}94.$$

По таблицам критерия Фишера:

$$F_{0,05}(7;51) = 2{,}26; F_{0,01}(7;51) = 3{,}14.$$

Имеем: $F_{Ad} > F_{0,01}$, то есть модель не является адекватной, гипотеза H_0 отклоняется.

Для построения доверительного чехла вдоль теоретической линии регрессии с надежностью 95% вычислим:

$$2\sigma_y \cdot \sqrt{\frac{1-r^2}{n-2}} = 2 \cdot \sqrt{2{,}383} \cdot \sqrt{\frac{1-0{,}448}{60-2}} = 0{,}301.$$

Получили вершины параллелограма ($\bar{x} \pm 0{,}254$; $\bar{y} \pm 0{,}301$).

Диагонали этого параллелограмма определяют границы 95%-го доверительного чехла (рис. 36.4).

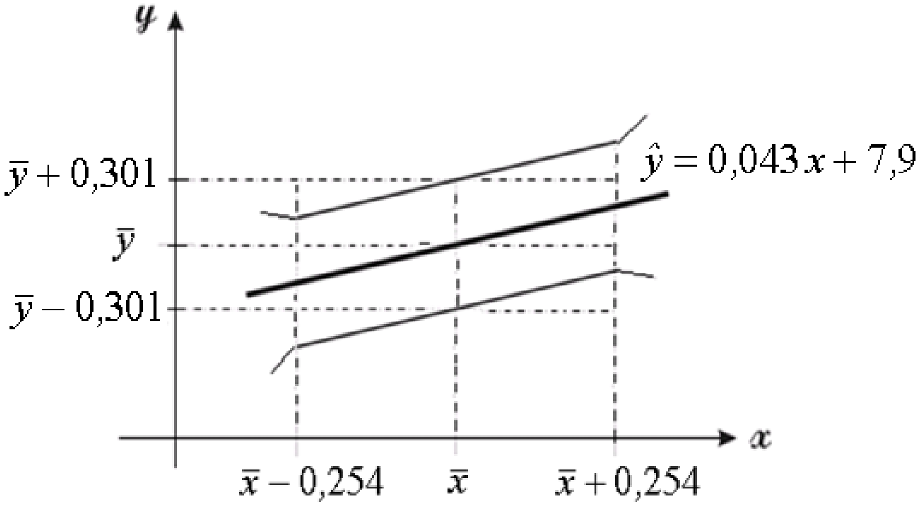


Рис. 36.4. **Доверительный чехол**

Адекватность модели также можно проверить по остаточной дисперсии

$$\sigma^2_{ост} = \frac{1}{n-l}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

где n – число наблюдений;

l – число параметров модели, которые определяются по методу наименьших квадратов ($l=2$ для уравнения парной регрессии, $l=m+1$, m – число факторов);

y_i – эмпирические значения;

$\hat{y}_i$ – теоретические значения.

Чем меньше остаточная дисперсия, тем лучше уравнение регрессии соответствует опытным данным.

Вопросы для самодиагностики

1. Что называется статистической зависимостью? В чем заключаются ее особенности по сравнению с функциональной зависимостью?
2. Что называется корреляционной зависимостью?
3. Как строится корреляционная таблица? Что такое корреляционное поле? Как вычисляются условные средние?
4. Что называется эмпирической линией регрессии?
5. Каковы свойства сопряженных эмпирических линий регрессии?

6. Описать свойства выборочного коэффициента корреляции.

7. Как вычисляется интервальная оценка коэффициента корреляции?

8. Описать свойства выборочного корреляционного отношения.

9. Как находят теоретические линии регрессии?

10. Какие статистические гипотезы используются в регрессионном анализе? Как проверить значимость модели и ее адекватность?

11. Как построить 95 -процентную доверительную полосу (чехол) для выборочной линии регрессии?

Упражнения

36.1. Проведено исследование зависимости себестоимости продукции (y) от затрат на автоматизацию труда (x).

Составить уравнение прямой регрессии Y на X, если $\sum x_i = 54{,}40$; $\sum y_i = 789{,}0$; $n = 20$; $\sum y_i^2 = 33803{,}36$; $\sum x_i y_i = 2322{,}26$; $\sum x_i^2 = 165{,}64$.

36.2. Составить уравнение зависимости y (стоимость основных фондов предприятия) и x (объем валовой продукции), если $r_{yx} = 0{,}68$, $\overline{y} = 250$, $\sigma_y = 38$, $\overline{x} = 47$, $\sigma_x = 12$; $n = 16$, оценить значимость коэффициента корреляции, пояснить результаты.

36.3. Результаты измерений значений двумерной случайной величины приведены в виде корреляционной таблицы:

Y \ X	1	2	3	4	5
4	2	3			
6		4	10	1	
8		3	12	4	
10			3	6	2

Вычислить коэффициент корреляции и проверить существенность корреляционной связи между случайными величинами X и Y.

36.4. Проверить линейную модель на значимость, если $n = 100$; $r_{xy} = 0{,}724$.

36.5. Составить уравнение парной регрессии и оценить тесноту связи по данным $\sum xy = 627{,}3$; $\sum x = 229{,}6$; $\sum y = 269$; $n = 100$; $\sigma_x^2 = 4{,}75$; $\sigma_y^2 = 10{,}208$. Проверить значимость и адекватность модели. Построить доверительный чехол для теоретической регрессии.

36.6. Найти корреляционное отношение $\eta_{y/x}$ и пояснить результат, если $\bar{y} = 33{,}5, \sigma_y^2 = 98{,}75, r = 0{,}67$ и условные средние значения фактора Y для постоянного значения фактора X приведены в таблице:

$\bar{y}_{(X=x_i)}$	11	18,14	30	33,8	42,05	51
m_{x_i}	4	7	10	57	19	3

Проверить линейность корреляционной связи с уровнем значимости $\alpha = 0{,}05$.

36.7. Найти корреляционное отношение $\eta_{y/x}$ и истолковать его, если $\bar{y} = 0{,}36$; $\sum \overline{y_x^2} = 136{,}94$; $s_y^2 = 1{,}985$; $n = 10$.

36.8. Зависимость между переменными X и Y задана корреляционной таблицей сгруппированных данных:

$y_k \backslash x_i$	1	2	3	4	5	$l_к$
1	2					2
2	1	7	2			10
3		2	5	1		8
4			1	2	2	5
h_i	3	9	8	3	2	$n = 25$

Построить поле корреляции и эмпирические линии регрессии.

Составить уравнения линий регрессий X на Y и Y на X.

Вычислить коэффициент корреляции и корреляционное отношение.

Построить доверительный чехол для теоретической регрессии.

Дать пояснение полученным результатам.

Раздел 4. Математическое программирование. Исследование операций

37. Предмет математического программирования

37.1. Задачи оптимального управления

Управление и планирование являются наиболее сложными функциями в работе предприятий, фирм, служб администраций всех уровней.

Задача оптимального управления может быть сформулирована следующим образом:

имеется некоторый объект, состояние которого характеризуется двумя видами параметров – параметрами состояния и параметрами управления, причем в зависимости от выбора последних процесс управления объектом протекает тем или иным образом. Качество процесса управления оценивается с помощью некоторой числовой функции, на основе чего ставится *задача*: найти такую последовательность значений управляющих параметров, для которой данная функция принимает экстремальное значение.

Существуют различные подходы к классификации задач оптимального управления. Прежде всего, их можно классифицировать в зависимости от объекта управления: *задачи управления с* сосредоточенными параметрами (управление самолетом как единым целым); *задачи управления объектами с* распределенными параметрами (управление непрерывным технологическим процессом). В зависимости от типа исходов, к которым приводят применяемые управления, выделяют *детерминированные* и *стохастические* задачи. В последнем случае результатом управления является множество исходов, описываемых вероятностями их наступления. По характеру изменения управляемой системы во времени различают задачи: с *дискретно* изменяющимся *временем*; с *непрерывно* изменяющимся *временем*.

Важной группой математических методов при решении задач управления в экономике являются *оптимизационные методы*, поскольку перед менеджерами, экономистами, работниками системы управления, инженерами разного уровня возникает проблема принятия решения.

Примерами оптимизационных задач являются задачи оптимального планирования, в которых выделяют *переменные* и *параметры* (количество продуктов, которое покупается, количество произведенной продукции, количество груза, который перевозится), цель, которую необходимо достичь (*функция цели*), что следует оптимизировать (минимизировать расходы на потребление, максимизировать прибыль предприятия, минимизировать стоимость перевозок) и *ограничения*, то есть условия, которые ограничивают возможности достижения цели (в рационе должны быть определены компоненты, ограниченные ресурсы предприятия, количество товара, который перевозится). Целевая функция имеет смысл ожидаемой «ценности», или «полезности». Задача оптимального планирования также называется оптимизационной, или экстремальной задачей.

37.2. Общая постановка оптимизационной задачи, ее структура: целевая функция, ограничения как способ описания множества допустимых решений

В задачах оптимизации должны быть выделены характеристики объекта (объектов), которыми можно или нужно варьировать для достижения цели. Такие характеристики называются ***управляемыми переменными***, или ***управляемыми параметрами***. Существуют также ***неуправляемые переменные***, изменение значений которых не зависит от управляющего субъекта.

Набор значений управляемых переменных в задачах оптимизации называется ***решением***. Значения управляемых переменных могут быть ограниченными. Решение, которое удовлетворяет выдвинутым ограничениям, называют ***допустимым решением (планом)***. ***Оптимальным*** называется допустимое решение, которое наилучшим образом соответствует поставленной задаче, например решение, при котором целевая функция экстремальна.

Оптимизационная модель состоит из переменных, ограничений, которые на них налагаются, и формулировки цели. Цель определяет целевую функцию, которая задается на множестве допустимых решений D. Множество D выражает меру реализации цели: если D пустое, то решения не существует; если D содержит одну точку, то эта точка является единственным допустимым решением задачи и такая задача не представляет интерес для исследования; если D содержит больше чем одно решение, тогда задача оптимизации предусматривает выделение оптимального решения из множества допустимых решений. При этом, если D конечное, то оптимальное решение может быть найдено в результате простого перебора всех точек D и вычисление в них функции цели. Если D бесчисленное или является континуумом, то оптимальное решение следует искать на бесконечном множестве допустимых решений.

Математическую модель задачи оптимизации записывают так:

найти точку $X = (x_1, x_2, ..., x_n)$, в которой ***функция цели*** $f(x_1, x_2, ..., x_n)$ достигнет своего максимума (минимума) при условиях (***ограничениях***):

$$g_i(x_1, x_2, ..., x_n) \le (\ge, =) b_i,$$
$$x_j \ge 0,\ j = \overline{1,n},\ i = \overline{1,m},$$

где $f(X), g_i(X)$ – некоторые функции;

$X = (x_1, x_2, ..., x_n)$ – набор параметров управления.

Раздел математики, занимающийся изучением экстремальных (оптимизационных) задач управления, планирования и разработкой методов их решения, получил название ***математического программирования***.

В зависимости от вида функции цели и ограничений математическое программирование делится на линейное и нелинейное. Если функция цели и система ограничений является линейной, то говорят о *линейном программировании*, в противоположном случае, то есть, если есть нелинейность или в функции цели, или в системе ограничений, или и там, и там, то возникает задача *нелинейного программирования*.

В задачах линейного программирования возможны случаи, когда параметры управления могут принимать лишь целые дискретные значения. При решении подобных задач используется *целочисленное программирование*. В некоторых случаях исходные параметры задачи могут изменяться в некоторых пределах, для их решения применяется *параметрическое программирование*.

В случае квадратичной функции цели и линейной системы ограничений или наоборот задачу оптимизации называют задачей *квадратичного программирования*.

Если функцию цели можно представить в виде суммы таких функций, каждая из которых зависит только от одной переменной, то рассматривают задачу *сепарабельного программирования*.

По информационным характеристикам оптимизационные задачи разделяются на статические и динамические. Если субъект в процессе принятия решения не изменяет свое информационное состояние, то принятие решения рассматривается как мгновенный факт и такие задачи называются *статическими*. Если субъект в процессе принятия решения изменяет свое информационное состояние, то решение принимается поэтапно и задача называется *динамической*.

На практике часто приходится сталкиваться с ситуациями, в которых необходимо принимать решения при наличии двух или более сторон, имеющих различные цели. Результаты любого действия каждой из сторон зависят от решений партнеров. В экономике подобные ситуации встречаются довольно часто. Для решения задач с конфликтными ситуациями используют математические методы *теории игр*.

Сетевые модели, в основе которых лежит теория графов, позволяют проводить их оптимизацию, а также совокупность расчетных и организационных мероприятий по управлению комплексами работ при создании новых изделий и технологий.

Цель изучения *системы массового обслуживания* состоит в том, чтобы контролировать ее характеристики для проведения оптимизации системы в целом. Рассмотрение моделей *управления запасами* преследует цель выбора для предприятий оптимальных расходов на доставку, хранение комплектующих материалов и ресурсов, необходимых для изготовления изделий.

Кроме приведенных моделей, к оптимизационным моделям иногда относят имитационные модели и эвристические. *Имитационные* модели позволяют имитировать поведение сложных систем, для которых невозможно построить математические модели и получить решение. Здесь возникают сложности, связанные с разработкой эксперимента, проверкой статистической значимости его результатов, а также сам процесс оптимизации является сложным. Если задача оптимизации очень сложная, однако есть предположение, что решение существует, то используют различные *эвристические* методы, которые основаны на интуиции исследователя или на эмпирических данных.

37.3. Примеры моделей задач математического программирования в экономике

Простейшая задача производственного планирования

Пусть имеется некоторый экономический объект (предприятие, цех, бригада и т. п.), который может производить некоторую продукцию n видов. В процессе производства допустимо использование m видов ресурсов (сырья). Применяемые технологии характеризуются нормами затрат единицы сырья на единицу производимого продукта.

Обозначим через a_{ij} количество i-го ресурса ($i = \overline{1,m}$), которое необходимо для производства единицы j-го продукта ($j = \overline{1,n}$).

Весь набор технологических затрат в производстве j-го продукта можно представить в виде вектора-столбца:

$$A_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix},$$

а технологию рассматриваемого предприятия (объекта) – в виде прямоугольной матрицы размерности $m \times n$:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix}.$$

Если j-й продукт производится в количестве x_j, то в рамках описанных выше технологий мы должны потратить $a_{1j}x_j$ первого ресурса, $a_{2j}x_j$ – второго и т. д., $a_{mj}x_j$ – m-го.

Сводный план производства по всем продуктам может быть представлен в виде n-мерного вектора-строки $X = (x_1, x_2, \dots, x_j, \dots, x_n)$.

Тогда общие затраты по i-му ресурсу на производство всех продуктов можно выразить в виде суммы $\sum_{j=1}^{n} a_{ij}x_j$, представляющей собой скалярное произведение векторов A_j и X.

Очевидно, что всякая реальная производственная система имеет ограничения на ресурсы, которые она тратит в процессе производства. В рамках излагаемой модели эти ограничения порождаются m-мерным вектором $b = (b_1, b_2, ..., b_m)$, где b_i – максимальное количество i-го ресурса, которое можно потратить в производственном процессе.

В математической форме данные ограничения представляются в виде системы m неравенств:

$$\begin{aligned}
&a_{11}x_1 + a_{12}\, x_2 + ... + a_{1n}\, x_n \le b_1, \\
&a_{21}x_1 + a_{22}\, x_2 + ... + a_{2n}\, x_n \le b_2, \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&a_{m1}x_1 + a_{m2}\, x_2 + ... + a_{mn}\, x_n \le b_m.
\end{aligned} \tag{37.1}$$

Применяя правила матричной алгебры, систему (37.1) можно записать в краткой форме, представив левую часть как произведение матрицы A на вектор X, а правую – как матрицу b:

$$AX \le b. \tag{37.2}$$

К системе (37.2) также должны быть добавлены естественные ограничения на неотрицательность компонент плана производства: $x_1 \ge 0$, $x_2 \ge 0, ..., x_j \ge 0, ..., x_n \ge 0$, или, что то же самое,

$$X \ge 0. \tag{37.3}$$

Обозначив через c_j прибыль от реализации единицы j-го продукта, получим выражение суммарного дохода от выполнения плана производства, задаваемого вектором X:

$$Z = \sum_{j=1}^{n} c_j x_j = cX. \tag{37.4}$$

Формулы (37.2) – (37.4) являются простейшей математической моделью, описывающей отдельные стороны функционирования некоторого экономического объекта, поведением которого мы хотим управлять.

В рамках данной модели можно поставить различные задачи, но, скорее всего, самой «естественной» будет задача поиска такого плана производства $X \in R^n$, который дает наибольшее значение суммарного дохода, то есть функции (37.4), и одновременно удовлетворяет системе ограничений (37.2) – (37.3).

Кратко такую задачу можно записать в следующем виде:

$$Z = cX \to \max, \text{ где } X \in D = \{X \in R^n | AX \le b, X \ge 0\}. \quad (37.5)$$

Управление портфелем активов

Рассмотрим проблему принятия инвестором решения о вложении имеющегося у него капитала. Набор характеристик потенциальных объектов для инвестирования, имеющих условные имена от A до F, задается в табл. 37.1.

Таблица 37.1

Характеристики объектов инвестирования

Название	Доходность (в %)	Срок выкупа (год)	Надежность (в баллах)
A	5,5	2010	5
B	6,0	2016	4
C	8,0	2019	2
D	7,5	2011	3
E	5,5	2009	5
F	7,0	2012	4

Предположим, что при принятии решения о приобретении активов должны быть соблюдены условия:

a) суммарный объем капитала, который должен быть вложен, составляет $100 000;

b) доля средств, вложенная в один объект, не может превышать четверти от всего объема;

c) более половины всех средств должны быть вложены в долгосрочные активы (допустим, на рассматриваемый момент к таковым относятся активы со сроком погашения после 2015 г.);

d) доля активов, имеющих надежность менее чем 4 балла, не может превышать трети от суммарного объема.

Приступим к составлению экономико-математической модели для данной ситуации. Целесообразно начать процесс с определения структуры управляемых переменных. В рассматриваемом примере в качестве таких переменных выступают объемы средств, вложенные в активы той или иной фирмы. Обозначим их как $x_A, x_B, x_C, x_D, x_E, x_F$.

Тогда суммарная прибыль от размещенных активов, которую получит инвестор, может быть представлена в виде:

$$P = 0{,}055x_A + 0{,}06\,x_B + 0{,}08x_C + 0{,}075x_D + 0{,}055x_E + 0{,}07x_F. \quad (37.6)$$

На следующем этапе моделирования нужно формально описать перечисленные выше ограничения (a – d) на структуру портфеля:

a) ограничение на суммарный объем активов:

$$x_A + x_B + x_C + x_D + x_E + x_F \le 100000; \quad (37.7)$$

b) ограничение на размер доли каждого актива:

$$\begin{aligned} &x_A \le 25000, x_B \le 25000, x_C \le 25000, \\ &x_D \le 25000, x_E \le 25000, x_F \le 25000; \end{aligned} \quad (37.8)$$

c) ограничение, связанное с необходимостью вкладывать половину средств в долгосрочные активы:

$$x_B + x_C \ge 50000; \quad (37.9)$$

d) ограничение на долю ненадежных активов:

$$x_C + x_D \le 30000. \quad (37.10)$$

Наконец, система ограничений в соответствии с экономическим смыслом задачи должна быть дополнена условиями неотрицательности для искомых переменных:

$$x_A \ge 0, x_B \ge 0, x_C \ge 0, x_D \ge 0, x_E \ge 0, x_F \ge 0. \quad (37.11)$$

Выражения (37.6) – (37.11) образуют математическую модель поведения инвестора. В рамках этой модели может быть поставлена задача поиска таких значений переменных $x_A, x_B, x_C, x_D, x_E, x_F$, при которых достигается наибольшее значение прибыли (37.6) и одновременно выполняются ограничения на структуру портфеля активов (37.7) – (37.11).

Транспортная задача

Рассмотрим проблему организации перевозки некоторого продукта между пунктами его производства, количество которых равно m, и n пунктами потребления. Каждый i-й пункт производства ($i=\overline{1,m}$) характеризуется запасом продукта $a_i \ge 0$, а каждый j-й пункт потребления ($j=\overline{1,n}$) – потребностью в продукте $b_j \ge 0$. Сеть дорог, соединяющая систему рассматриваемых пунктов, моделируется с помощью матрицы C размерности $m \times n$, элементы которой c_{ij} представляют собой нормы затрат на перевозку единицы груза из пункта производства i в пункт потребления j. План перевозки груза в данной транспортной сети представляется в виде массива элементов размерности $m \times n$:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{ij} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mj} & \dots & x_{mn} \end{pmatrix}. \qquad (37.12)$$

Если реальная перевозка между пунктами i и j отсутствует, то полагают $x_{ij} = 0$.

Ограничения на возможные значения $X \in R^{mn}$ имеют вид:

1. Ограничение на удовлетворение потребностей во всех пунктах потребления:

$$\sum_{i=1}^{m} x_{ij} \ge b_j, \; j=\overline{1,n}. \qquad (37.13)$$

2. Ограничения на возможности вывоза запасов из всех пунктов производства:

$$\sum_{j=1}^{n} x_{ij} \le a_i, \; i=\overline{1,m}. \qquad (37.14)$$

3. Условия неотрицательности компонент вектора плана:

$$x_{ij} \geq 0, i = \overline{1,m}, \ j = \overline{1,n}. \tag{37.15}$$

Существенной характеристикой описываемой модели является соотношение параметров a_i и b_j. Если суммарный объем производства равен суммарному объему потребления, а именно $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то система называется *сбалансированной* (*закрытой*). При выполнении условия сбалансированности разумно накладывать такие ограничения на суммарный ввоз и вывоз груза, при которых полностью вывозится весь груз и не остается неудовлетворенных потребностей, то есть условия (37.13) и (37.14) приобретают форму равенств.

По аналогии с задачей производственного планирования предположим, что затраты на перевозку прямо пропорциональны количеству перевозимого груза. Тогда суммарные затраты на перевозку в системе примут вид:

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}. \tag{37.16}$$

Функция (37.16) и описанные выше ограничения, записанные в форме

$$D = \left\{ X \in R^{mn} \middle| \sum_{i=1}^{m} x_{ij} = a_i, i = \overline{1,m}, \ \sum_{j=1}^{n} x_{ij} = b_j, \ j = \overline{1,n}, \ X \geq 0 \right\}, \tag{37.17}$$

задают *транспортную модель*. На ее основе может быть сформулирована задача минимизации суммарных затрат на перевозки:

$$Z = cX \to \min, x \in D, \tag{37.18}$$

которая имеет название *транспортной задачи в матричной форме*.

Вопросы для самодиагностики

1. Как классифицируют задачи оптимального управления?
2. Сформулировать общую постановку оптимизационной задачи.
3. Привести примеры экономических оптимизационных задач.

38. Линейное программирование

38.1. Общая постановка задачи линейного программирования. Экономические примеры использования моделей линейного программирования

Линейное программирование – наука о методах исследования и отыскания экстремальных (наибольших и наименьших) значений линейной функции, на неизвестные которой наложены линейные ограничения.

Эта линейная функция называется ***функцией цели***, а ограничения, которые математически записываются в виде уравнений или неравенств, называются ***системой ограничений***.

Математическое выражение функции цели и ее ограничений называется ***математической моделью экономической задачи***.

В общем виде математическая модель задачи линейного программирования (ЗЛП) записывается как

$$Z = \sum_{j=1}^{n} c_j x_j \to \max\ (\min)$$

при ограничениях:

$$\begin{cases} a_{11}x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n \le (\ge, =)\, b_1, \\ a_{21}x_1 + a_{22}\,x_2 + \ldots + a_{2n}\,x_n \le (\ge, =)\, b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}\,x_2 + \ldots + a_{mn}\,x_n \le (\ge, =)\, b_m, \\ x_j \ge 0, i = \overline{1,m},\ j = \overline{1,n}, \end{cases} \quad \text{или} \quad \begin{cases} \sum_{j=1}^{n} a_{ij}x_j \le (\ge, =)\, b_i, \\ x_j \ge 0, i = \overline{1,m},\ j = \overline{1,n}, \end{cases}$$

где x_j – неизвестные;

a_{ij}, b_i, c_j – заданные постоянные величины.

Математическая модель ЗЛП может быть задана в канонической и неканонической формах. Если все ограничения системы заданы уравнениями, то такая форма модели задачи называется ***канонической.*** Если хотя бы одно ограничение является неравенством, то форма модели задачи ЛП является ***неканонической***.

Чтобы перейти от неканонической формы модели к канонической, необходимо в каждое неравенство ввести ***балансовую*** (дополнительную) переменную x_{n+i}. Если знак неравенства ≤, то балансовая переменная вводится со знаком плюс, если знак неравенства ≥, то – минус. В целевую функцию балансовые переменные не вводятся.

Чтобы составить математическую модель ЗЛП, необходимо:

ввести обозначения переменных;

исходя из цели экономических исследований, составить целевую функцию;

учитывая ограничения в использовании экономических показателей задачи и их количественные закономерности, записать систему ограничений.

Рассмотрим основные экономические задачи, демонстрирующие использование модели линейного программирования.

1. Задача планирования производства: необходимо определить план производства одного или нескольких видов продукции, который обеспечивает наиболее рациональное использование имеющихся материальных, финансовых и других видов ресурсов. Такой план должен быть оптимальным с точки зрения выбранного критерия – максимума прибыли, минимума затрат на производство и т. д.

2. *Задача о смесях*: необходимо выбрать наилучший способ смешения исходных ингредиентов для получения смеси с заданными свойствами. Смесь должна иметь требуемые свойства, которые определяются количеством компонентов, входящих в состав исходных ингредиентов. Как правило, известны стоимостные характеристики ингредиентов и искомую смесь требуется получить с наименьшими затратами. Для многопродуктовых задач, в которых требуется получить несколько смесей, характерным является критерий минимизации затрат. Задачи оптимального смешения встречаются во многих отраслях промышленности (металлургия, парфюмерия, пищевая промышленность, фармакология, сельское хозяйство). Примерами задач о смесях могут служить определение кормового рациона скота на животноводческих фермах, составление рецептуры шихты на металлургическом производстве и т. п.

3. Задача оптимального раскроя: необходимо выбрать один или несколько способов раскроя материала и определить, какое количество материала следует раскраивать, применяя каждый из выбранных способов. Задачи такого типа возникают в металлургии и машиностроении, лесной, лесообрабатывающей, легкой промышленности. Выделяют два этапа решения задачи оптимального раскроя. На первом этапе определяются рациональные способы раскроя материала, на втором – решается задача линейного программирования для определения интенсивности использования рациональных способов раскроя.

4. *Задача планирования финансов*: требуется выбрать такие способы вложения денег под проценты, совокупность которых позволяет минимизировать первоначальный вклад, необходимый для выплаты займа, или максимизировать доход. При решении задач финансового планирования можно учитывать риск и другие факторы, влияющие на выбор способов вложения денег.

Различают две модели задачи планирования финансов.

Модель А минимизации целевого фонда. Предположим, что в определенные моменты времени необходимо выплачивать известные суммы денег по взятому ранее займу. Чтобы накопить эти суммы, можно заранее создать целевой фонд, а средства из этого фонда использовать для срочных вкладов. Каждый срочный вклад характеризуется моментом времени вложения, сроком погашения и доходностью. Задача состоит в том, чтобы определить минимальный размер целевого фонда и выбрать те виды срочных вкладов, которые следует использовать, чтобы сделать выплату по займу.

Модель В максимизации дохода. Предположим теперь, что вкладчик собирается делать вклады для того, чтобы через определенный период времени получить максимальный доход. Задача состоит в том, чтобы определить величину максимального дохода при фиксированном размере целевого фонда и выбрать те виды срочных вкладов, которые следует использовать.

5. *Транспортная задача*: необходимо перевезти продукцию, которая в определенных количествах предлагается различными производителями. Известны потребности нескольких потребителей в этой продукции. Требуется определить, от каких производителей и в каких объемах должны получать продукцию потребители. Поставки должны осуществляться таким образом, чтобы совокупные издержки на транспортировку продукции были минимальными.

6. *Задача о назначениях*: назначения исполнителей на различные виды работ, например: подбор кадров и назначение кандидатов на вакантные должности, распределение источников капитальных вложений

между различными проектами научно-технического развития, распределение экипажей самолетов между авиалиниями. При этом требуется так распределить кадры (средства или виды работ), чтобы общие затраты на выполнение были минимальными.

38.2. Графический метод решения задач линейного программирования

Допустимым решением (планом) ЗЛП называется вектор $\overline{X} = (x_1, x_2, ..., x_j, ..., x_n)$, удовлетворяющий системе ограничений.

Множество допустимых решений образует ***область допустимых решений*** (ОДР).

Опорным решением (планом) называется решение, которое удовлетворяет условиям неотрицательности.

Допустимое решение, при котором целевая функция достигает своего экстремального значения, называется ***оптимальным решением*** (оптимальным планом) ЗЛП и обозначается $\overline{X}_{опт}$.

Наиболее простым и наглядным методом линейного программирования является *графический метод*. Он применяется для решения ЗЛП с двумя переменными, заданными в неканонической форме, и с многими переменными в канонической форме при условии, что они содержат не более двух свободных переменных ($n-m=2$, где n – количество переменных, а m – ранг матрицы системы ограничений).

Теоретической основой графического метода являются некоторые понятия аналитической геометрии. Множество точек n-мерного пространства, содержащее вместе с любыми двумя точками A и B и их выпуклую линейную комбинацию, то есть все точки $C = \lambda_1 A + \lambda_2 B$ *(*$\lambda_1 + \lambda_2 = 1$*)* отрезка AB*,* называется ***выпуклым телом (областью****,* ***фигурой).*** Примеры плоских выпуклых фигур приведены на рис. 38.1.

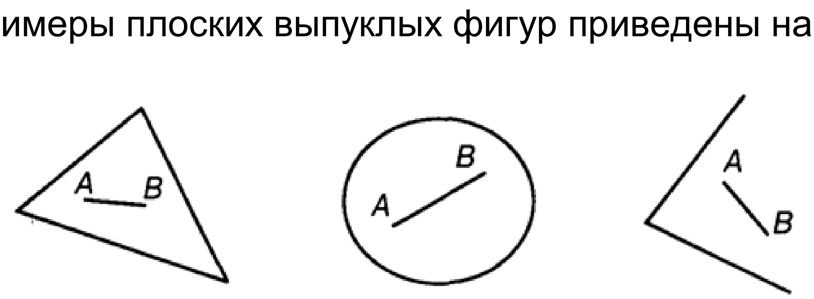


Рис. 38.1. **Плоские выпуклые фигуры**

Примеры невыпуклых фигур приведены на рис. 38.2.

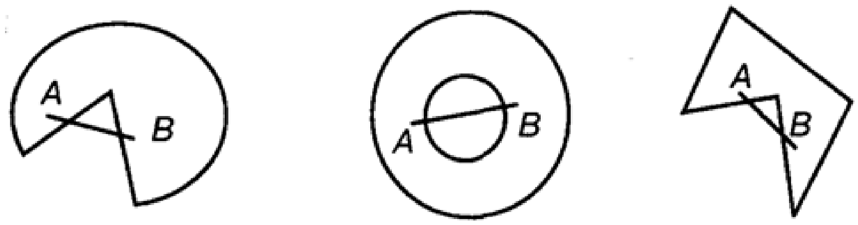


Рис. 38.2. **Плоские невыпуклые фигуры**

Точка A называется ***внутренней точкой выпуклой области***, если в сколь угодно малой окрестности этой точки содержатся только точки этой области (рис. 38.3).

Точка B называется ***граничной точкой выпуклой области***, если в сколь угодно малой окрестности этой точки содержатся как точки данной области, так и не принадлежащие ей (рис. 38.3).

Точка C называется ***угловой точкой выпуклой области***, если она является граничной и не лежит внутри отрезка, соединяющего две другие точки этой области (рис. 38.3).

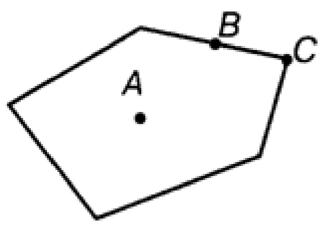


Рис. 38.3. **Типы точек выпуклой области**

Если область включает все свои граничные точки, то она называется ***замкнутой.***

Выпуклая область может быть ограниченной и неограниченной.

Ограниченной называется область, если существует такое число $M > 0$, что вектор $\overline{x}$, соединяющий любые две точки области, по абсолютной величине меньше M, то есть $|\overline{x}| \le M$. Для этой области все ее точки находятся на конечном расстоянии от начала координат. Если найдутся точки области, сколь угодно удаленные от начала координат, то область называется ***неограниченной***.

Выпуклая замкнутая ограниченная область, имеющая конечное число угловых точек, называется выпуклым n***-мерным многогранни-***

ком. На плоскости это выпуклый многоугольник. ***Пересечением выпуклых областей*** называется множество точек, являющееся общей частью этих областей.

Теорема 1. Пересечение выпуклых областей есть выпуклая область.

Теорема 2. Множество точек выпуклого n-мерного многогранника совпадает с множеством любых выпуклых линейных комбинаций его угловых точек.

Построение многоугольника планов

Известно, что любое уравнение $a_1x_1 + a_2x_2 = b$ с двумя переменными геометрически изображает прямую линию на плоскости в декартовой системе координат.

Каждое неравенство $a_1x_1 + a_2x_2 < b$ с двумя переменными геометрически определяет полуплоскость в декартовой системе с граничной прямой $a_1x_1 + a_2x_2 = b$, вторую часть полуплоскости определяет неравенство $a_1x_1 + a_2x_2 > b$.

Для того чтобы определить размещение соответствующих полуплоскостей относительно граничной прямой, надо подставить координаты любой точки (проще всего взять начало координат) в левую часть неравенства. Если координаты начала координат или выбранной точки удовлетворяют неравенству, то соответствующая полуплоскость включает начало координат или выбранную точку. Если координаты начала координат или выбранной точки не удовлетворяют неравенству, то соответствующая полуплоскость размещена по отношению к граничной по другую сторону, чем начало координат или выбранная точка. Общая часть плоскости, определяемая всеми ограничениями, называется ***многоугольником планов. Опорной прямой*** к многоугольнику планов называется прямая, имеющая с многоугольником, расположенным по одну сторону от нее, хотя бы одну общую точку.

Теорема 3. Если функция цели $F(X)$ принимает максимальное значение в некоторой точке множества допустимых планов D, то она принимает это значение в угловой точке данного множества.

Теорема 4. Если функция цели $F(X)$ принимает максимальное значение в нескольких точках множества D (например, в точках A_1, A_2), то она принимает это же значение в любой точке, являющейся их выпуклой линейной комбинацией, то есть $A=\lambda_1 A_1+\lambda_2 A_2$, где $\lambda_1+\lambda_2=1$.

С геометрической точки зрения в ЗЛП требуется найти такую угловую точку или набор угловых точек из множества допустимых решений, в которых достигается экстремальное значение функции цели.

Для нахождения экстремального значения целевой функции $F(X)$ при графическом решении ЗЛП используют ***вектор-градиент*** $\overline{grad}\,F(X)$ на плоскости X_1OX_2, который обозначим $\overline{N}$. Этот вектор показывает направление наискорейшего изменения целевой функции, он равен:

$$\overline{grad}\,F(X)=\left(\frac{\partial F}{\partial x_1},\frac{\partial F}{\partial x_2}\right).$$

Координатами вектора $\overline{N}$ являются коэффициенты целевой функции $F(X)$. Перпендикулярно вектору $\overline{N}$ проходит ***линия уровня***, которую перемещают вдоль вектора $\overline{N}$, параллельно себе, пока она не станет опорной к многоугольнику планов.

Алгоритм решения ЗЛП графическим методом

1. Находим область допустимых решений системы ограничений задачи.

2. Строим вектор-градиент $\overline{N}$.

3. Проводим линию уровня L_0, которая перпендикулярна вектору $\overline{N}$.

4. Линию уровня перемещаем по направлению вектора $\overline{N}$ параллельно себе пока она не станет опорной к многоугольнику планов. Точка входа в область – точка минимума функции, точка выхода из области – точка максимума.

Если только одна общая точка с областью допустимых решений, то эта точка будет точкой экстремума и решение ЗЛП единственное.

Если окажется, что линия уровня проходит через одну из сторон ОДР, то в таком случае экстремум достигается во всех точках соответствующей стороны, а ЗЛП будет иметь бесчисленное множество решений. Говорят, что такая ЗЛП имеет ***альтернативный оптимум***, и ее решение находят в виде выпуклой линейной комбинации решений $\overline{X}_1^*, \overline{X}_2^*$:

$$\overline{X}_{опт} = \lambda_1 X_1^* + \lambda_2 X_2^*,$$

где $\lambda_1 + \lambda_2 = 1.$

ЗЛП может быть неразрешима, когда определяющие ее ограничения окажутся противоречивыми (то есть область допустимых решений окажется пустым множеством).

5. Находим координаты точки экстремума и значение целевой функции в этой точке.

Пример 38.1. Найти максимум функции $Z = 5x_1 + 4x_2$, если

$$\begin{cases} x_1 + x_2 \ge 10, \\ -3x_1 + 2x_2 \le 6, \\ -3x_1 + 10x_2 \ge 30, \\ 3x_1 + 5x_2 \le 60, \\ x_1 \ge 0, \\ x_2 \ge 0. \end{cases}$$

Решение.

Определим вначале многоугольник решений – ОДР.

Для этого построим прямые, которые ограничивают этот многоугольник:

I: $x_1 + x_2 = 10.$ II: $-3x_1 + 2x_2 = 6.$

III: $-3x_1 + 10x_2 = 30.$ IV: $3x_1 + 5x_2 = 60.$

Каждая прямая отсекает отрезки на осях координат: I – (10;10); II – (-2;3); III – (-10;3); IV – (20;12).

Проверим справедливость каждого неравенства по одной точке (начало координат).

Рассмотрим первое неравенство: $x_1 + x_2 \ge 10$. Начало координат (0;0) не удовлетворяет этому неравенству, поэтому область решения лежит по другую сторону от прямой I; это показано на рис. 38.4 стрелками. Для второго неравенства $-3x_1 + 2x_2 \le 6$ начало координат (0;0) удовлетворяет этому неравенству, поэтому область решения лежит по ту же сторону от прямой II, что и начало координат.

Рассмотрим неравенство $-3x_1 + 10x_2 \geq 30$. Начало координат (0;0) не удовлетворяет этому неравенству, поэтому область решения лежит по другую сторону от прямой III.

Рассмотрим неравенство $3x_1 + 5x_2 \leq 60$. Начало координат (0;0) удовлетворяет этому неравенству, поэтому область решения лежит по ту же сторону от прямой IV, что и начало координат.

Замечание: если ограничивающая прямая проходит через начало координат, то вместо точки (0;0) можно взять любую другую точку.

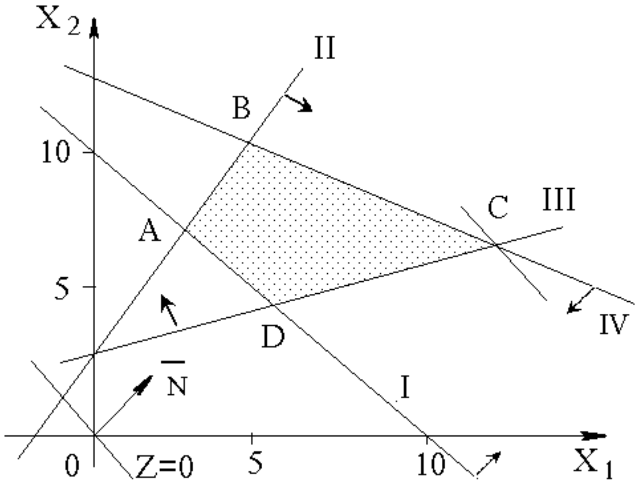


Рис. 38.4. **Многоугольник решений**

Образовалась область решения всех неравенств – это многоугольник *АВСД* – многоугольник планов.

Теперь строим градиент целевой функции. Это вектор: $\overline{N} = 5;4$.

Далее строим линию уровня, соответствующую $Z = 0$; она проходит через начало координат и перпендикулярна $\overline{N}$. Перемещая ее в направлении вектора $\overline{N}$, видим, что наиболее удаленной вершиной многоугольника является вершина *С*.

Вычислим координаты этой вершины. Точка *С* лежит на пересечении прямых III и IV. Решим систему уравнений этих прямых:

$$\begin{cases} -3x_1 + 10x_2 = 30, \\ 3x_1 + 5x_2 = 60. \end{cases}$$

Откуда $x_1 = 10, \quad x_2 = 6.$

Следовательно, $Z_{max} = 5 \cdot 10 + 4 \cdot 6 = 74.$

Пример 38.2. Найти минимум функции $Z=-3x_1+10x_2$, если

$$\begin{cases} x_1+x_2\le 10, \\ -3x_1+2x_2\le 6, \\ -3x_1+10x_2\ge 30, \\ 3x_1+5x_2\ge 60, \quad x_1\ge 0, \quad x_2\ge 0. \end{cases}$$

Решение.

Для построения многоугольника решений можно использовать пример 38.1. Далее строим градиент целевой функции (рис. 38.5).

Это вектор: $\overline{N}=\{-3;10\}$.

Строим линию уровня $Z=0$, она проходит через начало координат и перпендикулярна $\overline{N}$. Перемещая ее в направлении вектора $\overline{N}=\{-3;10\}$, видим, что вектор $\overline{N}$ является нормальным вектором прямой III, значит линия уровня и прямая III параллельны, то есть при входе в область линия уровня совпадает с прямой *DC*.

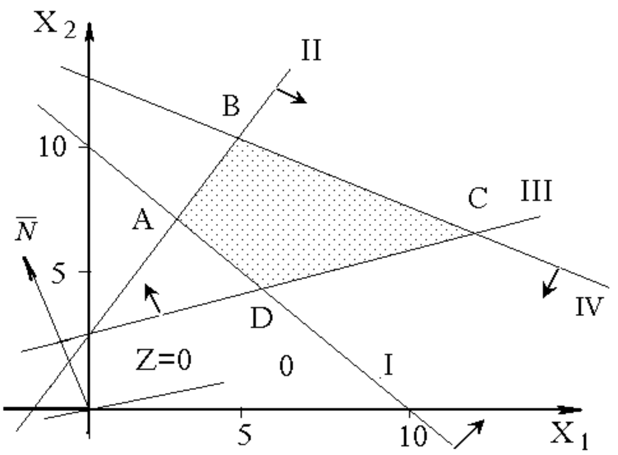


Рис. 38.5. **Многоугольник решений**

Вычислим координаты точек *D* и *C*, а также значение функции цели. Точка *D* лежит на пересечении прямых I и III. Решим систему уравнений:

$$\begin{cases} x_1+x_2=10, \\ -3x_1+10x_2=30. \end{cases}$$

Откуда $x_1=\dfrac{70}{13}; \quad x_2=\dfrac{60}{13}, \quad \overline{X}_1\left(\dfrac{70}{13},\dfrac{60}{13}\right).$

Тогда $Z_D=-3\cdot\dfrac{70}{13}+10\cdot\dfrac{60}{13}=30.$

Точка *C* лежит на пересечении прямых III и IV.

Решим систему уравнений:

$$\begin{cases} -3x_1+10x_2=30, \\ 3x_1+5x_2=60. \end{cases}$$

Откуда $x_1=10;\quad x_2=6,\ \overline{X}_2(10,6),\ Z_C=-3\cdot10+10\cdot6=30.$

Следовательно, минимум будет в любой точке отрезка *DC*:

$$\overline{X}_{\min}=\lambda_1\overline{X}_1+\lambda_2\overline{X}_2, \text{ где } \lambda_1+\lambda_2=1.$$

$$\overline{X}_{\min}=\left(\frac{70}{13}\lambda_1+10\lambda_2;\frac{60}{13}\lambda_1+6\lambda_2\right).$$

38.3. Исследование задачи линейного программирования: понятие опорного плана, теоремы о существовании оптимального опорного плана

Известно, что если ранг системы векторов $\overline{A}_1,\overline{A}_2,\ldots,\overline{A}_n$ равен $r<n$, то переменные, соответствующие базисным векторам, называются ***базисными переменными***, а остальные переменные называюся свободными переменными и решение $X=(x_1,x_2,\ldots,x_n)$ системы линейных уравнений $x_1\overline{A}_1+x_2\overline{A}_2+\ldots+x_n\overline{A}_n=\overline{B}$, которое содержит только базисные переменные (свободные переменные равны нулю), называется ***базисным решением системы***. ***Опорным решением*** называется базисное неотрицательное решение.

В канонической форме ограничения ЗЛП представлены в виде системы линейных уравнений, число которых не превышает числа неизвестных. Эта система может не иметь ни одного решения, иметь только одно решение, но чаще всего такая система имеет бесчисленное множество решений.

Общее число неизвестных в канонической форме обозначим через n, число уравнений – через m, где $m<n$. Возможно, что не все уравнения системы являются независимыми, не исключено, что некоторые ограничения дублируют ранее введенные или являются их комбинацией.

Обозначим через r – ранг матрицы системы ограничений – число линейно независимых уравнений ($r\le m$).

Неоднородная система линейных уравнений может быть решена относительно некоторых r неизвестных, которые будем называть ***базисными***. Другие $(n-r)$ неизвестных называются ***свободными***. При этой операции все лишние ограничения будут отброшены. Базисные неизвестные выражены через свободные, они являются функциями свободных неизвестных. Такая форма представления системы линейных уравнений называется ее ***общим решением***.

Если в исходную систему подставить выражения базисных неизвестных через свободные, то все уравнения системы превратятся в тождества. Отметим, что общее решение – это не числа, а формулы. Свободные неизвестные могут принимать произвольные значения из некоторого интервала, базисные неизвестные вычисляются для каждого конкретного набора числовых значений свободных неизвестных. Такие числовые решения называются ***частными решениями*** системы. Одно из этих частных решений, когда все свободные неизвестные приняты равными нулю, называется ***базисным***. Если полученное базисное решение неотрицательно, его называют ***опорным решением***, или ***базисным планом***.

При этом ***опорным планом*** называется неотрицательный набор неизвестных, которые удовлетворяют всем ограничениям задачи. ***Оптимальным планом*** ЗЛП называется опорный план, который придает функции цели экстремальное значение.

Теорема 1. Если множество планов ЗЛП не пусто, то среди них имеется хотя бы один опорный план.

Теорема 2. Если множество планов ЗЛП не пусто и целевая функция ограничена снизу (сверху) на этом множестве, то задача имеет хотя бы один оптимальный план.

38.4. Теоретические основы симплекс-метода решения задачи линейного программирования

Классическим методом решения ЗЛП стал ***симплекс-метод***, получивший также название *метода последовательного улучшения плана*, разработанный в 1949 г. американским математиком Д. Данцигом (это обобщение метода разрешающих множителей, разработанного в 1939 г. отечественным математиком Л. Канторовичем).

Метод является универсальным, так как позволяет решить практически любую задачу линейного программирования, записанную в канонической форме.

Идея симплекс-метода заключается в том, что начиная с некоторого исходного опорного решения осуществляется последовательно направленное перемещение по опорным решениям задачи к оптимальному. Значение целевой функции при этом перемещении для задач на максимум не убывает, на минимум – не возрастает. Так как число опорных решений конечно, то через конечное число шагов получается оптимальное решение.

Применение симплекс-метода возможно, если система ограничений записана в канонической форме и в системе ограничений выделен единичный базис таким образом, чтобы правые части системы остались неотрицательными.

Не уменьшая общности, для простоты изложения рассмотрим функцию четырех переменных.

Пусть задана функция цели, которая исследуется на минимум, система ограничений приведена к канонической форме и выделен единичный базис.

Модель задачи имеет вид:

$$Z = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \ (\min),$$

$$\begin{cases} x_1 + a_{14}x_4 = b_1, \\ x_2 + a_{24}x_4 = b_2, \quad x_j \geq 0, \quad (j = \overline{1,4}). \\ x_3 + a_{34}x_4 = b_3, \end{cases}$$

Исходное базисное решение (план) – $\overline{X_0} = (b_1, b_2, b_3, 0)$, тогда

$$Z(\overline{X_0}) = c_1b_1 + c_2b_2 + c_3b_3.$$

Запишем систему ограничений в векторной форме:

$$\overline{A}_1x_1 + \overline{A}_2x_2 + \overline{A}_3x_3 + \overline{A}_4x_4 = \overline{A}_0,$$

где $\overline{A_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$; $\overline{A_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$; $\overline{A_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$; $\overline{A_4} = \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix}$; $\overline{A_0} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$.

Координаты векторов $\overline{A_0}, \overline{A_4}$ – это координаты разложения этих векторов в единичном базисе $\overline{A_1}, \overline{A_2}, \overline{A_3}$, то есть

$$\overline{A_0} = b_1\overline{A_1} + b_2\overline{A_2} + b_3\overline{A_3},$$
$$\overline{A_4} = a_{14}\overline{A_1} + a_{24}\overline{A_2} + a_{34}\overline{A_3}.$$

Введем обозначения:

1. $\overline{C_{баз}} = (c_1, c_2, c_3)$.
2. $z_j = \overline{C_{баз}} \cdot \overline{A_j}$.
3. $\Delta_j = z_j - c_j$, где c_j – коэффициент функции цели при x_j.

Вычислим значения z_j и оценки Δ_j для всех векторов:

$$z_1 = \overline{C_{баз}} \cdot \overline{A_1} = c_1, \quad \Delta_1 = z_1 - c_1 = 0,$$

$$z_2 = \overline{C_{баз}} \cdot \overline{A_2} = c_2, \quad \Delta_2 = z_2 - c_2 = 0,$$

$$z_3 = \overline{C_{баз}} \cdot \overline{A_3} = c_3, \quad \Delta_3 = z_3 - c_3 = 0,$$

$$z_4 = \overline{C_{баз}} \cdot \overline{A_4} = c_1 a_{14} + c_2 a_{24} + c_3 a_{34},$$
$$\Delta_4 = z_4 - c_4 = c_1 a_{14} + c_2 a_{24} + c_3 a_{34} - c_4 \neq 0.$$

Выполним симплекс-преобразование однократного замещения базиса. Умножим $\overline{A_4}$ на некоторое $\theta > 0$ и вычтем из $\overline{A_0}$, получим

$$\overline{A_0} = (b_1 - \theta a_{14})\overline{A_1} + (b_2 - \theta a_{24})\overline{A_2} + (b_3 - \theta a_{34})\overline{A_3} + \theta\overline{A_4}.$$

Трехмерный вектор $\overline{A_0}$ разложен по четырем векторам. Это значит, что одна координата равна нулю.

Выполним симплекс-преобразование, а именно выберем

$$\theta = \min\left(\frac{b_1}{a_{14}}, \frac{b_2}{a_{24}}, \frac{b_3}{a_{34}}\right).$$

Пусть $\theta = \dfrac{b_3}{a_{34}}$, тогда $\overline{A_0} = (b_1 - \theta a_{14})\overline{A_1} + (b_2 - \theta a_{24})\overline{A_2} + \theta\overline{A_4}$.

Таким образом, в результате симплекс-преобразования получили неотрицательные координаты вектора и новый опорный план

$$\overline{X_1} = (b_1 - \theta a_{14}, b_2 - \theta a_{24}, 0, \theta),\ Z(\overline{X_1}) = c_1(b_1 - \theta a_{14}) + c_2(b_2 - \theta a_{24}) + c_4\theta.$$

Так как функция цели исследуется на минимум, то $Z(\overline{X_1})$ должно быть меньше $Z(\overline{X_0})$.

Поскольку $\theta = \dfrac{b_3}{a_{34}}$, то $b_3 = \theta a_{34}$ и

$$\begin{aligned} Z(\overline{X_1}) &= c_1 b_1 + c_2 b_2 + c_3 b_3 - c_3 b_3 - c_1\theta a_{14} - c_2\theta a_{24} + c_4\theta = \\ &= Z(\overline{X_0}) - \theta(c_1 a_{14} + c_2 a_{24} + c_3 a_{34} - c_4) = Z(\overline{X_0}) - \theta\Delta_4. \end{aligned}$$

По условию $\theta > 0$. Тогда, если $\Delta_4 > 0$, то выполняется условие $Z(\overline{X_1}) < Z(\overline{X_0})$, то есть идем к оптимальному плану, если $\Delta_4 \le 0$, то $Z(\overline{X_1}) \ge Z(\overline{X_0})$ и исходный план является оптимальным.

Итак, можно сформулировать ***критерий оптимальности решения ЗЛП***: если функция цели исследуется на минимум, то при $\Delta_j \le 0$ план является оптимальным; если функция цели исследуется на максимум, то при $\Delta_j \ge 0$ план является оптимальным.

Рассмотрим ЗЛП в общем виде.

Найдем минимум (максимум) функции $Z = c_1x_1 + c_2x_2 + ... + c_nx_n$, если

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + c_n x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + c_{1n}x_n = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m, \end{cases} \quad x_j \ge 0, \quad (j = \overline{1,n}).$$

Предположим, что система ограничений приведена к единичному базису.

Векторная форма задачи имеет такой вид: найти минимум (максимум) функции $Z = CX$, если

$$x_1\overline{P_1}+x_2\overline{P_2}+...+x_n\overline{P_n}=\overline{P_0},\ x_j\geq 0,\quad \left(j=\overline{1,n}\right),$$

где

$$\overline{P_1}=\begin{pmatrix}1\\0\\.\\.\\.\\0\end{pmatrix};\ \overline{P_2}=\begin{pmatrix}0\\1\\.\\.\\.\\0\end{pmatrix};...;\overline{P_m}=\begin{pmatrix}0\\0\\.\\.\\.\\1\end{pmatrix};\overline{P_{m+1}}=\begin{pmatrix}a_{1m+1}\\a_{2m+1}\\.\\.\\.\\a_{mm+1}\end{pmatrix};\ ...;\ \overline{P_n}=\begin{pmatrix}a_{1n}\\a_{2n}\\.\\.\\.\\a_{mn}\end{pmatrix};\ \overline{P_0}=\begin{pmatrix}b_1\\b_2\\.\\.\\.\\b_m\end{pmatrix}.$$

$\overline{X}=\left(b_1,b_2,...,b_m,0,...,0\right)$ – исходный опорный план.

Система единичных векторов $\overline{P_1},\overline{P_2},...,\overline{P_m}$ образует базис m-мерного пространства, поэтому каждый из векторов $\overline{P_1},\overline{P_2},...,\overline{P_m},...,\overline{P_n}$, а также $\overline{P_0}$ может быть представлен в виде линейной комбинации векторов данного базиса.

Пусть $\overline{C_{баз}}=\left(c_1,c_2,...,c_m\right)$, где $c_1,c_2,...,c_m$ – коэффициенты функции цели для базисных векторов; $z_j=\overline{C_{баз}}\cdot\overline{P_j}$; $\Delta_j=z_j-c_j$, тогда

$$\overline{P_j}=\sum_{i=1}^{m}a_{ij}\overline{P_i}\ \left(j=\overline{0,n}\right);\ z_j=\sum_{i=1}^{m}c_ia_{ij};\ \Delta_j=\sum_{i=1}^{m}c_ia_{ij}-c_j\ \left(j=\overline{1,n}\right).$$

Критерий оптимальности можно сформулировать следующим образом:

Если функция цели исследуется на минимум и для некоторого вектора, не входящего в базис, выполняется условие $\Delta_j=z_j-c_j>0$ $\left(j=\overline{1,n}\right)$, то можно получить новый опорный план, для которого значение целевой функции будет меньше исходного. Если для всех векторов выполняется условие $\Delta_j=z_j-c_j\leq 0$ $\left(j=\overline{1,n}\right)$, то полученный план является оптимальным.

Для исследования функции на максимум оптимальность плана требует выполнения условия $\Delta_j=z_j-c_j\geq 0$ $\left(j=\overline{1,n}\right)$, а условие $\Delta_j=z_j-c_j<0$ означает, что полученный план можно улучшить.

При этом могут быть два случая:

если все координаты вектора, подлежащего вводу в базис, неположительные, то ЗЛП не имеет решения;

если имеется хотя бы одна положительная координата у вектора, подлежащего вводу в базис, то можно получить новый опорный план.

Алгоритм решения ЗЛП симплекс-методом

1. Записать систему ограничений задачи с положительными свободными членами в канонической форме, то есть в виде уравнений.

2. Выделить в системе ограничений единичный базис, сохранив свободные члены положительными, то есть выполнить симплекс-преобразование.

3. Найти исходный опорный план и значение функции цели.

4. Проверить исходный план на оптимальность. Для этого необходимо вычислить значения $z_j = \overline{C_{баз}} \cdot \overline{P_j}$ и оценки $\Delta_j = z_j - c_j$.

Если все оценки $\Delta_j \geq 0$ (в случае максимизации целевой функции) или все оценки $\Delta_j \leq 0$ (в случае минимизации целевой функции), то исходный опорный план является оптимальным и задача решена.

Если нарушен критерий оптимальности, то есть среди оценок есть хотя бы одна отрицательная (в случае максимизации целевой функции) или положительная (в случае минимизации целевой функции), то план необходимо улучшить.

5. Улучшение исходного опорного плана:

а) если критерий оптимальности нарушен для одного j-го вектора, то для него вычисляем параметр:

$$\theta_j = \min\left(\frac{b_i}{a_{ij}}\right), \text{ где } a_{ij} > 0,$$

и соответствующий вектор вводим в базис, используя однократное замещение базиса;

b) если критерий оптимальности нарушен для нескольких векторов, то для каждого соответствующего вектора вычисляем $\left|\Delta_j \theta_j\right|$ и в базис вводим вектор с максимальным значением этого произведения ($\max\left|\Delta_j \theta_j\right|$), то есть направляющий столбец $\overline{P_j}$ определяется максимальным значением абсолютной величины произведения $|\Delta_j \theta_j|$, а направляющая строка – минимальным отношением $\frac{b_i}{a_{ij}}$.

6. Новый опорный план проверяем на оптимальность. Процесс продолжаем до тех пор, пока не будет выполнен критерий оптимальности.

Замечание 1: если для j-го вектора нарушен критерий оптимальности, а координаты j-го вектора отрицательные, то план не является оптимальным, улучшить его невозможно, ЗЛП решения не имеет.

Замечание 2 (проблема вырождения): при применении симплекс-метода монотонный рост целевой функции при исследовании на максимум (или монотонное убывание при исследовании на минимум) имеет место только тогда, когда при каждой итерации мы получаем невырожденный опорный план. Встречаются задачи, когда основная система ограничений в правой части содержит один или несколько нулей. Аналогичная ситуация может возникать и во время решения задачи.

Рассмотрим, как изменяется целевая функция, если при очередном преобразовании по методу Жордана – Гаусса разрешающий элемент содержится в строке с нулевой правой частью. Известно, что при каждой итерации целевая функция изменяется на величину $\Delta Z = |\theta \cdot \Delta|$. Поскольку $\theta = 0$, то введение в базис вектора, которому отвечает нулевое симплексное отношение, не сопровождается изменением целевой функции. Как правило, после нескольких преобразований с использованием нулевого симплексного отношения можно получить план, который уже был раньше, то есть симплекс-метод *зацикливается*.

Замечание 3: если при вычислении $\theta_j = \min\left(\frac{b_i}{a_{ij}}\right)$ получают несколько одинаковых минимальных значений, то вычисляют вспомогательное значение $\theta_j^* = \min\left(\frac{a_{ij}^*}{a_{ij}}\right)$, где a_{ij}^* – координаты свободного вектора, при этом a_{ij}^* включают с учетом знака (то есть и $a_{ij}^* < 0$).

Рассмотрим это на примере.

Пусть имеем векторы в таблице (табл. 38.1).

Таблица 38.1

Векторы системы ограничений ЗЛП

P_0	P_1	P_2	P_3	P_4	P_5	P_6
6	–2	1	1	0	0	0
9	–1	3/2	0	1	0	0
30	–1	5	0	0	1	0
12	1	1	0	0	0	1

Вычислим $\theta_2 = \min\left(\frac{6}{1};\frac{9}{3/2};\frac{30}{5};\frac{12}{1}\right) = \min(6;6;6;12) = 6$,

$\theta_2^* = \min\left(\frac{-2}{1};\frac{-1}{3/2};\frac{-1}{5};\frac{1}{1}\right) = \min\left(-2;-\frac{2}{3};-\frac{1}{5};1\right) = -2$, то есть разрешающий элемент 1 в первой строке вектора P_2.

Замечание 4 (единственность решения): если план оптимальный, а некоторые оценки, соответствующие свободным векторам, равны нулю ($\Delta_j = 0$), то, если эти векторы вводить в базис, можно получить новый оптимальный план, но значение функции цели не изменится, то есть задача имеет множество решений, их выпуклая линейная комбинация также является решением ЗЛП (геометрически, это две угловые точки многоугольника планов и отрезок, их соединяющий, тоже являются решением ЗЛП). Такое решение называется ***альтернативным оптимумом***.

Пример 38.3. Найти минимум функции

$$Z = x_1 - x_2 + x_3 + x_4 + x_5 - x_6,$$

если

$$\begin{cases} x_1 + x_4 + 6x_6 = 9, \\ 3x_1 + x_2 - 4x_3 + 2x_6 = 2, \\ x_1 + 2x_3 + x_5 + 2x_6 = 6, \end{cases}$$

$$x_j \ge 0 \quad (j = \overline{1,6}).$$

Решение.

Запишем матрицу ограничений: $\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 6 & 9 \\ 3 & 1 & -4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 2 & 0 & 1 & 2 & 6 \end{array}\right)$.

Единичные векторы P_2, P_4, P_5 принимаем за базис.

Свободным переменным присваиваем значения $x_1 = x_3 = x_6 = 0.$ Тогда $x_2 = 2,\ x_4 = 9,\ x_5 = 6.$

Следовательно, опорный план: $\overline{X}_0 = (0; 2; 0; 9; 6; 0)$.

Значение целевой функции: $Z_0 = -2 + 9 + 6 = 13.$

Проверим начальный опорный план на оптимальность. Для этого составим симплекс-таблицу (табл. 38.2).

Таблица 38.2

Симплекс-таблица (исходные данные – 1-я итерация)

Базис	$C_{баз}$	c_j	1	–1	1	1	1	–1
		P_0	P_1	P_2	P_3	P_4	P_5	P_6
P_4	1	9	1	0	0	1	0	6
P_2	–1	2	3	1	–4	0	0	2
$\leftarrow P_5$	1	6	1	0	2	0	1	2
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$		13	–1	–1	6	1	1	6
$\Delta_j = z_j - c_j$		–	–2	0	**5**	0	0	7

Последняя строка табл. 38.2 показывает, что начальный план не является оптимальным, так как имеем две положительные оценки: $\Delta_3 = 5,\ \Delta_6 = 7$. Поэтому переходим к новому опорному плану. Для этого вводим в базис вектор, для которого $\left|\Delta_j \theta_j\right|$ наибольшее, где

$$\theta_j = \min_{a_{ij} > 0} \frac{b_i}{a_{ij}}.$$

$\Delta_3 = 5$: так как соответствующий столбец имеет один положительный элемент, то $\theta_3 = \dfrac{6}{2} = 3$. Тогда $\left|\Delta_3\theta_3\right| = 5 \cdot 3 = 15.$

$\Delta_6 = 7$: $\theta_6 = \min\left\{\dfrac{9}{6}, \dfrac{2}{2}, \dfrac{6}{2}\right\} = 1$, тогда $\left|\Delta_6\theta_6\right| = 7 \cdot 1 = 7.$

Так как $\left|\Delta_3\theta_3\right| > \left|\Delta_6\theta_6\right|$, то в базис вводим вектор P_3, разрешающий элемент 2 в последней строке.

Из базиса выводим вектор P_5, у которого единица в последней строке. Составляем новую таблицу (табл. 38.3). Пояснения к ее заполнению приведены ниже.

Таблица 38.3

Симплекс-таблица (2-я итерация)

Базис	$C_{баз}$	c_j	1	–1	1	1	1	–1
		P_0	P_1	P_2	P_3	P_4	P_5	P_6
P_2	–1	14	5	1	0	0	2	6
$\leftarrow P_4$	1	9	1	0	0	1	0	6
$\rightarrow P_3$	1	3	½	0	1	0	½	1
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$		–2	–7/2	–1	1	1	–3/2	1
$\Delta_j = z_j - c_j$		–	–9/2	0	0	0	–5/2	2

Имеем новый базис: P_2, P_3, P_4.

В столбце $C_{баз}$ записаны коэффициенты при базисных неизвестных целевой функции: –1; 1; 1.

Заполняем строку вектора, который ввели в базис (P_3). Для этого элементы третьей строки табл. 38.2 делим на разрешающий элемент 2. Получили *разрешающую (ведущую) строку*.

В столбце P_3 необходимо получить единичный вектор. В первой строке табл. 38.2 имеем 0, поэтому эту строку перепишем без изменений. Во второй строке табл. 38.2 на месте элемента (–4) необходимо получить 0. Для этого каждый элемент ведущей строки табл. 38.3 умножим на 4 и прибавим к соответствующим элементам второй строки табл. 38.2.

Получим новый опорный план $\overline{X_1} = (0; 14; 3; 9; 0; 0)$. Проверим его на оптимальность. Для этого следует найти оценки векторов (Δ_j). Есть одно положительное число $\Delta_6 = 2$, план не оптимальный.

Переходим к новому опорному плану. Для этого вводим в базис вектор P_6, так как только $\Delta_6 = 2 > 0$.

Находим отношение свободных членов к коэффициентам последнего столбца: $\theta_6 = min\left\{\frac{14}{6}, \frac{9}{6}, \frac{3}{1}\right\} = \frac{9}{6}$, поэтому выводим из базиса вектор P_4.

Следовательно, разрешающим является столбец вектора P_6 и строка вектора P_4, а разрешающим элементом является 6.

Составляем новую таблицу (табл. 38.4).

Таблица 38.4

Симплекс-таблица (3-я итерация)

Базис	$C_{баз}$	c_j	1	–1	1	1	1	–1
		P_0	P_1	P_2	P_3	P_4	P_5	P_6
P_2	–1	5	4	1	0	–1	2	0
$\to P_6$	–1	3/2	1/6	0	0	1/6	0	1
P_3	1	3/2	1/3	0	1	–1/6	1/2	0
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$		– 5	–23/6	–1	1	2/3	–3/2	–1
$\Delta_j = z_j - c_j$			–29/6	0	0	–1/3	–5/2	0

Имеем новый базис: P_2, P_3, P_6. Заполняем строку вектора P_6. Для этого элементы второй строки табл. 38.3 делим на разрешающий элемент 6 и записываем в табл. 38.4.

Получим ведущую строку. В столбце P_6 надо получить единичный вектор. В первой строке табл. 38.3 требуется получить 0 на месте 1. Для этого каждый элемент ведущей строки табл. 38.4 надо умножить на (–6) и прибавить к соответствующим элементам первой строки табл. 38.3. Полученные элементы записываем в первую строку табл. 38.4.

В третьей строке табл. 38.3 на месте 1 требуется получить 0. Для этого каждый элемент второй строки табл. 38.4 надо умножить на (–1) и прибавить к соответствующим элементам третьей строки табл. 38.3. Полученные элементы записываем в третью строку табл. 38.4.

Получим новый опорный план: $\overline{X_2} = \left(0; 5; \frac{3}{2}; 0; 0; \frac{3}{2}\right)$.

Проверим его на оптимальность. Для этого вычислим все значения $z_j = \overline{C_{баз}} \cdot \overline{P_j}$ и оценки векторов $\Delta_j = z_j - c_j$. Все оценки $\Delta_j \le 0$. Следовательно, план $\overline{X_2} = \left(0; 5; \frac{3}{2}; 0; 0; \frac{3}{2}\right)$ оптимальный, так как свободные векторы P_4, P_5 имеют оценки $\Delta_j \ne 0$, то план единственный.

Для полученного плана $Z_{\min} = -5$.

Пример 38.4. Кирпичный завод выпускает две марки кирпича A_1 и A_2. Для этого используется глина трех видов B_1, B_2, B_3. Задан месячный запас глины 10, 30, 47 т соответственно. Для производства 1 тыс. шт. кирпича 1-й марки необходимо 1 т глины B_1 и 1 т глины B_3; для производства 1 тыс. шт. кирпича 2-й марки требуется 2 т глины B_2 и 2 т глины B_3. Прибыль от реализации 1 тыс. шт. кирпича 1-й марки – 40 грн, 2-й марки – 70 грн. Составить план выпуска кирпича, обеспечивающий максимальную прибыль.

Решение.

Составим таблицу исходных данных (табл. 38.5).

Таблица 38.5

Исходные данные задачи о выпуске кирпича

Марка кирпича / Вид глины	A_1	A_2	Запас глины, т
B_1	1	0	10
B_2	0	2	30
B_3	1	2	47
Прибыль от реализации 1 тыс. шт. кирпича, грн	40	70	–
Количество выпускаемого кирпича, тыс. шт.	x_1	x_2	–

Составим математическую модель задачи. Обозначим количество выпускаемого кирпича марки A_1 и A_2 – x_1, x_2 тыс. шт. соответственно. Тогда прибыль от реализации всего кирпича $Z = 40x_1 + 70x_2$, функция цели Z исследуется на максимум.

Составим систему ограничений: $1 \cdot x_1 + 0 \cdot x_2 \le 10$ – количество глины B_1, которое расходуется на изготовление всего кирпича, не превосходит имеющегося запаса.

Рассуждая аналогично, получим ограничения по глине B_2 и B_3: $0 \cdot x_1 + 2 \cdot x_2 \le 30,\ 1 \cdot x_1 + 2 \cdot x_2 \le 47.$

Добавим условия неотрицательности $x_1, x_2 \ge 0$.

Математическая модель задачи имеет вид:

$$Z = 40x_1 + 70x_2 \ (\max),$$

$$\begin{cases} x_1 \le 10, \\ 2x_2 \ge 30, \\ x_1 + 2x_2 \le 47, \end{cases}$$

$$x_j \ge 0 \quad (j = \overline{1,2}).$$

Приведем систему ограничений к канонической форме. Для этого к левой части неравенств системы добавим дополнительные переменные, тогда:

$$Z = 40x_1 + 70x_2 \ (\max),$$

$$\begin{cases} x_1 + x_3 = 10, \\ 2x_2 + x_4 = 30, \\ x_1 + 2x_2 + x_5 = 47, \ x_j \ge 0 \ (j = \overline{1,5}). \end{cases}$$

Решаем задачу с помощью симплекс-таблицы (табл. 38.6)

Таблица 38.6

Симплекс-таблица

<table>
<tr><td rowspan="2">№ п/п</td><td rowspan="2">Базис</td><td rowspan="2">$\overline{C_{баз}}$</td><td>c_j</td><td>40</td><td>70</td><td>0</td><td>0</td><td>0</td><td rowspan="2">Примечания</td></tr>
<tr><td>$\overline{A_0}$</td><td>$\overline{A_1}$</td><td>$\overline{A_2}$</td><td>$\overline{A_3}$</td><td>$\overline{A_4}$</td><td>$\overline{A_5}$</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>1</td><td>$\overline{A_3}$</td><td>0</td><td>10</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td><td rowspan="5">$\overline{X_0} = (0;0;10;30;47)$, $Z_0 = 0$, $\Delta_1, \Delta_2 < 0$ – план неоптимальный, выполним симплекс-преобразование.
$\theta_1 = \min\left(\frac{10}{1}; \frac{47}{1}\right) = 10$,
$\lvert\theta_1\Delta_1\rvert = \lvert 10 \cdot (-40)\rvert = 400$,
$\theta_2 = \min\left(\frac{30}{2}; \frac{47}{2}\right) = 15$,
$\lvert\theta_2\Delta_2\rvert = \lvert 15 \cdot (-70)\rvert = 1050$.
Так как $\lvert\theta_2\Delta_2\rvert > \lvert\theta_1\Delta_1\rvert$, то в базис вводим вектор $\overline{A_2}$</td></tr>
<tr><td>2</td><td>$\overline{A_4}$</td><td>0</td><td>30</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>3</td><td>$\overline{A_5}$</td><td>0</td><td>47</td><td>1</td><td>2</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td colspan="3">$z_j = \overline{C_{баз}} \cdot \overline{A_j}$</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="3">$\Delta_j = z_j - c_j$</td><td>–</td><td>–40</td><td>–70</td><td>0</td><td>0</td><td>0</td></tr>
</table>

Продолжение табл. 38.6

1	2	3	4	5	6	7	8	9	10
4	$\overline{A_3}$	0	10	1	0	1	0	0	[1]
5	$\overline{A_2}$	70	15	0	1	0	1/2	0	[2]/2
6	$\overline{A_5}$	0	17	1	0	0	– 1	1	[2]·(–1)+[3]
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			1050	0	70	0	35	0	$\overline{X_1} = (0;15;10;0;17)$, $Z_1 = 1050$, $\Delta_1 < 0$ – план неоптимальный. $\theta_1 = \min\left(\frac{10}{1}; \frac{17}{1}\right) = 10$, в базис вводим вектор $\overline{A_1}$
$\Delta_j = z_j - c_j$			–	–40	0	0	35	0	
7	$\overline{A_1}$	40	10	1	0	1	0	0	[4]
8	$\overline{A_2}$	70	15	0	1	0	1/2	0	[5]
9	$\overline{A_5}$	0	7	0	0	–1	–1	1	[4]·(–1)+[6]
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			1450	40	70	40	35	0	$\Delta_j \geq 0$, план оптимальный и единственный, так как для свободных векторов $\Delta_j \neq 0$. $\overline{X_{\max}} = (10;15;0;0;7)$, $Z_{\max} = 1450$
$\Delta_j = z_j - c_j$			–	0	0	40	35	0	

Так как в исходной задаче две переменные, то в ответе оставляем только x_1, x_2: $\overline{X_{\max}} = (10;15)$, $Z_{\max} = 1450$.

Подставим $\overline{X_{\max}}$ в систему ограничений:

$$\begin{cases} 10 = 10 \quad (\textit{дефицит}), \\ 2 \cdot 15 = 30 \quad (\textit{дефицит}), \\ 10 + 2 \cdot 15 = 40 < 47 \quad (\textit{запас}). \end{cases}$$

Экономическая интерпретация результата: для того чтобы получить максимальную прибыль в размере 1 450 грн, требуется выпускать 10 тыс. шт. кирпича 1-й марки и 15 тыс. шт. кирпича 2-й марки, при этом глина первого и второго видов израсходована полностью, остался запас глины третьего вида.

38.5. Понятие о модифицированном симплекс-методе

В основу модифицированного симплекс-метода положены такие особенности линейной алгебры, которые позволяют в ходе решения задачи работать с частью матрицы ограничений. Иногда метод называют *методом обратной матрицы*.

В процессе работы алгоритма происходит спонтанное обращение матрицы ограничений по частям, соответствующим текущим базисным векторам. Указанная способность делает весьма привлекательной машинную реализацию вычислений вследствие экономии памяти под промежуточные переменные и значительного сокращения времени счета. Этот метод хорош для ситуаций, когда число переменных n значительно превышает число ограничений m.

В целом, метод отражает традиционные черты общего подхода к решению задач линейного программирования, включающего в себя канонизацию условий задачи, расчет оценок плана, проверку условий оптимальности, принятие решений о коррекции базиса и исключение по схеме Жордана – Гаусса.

Особенности заключаются в наличии двух таблиц – основной и вспомогательной, порядке их заполнения и некоторой специфичности расчетных формул.

38.6. Метод искусственного базиса. Расширенная М-задача

Метод искусственного базиса применяется для решения задач, математическая модель которых представляется в канонической форме, но построить исходный опорный план достаточно сложно, а также для задач, представленных в общей форме.

Рассмотрим задачу, математическая модель которой представляется в общей форме:

$$Z = \sum_{j=1}^{n} c_j x_j \to extr(\max, \min)$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\ a_{k+11}x_1 + a_{k+12}x_2 + \ldots + a_{k+1n}x_n \ge b_{k+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{l1}x_1 + a_{l2}x_2 + \ldots + a_{ln}x_n \ge b_l, \\ a_{l+11}x_1 + a_{l+12}x_2 + \ldots + a_{l+1n}x_n = b_{l+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\ \quad x_j \ge 0, \quad j = \overline{1,n}. \end{cases}$$

Пусть первые k неравенств основной системы ограничений, которые содержат знак «$\le$», имеют неотрицательные правые части ($b_i \ge 0$, $i = \overline{1,k}$).

Добавив в левые части каждого из этих неравенств соответствующую дополнительную неотрицательную неизвестную: $x_{n+1},\ldots,x_{n+k}$, перейдем к уравнениям, а соответствующие дополнительным переменным единичные векторы $\overline{P}_{n+1},\ldots,\overline{P}_{n+k}$ войдут в базис пространства решений.

Если среди k неравенств основной системы ограничений есть такие, которые имеют отрицательную правую часть, их надо умножить на (– 1), в результате чего получим неравенства со знаком «$\ge$».

Неравенства со знаком «$\ge$» основной системы ограничений с неотрицательной правой частью (это неравенства, которые имеют номера от $k+1$ до l) можно преобразовать в уравнения, отняв в левых частях каждой из них соответствующие дополнительные неотрицательные переменные: $x_{n+k+1},\ldots,x_{n+l}$. Как и дополнительные неизвестные, которые добавлены в первых из k неравенств с положительной правой частью, эти неизвестные будут входить в целевую функцию с нулевыми коэффициентами, но соответствующие им векторы $\overline{P}_{n+k+1},\ldots,\overline{P}_{n+l}$ не войдут в базис пространства решений, поскольку имеют отрицательные координаты.

Итак, благодаря введению дополнительных неизвестных, получена математическая модель исходной задачи в канонической форме:

$$Z = c_1x_1 + c_2x_2 + c_nx_n + 0 \cdot x_{n+1} + ... + 0 \cdot x_{n+l} \quad \to \quad extr$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + x_{n+1} = b_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + ... + a_{kn}x_n + x_{n+k} = b_k, \\ a_{k+11}x_1 + a_{k+12}x_2 + ... + a_{k+1n}x_n - x_{n+k+1} = b_{k+1}, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{l1}x_1 + a_{l2}x_2 + ... + a_{l\,n}x_n - x_{n+l} = b_l, \\ a_{l+11}x_1 + a_{l+12}x_2 + ... + a_{l+1n}x_n = b_{l+1}, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m, \\ \quad x_j \ge 0, \quad j = \overline{1, n+l}. \end{cases}$$

При этом дополнительные векторы $\overline{P}_{n+1}, ..., \overline{P}_{n+l}$ не образуют единичный базис. Можно попробовать найти базис с помощью преобразований по методу Жордана – Гаусса, но это достаточно сложная задача, намного проще воспользоваться *методом искусственного базиса*.

Он заключается в том, что в левую часть уравнения вводится ***фиктивная переменная***. Если целевая функция исследуется на минимум, то фиктивная переменная включается в нее с достаточно большим коэффициентом M – штрафом (M считается значительно больше любого другого из коэффициентов целевой функции и их линейной комбинации). Если целевая функция исследуется на максимум, то коэффициент M, с которым входит в нее фиктивная переменная, вводится с отрицательным знаком.

Векторы, которые соответствуют фиктивным переменным, вместе с векторами, которые соответствуют дополнительным переменным, образуют единичный базис пространства решений. Базис, который содержит вектор, соответствующий фиктивной переменной, называется *ис-*

кусственным, а сама задача линейного программирования, для решения которой используется искусственный базис, называется *расширенной*, или *М-задачей*. Для решения расширенной задачи используется симплекс-метод.

Теорема (*условие решения М-задачи).* Если оптимальный план расширенной М-задачи не содержит фиктивных переменных, то этот план является оптимальным и для исходной задачи, если же фиктивные переменные входят в базис, то исходная задача решения не имеет.

Пример 38.5. Найти минимум функции $Z = 2x_1 + 3x_2 \to \min$ при ограничениях

$$\begin{cases} 3x_1 + 2x_2 \geq 6, \\ x_1 + 4x_2 \geq 4,\ x_1, x_2 \geq 0. \end{cases}$$

Решение.

Вводим в систему ограничений фиктивные базисные неизвестные:

$$\begin{cases} 3x_1 + 2x_2 - x_3 + x_5 = 6, \\ x_1 + 4x_2 - x_4 + x_6 = 4. \end{cases}$$

Новые базисные неизвестные x_5 и x_6 вводим в целевую функцию с коэффициентом M (штрафом), где M – достаточно большое число:

$$Z = 2x_1 + 3x_2 + Mx_5 + Mx_6 \to \min.$$

Далее решаем задачу обычным симплекс-методом. В процессе улучшения плана фиктивные переменные надо выводить из базиса в первую очередь. Если будет получен оптимальный план с базисной фиктивной переменной, то это означает, что исходная система ограничений несовместна, задача не имеет решения. Заполняем симплекс-таблицу (табл. 38.7).

Коэффициенты целевой функции записаны в самой верхней строке. В системе ограничений выделен фиктивный базис $P_5,\ P_6$.

Заполняем столбец $C_{баз}$, вычисляем z_j и оценки Δ_j. Так как значение M велико, то оценки Δ_1 и Δ_2 – положительные.

Таблица 38.7

Симплекс-таблица (исходные данные – 1-я итерация)

№	Базис	$C_{баз}$	c_j	2	3	0	0	M	M
			P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_5	M	6	**3**	2	–1	0	1	0
2	P_6	M	4	1	4	0	–1	0	1
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$			$10M$	$4M$	$6M$	$-M$	$-M$	M	M
$\Delta_j = z_j - c_j$			$10M$	$4M-2$	$6M-3$	$-M$	$-M$	0	0

Ответим на вопрос, какой вектор в первую очередь ввести в базис. Разрешающий элемент определяем с помощью симплекс-преобразования однократного замещения базиса.

$$\theta_1 = \min\left\{\frac{6}{3}, \frac{4}{1}\right\} = 2, \ \left|\theta_1\Delta_1\right| = 2(4M-2),$$

$$\theta_2 = \min\left\{\frac{6}{2}, \frac{4}{4}\right\} = 1, \ \left|\theta_2\Delta_2\right| = 1(6M-3).$$

Так как $\left|\theta_1\Delta_1\right| > \left|\theta_2\Delta_2\right|$, то в базис вводим вектор P_1. Чем больше $\left|\theta_j\Delta_j\right|$, тем быстрее идем к минимуму.

Получаем новую таблицу (табл. 38.8).

Таблица 38.8

Симплекс-таблица (2-я итерация)

№	Базис	$C_{баз}$	c_j	2	3	0	0	M	M	*Примечания*
			P_0	P_1	P_2	P_3	P_4	P_5	P_6	
3	P_1	2	2	1	$^2/_3$	$-^1/_3$	0	$^1/_3$	0	[1] : 3
4	P_6	M	2	0	$^{10}/_3$	$^1/_3$	–1	$-^1/_3$	1	[3]·(–1) + [2]
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$			$2M+4$	2	$^{10}/_3 M+^4/_3$	$^1/_3 M-^2/_3$	$-M$	$-^1/_3 M+^1/_3$	M	
$\Delta_j = z_j - c_j$			$2M+4$	0	$^{10}/_3 M-^5/_3$	$^1/_3 M-^2/_3$	$-M$	$-^4/_3 M+^1/_3$	0	

Имеем две положительные оценки Δ_2 и Δ_3.

Вводим в базис P_2 . Разрешающий элемент $^{10}/_3$ выделен в таблице (см. табл. 38.8).

$$\theta_2 = \min\left\{\frac{2}{2/3}, \frac{2}{10/3}\right\} = \frac{3}{5},\ |\theta_2\Delta_2| = \frac{3}{5}\left(\frac{10}{3}M - \frac{5}{3}\right) \approx 2M,$$

$$\theta_3 = \min\left\{\frac{2}{1/3}\right\} = 6,\ |\theta_3\Delta_3| = 6\left(\frac{1}{3}M - \frac{2}{3}\right) \approx 2M.$$

Выгодно ввести в базис P_2, так как x_2 – исходная переменная.

Получаем табл. 38.9.

Таблица 38.9

Симплекс-таблица (3-я итерация)

№	Базис	$C_{баз}$	c_j	2	3	0	0	M	M	Примечания
			P_0	P_1	P_2	P_3	P_4	P_5	P_6	
5	P_2	3	$^3/_5$	0	1	$^1/_{10}$	$-^3/_{10}$	$-^1/_{10}$	$^3/_{10}$	[4] : $^{10}/_3$
6	P_1	2	$^8/_5$	1	0	$-^2/_5$	$^1/_5$	$^2/_5$	$-^1/_5$	[5]·(-2/3)+ [3]
$z_j = \overline{C_{баз}} \cdot \overline{P_j}$			5	2	3	$-^1/_2$	$-^1/_2$	$^1/_2$	$^1/_2$	
$\Delta_j = z_j - c_j$			5	0	0	$-^1/_2$	$-^1/_2$	$-M+^1/_2$	$-M+^1/_2$	

После трех итераций получено оптимальное решение – все оценки $\Delta_j \le 0$, фиктивные неизвестные выведены из базиса.

Найденное базисное решение $x_1 = {}^8/_5$; $x_2 = {}^3/_5$ является оптимальным. Минимальное значение функции цели $Z_{\min} = 5$, план единственный, так как свободные векторы P_3, P_4, P_5, P_6 имеют оценки $\Delta_j \ne 0$.

Вопросы для самодиагностики

1. Сформулировать задачу линейного программирования.

2. Дать определения следующих понятий: план, опорный план, оптимальный план, решение задачи.

3. Чем отличается общая задача линейного программирования от канонической? Всегда ли общую задачу линейного программирования можно привести к канонической форме?

4. Какая точка выпуклого множества называется угловой?

5. Какой план называется базисным? Как связаны базисные планы и угловые точки области определения задачи линейного программирования?

6. Как с точки зрения геометрической интерпретации можно представить процесс поиска оптимального плана в задаче линейного программирования?

7. Сформулировать критерий оптимальности в симплекс-методе.

8. Сформулировать основные этапы симплекс-метода.

9. Для чего применяется преобразование Жордана – Гаусса?

10. Какой элемент симплекс-таблицы называется ведущим?

11. При каких условиях делается вывод о неограниченности целевой функции в решаемой задаче?

12. В чем особенности метода искусственного базиса?

Упражнения

38.1. Найти решения ЗЛП с помощью графического метода:

1. $z = x_1 - 2x_2 \quad (\min),$
$$\begin{cases} x_1 - x_2 \le 1, \\ x_1 + x_2 \ge 2, \\ x_1 - 2x_2 \le 0, \end{cases}$$
$x_1 \ge 0, \quad x_2 \ge 0.$

2. $z = x_1 + x_2 \quad (\max),$
$$\begin{cases} x_1 + 2x_2 \le 10, \\ x_1 + 2x_2 \ge 2, \\ 2x_1 + x_2 \le 10, \end{cases}$$
$x_1 \ge 0, \quad x_2 \ge 0.$

3. $z = x_1 + 3x_2 \quad (\max),$
$$\begin{cases} x_1 - x_2 \le 1, \\ 2x_1 + x_2 \le 2, \\ x_1 - x_2 \ge 0, \end{cases}$$
$x_1 \ge 0; \quad x_2 \ge 0.$

4. $z = x_1 + x_2 \quad (\max),$
$$\begin{cases} x_1 - 4x_2 - 4 \le 0, \\ 3x_1 - x_2 \ge 0, \\ x_1 + x_2 - 4 \ge 0, \end{cases}$$
$x_1 \ge 0, x_2 \ge 0.$

5. $z = x_1 - x_2 \quad (\max),$
$$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le -8, \\ x_1 + x_2 \le 5, \end{cases}$$
$x_1 \ge 0, x_2 \ge 0.$

6. $z = 2x_1 - x_2 \quad (\min),$
$$\begin{cases} x_1 + x_2 \ge 4, \\ -x_1 + 2x_2 \le 2, \\ x_1 + 2x_2 \le 10, \end{cases}$$
$x_1 \ge 0, x_2 \ge 0.$

38.2. Решить симплексным методом следующие ЗЛП:

1. $z = 4x_1 + 2x_2 \to \max$
$$\begin{cases} x_1 - 2x_2 \le 2, \\ -3x_1 + 2x_2 \le 2, \\ 4x_1 + x_2 \le 8, \end{cases}$$
$x_j \ge 0, j = \overline{1,2}.$

2. $z = 11x_1 + 3x_2 \to \max$
$$\begin{cases} 7x_1 - 3x_2 \le 21, \\ -5x_1 + 6x_2 \le 12, \\ 9x_1 + 10x_2 \le 90, \end{cases}$$
$x_j \ge 0, j = \overline{1,2}.$

3. $z = x_1 + x_2 + x_3 \to \min$

$$\begin{cases} x_1 - x_4 - 2x_6 = 5, \\ x_2 - 2x_4 - 3x_5 + x_6 = 3, \\ x_3 + 2x_4 - 5x_5 + 6x_6 = 5, \end{cases}$$

$x_j \geq 0,\ j = \overline{1,6}.$

4. $z = -x_1 + 3x_2 + 2x_3 \to \min$

$$\begin{cases} x_1 + x_2 + 2x_3 \geq -5, \\ 2x_1 - 3x_2 + x_3 \leq 3, \\ 2x_1 - 5x_2 + 6x_3 \leq 5, \end{cases}$$

$x_j \geq 0,\ j = \overline{1,3}.$

38.3. На швейной фабрике для изготовления четырех видов изделий может быть использована ткань трех артикулов. Нормы затрат ткани всех артикулов для пошива одного изделия приведены в табл. 38.10.

В таблице также указано общее количество ткани каждого артикула и цена одного изделия определенного вида, который есть в распоряжении фабрики.

Таблица 38.10

Исходные данные

Артикул ткани	Норма затрат ткани (м) на один вид изделия				Общее количество ткани
	1	2	3	4	
I	1	–	2	1	180
II	–	1	3	2	210
III	4	2	–	4	800
Цена одного изделия (грн)	9	6	4	7	

С помощью симплекс-метода определить, сколько изделий каждого вида должна изготовить фабрика, чтобы стоимость изготовленной продукции была минимальной.

38.4. Предприятие выпускает четыре вида продукции и использует три типа основного оборудования: токарное, фрезеровочное и шлифовальное. Затраты времени на изготовление единицы продукции для каждого из типов оборудования, общий фонд рабочего времени каждого из типов оборудования, прибыль от реализации одного изделия данного вида приведены в табл. 38.11.

Определить объем выпуска каждого из изделий, при котором общая прибыль от их реализации будет максимальной.

Таблица 38.11

Исходные данные

Тип оборудования	Затраты времени (станко-час) на единицу продукции вида				Общий фонд рабочего времени
	1	2	3	4	
Токарное	2	1	1	3	200
Фрезеровальное	1	–	2	1	200
Шлифовальное	1	2	1	–	200
Прибыль от реализации единицы продукции (грн)	8	3	2	1	

38.5. Для производства столов и шкафов мебельная фабрика использует необходимые ресурсы. Нормы затрат на одно изделие данного вида, прибыль от реализации одного изделия и общее количество имеющихся ресурсов каждого вида приведены в табл. 38.12.

Таблица 38.12

Исходные данные

Ресурсы	Нормы затрат ресурсов на одно изделие		Общее количество ресурсов
	стол	шкаф	
Древесина (м^3) I вида	0,2	0,1	40
II вида	0,1	0,3	60
Трудоемкость (человеко-час)	1,2	1,5	371,4
Прибыль от реализации одного изделия (грн)	6	8	

Определить, сколько столов и шкафов фабрике необходимо изготавливать, чтобы прибыль от реализации была максимальной.

38.6. Решить с помощью метода искусственного базиса ЗЛП:

$$Z(\mathbf{X}) = 800x_1 + 400x_2 + 600x_3 + 500x_4 \quad \to \quad \min$$

при ограничениях $\begin{cases} 10x_1 + 3x_2 + 8x_3 + 2x_4 \geq 5; \\ 90x_1 + 150x_2 + 75x_3 + 175x_4 \geq 100; \\ 45x_1 + 25x_2 + 20x_3 + 37x_4 \geq 30; \\ x_1 + x_2 + x_3 + x_4 = 1; \; x_j \geq 0, \quad j = \overline{1,4}. \end{cases}$

39. Двойственность в линейном программировании

39.1. Теория двойственности для пары взаимно двойственных задач

Произвольную задачу линейного программирования можно определенным образом сопоставить с другой задачей линейного программирования, называемой ***двойственной***. Первоначальная задача по отношению к двойственной является исходной. Эти две задачи тесно связаны между собой и образуют единую двойственную пару.

Сформулируем эти задачи.

Прямая задача

Для производства n видов изделий $A_1, A_2, ..., A_n$ используется сырье m видов $B_1, B_2, ..., B_m$. Заданы запасы сырья $b_1, b_2, ..., b_m$, нормы затрат сырья на изготовление единицы изделия (матрица $A_{m\times n} = \left(a_{ij}\right)$, где a_{ij} – количество i-го сырья, необходимого для изготовления единицы j-го изделия), прибыль $c_1, c_2, ..., c_n$ от реализации единицы изделия. Требуется составить план производства изделий, обеспечивающий максимальную прибыль.

Матрица исходных данных задачи представлена в табл. 39.1.

Таблица 39.1

Исходные данные

Изделия / Сырье	A_1	A_2	…	A_n	Запас	Условные цены ресурсов
B_1	a_{11}	a_{12}	…	a_{1n}	b_1	y_1
B_2	a_{21}	a_{22}	…	a_{2n}	b_2	y_2
…	…	…	…	…	…	…
B_m	a_{m1}	a_{m2}	…	a_{mn}	b_m	y_m
Прибыль	c_1	c_2	…	c_n	–	–
Количество выпускаемой продукции	x_1	x_2	…	x_n	–	–

Математическая модель исходной задачи имеет вид:

$$Z = \sum_{j=1}^{n} c_j x_j \to \max,$$

$$\begin{cases} a_{11}x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n \le b_1, \\ a_{21}x_1 + a_{22}\,x_2 + \ldots + a_{2n}\,x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}\,x_2 + \ldots + a_{mn}\,x_n \le b_m, \\ x_j \ge 0, i = \overline{1,m},\ j = \overline{1,n}. \end{cases}$$

Составим двойственную задачу.

Пусть некоторая организация решила закупить ресурсы $B_1, B_2, \ldots, B_m$ данного предприятия. Необходимо установить оптимальные цены на ресурсы $y_1, y_2, \ldots, y_m$. При этом требуется удовлетворить следующие условия:

1. Покупающая организация стремится минимизировать стоимость всех ресурсов $f = \sum_{i=1}^{m} b_i y_i \to \min$.

2. За каждый вид ресурсов необходимо заплатить не менее той суммы, которую предприятие могло бы получить в виде прибыли при переработке сырья в готовую продукцию, то есть

$$a_{j1}y_1 + a_{j2}\,y_2 + \ldots + a_{jm}y_m \ge c_j,\ j = \overline{1,n}.$$

Получим математическую модель двойственной задачи:

$$f = \sum_{i=1}^{m} b_i y_i \to \min,$$

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \ge c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \ge c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \ge c_n, \\ y_i \ge 0, i = \overline{1,m}. \end{cases}$$

Сравним математические модели исходной и двойственной задач:

число неизвестных одной задачи равно числу неравенств другой задачи;

матрицу коэффициентов системы ограничений одной задачи получаем путем транспонирования матрицы коэффициентов системы ограничений другой задачи;

неравенства в системах ограничений противоположного смысла, однако условие неотрицательности переменных сохраняется;

свободные члены системы ограничений одной задачи являются коэффициентами функции цели другой задачи;

если одна задача исследует минимум функции цели, то другая – максимум и наоборот.

Для первой задачи вторая является двойственной, но для второй задачи первая тоже является двойственной. Поэтому речь идет о паре взаимно двойственных задач.

Различают симметричные и несимметричные двойственные задачи.

Модели симметричных двойственных задач

Исходная задача:

1. $Z(X) = \sum_{j=1}^{n} c_j x_j \to \max,$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i, \\ x_j \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases}$$

2. $Z(X) = \sum_{j=1}^{n} c_j x_j \to \min,$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \ge b_i, \\ x_j \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases}$$

Двойственная задача:

1. $f(Y) = \sum_{i=1}^{m} b_i y_i \to \min,$

$$\begin{cases} \sum_{i=1}^{m} a_{ij} y_i \ge c_j, \\ y_i \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases}$$

2. $f(Y) = \sum_{i=1}^{m} b_i y_i \to \max,$

$$\begin{cases} \sum_{i=1}^{m} a_{ij} y_i \le c_j, \\ y_i \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases}$$

Исходная задача:

1. $Z(X)=\sum\limits_{j=1}^{n}c_jx_j\rightarrow\max,$

$$\begin{cases}\sum\limits_{j=1}^{n}a_{ij}x_j=(\geq)b_i,\\ x_j\geq 0,\ i=\overline{1,m},\ j=\overline{1,n}.\end{cases}$$

2. $Z(X)=\sum\limits_{j=1}^{n}c_jx_j\rightarrow\min,$

$$\begin{cases}\sum\limits_{j=1}^{n}a_{ij}x_j=(\leq)b_i,\\ x_j\geq 0,\ i=\overline{1,m},\ j=\overline{1,n}.\end{cases}$$

Для составления математической модели двойственной задачи пользуются тем же правилом, что и для составления симметричной задачи, с учетом следующих особенностей:

если система ограничений исходной задачи содержит неравенства, то их надо привести к одному смыслу (стандартной форме), то есть, если $Z_{\max}$, то ограничения типа $\leq$, если $Z_{\min}$ – $\geq$ (для этого следует умножить соответствующее неравенство на (-1));

если в исходной задаче функция цели исследуется на $\max$, то в двойственной – на $\min$ и система ограничений должна состоять из неравенств типа $\geq$; если в исходной задаче функция цели исследуется на $\min$, то в двойственной – на $\max$ и система ограничений должна состоять из неравенств типа $\leq$;

условие неотрицательности y_i не соблюдается.

Математическая модель двойственной задачи имеет вид:

1. $f(Y)=\sum\limits_{i=1}^{m}b_iy_i\rightarrow\min,$

$$\begin{cases}\sum\limits_{i=1}^{m}a_{ij}y_i\geq c_j,\\ i=\overline{1,m},\ j=\overline{1,n}.\end{cases}$$

2. $f(Y)=\sum\limits_{i=1}^{m}b_iy_i\rightarrow\max,$

$$\begin{cases}\sum\limits_{i=1}^{m}a_{ij}y_i\leq c_j,\\ i=\overline{1,m},\ j=\overline{1,n}.\end{cases}$$

Пример 39.1. Составить математическую модель двойственной задачи к исходной:

$$Z=3x_1+2x_2\rightarrow\max,$$

$$\begin{cases} 9x_1 + 5x_2 \le 1431, \\ 7x_1 + 8x_2 \le 1224, \\ 4x_1 + 16x_2 \le 1328, \end{cases}$$
$$x_1, x_2 \ge 0.$$

Решение.

Пусть y_1, y_2, y_3 – переменные двойственной задачи (их количество определено исходя из числа ограничений исходной задачи).

Тогда функция цели двойственной задачи имеет вид:

$$f = 1434y_1 + 1224y_2 + 1328y_3 \to \min.$$

Матрица системы ограничений двойственной задачи является транспонированной матрицей по отношению к соответствующей матрице системы ограничений исходной задачи.

То есть, если $A = \begin{pmatrix} 9 & 5 \\ 7 & 8 \\ 4 & 16 \end{pmatrix}$, то матрица $A^T = \begin{pmatrix} 9 & 7 & 4 \\ 5 & 8 & 16 \end{pmatrix}$.

Таким образом, система ограничений двойственной задачи имеет вид:

$$\begin{cases} 9y_1 + 7y_2 + 4y_3 \ge 3, \\ 5y_1 + 8y_2 + 16y_3 \ge 2, \\ y_i \ge 0,\ i = \overline{1,3}. \end{cases}$$

Пример 39.2. Составить математическую модель двойственной задачи к исходной:

$$Z = -3x_1 + x_2 \to \min$$
$$\begin{cases} -2x_1 + x_2 \le 2, \\ 2x_1 + x_2 \ge 3, \\ 2x_1 + x_2 \le 6, \\ -x_1 + x_2 \ge 0, \end{cases}$$
$$x_1, x_2 \ge 0.$$

Решение.

Приведем систему ограничений к стандартной форме. Так как функция цели исходной задачи исследуется на $\min$, то ограничения приводим к неравенствам типа $\geq$:

$$\begin{cases} 2x_1 - x_2 \geq -2, \\ 2x_1 + x_2 \geq 3, \\ -2x_1 - x_2 \geq -6, \\ -x_1 + x_2 \geq 0. \end{cases}$$

Переменные двойственной задачи – y_1, y_2, y_3, y_4.

Тогда математическая модель двойственной задачи имеет вид:

$$f = -2y_1 + 3y_2 - 6y_3 \to \max,$$

$$\begin{cases} 2y_1 + 2y_2 - 2y_3 - y_4 \leq -3, \\ -y_1 + y_2 - y_3 + y_4 \leq 1. \end{cases}$$

39.2. Первая и вторая теоремы двойственности

Фундаментальные свойства, которыми обладают двойственные задачи линейного программирования, могут быть сформулированы в виде приводимых ниже утверждений. Их обычно называют *теоремами двойственности*.

Теорема 1. Если одна из двойственных задач имеет оптимальный план $\overline{X}^*$, то и другая также имеет оптимальный план $\overline{Y}^*$, причем значения функций цели для оптимальных планов $\overline{X}^*$ и $\overline{Y}^*$ совпадают, то есть выполняется равенство

$$Z\left(\overline{X}^*\right)_{\max} = f\left(\overline{Y}^*\right)_{\min} \text{ или } Z\left(\overline{X}^*\right)_{\min} = f\left(\overline{Y}^*\right)_{\max}.$$

Если одна из двойственных задач неразрешима так как $Z\left(\overline{X}^*\right)_{\max} \to \infty$ (или $Z\left(\overline{X}^*\right)_{\min} \to -\infty$), то другая задача не имеет решений.

Теорема 2. Если при подстановке компонент оптимального плана в систему ограничений исходной задачи i-е ограничение выполняется как строгое неравенство, то i-я компонента оптимального плана двойственной задачи равна нулю ($y_i = 0$); если i-е ограничение выполняется как строгое равенство, то i-я компонента оптимального плана двойственной задачи отлична от нуля ($y_i \neq 0$).

Верно и обратное, то есть, если j-я компонента оптимального плана исходной задачи равна нулю ($x_j = 0$), то j-е ограничение двойственной задачи выполняется как строгое неравенство, если j-я компонента оптимального плана исходной задачи не равна нулю ($x_j \neq 0$), то j-е ограничение двойственной задачи выполняется как строгое равенство.

Практическое значение теорем двойственности состоит в том, что они позволяют заменить процесс решения основной задачи на решение двойственной, которое в определенных случаях может оказаться более простым. Например, задача, область допустимых значений которой описывается двумя уравнениями, связывающими шесть переменных ($m = 2, n = 6$), не может быть решена графическим методом. Однако данный метод может быть применен для решения двойственной к ней задачи, которая имеет только две переменные.

Пример 39.3. Найти решение двойственной задачи с помощью оптимального решения: $\overline{X}^* = (44; 27)$, $Z_{\max} = 486$ исходной задачи:

$$Z = 3x_1 + 2x_2 \to \max,$$

$$\begin{cases} 9x_1 + 5x_2 \leq 1431, \\ 7x_1 + 8x_2 \leq 1224, \\ 4x_1 + 16x_2 \leq 1328, \end{cases}$$

$$x_1, x_2 \geq 0.$$

Решение.

Составим математическую модель двойственной задачи:

$$f = 1434y_1 + 1224y_2 + 1328y_3 \to \min,$$

$$\begin{cases} 9y_1 + 7y_2 + 4y_3 \geq 3, \\ 5y_1 + 8y_2 + 16y_3 \geq 2, \\ y_i \geq 0,\ i = \overline{1,3}. \end{cases}$$

По первой теореме двойственности: $f_{\min} = Z_{\max} = 486.$

По второй теореме двойственности:

1) первая и вторая компоненты оптимального плана (решения) исходной задачи $x_1 = 144 \neq 0,\ x_2 = 27 \neq 0$, поэтому первое и второе ограничения двойственной задачи выполняются как равенства;

2) если компоненты оптимального плана исходной задачи подставить в систему ограничений, то:

$$\begin{cases} 9 \cdot 144 + 5 \cdot 27 = 1431 \Rightarrow y_1 \neq 0, \\ 7 \cdot 144 + 8 \cdot 27 = 1224 \Rightarrow y_2 \neq 0, \\ 4 \cdot 144 + 16 \cdot 27 = 1008 < 1328 \Rightarrow y_3 = 0. \end{cases}$$

Таким образом, имеем: $\begin{cases} 9y_1 + 7y_2 + 4y_3 = 3, \\ 5y_1 + 8y_2 + 16y_3 = 2, \\ y_3 = 0. \end{cases}$

Решаем эту систему:

$$\begin{cases} 9y_1 + 7y_2 = 3, \\ 5y_1 + 8y_2 = 2, \\ y_3 = 0, \end{cases} \Rightarrow \begin{cases} y_1 = \dfrac{10}{37}, \\ y_2 = \dfrac{3}{37}, \\ y_3 = 0. \end{cases}$$

Итак, получили: $Y_{опт} = \begin{pmatrix} \dfrac{10}{37} & \dfrac{3}{37} & 0 \end{pmatrix}.$

Рассмотрим первую теорему двойственности более детально.

Запишем исходную и двойственную задачи в матричной форме.

Пусть $\underset{m\times n}{A} = (a_{ij})$ – матрица системы ограничений,

$A_0 = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$, $X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$ – столбцы свободных членов системы ограничений и переменных исходной задачи соответственно,

$C = (c_1, c_2, ..., c_n)$ – строка коэффициентов функции цели исходной задачи,

$Y = (y_1, y_2, ..., y_m)$ – строка двойственных переменных.

Тогда исходная задача: $Z = C \cdot X \ (\max),\ AX = A_0,\ X \geq 0,\ A_0 \geq 0.$

Двойственная задача: $f = Y \cdot A_0 \ (\min),\ A^T Y = C.$

Предположим, что при решении симплекс-методом исходной задачи получено m базисных векторов, а именно $A_1, A_2, ..., A_m$ в оптимальном плане, то есть практически в симплекс-таблице решается матричное уравнение $D \cdot X = A_0$, где D – матрица из элементов исходной матрицы A при базисных неизвестных.

Решим матричное уравнение:

$$D \cdot X = A_0,$$

$$D^{-1} \cdot D \cdot X = D^{-1} \cdot A_0,$$

$$X_{\max} = D^{-1} \cdot A_0 \text{ – оптимальное решение.}$$

Оптимальное решение двойственной задачи можно найти в следующем виде:

$$Y_{\min} = C \cdot D^{-1}.$$

Проверим первую теорему двойственности:

$$f_{\min} = Y \cdot A_0 = C \cdot D^{-1} \cdot A_0 = C \cdot X_{\max} = Z_{\max}, \text{ то есть } f_{\min} = Z_{\max}.$$

Решение $Y_{\min} = C \cdot D^{-1}$ находится в строке z_j последней итерации симплекс-таблицы.

Таким образом:

если известен оптимальный план исходной задачи, то оптимальный план двойственной задачи может быть найден или с помощью обратной матрицы $Y_{опт} = C \cdot D^{-1}$ или с использованием второй теоремы двойственности;

если в системе ограничений исходной задачи есть исходный единичный базис, то решение двойственной задачи находится в строке z_j, соответственно расположению единиц в исходном единичном базисе;

если при решении исходной задачи симплекс-методом единичный базис выделяется путем преобразования исходной матрицы, то решение двойственной задачи находят с помощью второй теоремы двойственности;

если исходная задача решается методом искусственного базиса, то решение двойственной задачи находится в строке z_j напротив единичных векторов искусственного базиса.

39.3. Экономическая интерпретация теорем двойственности

Традиционная экономическая интерпретация двойственной ЗЛП базируется на модели *простейшей задачи производственного планирования.*

Пусть $\overline{X}^* = (x_1^*, x_2^*, \ldots, x_n^*)$, $\overline{Y}^* = (y_1^*, y_2^*, \ldots, y_m^*)$ – оптимальные планы прямой и двойственной задач.

Тогда $Z_{\max} = c_1 x_1^* + c_2 x_2^* + \ldots + c_n x_n^*$ – максимальная прибыль реализации продукции, $f_{\min} = b_1 y_1^* + b_2 y_2^* + \ldots + b_m y_m^*$ – минимальные затраты на сырье.

По первой теореме двойственности $f_{\min} = Z_{\max}$, то есть весь тот расход, который вложен в ресурсы, компенсируется полученной прибылью и производство может функционировать.

Так как $f_{\min} = Z_{\max}$, то

$$Z_{\max} = b_1 y_1^* + b_2 y_2^* + \ldots + b_m y_m^*, \quad \frac{\partial Z_{\max}}{\partial b_i} = y_i^*,$$

$$\text{то есть } \frac{\Delta Z_{\max}}{\Delta b_i} \approx y_i^*, \ \Delta Z_{\max} \approx \Delta b_i \cdot y_i^*,$$

где Δb_i – увеличение запаса i-го ресурса;

$\Delta Z_{\max}$ – изменение прибыли при увеличении запаса i-го ресурса.

Если $\Delta b_i = 1$, то $\Delta Z_{\max} \approx y_i^*$, то есть y_i^* показывает, на сколько возрастает прибыль в оптимальном плане при увеличении уровня запаса i-го ресурса на единицу:

$$Z_{нов\max} = Z_{исх\max} + \Delta Z_{\max}.$$

Если y_i^* мало, то при значительном увеличении i-го ресурса будет получено небольшое увеличение оптимальной прибыли и поэтому ценность i-го ресурса небольшая.

Если $y_i^* = 0$, то при увеличении i-го ресурса оптимальная прибыль остается неизменной и ценность i-го ресурса равна нулю. То есть сырье, запасы которого превышают потребности в нем, не представляет ценности для производства и поэтому оценку его ценности можно принять равной нулю.

Если y_i^* велико, то при незначительном увеличении i-го ресурса будет получено существенное увеличение оптимальной прибыли и ценность i-го ресурса высокая. Уменьшение ресурса ведет к существенному уменьшению выпуска продукции.

y_i^* можно считать некоторой характеристикой ценности i-го ресурса (их называют ***двойственными оценками***, или *теневыми ценами*).

Таким образом, имея решение двойственной задачи, можем решать задачу перераспределения ресурсов, оценивая прибыль от реализации дополнительных ресурсов.

Рассмотрим вторую теорему двойственности.

В прямой задаче:

1) $a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n < b_i \Rightarrow y_i = 0;$ (39.1)

2) $a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = b_i \Rightarrow y_i \neq 0.$ (39.2)

В левых частях (39.1) и (39.2) стоит расход i-го ресурса на изготовление всей продукции. Если выполняется строгое неравенство (39.1), то расход i-го ресурса меньше запаса, то есть закупать его не следует, поэтому ценность этого сырья равна нулю. Если выполняется равенство (39.2), то i-й ресурс израсходован полностью, следует запланировать закупку i-го ресурса на следующий плановый период, то есть ценность этого вида сырья $y_i \neq 0$.

В двойственной задаче:

1) $a_{1j}y_1 + a_{2j}y_2 + ... + a_{mj}y_m > c_j \Rightarrow x_j = 0$; (39.3)

2) $a_{1j}y_1 + a_{2j}y_2 + ... + a_{mj}y_m = c_j \Rightarrow x_j \neq 0$. (39.4)

В левых частях (39.3) и (39.4) стоит расход на сырье, необходимое для изготовления единицы j-го вида продукции. Если выполняется строгое неравенство (39.3), то расход на сырье больше прибыли от реализации единицы j-го вида продукции, то есть предприятие работает нерентабельно, нет самоокупаемости, значит j-й вид продукции не следует выпускать на этом предприятии ($x_j = 0$). Если выполняется равенство (39.4), то расход на ресурсы компенсируется полученной прибылью, то есть j-й вид продукции выпускать следует $x_j \neq 0$.

Устойчивость двойственных оценок

Так как $\Delta Z_{\max} \approx \Delta b_i \cdot y_i^*$, то требуется определить, в каких же пределах можно изменять Δb_i.

Пусть найдено

$$X_{опт} = D^{-1} \cdot A_0,$$

где D – исходная матрица коэффициентов для оптимального базиса; $D^{-1} = \begin{pmatrix} a_{1,n-m}^* & ... & a_{1n}^* \\ a_{2,n-m}^* & ... & a_{2n}^* \\ ... & ... & ... \\ a_{m,n-m}^* & ... & a_{mn}^* \end{pmatrix}$ – матрица в последней итерации симплекс-таблицы на месте векторов исходного единичного базиса;

A_0 – столбец свободных членов системы ограничений.

Изменим A_0 и получим столбец $A_0^I = \begin{pmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \dots\dots\dots \\ b_m + \Delta b_m \end{pmatrix}$,

тогда $X_{опт}^I = D^{-1} \cdot A_0^I = \begin{pmatrix} a_{1,n-m}^* (b_1 + \Delta b_1) + \dots + a_{1n}^* (b_1 + \Delta b_1) \\ a_{2,n-m}^* (b_2 + \Delta b_2) + \dots + a_{2n}^* (b_2 + \Delta b_2) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m,n-m}^* (b_m + \Delta b_m) + \dots + a_{mn}^* (b_m + \Delta b_m) \end{pmatrix}$.

Так как $X_{опт}^I \geq 0$ (все $x_j \geq 0$ в оптимальном плане), то получаем систему неравенств для определения Δb_i:

$$\begin{cases} a_{1,n-m}^* (b_1 + \Delta b_1) + \dots + a_{1n}^* (b_1 + \Delta b_1) \geq 0, \\ a_{2,n-m}^* (b_2 + \Delta b_2) + \dots + a_{2n}^* (b_2 + \Delta b_2) \geq 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m,n-m}^* (b_m + \Delta b_m) + \dots + a_{mn}^* (b_m + \Delta b_m) \geq 0. \end{cases} \qquad (39.5)$$

Считая, что все $\Delta b_i = 0$, кроме одного, находим пределы возможных изменений величины ресурсов. Если запасы ресурсов изменять одновременно и эти конкретные изменения удовлетворяют системе (39.5), то $\overline{X}_{опт} \geq 0$ и оптимальное решение двойственной задачи остается прежним, поэтому изменение максимальной прибыли можно найти по формуле $\Delta Z_{\max} = \Delta b_1 y_1^* + \Delta b_2 y_2^* + \dots + \Delta b_m y_m^*$.

39.4. Понятие о двойственном симплекс-методе

Непосредственное приложение теории двойственности к вычислительным алгоритмам линейного программирования позволило разработать еще один метод решения ЗЛП, получивший название ***двойственного симплекс-метода***, или *метода последовательного уточнения оценок*.

Следует подчеркнуть, что двойственный симплекс-метод непосредственно используется для поиска решения прямой задачи.

Рассмотрим исходную задачу минимизации:

$$Z = \sum_{j=1}^{n} c_j x_j \to \min, \tag{39.6}$$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \ge b_i, \\ x_j \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases} \tag{39.7}$$

Приведем задачу к канонической форме, введя дополнительные переменные $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$:

$$Z = \sum_{j=1}^{n} c_j x_j \to \min, \tag{39.8}$$

$$\begin{cases} -\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = -b_i, \\ x_j \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases} \tag{39.9}$$

Задача, двойственная к (39.8) и (39.9), будет иметь вид:

$$f = \sum_{i=1}^{m} b_i y_i \to \max, \tag{39.10}$$

$$\begin{cases} \sum_{i=1}^{m} a_{ij} y_i \le c_j, \\ y_i \ge 0,\ i = \overline{1,m},\ j = \overline{1,n}. \end{cases} \tag{39.11}$$

Исходный базис образуют векторы $P_{n+1}, P_{n+2}, \ldots, P_{n+m}$, причем не все $b_i > 0$, есть и $b_i < 0$.

При этом начальный план равен

$$\begin{cases} x_j = 0,\ j = \overline{1,n}; \\ x_{n+i} = -b_i,\ i = \overline{1,m}. \end{cases}$$

Критерий оптимальности плана $\overline{X}$ в двойственном симплекс-методе заключается в том, что $\overline{X}$ одновременно должен являться допустимым планом прямой задачи, то есть все его компоненты должны быть неотрицательны ($x_j \geq 0$). Обратно, если хотя бы одна из компонент плана является отрицательной, то процесс улучшения значения целевой функции должен быть продолжен (по схеме симплекс-метода).

Двойственным симплекс-методом можно решать задачи линейного программирования, у которых при положительном базисе свободные члены системы ограничений могут быть отрицательными, то есть метод позволяет не выполнять дополнительных преобразований и уменьшает размеры симплекс-таблицы.

Пример 39.4. Найти оптимальное решение задачи:

$$Z = -2x_1 + x_2 + 5x_3 \to \min,$$

$$\begin{cases} x_1 + x_2 - x_3 \leq 4, \\ x_1 - 5x_2 + x_3 \geq 5, \; x_j \geq 0, \; j = \overline{1,3}. \end{cases}$$

Решение.

Умножим второе неравенство на (-1) и приведем систему к канонической форме:

$$\begin{cases} x_1 + x_2 - x_3 + x_4 = 4, \\ -x_1 + 5x_2 - x_3 + x_5 = -5, \; x_j \geq 0, \; j = \overline{1,5}. \end{cases}$$

Далее решаем задачу с помощью симплекс-таблицы (табл. 39.2).

Таблица 39.2

Симплекс-таблица двойственного симплекс-метода

№ п/п	Базис	$\overline{C_{баз}}$	c_j	–2	1	5	0	0	Примечания
			$\overline{A_0}$	$\overline{A_1}$	$\overline{A_2}$	$\overline{A_3}$	$\overline{A_4}$	$\overline{A_5}$	
1	2	3	4	5	6	7	8	9	10
1	$\overline{A_4}$	0	4	1	1	–1	1	0	$\Delta_1 > 0$ – нарушен критерий оптимальности, введем в базис $\overline{A_1}$
2	$\overline{A_5}$	0	–5	–1	5	–1	0	1	
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			0	0	0	0	0	0	$\theta_1 = \min\left(\frac{4}{1}; \frac{-5}{-1}\right) = 4$
$\Delta_j = z_j - c_j$			–	2	–1	–5	0	0	

Продолжение табл. 39.2

1	2	3	4	5	6	7	8	9	10
3	$\overline{A}_1$	–2	4	1	1	–1	1	0	[1]
4	$\overline{A}_5$	0	–1	0	6	–2	1	1	[1]+[2]
$z_j = \overline{C}_{баз} \cdot \overline{A}_j$			–8	–2	–2	2	–2	0	$\Delta_j \le 0$, но $x_5 = -1 < 0$. Выберем (–2) ведущим элементом
$\Delta_j = z_j - c_j$			–	0	–3	–3	–2	0	
5	$\overline{A}_1$	–2	9/2	1	–2	0	1/2	–1/2	[6]+[3]
6	$\overline{A}_3$	5	1/2	0	–3	1	– 1/2	–1/2	[4]/(–2)
$z_j = \overline{C}_{баз} \cdot \overline{A}_j$			–13/2	–2	–11	5	–7/2	–3/2	$\Delta_j \le 0$, $x_j > 0$
$\Delta_j = z_j - c_j$			–	0	–12	0	–7/2	–3/2	

Оптимальный план $\overline{X}_{min} = (9/2; 0; 1/2)$, $Z_{min} = -13/2$.

Таким образом, при решении задач вида (39.6) и (39.7) двойственный симплекс-метод имеет несомненные преимущества по сравнению с прямым.

Еще одно важное направление использования двойственного симплекс-метода связано с поиском оптимальных планов в тех задачах, условия которых претерпели некоторые изменения после того, как они уже были решены с помощью стандартной симплекс-процедуры.

Типичными примерами таких изменений являются:

изменение компонент вектора ограничений b, что, допустим, может быть интерпретировано как корректировка объемов доступных ресурсов в процессе управления экономическим объектом (тогда ранее найденный оптимальный базис можно использовать в качестве исходного базиса при продолжении решения);

добавление новых ограничений к системе условий задачи, что достаточно часто случается при совершенствовании используемой экономико-математической модели (в задачах целочисленного программирования, которые будут рассмотрены позже).

Вопросы для самодиагностики

1. Как построить математическую модель двойственной задачи по модели исходной для симметричной пары сопряженных задач?

2. Как построить математическую модель двойственной задачи по модели исходной для несимметричной пары сопряженных задач?

3. Как найти решение двойственной задачи по оптимальному плану исходной с помощью теорем двойственности?

4. Привести экономическое толкование двойственных оценок в задачах линейного программирования.

5. Как провести анализ устойчивости оптимального плана относительно рыночной цены готовой продукции?

6. Как определить состояние ресурсов прямой задачи и интервалов устойчивости двойственных оценок относительно изменений запасов дефицитных ресурсов?

7. В чем суть двойственного симплекс-метода?

Упражнения

39.1. К данным задачам составить двойственные:

1. $z = x_1 - 2x_2 + 3x_3 - x_4 \quad \max$,

$$\begin{cases} 2x_1 - x_2 + 2x_3 - 3x_4 \le 5, \\ x_1 + 2x_2 - x_3 + x_4 \le 3, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,4}.$

2. $z = 3x_2 - x_4 \quad \max$,

$$\begin{cases} x_1 - 2x_2 \quad + x_4 = 8, \\ \qquad x_2 + x_3 - 3x_4 = 6, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,4}.$

3. $z = -x_1 + x_2 + x_3 \quad \max$,

$$\begin{cases} x_1 - 3x_2 + x_3 \ge 4, \\ 2x_1 + x_2 - 2x_3 \ge 1, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,3}.$

4. $z = 2x_1 + 4x_2 + x_3 \quad \min$,

$$\begin{cases} x_1 - 2x_2 + x_3 \le 4, \\ 2x_1 + 3x_2 - x_3 \ge 2, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,3}.$

39.2. Сформулировать двойственные задачи по отношению к данным и найти их решение:

1. $F = 6x_1 + 8x_2 + x_3 \to \max,$

$$\begin{cases} x_1 + x_2 + x_3 \le 3, \\ x_1 + 2x_2 \le 4, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,3}.$

2. $F = 27x_1 + 10x_2 + 15x_3 + 28x_4 \to \max,$

$$\begin{cases} 3x_1 + 2x_2 + x_3 + 2x_4 \le 2, \\ 3x_1 + x_2 + 3x_3 + 4x_4 \le 5, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,4}.$

3. $z = -x_1 + x_2 + x_3 \quad \max$,

$$\begin{cases} x_1 - 3x_2 + x_3 \ge 4, \\ 2x_1 + x_2 - 2x_3 \ge 1, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,3}.$

4. $z = 2x_1 + 4x_2 + x_3 \quad \min$,

$$\begin{cases} x_1 - 2x_2 + x_3 \le 4, \\ 2x_1 + 3x_2 - x_3 \ge 2, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,3}.$

5. $z = x_1 - 2x_2 + 3x_3 - x_4 \quad \max$,

$$\begin{cases} 2x_1 - x_2 + 2x_3 - 3x_4 \le 5, \\ x_1 + 2x_2 - x_3 + x_4 \le 3, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,4}.$

6. $z = 3x_2 - x_4 \quad \max$,

$$\begin{cases} x_1 - 2x_2 \quad + x_4 = 8, \\ \quad x_2 + x_3 - 3x_4 = 6, \end{cases}$$

$x_j \ge 0,\ j = \overline{1,4}.$

39.3. Определить являются ли данные векторы $\overline{X}^*$ и $\overline{Y}^*$ оптимальными решениями данной задачи и двойственной к ней:

1. $z = x_1 + 10x_2 + 8x_3 \quad \max$,

$$\begin{cases} x_1 + 4x_2 + x_3 = 2, \\ x_1 + 2x_2 - x_3 = 0, \end{cases}$$

$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0,$

$\overline{X}^* = (1, 0, 1),\ \overline{Y}^* = \left(\frac{9}{2}, -\frac{7}{2}\right).$

2. $z = x_1 + 4x_2 + x_3 \quad \max$,

$$\begin{cases} 5x_1 + 12x_2 + 2x_3 = 9, \\ 3x_1 + 10x_2 + 4x_3 = 11, \end{cases}$$

$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0,$

$\overline{X}^* = (1, 0, 2),\ \overline{Y}^* = \left(\frac{3}{14}, \frac{1}{14}\right).$

39.4. Решить следующие задачи двойственным симплекс-методом:

1. $z = x_1 - 2x_2 \ (\max)$,

$$\begin{cases} 5x_1 - 2x_2 \le 3; \\ x_1 + x_2 \ge 1; \\ -3x_1 + x_2 \le 3; \end{cases}$$

$x_j \ge 0,\ j = \overline{1,2}.$

2. $z = -7x_1 + 2x_2 \ (\min)$,

$$\begin{cases} x_1 + x_2 \ge 1; \\ 5x_1 + x_2 \ge 3; \\ -3x_1 + x_2 \le 3; \\ 2x_1 + x_2 \le 4; \end{cases}$$

$x_j \ge 0,\ j = \overline{1,2}.$

40. Методика решения транспортной задачи

40.1. Постановка транспортной задачи, поиск исходного опорного плана

Транспортная задача (ТЗ) – это целый класс оптимизационных задач, связанных с распределением ресурсов при существовании определенных ограничений относительно их количества или условий распределения. ТЗ используется при построении математических моделей, которые связаны с управлением финансами и планированием производства, а также задач управления, которые возникают на уровне регионов и даже на государственном уровне.

Транспортная задача – одна из распространенных задач линейного программирования. Ее цель – разработка наиболее рациональных путей и способов транспортирования товаров, устранение чрезмерно дальних, встречных, повторных перевозок. Все это сокращает время продвижения товаров, уменьшает затраты предприятий, фирм, связанные с осуществлением процессов снабжения сырьем, материалами, топливом, оборудованием и т. д.

В общем виде задачу можно представить следующим образом: в m пунктах производства $A_1, A_2, ..., A_m$ имеется однородный груз в количестве соответственно $a_1, a_2, ..., a_m$. Этот груз необходимо доставить в n пунктов назначения $B_1, B_2, ..., B_n$ в количестве соответственно $b_1, b_2, ..., b_n$. Стоимость перевозки единицы груза (тариф) из пункта A_i в пункт B_j равна c_{ij}.

Требуется составить план перевозок, позволяющий вывезти весь груз, удовлетворить всех потребителей с минимальными затратами.

В зависимости от соотношения между суммарными запасами груза и суммарными потребностями в нем транспортные задачи могут быть закрытыми и открытыми.

Если $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, транспортная задача называется *закрытой*, если $\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j$ – *открытой*.

Для того, чтобы транспортная задача имела решение, необходимо и достаточно, чтобы ее модель была закрытой. При решении открытой задачи вводят фиктивного поставщика (если $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$) или потребителя (если $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$). Открытая транспортная задача становится закрытой и решается по алгоритму для закрытых транспортных задач, причем тарифы, соответствующие фиктивному поставщику или потребителю, считают равными нулю. В целевой функции фиктивный поставщик или потребитель не учитывается.

Обозначим через x_{ij} количество груза, перевозимого из пункта A_i в пункт B_j. Рассмотрим закрытую транспортную задачу. Ее условия запишем в таблицу исходных данных, которую будем использовать для нахождения решения (табл. 40.1).

Таблица 40.1

Матрица исходных данных

A_i \ B_j	B_1	B_2	...	B_j	...	B_n	a_i
A_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1j} x_{1j}	...	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	...	c_{2j} x_{2j}	...	c_{2n} x_{2n}	a_2
...	...	...	...	...	...	...	...
A_i	c_{i1} x_{i1}	c_{i2} x_{i2}	...	c_{ij} x_{ij}	...	c_{in} x_{in}	a_i
...	...	...	...	...	...	...	...
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	...	c_{mj} x_{mj}	...	c_{mn} x_{mn}	a_m
b_j	b_1	b_2	...	b_j	...	b_n	$\sum a_i$ \ $\sum b_j$

Математическая модель закрытой транспортной задачи имеет вид:

$$F(X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\to\min \qquad (40.1)$$

при ограничениях:

$$\sum_{i=1}^{m}x_{ij}=b_j, \qquad (40.2)$$

$$\sum_{j=1}^{n}x_{ij}=a_i, \qquad (40.3)$$

$$x_{ij}\ge 0,\ i=\overline{1,m},\ j=\overline{1,n}.$$

Совокупность значений x_{ij} $(i=\overline{1,m},\ j=\overline{1,n})$, которая удовлетворяет системе ограничений ТЗ, называется ***планом решения ТЗ***. ***Оптимальным решением*** транспортной задачи является матрица $X_{опт}=(x_{ij})_{m\times n}$, удовлетворяющая системе ограничений и доставляющая минимум целевой функции.

Просуммируем ограничения (40.2):

$$\sum_{j=1}^{n}\sum_{i=1}^{m}x_{ij}=\sum_{j=1}^{n}b_j$$

и ограничения (40.3):

$$\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\sum_{i=1}^{m}a_i.$$

Так как $\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j$, то это значит, что система линейно зависима. Достаточно убрать одно из ограничений, чтобы получить независимую систему. Число ограничений $m+n$, поэтому ранг матрицы системы ТЗ: $rang=m+n-1$.

План решения, который содержит $m+n-1$ базисных перевозок $(x_{ij}\neq 0)$, называется ***невырожденным***. Если число перевозок $r>m+n-1$ или $r<m+n-1$, то план называется ***вырожденным***.

Число базисных перевозок – это число занятых клеток в матрице исходных данных ($ЧЗкл$), то есть, если $ЧЗкл=rang$ – план невырожденный, если $ЧЗкл\neq rang$ – план вырожденный.

Особенности транспортной задачи

1. Число уравнений системы ораничений меньше числа неизвестных $m+n<m\cdot n$.

2. Каждая переменная встречается в уравнениях системы ограничений только два раза.

3. Коэффициенты при переменных во всех уравнениях системы ограничений равны 1.

4. Система ограничений задана в виде равенств.

5. Все элементы матрицы системы ограничений равны 0 или 1.

Транспортная задача как задача линейного программирования может быть решена симплекс-методом, однако наличие большого числа переменных и ограничений делает вычисления громоздкими. Поэтому для решения транспортных задач разработан специальный метод, имеющий те же этапы, что и симплексный метод.

Алгоритм решения ТЗ

Алгоритм решения ТЗ состоит в последовательности таких действий:

определение исходного опорного плана;

проверка плана на оптимальность;

улучшение исходного опорного плана, если найденный план не является оптимальным, то есть переход к новому опорному плану с меньшим значением целевой функции;

новый план также проверяют на оптимальность;

если план оптимальный, то он и является решением ТЗ.

Методы построения исходного опорного плана.
Метод северо-западного угла (диагональный метод)

Распределение груза между потребителями начинают с левой верхней клетки (северо-западного угла) таблицы перевозок. Распределяем запасы первого поставщика A_1. Вначале удовлетворяем потребности потребителя B_1 за счет A_1. В клетку x_{11} записываем меньшее из чисел a_1 и b_1, то есть $x_{11} = \min\{a_1; b_1\}$.

Пусть $x_{11} = a_1$, тогда первый поставщик израсходовал весь ресурс и в дальнейшем распределении груза участия не принимает, то есть этот поставщик исключается из рассмотрения.

Если спрос первого потребителя не удовлетворили в полном объеме, тогда в клетку, которая находится во второй строке первого столбца, направляют груз в количестве: $x_{21} = \min(a_2; b_1 - x_{11})$.

Наоборот, если $x_{11} = b_1$, то первый потребитель получил нужный ему груз, тогда из оставшихся ресурсов первого поставщика удовлетворяют нужды второго потребителя, решая при этом следующую задачу: $x_{12} = \min(a_1 - x_{11}; b_2)$ и т. д. Такое распределение проводят до тех пор, пока вся таблица не будет заполнена.

Клетки соответствующего столбца или строки, в которых потребности потребителя удовлетворены и запасы израсходованы, заполняют нулями или прочерками. При определении опорного плана методом северо-западного угла не учитывается стоимость перевозки единицы груза, поэтому этот план может быть далеким от оптимального.

Метод минимальной стоимости

Данный метод позволяет построить опорное решение, которое является достаточно близким к оптимальному. Он состоит из ряда однотипных шагов, на каждом из которых заполняется только одна клетка, которая отвечает минимальной стоимости c_{ij}.

Построение исходного опорного плана начинают с определения клетки, которая имеет наименьшую стоимость перевозок c_{ij}, и туда делают поставку объемом $x_{ij} = \min(a_i; b_j)$.

После этого исключают строку, то есть заполняют нулями, если запасы поставщика израсходованы, или исключают столбец, то есть заполняют нулями, если потребности потребителя полностью удовлетворены. Далее снова выбирают свободную клетку с наименьшей стоимостью и процесс продолжают до тех пор, пока все запасы не будут израсходованы, а потребности удовлетворены.

Если m и n достаточно велики, то применяют:

метод минимальной стоимости в строке, то есть распределяют груз a_1 последовательно по минимальной стоимости первой строки, затем груз a_2 – по минимальной стоимости второй строки и т. д.;

метод минимальной стоимости в столбце, то есть распределяют груз b_1 по минимальной стоимости в первом столбце, затем груз b_2 – по минимальной стоимости во втором столбце и т. д.;

метод двойного предпочтения – выделяют клетки с минимальной стоимостью в каждой строке и клетки с минимальной стоимостью в столбце. Вначале заполняют клетки с двойной отметкой, затем – с одной отметкой. Остальные клетки заполняют любым другим методом.

Как правило, наиболее близкое к оптимальному решению дает метод минимальной стоимости и метод двойного предпочтения.

Пример 40.1. Составить опорный план перевозки однородного груза от пунктов производства A_1, A_2 и A_3 к пунктам потребления B_1, B_2, B_3 и B_4, если возможности поставщиков соответственно равны 180, 400, 280 условных единиц, а потребности каждого из потребителей составляют 240, 320, 120 и 180 условных единиц соответственно.

Тарифы перевозок заданы в виде матрицы стоимости:

$$C = \begin{pmatrix} 22 & 15 & 40 & 18 \\ 9 & 12 & 32 & 16 \\ 11 & 38 & 10 & 14 \end{pmatrix}.$$

Решение.

Общие запасы груза всех поставщиков составляют:

$$\sum_{i=1}^{3} a_i = 180 + 400 + 280 = 860.$$

Это совпадает с общими потребностями всех потребителей:

$$\sum_{j=1}^{4} b_j = 240 + 320 + 120 + 180 = 860.$$

Таким образом, данная транспортная задача является закрытой.

Оформим исходные данные в виде таблицы и найдем опорный план задачи методом северо-западного угла (табл. 40.2).

Общая стоимость перевозок по этому плану составляет:

$$Z(X_0) = 22 \cdot 180 + 9 \cdot 60 + 12 \cdot 320 + 32 \cdot 20 + 10 \cdot 100 + 14 \cdot 180 = 12500.$$

Таблица 40.2

Исходный опорный план по методу северо-западного угла

Поставщики	Потребители				Запасы, a_i
	B_1	B_2	B_3	B_4	
A_1	22 180	15 0	40 0	18 0	180
A_2	9 60	12 320	32 20	16 0	400
A_3	11 0	38 0	10 100	14 180	280
Потребности, b_j	240	320	120	180	860 860

Решим задачу методом минимальной стоимости (табл. 40.3).

Таблица 40.3

Исходный опорный план по методу минимальной стоимости

Поставщики	Потребители				Запасы, a_i
	B_1	B_2	B_3	B_4	
A_1	22 0	15 160	40 0	18 20	180
A_2	9 240	12 160	32 0	16 0	400
A_3	11 0	38 0	10 120	14 160	280
Потребности, b_j	240	320	120	180	860 860

$$Z(X_0) = 160 \cdot 15 + 20 \cdot 18 + 240 \cdot 9 + 160 \cdot 12 + 120 \cdot 10 + 160 \cdot 14 = 10280.$$

Сравнивая полученные результаты решения задачи двумя методами, можем сделать вывод, что метод минимальной стоимости быстрее ведет к оптимальному плану.

40.2. Метод потенциалов

Одним из наиболее эффективных и распространенных методов определения оптимальности плана ТЗ является *метод потенциалов*. Впервые он был предложен в 1949 г. Канторовичем Л. В. и Гавуриным М. К. Этот метод позволяет проверить, является ли данный опорный план оптимальным, и если нет, то определить такой путь улучшения этого плана, по которому каждая следующая итерация дает опорный план, которому отвечает меньшее значение целевой функции.

Следует заметить, что при решении транспортной задачи может оказаться, что $ЧЗкл < rang = m+n-1$. В этом случае задача имеет ***вырожденное решение***. Для возможного исключения вырожденного решения целесообразно нулевую клетку с наименьшим тарифом считать условно занятой. Условно занятой выбирают такую клетку, чтобы в каждой строке и каждом столбце было не менее одной занятой клетки и не образовывался квадрат из занятых клеток.

Если $ЧЗкл > rang$, то составляют исходный опорный план другим методом.

Действительно, составим двойственную к транспортной задаче (40.1) – (40.3). Для этого ограничениям (40.2) поставим в соответствие u_i, ограничениям (40.3) – v_j. Составим математическую модель двойственной задачи:

$$f = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \; (\max),$$

$$u_i + v_j \le c_{ij}, \; i = \overline{1,m}, \; j = \overline{1,n}.$$

По второй теореме двойственности:

если $x_{ij} \ne 0 \Rightarrow u_i + v_j = c_{ij}$;

если $x_{ij} = 0 \Rightarrow u_i + v_j < c_{ij}$.

Найденное исходное опорное решение проверяется на оптимальность методом потенциалов по следующему *критерию*: если опорное решение транспортной задачи является оптимальным, то ему соответствует система $m+n$ действительных чисел u_i и v_j, удовлетворяющих условиям $u_i + v_j = c_{ij}$ для занятых клеток и $u_i + v_j \le c_{ij}$ для свободных клеток. Числа u_i и v_j называют ***потенциалами***, u_i – потенциал поставщика, v_j – потенциал потребителя.

В распределительную таблицу добавляют строку v_j и столбец u_i. Потенциалы u_i и v_j находят из равенства $u_i + v_j = c_{ij}$, справедливого для занятых клеток, так как $rang = m + n - 1$. Одному из потенциалов присваивается произвольное значение, например $u_1 = 0$, тогда остальные потенциалы определяются однозначно. Так, если известен потенциал u_i, то $v_j = c_{ij} - u_i$; если известен потенциал v_j, то $u_i = c_{ij} - v_j$.

Обозначим $\Delta_{ij} = u_i + v_j - c_{ij}$. Эту оценку называют ***оценкой клеток***. Если $\Delta_{ij} \le 0$, то опорное решение является оптимальным. Если хотя бы одна из оценок $\Delta_{ij} > 0$, то опорное решение не является оптимальным и его можно улучшить, перейдя от одного опорного решения к другому.

Для этого находят клетку таблицы, которой соответствует самая большая положительная оценка $\max\{\Delta_{ij}\} = \Delta_{lk}$; строится цикл, который включает в свой состав данную клетку и часть клеток, занятых опорным решением. Определяются ***цикл*** (***контур***) ***перераспределения*** и объем груза, который можно перераспределить в соответствии с этим циклом. Клетка, которой отвечает самая большая положительная оценка, вместе с другими из занятых клеток должна образовывать замкнутый контур. Поворот можно осуществлять под прямым углом только в занятых клетках (можно занимать пустую клетку, если опорный план вырожденный) (рис. 40.1).

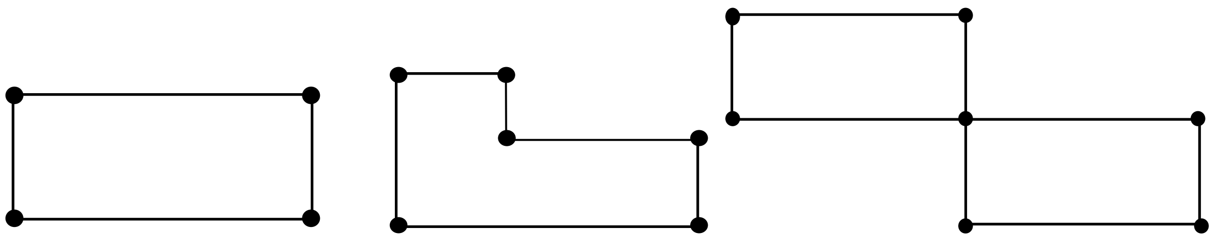


Рис. 40.1. **Типы контуров перераспределения груза**

Всем угловым клеткам цикла присваивают соответствующий знак.

Свободную клетку, в которую необходимо осуществить поставку, обозначают знаком «+», следующую угловую клетку цикла – знаком «–», дальше – знаком «+». Знаки расставляем до тех пор, пока не вернемся к исходной клетке.

Для определения объема груза θ, который можно перераспределить по этому циклу, рассматривают те угловые клетки, которые вошли в цикл перевозок со знаком «–», и по ним определяют наименьший объем груза. Это и есть то количество груза, которое необходимо перераспределить по данному циклу.

Количество груза в каждой угловой клетке цикла меняют на величину θ, причем так, что перевозка во всех клетках со знаком «–» уменьшается на θ, а во всех клетках со знаком «+» увеличивается на θ (в клетках, которые не обозначены никакими знаками, объем груза не меняется).

Вследствие перераспределения груза получают новый опорный план, по которому целевая функция имеет значение, меньше предыдущего на $\Delta Z = \left|\Delta_{ij} \cdot \theta\right|$. Этот план снова проверяют на оптимальность методом потенциалов. Процесс продолжают до тех пор, пока не будет получен оптимальный план.

Пример 40.2. Проверить опорный план, полученный в табл. 40.3, на оптимальность. Найти оптимальный план.

Решение.

Полученный план является невырожденным, так как число занятых клеток равно $rang = m + n - 1 = 6$, где m – число поставщиков ($m = 3$); n – число потребителей ($n = 4$).

Проверим найденный план на оптимальность с помощью метода потенциалов. Для этого каждому поставщику поставим в соответствие потенциал u_i, а каждому потребителю – потенциал v_j. Для каждой занятой клетки таблицы должно выполняться условие $u_i + v_j = c_{ij}$, поэтому ставим в соответствующую клетку c_{ij}, а для каждой свободной – $u_i + v_j \le c_{ij}$ ($i = \overline{1,3}$, $j = \overline{1,4}$), поэтому в нижнюю часть клетки ставим c_{ij}, в верхнюю – $u_i + v_j$.

Составим таблицу потенциалов для табл. 40.3, получим табл. 40.4.

Таблица 40.4

Таблица потенциалов для плана X_0

u_i \ v_j	$v_1=12$	$v_2=15$	$v_3=14$	$v_4=18$
$u_1=0$	12 / 22	15	14 / 40	18
$u_2=-3$	9	12	11 / 32	15 / 16
$u_3=-4$	8 / 11	11 / 38	10	14

Все оценки Δ_{ij} положительны. Следовательно, получен оптимальный план. Общая стоимость перевозок по этому плану минимальна и равна $Z_{\min}=10\,280$.

Можем сделать вывод, что найденный план является оптимальным и задача решена. Следовательно, минимальная стоимость перевозки груза в размере 10 280 достигается при $X_{опт.}=\begin{pmatrix} 0 & 160 & 0 & 20 \\ 240 & 160 & 0 & 0 \\ 0 & 0 & 120 & 160 \end{pmatrix}$.

Пример 40.3. Завод имеет три цеха A_1, A_2, A_3 и четыре склада B_1, B_2, B_3, B_4. Каждый цех производит за смену соответственно 70, 40, 40 тыс. шт. деталей. Пропускная способность складов характеризуется следующими показателями соответственно 35, 25, 43, 47 тыс. шт. деталей.

Задана матрица стоимости перевозок из цеха на склад

$$C_{3\times 4}=\begin{pmatrix} 3 & 5 & 4 & 2 \\ 1 & 6 & 3 & 4 \\ 3 & 3 & 1 & 5 \end{pmatrix}.$$

Составить план перевозок с минимальными затратами.

Решение.

Составим матрицу исходных данных и распределим груз по методу северо-западного угла (табл. 40.5).

Таблица 40.5

Распределение груза по методу северо-западного угла

Цех \ Склад	B_1	B_2	B_3	B_4	a_i
A_1	3 35	5 25	4 10	2 0	70
A_2	1 0	6 0	3 33	4 7	40
A_3	3 0	3 0	1 0	5 40	40
b_j	35	25	43	47	150 \ 150

$\sum_{i=1}^{3} a_i = \sum_{j=1}^{4} b_j = 150$ – модель закрытая.

$Z_{исх} = 35 \cdot 3 + 25 \cdot 5 + 10 \cdot 4 + 33 \cdot 3 + 7 \cdot 4 + 40 \cdot 5 = 597$ грн.

Распределим груз по методу минимальной стоимости (табл. 40.6).

Таблица 40.6

Распределение груза по методу минимальной стоимости

Цех \ Склад	B_1	B_2	B_3	B_4	a_i
A_1	3 0	5 23	4 0	2 47	70
A_2	1 35	6 2 –	3 3 +	4 0	40
A_3	3 0	3 0 +	1 40 –	5 0	40
b_j	35	25	43	47	150 \ 150

$Z_{исх} = 23 \cdot 5 + 47 \cdot 2 + 35 \cdot 1 + 2 \cdot 6 + 3 \cdot 3 + 40 \cdot 1 = 305$ грн.

Исходное значение функции цели в двух методах резко отличается. Метод северо-западного угла – это тот метод, которым распределяет груз оператор в сети (кладовщик) на предприятии.

Распределение груза по методу минимальной стоимости дает экономию ≈ 300 грн.

Проверим более экономный план на оптимальность.

Имеем: $rang = 3+4-1 = 6, ЧЗкл = 6$, то есть план невырожденный.

Составим таблицу потенциалов (табл. 40.7).

Таблица 40.7

Таблица потенциалов для исходного плана

u_i \ v_j	$v_1 = 0$	$v_2 = 5$	$v_3 = 2$	$v_4 = 2$
$u_1 = 0$	0 / 3	5	2 / 4	2
$u_2 = 1$	1	6	3	3 / 4
$u_3 = -1$	–1 / 3	**4 / 3**	1	1 / 5

Критерий оптимальности нарушен: $\Delta_{32} = 4-3 = 1 > 0$, то есть в эту клетку необходимо отправить груз и провести его по замкнутому циклу в табл. 40.6. Количество перераспределяемого груза определяем с помощью $\theta = \min(2; 40) = 2$.

Выполнив перегрузку, получим табл. 40.8.

Таблица 40.8

Улучшение плана перевозки груза

Цех \ Склад	B_1	B_2	B_3	B_4	a_i
A_1	3 / 0	5 / 23	4 / 0	2 / 47	70
A_2	1 / 35	6 / 0	3 / 5	4 / 0	40
A_3	3 / 0	3 / 2	1 / 38	5 / 0	40
b_j	35	25	43	47	150 / 150

Имеем: $rang = 3+4-1 = 6, ЧЗкл = 6$, то есть план невырожденный.

Проверим его на оптимальность (табл. 40.9).

Таблица 40.9

Таблица потенциалов для улучшенного плана

u_i \ v_j	$v_1 = 0$	$v_2 = 4$	$v_3 = 2$	$v_4 = 1$
$u_1 = 1$	1 / 3	5	3 / 4	2
$u_2 = 1$	1	5 / 6	3	2 / 4
$u_3 = -1$	–1 / 3	3	1	0 / 5

Все оценки $\Delta_{ij} = u_i + v_j - c_{ij} \le 0$ – план оптимальный.

$$Z_{\min} = Z_{исх} - \theta \cdot \Delta_{32} = 305 - 2 \cdot 1 = 303 \text{ (грн).}$$

$$X_{\min} = \begin{pmatrix} 0 & 23 & 0 & 47 \\ 35 & 0 & 5 & 0 \\ 0 & 2 & 38 & 0 \end{pmatrix}.$$

Экономическая интерпретация потенциалов

Рассмотрим исходную задачу (40.1) – (40.3).

Составим математическую модель двойственной к ней задачи:

$$f = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \ (\max),$$

$$u_i + v_j \le c_{ij},$$

условие отрицательности не соблюдается, так как задачи несимметричны.

По первой теореме двойственности $Z_{\min} = f_{\max}$, то есть, если $Y_{опт} = (u_i^*; v_j^*)$, $i = \overline{1,m}$, $j = \overline{1,n}$, то $Z_{\min} = \sum_{i=1}^{m} a_i u_i^* + \sum_{j=1}^{n} b_j v_j^*$.

Тогда $\dfrac{\partial Z_{\min}}{\partial a_i} = u_i^*$, $\dfrac{\partial Z_{\min}}{\partial b_j} = v_j^*$, то есть

$$\Delta Z_{\min} \approx u_i^* \Delta a_i, \; \Delta Z_{\min} \approx v_j^* \Delta b_j.$$

Это значит, что если задать $\Delta a_i = 1$ или $\Delta b_j = 1$, то соответственно $\Delta Z_{\min} \approx u_i^*$ или $\Delta Z_{\min} \approx v_j^*$, то есть потенциалы u_i^* и v_j^* показывают, на сколько изменится функция цели, если запасы i-го поставщика или потребности j-го потребителя изменить на единицу:

$$Z_{\min} = Z_{исх} + \Delta Z_{\min}.$$

Если i-му поставщику добавить груз Δa за счет k-го поставщика, то

$$\Delta Z_{\min} = u_i \Delta a - u_k \Delta a = (u_i - u_k)\Delta a$$

и

$$Z_{\min} = Z_{исх} + (u_i - u_k)\Delta a.$$

Аналогично с потребителями: j-му потребителю добавить груз Δb за счет r-го потребителя, то

$$\Delta Z_{\min} = v_j \Delta b - v_r \Delta b = (v_j - v_r)\Delta b$$

и

$$Z_{\min} = Z_{исх} + (v_j - v_r)\Delta b.$$

Следовательно, в целях наибольшего снижения суммарных затрат на перевозки следует увеличивать объем склада с наименьшим потенциалом и сокращать объем склада с наибольшим потенциалом.

Таким образом, потенциал u_i можно интерпретировать как условную цену продукции на i-м месте ее производства, а v_j – как условную цену продукции на j-м месте ее потребления.

Вопросы для самодиагностики

1. Сформулировать экономическую постановку и построить математическую модель транспортной задачи.
2. В чем заключается необходимое и достаточное условие существования решения транспортной задачи?
3. Какие основные этапы решения транспортной задачи?

4. Как осуществить переход от одного опорного плана ТЗ к другому?

Упражнения

Решить ТЗ:

40.1

B_j / A_i	B_1	B_2	B_3	B_4	a_i
A_1	2	5	8	1	9
A_2	8	3	9	2	16
A_3	7	4	6	3	5
b_j	11	7	8	4	

40.2

B_j / A_i	B_1	B_2	B_3	B_4	B_5	a_i
A_1	4	6	8	3	2	7
A_2	5	3	4	6	4	13
A_3	3	2	5	7	5	20
b_j	10	10	5	8	7	

40.3

B_j / A_i	B_1	B_2	B_3	B_4	a_i
A_1	1	3	4	1	100
A_2	5	2	2	7	200
A_3	4	4	3	6	400
A_4	7	2	5	3	200
b_j	100	200	200	300	

40.4

B_j / A_i	B_1	B_2	B_3	B_4	a_i
A_1	1	6	9	3	200
A_2	3	2	2	4	400
A_3	4	5	4	7	400
A_4	1	4	3	9	800
b_j	200	400	600	200	

40.5

B_j / A_i	B_1	B_2	B_3	B_4	a_i
A_1	3	4	3	1	300
A_2	2	3	5	6	200
A_3	1	2	3	3	100
A_4	4	5	7	9	200
b_j	300	200	300	100	

40.6

B_j / A_i	B_1	B_2	B_3	B_4	a_i
A_1	1	3	4	2	200
A_2	1	2	4	1	200
A_3	3	4	5	9	300
A_4	6	3	7	6	300
b_j	200	300	400	200	

41. Целочисленное программирование

41.1. Сущность и классификация задач целочисленного программирования

Многие экономические задачи характеризуются тем, что объемы управляемых ресурсов (в силу тех или иных объективных свойств) могут принимать только целые значения. К ним относятся задачи производства и распределения неделимой продукции (загрузка оборудования, распределение автобусов, судов, самолетов по рейсам и т. д.). Математическая формализация данных ситуаций приводит к моделям дискретного (***целочисленного***) программирования.

В общем виде математическая модель задачи целочисленного программирования имеет вид:

$$Z(x) = \sum_{j=1}^{n} c_j x_j \to \max(\min)$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \le (\ge, =)\, b_i,\ i = \overline{1,m},\ x_j \ge 0,\ x_j - \textit{целые},\ j = \overline{1,n}.$$

В литературе, как правило, выделяют следующие классы целочисленных оптимизационных задач:

задачи с неделимостями;

экстремальные комбинаторные задачи;

задачи с разрывными целевыми функциями;

задачи на невыпуклых областях и др.

В подавляющем большинстве случаев наличие условий неделимости определяется физическими свойствами моделируемых объектов. Так, например, они могут появиться в качестве дополнительных ограничений в уже рассматривавшейся нами выше задаче производственного планирования, если в ней осуществляется управление выпуском крупной штучной продукции.

Классическим представителем задач данного класса стала так называемая *задача о рюкзаке*. Ее фабула носит достаточно условный характер и состоит в том, что турист, собирающийся в поход, может нести груз весом не более W кг.

Этот груз может состоять из набора предметов n типов, каждый предмет типа j весит w_j кг и характеризуется некоторой «полезностью» u_j, $j=\overline{1,n}$. В рамках описанной ситуации вполне естественным представляется вопрос: сколько предметов каждого вида нужно положить в рюкзак, чтобы его суммарная полезность была максимальной?

Если в качестве компонент плана x_j принять количество укладываемых предметов типа j, то данную задачу можно записать:

$$Z(X)=\sum_{j=1}^{n}u_jx_j \to \max,$$

$$\sum_{j=1}^{n}w_jx_j \le W, \quad x_j \ge 0, \quad x_j - \textit{целые}.$$

Как нетрудно заметить, представленная математическая модель носит универсальный характер, и к ней могут быть сведены многие экономические задачи. Ярким подтверждением этому служит и тот факт, что в литературе она также известна как задача *о загрузке судна*.

К классу *комбинаторных задач* относятся задачи оптимизации функции, заданной на конечном множестве, элементами которого служат выборки из n объектов. Классическим представителем математических проблем такого рода стала *задача о коммивояжере*. Она состоит в составлении маршрута посещения торговым агентом, находящимся в некотором начальном пункте, n других городов при условии, что задана матрица стоимости переездов из города в город (с учетом начального). Причем допустимым является такой маршрут, который предусматривает *однократное* посещение *всех* городов и возвращение в исходный пункт. Очевидно, что наилучший маршрут должен минимизировать суммарную стоимость переездов.

Многие экономические системы характеризуются наличием так называемых постоянных затрат, которые должны быть произведены независимо от объема производства. Учет в моделях этих и подобных факторов приводит к появлению в них целевых функций, *не обладающих свойством непрерывности*. В качестве примера может быть приведена *транспортная задача с фиксированными доплатами*. Она отличается от транспортной задачи в матричной постановке тем, что в ней затраты по перевозке груза из i-го пункта производства в j-й пункт потребления определяются как

$$c_{ij}(x_{ij}) = \left[\begin{array}{ll} 0, & x_{ij} = 0; \\ c_{ij}x_{ij} + d_{ij}, & x_{ij} > 0, \end{array}\right.$$

где c_{ij} – издержки на перевозку единицы груза;

d_{ij} – фиксированная доплата за аренду транспортных средств.

При таких предпосылках целевая функция суммарных затрат на перевозку $Z(X) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}(x_{ij})$ содержит «скачкообразные» разрывы, что существенно затрудняет ее минимизацию, поэтому стандартный метод решения основан на следующем преобразовании.

Если ввести вспомогательные переменные y_{ij}, такие, что

$$y_{ij} = 0 \text{ или } y_{ij} = 1,$$

$$x_{ij} \le \left(\min\{a_i, b_j\}\right) y_{ij},$$

то целевая функция примет вид

$$Z(X) = \sum_{i=1}^{m}\sum_{j=1}^{n} (c_{ij}x_{ij} + d_{ij}y_{ij}).$$

В силу характера ограничений задача является задачей частично-целочисленного программирования.

41.2. Методы решения задач целочисленного программирования

Оптимальное решение задачи, найденное симплекс-методом, в общем случае не является целочисленным. Его можно округлить до ближайших целых чисел. Однако такое округление может дать решение, не лучшее среди целочисленных решений, или может привести к решению, которое не удовлетворяет системе ограничений.

Поэтому для нахождения целочисленного решения нужен особый алгоритм.

Графический метод решения задачи целочисленного программирования

При наличии в задаче линейного программирования двух переменных задача может быть решена графическим методом.

В системе координат X_1OX_2 находят область допустимых решений, строят вектор $\overline{gradZ} = \overline{N}$ и линию уровня. Перемещая линию уровня по направлению вектора $\overline{N}$ для задач на максимум, определяют точку выхода из области и ее координаты.

В случае, когда координаты этой точки не целые, в области допустимых решений строят целочисленную решетку, то есть линии, соответствующие точкам с целыми координатами, и находят на границах их пересечения такие целые числа x_1 и x_2, которые удовлетворяют системе ограничений и при которых значение целевой функции наиболее близкое к оптимальному. Координаты такой вершины и является целочисленным решением.

Аналогично решается задача на минимум. Целочисленному минимуму целевой функции будет соответствовать координата вершины целочисленной области, которая лежит в области допустимых решений, более близкой к началу вектора $\overline{N}$.

Пример 41.1. Решить задачу целочисленного программирования

$$Z(x) = 2x_1 + 4x_2 \to \max$$

при ограничениях:

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \end{cases}$$
$$x_j \ge 0,\ x_j - целые,\ j = 1,2.$$

Решение.

Построим многоугольник планов (рис. 41.1).

$OABC$ – область допустимых решений. Оптимального решения задача достигает в точке $B\ 9/5; 41/15$, при этом максимальное значение целевой функции составляет 218/15 ед.

Полученное оптимальное решение не является целочисленным.

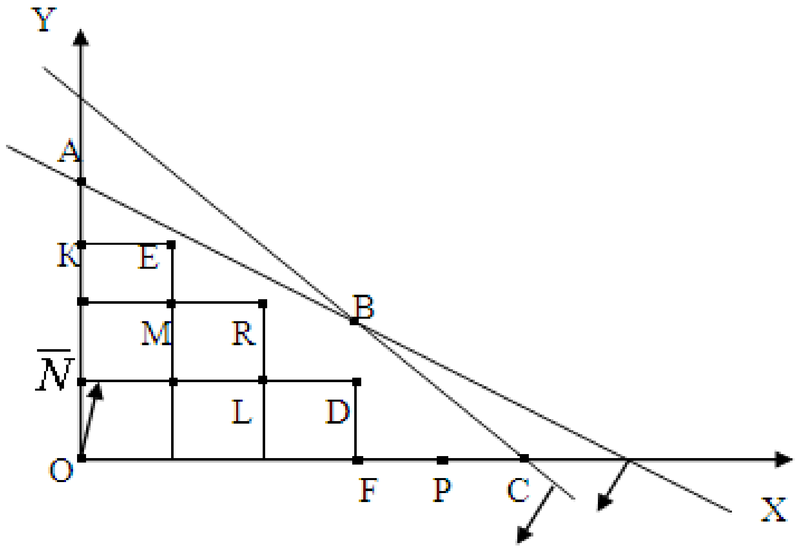


Рис. 41.1. **Область допустимых решений**

Условию целочисленности переменных удовлетворяют координаты 14 точек, которые принадлежат области допустимых решений. Чтобы найти точку, координаты которой определяют решение исходной задачи, заменим многоугольник $OABC$ многоугольником $OKEMRLDFP$, который содержит все допустимые точки с целочисленными координатами.

Строим вектор $\overline{N}$ (2, 4) и, передвигая линию уровня по направлению $\overline{N}$, получим в точке $E(1; 3)$ максимальное значение целевой функции 14 ед. Полученное оптимальное решение целочисленное.

Метод Гомори

Целочисленное решение может быть найдено с использованием алгоритма, предложенного Р. Гомори, который состоит из таких этапов:

1. Симплексным методом находят оптимальное решение задачи.
2. Если решение целочисленное, то задача решена.
3. Если же оно не является целочисленным и содержит хотя бы одну дробную координату, то накладывают дополнительно ограничение и вычисления продолжают до получения нового решения. Если дробных координат несколько, то дополнительное ограничение строят для той координаты, у которой дробная часть большая.
4. Если новое решение не является целочисленным, то снова строят дополнительно ограничение. Вычисления продолжают до тех пор, пока не будет получено целочисленное решение или показано, что задача не имеет целочисленного решения.

Пусть полученное оптимальное решение $X^* = (x_1, x_2, ..., x_n)$ не является целочисленным, то есть хотя бы одна компонента оптимального плана x_i дробная.

Последний шаг симплексной таблицы представлен в табл. 41.1.

Таблица 41.1

Последняя итерация симплекс-таблицы

Б	$C_б$	A_0	A_1	A_2	...	A_i	...	A_m	A_{m+1}	...	A_n
A_1	a_1	b_1	1	0	…	0	…	0	$\alpha_{1,m+1}$	…	α_{1n}
A_2	a_2	b_2	0	1	…	0	…	0	$\alpha_{2,m+1}$	…	α_{2n}
…	…	...	…	…	…	…	…	…	...	…	...
A_i	a_i	b_i	0	0	…	1	…	0	$\alpha_{i,m+1}$	…	α_{in}
…	…	...	…	…	…	…	…	…	...	…	...
A_m	a_m	b_m	0	0	…	0	…	1	$\alpha_{m,m+1}$	…	α_{mn}

Выделим целые и дробные части всех дробных x_i.

Целой частью числа x_i (обозначается $[x_i]$) называется целое число, ближайшее к данному, но расположенное левее данного числа.

Дробную часть x_i обозначим $\{x_i\}$, она определяется следующим образом:

$$\{x_i\} = x_i - [x_i].$$

Если b_i и хотя бы одно α_{ij} дробные, то с учетом введенных обозначений целых и дробных чисел дополнительное ограничение, которое должно быть составлено, имеет вид:

$$\sum_{j=m+1}^{n} \{\alpha_{ij}\} x_j \geq \{b_i\}.$$

Замечание 1: если b_i дробное, а все α_{ij} целые, то задача не имеет целочисленного решения.

Замечание 2: ограничение целочисленности может быть наложено не на все переменные, а лишь на их часть. В таком случае задача является частично целочисленной.

Пример 41.2. Решить задачу примера 41.1 методом Гомори.

Решение.

Математическая модель задачи: $Z(x) = 2x_1 + 4x_2 \to \max$

при ограничениях:

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \end{cases}$$
$$x_j \ge 0,\ x_j - целые,\ j = 1,2.$$

В табл. 41.2 представлено решение задачи симплексным методом.

Таблица 41.2

Решение задачи симплекс-методом

№ п/п	Баз.	$\overline{C_{баз}}$	c_j	2	4	0	0	Примечания
			$\overline{A_0}$	$\overline{A_1}$	$\overline{A_2}$	$\overline{A_3}$	$\overline{A_4}$	
1	2	3	4	5	6	7	8	9
1	$\overline{A_3}$	0	19/3	2	1	1	0	$\Delta_1, \Delta_2 < 0$ – план неоптимальный, $\theta_1 = \min\left(\frac{19/3}{2}; \frac{10}{1}\right) = \frac{19}{6}$, $\lvert\theta_1 \Delta_1\rvert = \left\lvert\frac{19}{6} \cdot 2\right\rvert = \frac{19}{3}$, $\theta_2 = \min\left(\frac{19/3}{1}; \frac{10}{3}\right) = \frac{10}{3}$, $\lvert\theta_2 \Delta_2\rvert = \left\lvert\frac{10}{3} \cdot 4\right\rvert = \frac{40}{3} - \max$
2	$\overline{A_4}$	0	10	1	3	0	1	
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			0	0	0	0	0	
$\Delta_j = z_j - c_j$			–	–2	–4	0	0	
3	$\overline{A_3}$	0	3	5/3	0	1	–1/3	$\Delta_1 < 0$ – план неоптимальный, $\theta_1 = \min\left(\frac{3}{5/3}; \frac{10/3}{1/3}\right) = \frac{3}{5/3} = \frac{9}{3}$
4	$\overline{A_2}$	4	10/3	1/3	1	0	1/3	
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			40/3	4/3	4	0	4/3	
$\Delta_j = z_j - c_j$			–	–2/3	0	0	4/3	

1	2	3	4	5	6	7	8	9
5	$\overline{A_1}$	2	9/5	1	0	3/5	–1/5	Все $\Delta_j \geq 0$ – план оптимальный
6	$\overline{A_2}$	4	41/15	0	1	–1/5	2/5	
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			218/15	2	4	2/5	6/5	
$\Delta_j = z_j - c_j$			–	0	0	2/5	6/5	

Получено $X_{опт} = (9/5; 41/15), \quad Z(X)_{\max} = 218/15$.

Это решение не является целочисленным.

Найдем целые и дробные части чисел 9/5 и 41/15:

$$\left[\frac{9}{5}\right] = 1; \left[\frac{41}{15}\right] = 2;$$

$$\left\{\frac{9}{5}\right\} = \frac{9}{5} - 1 = \frac{4}{5}; \quad \left\{\frac{41}{15}\right\} = \frac{41}{15} - 2 = \frac{11}{15}; \quad \frac{4}{5} = \frac{12}{15} > \frac{11}{15},$$

то есть сечение Гомори делаем к строке № 5.

Учитывая, что $\left[\frac{3}{5}\right] = 0; \left[-\frac{1}{5}\right] = -1$, то

$$\left\{\frac{3}{5}\right\} = \frac{3}{5} - 0 = \frac{3}{5}; \quad \left\{-\frac{1}{5}\right\} = -\frac{1}{5} + 1 = \frac{4}{5}.$$

Составляем дополнительное ограничение:

$$3/5x_3 + 4/5x_4 \geq 4/5,$$

или в каноническом виде:

$$3/5x_3 + 4/5x_4 - x_5 = 4/5.$$

Дальнейшие вычисления приведены в табл. 41.3.

Сравнивая полученное значение целевой функции целочисленного решения со значением оптимального решения исходной задачи, отмечаем, что поиск целочисленного решения приводит к уменьшению экстремального значения.

Таблица 41.3

Поиск целочисленного решения

№ п/п	Базис	$\overline{C_{баз}}$	c_j / $\overline{A_0}$	2 / $\overline{A_1}$	4 / $\overline{A_2}$	0 / $\overline{A_3}$	0 / $\overline{A_4}$	0 / $\overline{A_5}$	Примечания
5	$\overline{A_1}$	2	9/5	1	0	3/5	–1/5	0	$\theta_3 = \min\left(\frac{9/5}{3/5};\frac{4/5}{3/5}\right) = \frac{4}{3}$
6	$\overline{A_2}$	4	41/15	0	1	–1/5	2/5	0	
7	$\overline{A_5}$	0	4/5	0	0	3/5	4/5	–1	
9	$\overline{A_1}$	2	1	1	0	0	–1	1	[11]·(-3/5)+[5]
10	$\overline{A_2}$	4	3	0	1	0	2/3	–1/3	[11]·(1/5)+[6]
11	$\overline{A_3}$	0	4/3	0	0	1	4/3	–5/3	[7]/(3/5)
$z_j = \overline{C_{баз}} \cdot \overline{A_j}$			14	2	4	0	2/3	2/3	Все $\Delta_j \geq 0$ – план оптимальный
$\Delta_j = z_j - c_j$			–	0	0	0	2/3	2/3	

Ответ: $X_{опт} = (1,3)$; $Z(X)_{\max} = 14$.

Метод ветвей и границ

Метод ветвей и границ также предусматривает, что поиск оптимального плана начинается с решения задачи симплексным или графическим методом без учета целочисленности.

Если мы получим оптимальный целочисленный план, то он также будет оптимальным планом задачи целочисленного программирования.

Если план не будет целочисленным, то вводят дополнительные ограничения, по которым множество возможных решений (планов) определенным образом разделяется на подмножества, каждое из которых исследуется на поиск оптимального плана в целых числах. Если для этих подмножеств такой план не был найден, каждое из них тем же способом опять разбивается на подмножества. Процесс длится до тех пор, пока мы не получим целочисленное решение исходной задачи. Рассмотрим, как само множество планов разделяется на подмножества для поиска целочисленного решения.

Пусть ограничение представлено в виде: $m_j \leq x_j \leq M_j \quad (j = \overline{1,n})$, то есть известны нижняя и верхняя границы для каждой из нецелочисленных переменных x_j.

Предположим, что исходная задача, которую мы обозначим номером 1, решена симплексным методом без учета целочисленности, а также по исходному плану вычислен нижний предел целевой функции $Z_{\min} = Z_0$. То есть по любому произвольному плану X, принадлежащему многоугольнику планов, значения целевой функции будут не ниже нижнего предела: $Z(X) \geq Z_0$.

Для определенности допустим, что именно компонента x_i^* оптимального плана задачи не соответствует условию целочисленности. Тогда из множества возможных решений исключается область:

$$[x_i^*] < x_i^* < [x_i^*] + 1,$$

где $[x_i^*]$ – целая часть числа x_i^*.

Следовательно, исходная задача разделяется на две (обозначим их номерами 2 и 3), которые отличаются одна от другой тем, что в первой из них система ограничений содержит условие для i-й компоненты: $m_i \leq x_i \leq [x_i^*]$, а вторая – условие: $[x_i^*] + 1 \leq x_i \leq M_i$.

Решаем задачи 2 и 3 (не важно, в какой последовательности) симплексным методом без учета целочисленности.

Если во время решения одной из задач получим не целочисленный оптимальный план, по которому значение целевой превышает минимальное: $Z(X^*) \geq Z_0$, то задача разделяется на две (задачи 4 и 5) для последующего поиска целочисленного плана; если получим, что $Z(X^*) < Z_0$, то данная задача удаляется из списка.

Если при решении задач 2 или 3 мы получили целочисленный оптимальный план X^*, для которого $Z_{\max}(X^*) > Z_0$, то экстремальное значение целевой функции, которая соответствует этому плану, используется при решении задач 4 и 5 вместо значения Z_0.

Процесс будет длиться до тех пор, пока все задачи из списка не будут рассмотрены.

Решением исходной задачи считается целочисленный план (среди других целочисленных планов задач по списку), которому соответствует наибольшее значение целевой функции.

Пример 41.3. Решить задачу целочисленного программирования методом ветвей и границ:

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 0 \le x_1 \le 5; \\ 0 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Решение.

Обозначим исходную задачу номером 1. Решим ее без учета целочисленности. Вначале приведем математическую модель к каноническому виду, введя дополнительные переменные x_3, x_4, x_5 и x_6.

Получим такую целевую функцию:

$$Z = 3x_1 + x_2 + 0 \cdot x_3 + 0 \cdot x_4 + 0 \cdot x_5 + 0 \cdot x_6 \to \max.$$

При этом система ограничений имеет вид:

$$\begin{cases} 4x_1 + 3x_2 + x_3 = 18; \\ x_1 + 2x_2 + x_4 = 6; \\ x_1 + x_5 = 5; \\ x_2 + x_6 = 4; \\ x_j \ge 0, \ \ j = \overline{1,6}, \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Построив симплексную таблицу, получим исходное базисное решение: $X_0 = (0; \ \ 0; \ \ 18; \ \ 6; \ \ 5; \ \ 4)$, которое является целочисленным. Ему соответствует минимальное значение целевой функции $Z_0 = 0$, следовательно, его мы и принимаем за нижнюю границу целевой функции.

Решив задачу 1 симплексным методом (предлагаем сделать это самостоятельно), получим решение:

$$X_1^* = (4{,}5; \ \ 0; \ \ 0; \ \ 1{,}5; \ \ 0{,}5; \ \ 4), \ Z_{\max}(X_1^*) = 13.$$

Оптимальный план задачи 1 не является целочисленным, так как $x_1^* \notin \mathrm{N}$. Однако значение целевой функции этого плана больше минимального, поэтому исследуем эту задачу далее. Исключим из многоугольника планов область $4 < x_1^* < 5$, то есть исходная задача (номер 1) разделяется на две, математические модели которых отличаются только ограничением для x_1:

Задача 2

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 0 \le x_1 \le 4; \\ 0 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Задача 3

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 5 \le x_1 \le 5; \\ 0 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Решим каждую из них симплексным методом без учета целочисленности (поскольку план X_1^* не является целочисленным, значение Z_0 не изменяется). Задача 3 не имеет решений, поскольку ее условия противоречивы. Задача 2 дает решение: $X_2^* = (4;\ 2/3;\ 0;\ 2/3;\ 0;\ 10/3)$, $Z_{\max}(X_2^*) = 12\,2/3$. План X_2^* не является целочисленным, поэтому значение Z_0 остается неизменным. Чтобы рассматривать задачу дальше, исключаем область дробных решений для компоненты x_2^* оптимального плана, то есть область: $0 < x_2^* < 1$. Следовательно, задача снова разделяется на две.

Задача 4

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 0 \le x_1 \le 4; \\ 0 \le x_2 \le 0; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Задача 5

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 0 \le x_1 \le 4; \\ 1 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Решим задачу 4, не учитывая целочисленности, и получим результат: $X_4^* = (4;\ \ 0;\ \ 2;\ \ 2;\ \ 0;\ \ 0)$, $Z_{\max}(X_4^*) = 12$. Этот план является целочисленным, но мы не можем считать это ответом, поскольку не рассмотрен весь список задач. При дальнейшем рассмотрении получим новое минимальное значение $Z_0' = 12$, которое соответствует целочисленному плану X_4^*. Не учитывая целочисленности, будем иметь такое решение задачи 5:

$$X_5^* = (3{,}75;\ \ 1;\ \ 0;\ \ 0{,}25;\ \ 0{,}25;\ \ 0;\ \ 3),\ Z_{\max}(X_5^*) = 12{,}25.$$

Следовательно, получили $Z_{\max}(X_5^*) > Z_0'$, то есть это перспективная «ветвь» задач, но план $\mathbf{X}_5^*$ не является целочисленным, поэтому предельное значение целевой функции $Z_0' = 12$ остается неизменным.

Задачи 6 и 7 образуются из задачи 5, если из многоугольника ее планов исключить область $3 < x_1^* < 4$. Имеем такие математические модели для этих задач:

<u>Задача 6</u>

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 0 \le x_1 \le 3; \\ 1 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

<u>Задача 7</u>

$$Z = 3x_1 + x_2 \to \max,$$

$$\begin{cases} 4x_1 + 3x_2 \le 18; \\ x_1 + 2x_2 \le 6; \\ 4 \le x_1 \le 4; \\ 1 \le x_2 \le 4; \\ x_1, x_2 \in \mathrm{N} \cup \{0\}. \end{cases}$$

Задача 7 имеет противоречивые условия, следовательно, не имеет решений.

Задача 6 дает решение:

$$X_6^* = (3;\ \ 1{,}5;\ \ 1{,}5;\ \ 0;\ \ 0;\ \ 0{,}5;\ \ 2{,}5),\ Z_{\max}(X_6^*) = 10{,}5.$$

Получили, что $Z_{\max}(X_6^*) < Z_0'$, то есть ветвь задач, которая связана с задачей 6, является неперспективной, и далее ее исследовать не имеет смысла.

Список задач исчерпан, решение исходной задачи найдено:

$$Z_{\max}(X^*) = 12 \text{ при } X^* = X_4^* = (4;\ \ 0;\ \ 2;\ \ 2;\ \ 0;\ \ 0).$$

Вопросы для самодиагностики

1. Какие основные проблемы возникают при решении дискретных задач?

2. Какие экономико-математические модели могут быть сведены к задаче о коммивояжере?

3. Привести пример моделей с разрывными целевыми функциями.

4. Какой принцип используется для построения правильного сечения в методе Гомори?

5. Перечислить основные этапы метода Гомори.

6. Как используется двойственный симплекс-метод при решении целочисленной линейной задачи методом Гомори?

7. Перечислить принципиальные идеи, лежащие в основе методов ветвей и границ.

8. Как производится построение отсечения при решении целочисленной линейной задачи методом ветвей и границ?

Упражнения

Решить задачи целочисленного программирования:

а) $$Z = 5x_1 - 3x_2 \to \max, \quad \begin{cases} x_1 - 3x_2 \le 2, \\ 5x_1 + x_2 \le 6, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые};$$

б) $$Z = 2x_1 \to \max, \quad \begin{cases} 3x_1 + x_2 \le 6, \\ 5x_1 - x_2 \le 9, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые};$$

в) $$Z = 6x_1 + 5x_2 \to \max, \quad \begin{cases} x_2 \le 4, \\ 2x_1 + 3x_2 \le 15, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые};$$

г) $$Z = 2x_1 + 5x_2 \to \max, \quad \begin{cases} 3x_1 - x_2 \le 3, \\ x_2 \le 4, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые};$$

д) $$Z = x_1 - x_2 \to \min, \quad \begin{cases} 2x_1 + 3x_2 \le 4, \\ x_2 \le 2, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые};$$

е) $$Z = 2x_1 + 3x_2 \to \max, \quad \begin{cases} 5x_1 + 3x_2 \le 15, \\ x_2 \le 3, \end{cases} \quad x_1,\ x_2 \ge 0,\ \textit{целые}.$$

42. Предмет и задачи исследования операций

42.1. Связь математических методов моделирования и исследования операций в экономике. Этапы решения задач с использованием математических методов

Под *исследованием операций* понимается научный подход к решению задач организационного управления. Термин «исследование операций» возник в годы Второй мировой войны и означал первоначально планирование боевых операций.

В современной более общей трактовке – это научная методология и технология нахождения рационально обоснованных решений в различных областях человеческой деятельности.

Объектом изучения исследования операций являются операции в экономике, представляющие собой совокупность действий, приводящих экономическую систему к некоторой цели. *Предметом* является исследование этих операций с помощью математических методов их моделирования с целью обоснования принимаемых решений по организации оптимального управления этими операциями. Поэтому под термином «исследование операций» обычно понимают комплекс научных методов для предварительного решения задач эффективного управления организационными системами.

При решении задачи управления в экономике применение математических методов и моделей предполагает: построение экономических и математических моделей для задач принятия решения в сложных ситуациях или в условиях неопределенности; изучение взаимосвязей, определяющих впоследствии принятие решений, и установление критериев эффективности, позволяющих оценивать преимущество того или иного варианта действия.

Природа организационных систем может быть самой различной, а их общие математические модели находят применение не только при решении производственных и экономических задач, но и в биологии, социологических исследованиях, в геологоразведке и других практических сферах. Таким образом, исследование операций охватывает разнообразные вопросы построения математических моделей ситуаций, определения наилучших решений и их нахождения, анализа и практического применения.

Несмотря на многообразие задач организационного управления, при их решении можно выделить некоторую общую последовательность этапов, через которые проходит любое операционное исследование. Как правило, это:

1. Постановка задачи.

2. Построение содержательной (вербальной) модели рассматриваемого объекта (процесса). На данном этапе происходит формализация цели управления объектом, выделение возможных управляющих воздействий, влияющих на достижение сформулированной цели, а также описание системы ограничений на управляющие воздействия.

3. Построение математической модели, то есть перевод вербальной модели в ту форму, в которой для ее изучения может быть использован математический аппарат.

4. Решение задач, сформулированных на базе построенной математической модели.

5. Проверка полученных результатов на их адекватность природе изучаемой системы, включая исследование влияния так называемых немодельных факторов, и возможная корректировка первоначальной модели.

6. Реализация полученного решения на практике.

42.2. Операции и их эффективность

Задачи исследования операций возникают в сферах организационного управления и связаны с нахождением управленческих решений в тех или иных конкретных ситуациях. ***Операция*** – любое управляемое мероприятие, направленное на достижение цели. Результат операции зависит от способа ее проведения, организации, иначе – от выбора некоторых параметров. Эти параметры называют ***управляющими переменными.*** Определенный выбор параметров называется *решением*. ***Оптимальными*** считают те решения, которые по тем или иным соображениям предпочтительнее других.

Все множество операций можно условно разделить на два основных типа. В операциях первого типа (*детерминированных*) принятие управленческого решения приводит к ожидаемому результату. Закупка картофеля у двух поставщиков в известных количествах даст определенную относительную прибыль. Долевое распределение заказа на изготовление продукции между двумя заводами-изготовителями повлечет конкретные производственные издержки и т. д. В операциях второго типа (*стохастических, игровых*) принятие и реализация управленческого решения не гарантирует, вообще говоря, получение ожидаемого результата. На него могут повлиять не контролируемые случайные факторы (например, природные) или целенаправленные действия других участников операции, заинтересованных в том или ином ее исходе.

Любая задача исследования операций имеет начальные «дисциплинирующие» условия, то есть такие исходные данные, которые фиксированы с самого начала и не могут быть нарушены. В своей совокупности они формируют так называемое *множество возможных решений*.

Чтобы сравнивать между собой по эффективности разные решения, нужно иметь количественный критерий, называемый показателем *эффективности* (или целевой функцией). Этот показатель выбирается так, чтобы отражать целевую направленность операции.

Часто выполнение операции сопровождается действием случайных факторов. Тогда в качестве показателя эффективности берется не сама величина, которую хотелось бы оптимизировать, а ее среднее значение (или математическое ожидание). Иногда операция, сопровождаемая случайными факторами, преследует такую цель A, которая может быть либо достигнута полностью, либо не достигнута совсем (типа «да – нет»). Тогда в качестве показателя эффективности выбирают вероятность достижения этой цели $p(A)$.

Неправильный выбор показателя эффективности может привести к неоправданным затратам и потерям.

42.3. Экономико-математическое моделирование. Математическая модель операции

Суть экономико-математического моделирования заключается в описании социально-экономических систем и процессов в виде экономико-математических моделей. При этом математические методы следует понимать как инструмент, а экономико-математические модели – как продукт процесса математического моделирования.

Процесс математического моделирования включает в себя три структурных элемента: объект исследования, субъект (исследователь) и модель, характеризующую отношения между субъектом и объектом.

Общая *схема процесса экономико-математического моделирования* состоит из шести этапов:

1. Постановка экономической проблемы и ее качественный анализ. На этом этапе требуется сформулировать сущность проблемы, принимаемые предпосылки и допущения. Необходимо выделить важнейшие свойства моделируемого объекта, изучить его структуру и взаимосвязь его элементов, предварительно сформулировать гипотезы, объясняющие поведение и развитие объекта.

2. Построение математической модели. Это этап формализации экономической проблемы, то есть выражения ее в виде конкретных математических зависимостей (функций, уравнений, неравенств и др.).

Построение модели подразделяется в свою очередь на несколько стадий. Вначале определяется тип экономико-математической модели, изучаются возможности ее применения в данной задаче, уточняется конкретный перечень переменных и параметров, форма связей. Для некоторых сложных объектов целесообразно строить несколько разноаспектных моделей; при этом каждая модель выделяет лишь некоторые стороны объекта, а другие стороны учитываются приближенно.

3. Математический анализ модели. На этом этапе математическими приемами исследования выявляются общие свойства модели и ее решений.

Важным моментом является доказательство существования решения сформулированной задачи. При аналитическом исследовании выясняется, единственно ли решение, какие переменные могут входить в решение, в каких пределах они изменяются, каковы тенденции их изменения и т. д.

Модели сложных экономических объектов с большим трудом поддаются аналитическому исследованию; в таких случаях переходят к численным методам исследования.

4. Подготовка исходной информации. Используются методы теории вероятностей, теоретической и математической статистики для организации выборочных обследований, оценки достоверности данных и т. д.

5. Численное решение. Этот этап включает разработку алгоритмов численного решения задачи, подготовку программ на компьютере и непосредственное проведение расчетов.

6. Анализ численных результатов и их применение. На этом этапе решается вопрос о правильности и полноте результатов моделирования и применимости их как в практической деятельности, так и в целях усовершенствования модели. Поэтому должна быть проведена проверка адекватности модели по тем свойствам, которые выбраны в качестве существенных (должны быть произведены верификация и валидация модели). *Верификация модели* – проверка правильности структуры (логики) модели; *валидация модели* – проверка соответствия данных, полученных на основе модели, реальному процессу.

Перечисленные этапы экономико-математического моделирования находятся в тесной взаимосвязи, могут иметь место возвратные связи этапов.

В составе экономико-математических методов можно выделить следующие разделы:

экономическая кибернетика: системный анализ экономики, теория экономической информации и теория управляющих систем;

математическая статистика: экономические приложения данной дисциплины – выборочный метод, дисперсионный анализ, корреляционный анализ, регрессионный анализ, многомерный статистический анализ, факторный анализ, теория индексов и др.;

математическая экономика, изучающая те же вопросы с количественной стороны; эконометрия: теория экономического роста, теория производственных функций, межотраслевые балансы, национальные счета, анализ спроса и потребления, региональный и пространственный анализ, глобальное моделирование и др.;

методы принятия оптимальных решений, в том числе исследование операций в экономике, включающее в себя оптимальное (математическое) программирование, сетевые методы планирования и управления, программно-целевые методы планирования и управления, теорию и методы управления запасами, теорию массового обслуживания, теорию игр, теорию и методы принятия решений, теорию расписаний;

методы экспериментального изучения экономических явлений: математические методы анализа и планирования экономических экспериментов, методы машинной имитации (имитационное моделирование), деловые игры, методы экспертных оценок, разработанные для оценки явлений, не поддающихся непосредственному измерению.

Выделяют следующие признаки классификации экономико-математических моделей:

по *общему целевому назначению*: теоретико-аналитические, используемые при изучении общих свойств и закономерностей экономических процессов, и прикладные, применяемые в решении конкретных экономических задач анализа, прогнозирования и управления;

по *степени агрегирования объектов моделирования*: макроэкономические и микроэкономические;

по *конкретному назначению*: балансовые модели, выражающие требование соответствия наличия ресурсов и их использования; трендовые модели, в которых развитие моделируемой экономической системы отражается через тренд (длительную тенденцию) ее основных показателей; оптимизационные модели, предназначенные для выбора наилучшего варианта из определенного числа вариантов производства, распределения или потребления; имитационные модели, предназначенные для использования в процессе машинной имитации изучаемых систем или процессов и др.;

по *типу информации, используемой в модели*: аналитические, построенные на априорной информации, и идентифицируемые, построенные на апостериорной информации;

по *учету фактора времени*: статические, в которых все зависимости отнесены к одному моменту времени, и динамические, описывающие экономические системы в развитии;

по *учету фактора неопределенности*: детерминированные, если в них результаты на выходе однозначно определяются управляющими воздействиями, и стохастические (вероятностные), если при задании на входе модели определенной совокупности значений на ее выходе могут получаться различные результаты в зависимости от действия случайного фактора;

по *типу математического аппарата, используемого в модели*: матричные модели, модели линейного и нелинейного программирования, корреляционно-регрессионные модели, модели теории массового обслуживания, модели сетевого планирования и управления, модели теории игр и т. д.;

по *типу подхода к изучаемым социально-экономическим системам*: дескриптивные, предназначенные для описания и объяснения фактически наблюдаемых явлений или для прогноза этих явлений, и нормативные модели, в которых интересуются не тем, каким образом устроена и развивается экономическая система, а как она должна быть устроена и как должна действовать в смысле определенных критериев.

Для применения количественных методов исследования операций требуется построить математическую модель операции. При построении модели операция упрощается, схематизируется и схема операции описывается с помощью того или иного математического аппарата.

Модель операции – это достаточно точное описание операции на подходящем математическом языке (функции, уравнения, системы уравнений и неравенств, графы и т. п.). Составление модели операции требует понимания сущности описываемого явления и знания математического аппарата одновременно.

При этом отправной точкой является исходная планово-экономическая ситуация.

Изучение и анализ детерминированных и стохастических или игровых операций проводится на основе *математических моделей*. В математической модели при упрощающих предположениях, в некотором приближении к действительности описываются с помощью математических соотношений те решения, которые могут принимать участники операции, и влияние этих решений на интересующие их результаты.

Наиболее простой по форме является *однокритериальная детерминированная модель* операции, в которой фигурируют множество D некоторых решений x и целевая функция $f(x)$ – характеристика или критерий «качества» решения x. В однокритериальной модели операции на множестве решений задается *отношение предпочтения*, которое определяет наилучшее в смысле выбранного критерия *оптимальное* решение. Тогда задача поиска оптимального решения сводится к задаче на условный экстремум $f(x) \to min\,(max),\quad x \in D.$

Стохастические, или игровые, модели имеют более сложную природу и соответствующее математическое описание. Отношение предпочтения на множестве ситуаций задается не столь просто, как в однокритериальной модели.

Игровая модель для двух лиц имеет следующий вид:

$$H_1(x,y) \to max, H_2(x,y) \to max, \quad x \in X, y \in Y,$$

где x, y – стратегии игроков 1 и 2;

X, Y – множества стратегий;

$H_1(x,y)$, $H_2(x,y)$ – функции выигрыша игроков, заданные на множестве ситуаций $X \times Y$.

Зависимость выигрыша одного игрока от стратегий другого игрока есть специфическая особенность игры. Выигрыши игроков становятся известными лишь после того, как ситуация (x,y) сформирована игроками полностью.

Задачи исследования операций делятся на две категории: прямые и обратные.

Прямые задачи отвечают на вопрос: чему будет равен показатель эффективности Z, если в заданных условиях $y \in Y$ будет принято некоторое решение $x \in X$. Для решения такой задачи строится математическая модель, позволяющая выразить показатель эффективности через заданные условия и решение, а именно: заданные факторы (исходные данные), управляющие переменные (решение), Z – показатель эффективности (целевая функция), F – функциональная зависимость между переменными.

Обратные задачи отвечают на вопрос: как при данных условиях выбрать решение, чтобы показатель эффективности Z достиг максимума (минимума). Такую задачу называют задачей оптимизации решения. Пусть прямая задача решена, то есть модель операции задана и вид зависимости F известен. Тогда обратная задача (то есть задача оптимизации) может быть сформулирована следующим образом: требуется найти такое решение, при котором показатель эффективности Z есть оптимальное значение, взятое по всем решениям, входящим в множество возможных решений X. Метод поиска экстремума показателя эффективности Z и связанного с ним оптимального решения должен всегда выбираться, исходя из особенностей функции F и вида ограничений, накладываемых на решение. (Например, классическая задача линейного программирования.)

После того, как операция представлена в виде математической модели и на множестве всех решений введено отношение предпочтения, начинается этап анализа модели. С помощью средств математики выясняются вопросы существования и единственности оптимальных решений, выявляются их отличительные признаки, разрабатываются методы и алгоритмы поиска решений, анализируется чувствительность решений к исходным данным – параметрам и коэффициентам модели.

Результатом анализа служат несколько приемлемых вариантов решений. Их нахождение, исследование и аргументированное обоснование составляют *конечную цель исследования операций*.

Вопросы для самодиагностики

1. Описать объект и предмет исследования операций.
2. Какие этапы операционального исследования?
3. Привести основные понятия исследования операций. Какие существуют типы операций?
4. Сформулировать общую постановку задачи исследования операций.
5. Какие типы задач исследования операций?
6. Охарактеризовать этапы экономико-математического моделирования.
7. Какие основные классификационные признаки экономико-математических моделей. Привести примеры моделей, входящих в ту или иную классификационную рубрику.

43. Оптимизационные задачи управления запасами

43.1. Задача о распределении инвестиционных ресурсов между объектами, ее представление в виде модели динамического программирования; алгоритм поиска оптимального плана

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.

Динамическое программирование – один из разделов оптимального программирования, в котором процесс принятия решения и управления может быть разбит на отдельные этапы (шаги).

Динамическое программирование позволяет свести одну сложную задачу со многими переменными к нескольким задачам с малым числом переменных.

Это значительно сокращает объем вычислений и ускоряет процесс принятия управленческого решения. В отличие от линейного программирования, в котором симплексный метод является универсальным методом решения, в динамическом программировании такого универсального метода не существует.

Одним из основных методов динамического программирования является *метод рекуррентных соотношений*, который основывается на использовании принципа оптимальности, разработанного американским математиком Р. Беллманом. Принцип состоит в том, что, каковы бы ни были начальное состояние на любом шаге и управление, выбранное на этом шаге, последующие управления должны выбираться оптимальными относительно состояния, к которому прийдет система в конце данного шага. Использование данного принципа гарантирует, что управление, выбранное на любом шаге, не локально лучше, а лучше с точки зрения процесса в целом.

Оптимальное распределение ресурсов

Пусть имеется некоторое количество ресурсов x, которое необходимо распределить между n различными предприятиями, объектами, работами и т. д. так, чтобы получить максимальную суммарную эффективность от выбранного способа распределения.

Введем обозначения:

x_i – количество ресурсов, выделенных i-му предприятию ($i=\overline{1,n}$);

$g_i(x_i)$ – функция полезности, в данном случае это величина дохода от использования ресурса x_i, полученного i-м предприятием;

$f_k(x)$ – наибольший доход, который можно получить при использовании ресурсов x от первых k различных предприятий.

Математическую модель сформулированной задачи можно записать в виде:

$$f_n(x)=max\sum_{i=1}^{n}g_i(x_i) \text{ при ограничениях: } \sum_{i=1}^{n}x_i=x, \quad x_i\geq 0, \quad i=\overline{1,n}.$$

Для решения задачи необходимо получить рекуррентное соотношение, связывающее $f_k(x)$ и $f_{k-1}(x)$.

Обозначим через x_k количество ресурса, используемого k-м способом ($0 \le x_k \le x$), тогда для $(k-1)$ способов остается величина ресурсов, равная ($x - x_k$). Наибольший доход, который получается при использовании ресурса $(x - x_k)$ от первых $(k-1)$ способов, составит $f_{k-1}(x - x_k)$. Для максимизации суммарного дохода от k-го и первых $(k-1)$ способов необходимо выбрать x_k таким образом, чтобы выполнялись соотношения:

$$f_1(x) = g_1(x),$$

$$f_k(x) = max\{g_k(x_k) + f_{k-1}(x - x_k)\}, \quad k = \overline{2,n}.$$

Пример 43.1 *(распределение инвестиций для эффективного использования потенциала предприятия).* Совет директоров инвестиционной компании рассматривает предложения по наращиванию производственных мощностей для увеличения выпуска однородной продукции на четырех предприятиях, принадлежащих компании. Для расширения производства совет директоров выделяет инвестиционные средства в объеме 100 млн грн с дискретностью 20 млн грн. Прирост выпуска продукции на предприятиях зависит от выделенной суммы, его значения представлены в табл. 43.1.

Таблица 43.1

Прирост выпуска продукции, млн. грн

Выделяемые средства, млн грн	Предприятия			
	№ 1	№ 2	№ 3	№ 4
20	8	10	12	11
40	12	20	22	20
60	24	26	25	28
80	36	38	37	35
100	42	45	48	50

Найти распределение средств между предприятиями, обеспечивающее максимальный прирост выпуска продукции, причем на одно предприятие можно осуществить не более одной инвестиции.

Решение.

Разобьем решение задачи на четыре этапа по количеству предприятий, на которых предполагается осуществить инвестиции.

Рекуррентные соотношения будут иметь вид:

для предприятия № 1:

$$f_1(x) = g_1(x_1),$$

для всех остальных предприятий:

$$f_k(x) = max\{g_k(x_k) + f_{k-1}(x - x_k)\}, \quad k = \overline{2,n}.$$

Решение будем проводить согласно рекуррентным соотношениям в четыре этапа.

1-й этап. Инвестиции производим только первому предприятию.

Тогда:

$$f_1(20) = 8, \quad f_1(40) = 12, \quad f_1(60) = 24, \quad f_1(80) = 36, \quad f_1(100) = 42.$$

2-й этап. Инвестиции выделяем первому и второму предприятиям.

Рекуррентное соотношение для 2-го этапа имеет вид:

$$f_2(x) = max\{g_2(x_2) + f_1(x - x_2)\}.$$

Тогда

при x = 20: $f_2(20) = max\{8+0; 0+10\} = max\{8; 10\} = 10$;

при x = 40: $f_2(40) = max\{12; 8+10; 20\} = max\{12; 18; 20\} = 20$;

при x = 60:

$$f_2(60) = max\{24; 12+10; 8+20; 26\} = max\{24; 22; 28; 26\} = 28;$$

при x = 80:

$$f_2(80) = max\{36; 24+10; 12+20; 8+26; 38\} =$$
$$= max\{36; 34; 32; 34; 38\} = 38;$$

при x = 100:

$$f_2(100) = \max\{g_2(0) + f_1(100); g_2(20) + f_1(80); g_2(40) + f_1(60);$$
$$g_2(60) + f_1(40); g_2(80) + f_1(20); g_2(100) + f_1(0)\} =$$
$$= \max\{42; 36+10; 24+20; 26+12; 8+38; 45\} =$$
$$= \max\{42; 46; 44; 38; 46; 45\} = 46.$$

Таким образом, если 100 млн грн распределять между двумя предприятиями, то выгодно выделить первому предприятию 80 млн грн, второму – 20 млн. грн или второму – 80 млн грн, а первому – 20 млн грн. При этом максимальный совокупный доход будет 46 млн грн.

3-й этап. Финансируем 2-й этап и третье предприятие.

Итак, $f_2(20)=10,\ f_2(40)=20,\ f_2(60)=28,\ f_2(80)=38,\ f_2(100)=46.$

Расчеты проводим по формуле: $f_3(x)=max\{g_3(x_3)+f_2(x-x_3)\}$.

Тогда при x = 20: $f_3(20)=max\{10; 12\}=12$;

при x = 40: $f_3(40)=max\{20; 10+12; 22\}=max\{20; 22; 22\}=22$;

при x = 60:

$$f_3(60)=max\{28; 20+12; 10+22; 25\}=max\{28; 32; 32; 25\}=32;$$

при x = 80:

$$f_3(80)=max\{38; 28+12; 20+22; 10+25; 37\}=$$
$$=max\{38; 40; 42; 35; 37\}=42;$$

при x = 100:

$$f_3(100)=\max\{g_3(0)+f_2(100); g_3(20)+f_2(80); g_3(40)+f_2(60);$$
$$g_3(60)+f_2(40); g_3(80)+f_2(20); g_3(100)+f_2(0)\}=$$
$$=\max\{46; 38+12; 28+22; 25+20; 37+10; 48\}=$$
$$=\max\{46; 50; 50; 45; 47; 48\}=50.$$

Таким образом, если 100 млн грн перераспределять между тремя предприятиями, то выгодно выделить только второму предприятию 80 млн грн и третьему – 20 млн грн или первому – 20 млн грн, второму – 40 млн грн и третьему – 40 млн грн. При этом максимальный совокупный доход будет 50 млн грн.

4-й этап. Инвестиции в объеме 100 млн. грн распределяем между 3-м этапом и четвертым предприятием.

Итак, $f_3(20)=12,\ f_3(40)=22,\ f_3(60)=32,\ f_3(80)=42,\ f_3(100)=50.$

При x = 100:

$$f_4(100)=\max\{g_4(0)+f_3(100); g_4(20)+f_3(80); g_4(40)+f_3(60);$$
$$g_4(60)+f_3(40); g_4(80)+f_3(20); g_4(100)+f_3(0)\}$$

$$f_4(100) = \max\{50;\ 42+11;\ 32+20;\ 22+28;\ 12+35;\ 50\} =$$
$$= \max\{50;\ 53;\ 52;\ 50;\ 47;\ 50\} = 53.$$

Получены условия управления от 1-го до 4-го этапов.

Максимальный прирост выпуска продукции в 53 млн грн получен на 4-м этапе как 42 млн грн + 11 млн грн, то есть 11 млн грн соответствуют выделению 20 млн грн четвертому предприятию (см. табл. 43.1).

Согласно 3-му этапу 42 млн грн получено как 20 млн грн + 22 млн. грн, то есть 22 млн грн соответствует выделению 40 млн грн третьему предприятию.

Согласно 2-этапу 20 млн грн получено при выделении 40 млн грн второму предприятию.

Таким образом, инвестиции в объеме 100 млн грн. целесообразно выделить второму и третьему предприятиям по 40 млн грн каждому и четвертому предприятию 20 млн грн, при этом прирост продукции будет максимальным и составит 53 млн грн.

Пример 43.2. Распределить инвестиции в размере 700 тыс. грн по 4 фирмам. Размер инвестиций кратен 100 тыс. грн. Эффект от направления i-й фирме инвестиций в размере m (сотен тыс. грн) выражается функцией $f_i(m)$ (табл. 43.2).

Таблица 43.2

Исходные данные

x	0	100	200	300	400	500	600	700
$f_1(x_1)$	0	28	45	65	78	90	102	113
$f_2(x_2)$	0	25	41	55	65	75	80	85
$f_3(x_3)$	0	15	25	40	50	62	73	82
$f_4(x_4)$	0	33	33	42	48	53	56	58

Решение.

Покажем еще один способ решения задачи динамического программирования.

Приходим к задаче: $\sum\limits_{i=1}^{4} f_i(x_i) \to max, \quad \sum\limits_{i=1}^{4} x_i \le 700,\ x_i \ge 0, \quad i = \overline{1,4},$

где x_i – неизвестный размер инвестиций i-й фирме.

Эта задача решается методом динамического программирования: последовательно ищется оптимальное распределение для $k = \overline{2,4}$ фирм.

Пусть первым двум фирмам выделено m инвестиций, обозначим $z_2(m)$ величину инвестиций 2-й фирме, при которой сумма

$$f_2(z_2(j)) + f_1(m - z_2(j)), 0 \le j \le m$$

максимальна, саму эту максимальную величину обозначим $F_2(m)$.

Далее действуем также: находим функции z_3 и F_3 и т. д.

На k-м шаге для нахождения $F_k(m)$ используем основное рекуррентное соотношение:

$$F_k(m) = max\{f_k(j) + F_{k-1}(m - j)\},\ 0 \le j \le 7.$$

Заполняем табл. 43.3.

Таблица 43.3

Промежуточные расчеты для двух фирм

x_2	$m - x_2$	0	100	200	300	400	500	600	700
	$F_1(m - x_2)$ / $f_2(x_2)$	0	28	45	65	78	90	102	113
0	0	**0**	**28**	45	65	78	90	102	113
100	25	25	**53**	**70**	**90**	103	115	127	
200	41	41	69	86	**106**	119	131		
300	55	55	83	100	**120**	**133**			
400	65	65	93	110	130				
500	75	75	103	120					
600	80	80	108						
700	85	85							

Значения $f_2(x_2)$ складываем со значениями

$$F_1(m - x_2) = f_2(m - x_2)$$

и на каждой диагонали находим наибольшее число, которое отмечаем и указываем соответствующее значение z_2.

В табл. 43.4 выделен максимальный суммарный эффект от выделения соответствующего размера инвестиций двум предприятиям.

Таблица 43.4

Максимальный суммарный эффект (два предприятия)

m	0	100	200	300	400	500	600	700
$F_2(m)$	0	28	53	70	90	106	120	133
$z_2(m)$	0	0	100	100	100	200	300	300

Продолжая процесс, табулируем функции $F_3(m)$ и $z_3(m)$ (табл. 43.5).

Таблица 43.5

Промежуточные расчеты для трех фирм

x_3	$m-x_3$	0	100	200	300	400	500	600	700
	$F_2(m-x_3)$ / $f_3(x_3)$	0	28	53	70	90	106	120	133
0	0	**0**	**28**	**53**	**70**	**90**	**106**	120	133
100	15	15	43	68	85	105	**121**	**135**	
200	25	25	53	78	95	115	131		
300	40	40	68	93	110	130			
400	50	50	78	103	120				
500	62	62	90	115					
600	73	73	101						
700	82	82							

В табл. 43.6 выделен максимальный суммарный эффект от выделения соответствующего размера инвестиций трем предприятиям.

В табл. 43.7 заполняем только диагональ для значения m = 700. В табл. 43.8 выделен максимальный суммарный эффект от выделения соответствующего размера инвестиций четырем предприятиям.

Таблица 43.6

Максимальный суммарный эффект (три предприятия)

m	0	100	200	300	400	500	600	700
$F_3(m)$	0	28	53	70	90	106	121	135
$z_3(m)$	0	0	0	0	0	0	100	100

Таблица 43.7

Промежуточные расчеты для четырех фирм

x_4	$m-x_4$	0	100	200	300	400	500	600	700
	$F_3(m-x_4)$ / $f_4(x_4)$	0	28	53	70	90	106	121	135
0	0	**0**	**28**	**53**	70	**90**	106	121	135
100	20	20	48	**73**	**90**	**110**	**126**	**141**	
200	33	33	61	86	103	123	139		
300	42	42	70	95	112	132			
400	48	48	76	101	118				
500	53	53	81	106					
600	56	56	84						
700	58	58							

Таблица 43.8

Максимальный суммарный эффект (четыре предприятия)

m	0	100	200	300	400	500	600	700
$F_4(m)$	0	28	53	73	90	110	126	141
$z_4(m)$	0	0	0	0	0; 100	100	100	100

Результаты сведены в табл. 43.9. $F_4(700)=141$ показывает максимальный суммарный эффект по всем 4-м фирмам, а $z_4(700)=100$ – размер инвестиций в 4-ю фирму для достижения максимального эффекта. На долю первых трех фирм осталось (700 – 100) и для достижения максимального суммарного эффекта по первым трем фирмам в третью надо вложить 100 и т. д. В табл. 43.9 отмечены оптимальные значения инвестиций по фирмам и значения эффектов от них.

Таблица 43.9

Итоги расчетов

m	0	100	200	300	400	500	600	700
$F_1(m) = f_1(x_1)$	0	28	45	**65**	78	90	102	113
$z_1 = x_1$	0	100	200	**300**	400	500	600	700
$F_2(m)$	0	28	53	70	90	**106**	120	133
$z_2(m)$	0	0	100	100	100	**200**	300	300
$F_3(m)$	0	28	53	70	90	106	**121**	135
$z_3(m)$	0	0	0	0	0	0	**100**	100
$F_4(m)$	0	28	53	73	90	110	126	**141**
$z_4(m)$	0	0	0	0	0; 100	100	100	**100**

Таким образом, наилучшим является следующее распределение капитальных вложений по предприятиям:

$$x_1^* = 300; \quad x_2^* = 200; \quad x_3^* = 100; \quad x_4^* = 100.$$

Оно обеспечивает производственному объединению наибольший возможный прирост прибыли 141 тыс. грн.

43.2. Сущность проблемы оптимального управления запасами. Классификация издержек

Возникновение теории управления запасами можно связать с работами Ф. Эджуорта и Ф. Харриса, в которых исследовалась простая оптимизационная модель для определения экономичного размера партии поставки для складской системы с постоянным равномерным расходом и периодическим поступлением хранимого продукта.

Запасом называется любой ресурс, который хранится для удовлетворения будущих нужд. Примерами запасов могут стать полуфабрикаты, готовые изделия, материалы, различные товары, а также денежная наличность, находящаяся в хранилище.

Существуют причины, побуждающие предприятие создавать запасы:

1) дискретность поставок при непрерывном потреблении;

2) упущенная прибыль в случае отсутствия запаса;

3) случайные колебания:

а) спроса за период между поставками;

б) объема поставок;

в) длительности интервала между поставками;

4) предполагаемые изменения конъюнктуры:

а) сезонность спроса;

б) сезонность производства.

Существуют также причины, побуждающие предприятия стремиться к минимизации запасов на складах:

1) плата за хранение запаса;

2) физические потери при хранении;

3) моральный износ продукта.

Задача управления запасами состоит в выборе для предприятия целесообразного решения.

Основные понятия теории управления запасами

Издержки выполнения заказа (издержки заказа) – это накладные расходы, связанные с оформлением заказа. В промышленном производстве такими издержками являются затраты на переналадку оборудования и подготовительные операции.

Издержки хранения – расходы, связанные с физическим содержанием товаров на складе, плюс возможные проценты на капитал, вложенный в запасы. Обычно они выражены в абсолютных единицах или в процентах от закупочной цены и связаны с определенным промежутком времени.

Упущенная прибыль (издержки дефицита) – издержки, связанные с неудовлетворенным спросом, возникающим из-за отсутствия продукта на складе.

Совокупные издержки за период представляют собой сумму издержек заказа, издержек хранения и упущенной прибыли. Иногда к ним прибавляются издержки на закупку товара.

Срок выполнения заказа – время с момента заказа до момента его выполнения.

Точка восстановления – уровень запаса, при котором делается новый заказ.

В качестве критерия эффективности стратегии управления запасами выступает ***функция затрат (издержек),*** представляющая суммарные издержки.

Если все параметры модели управления запасами не меняются во времени, она называется *статической*, в противном случае – *динамической.* Если интенсивности пополнения, спроса и расхода запасов не случайные величины, то модель управления запасами считается *детерминированной*, если хотя бы одна из этих характеристик носит случайный характер – *стохастической*.

43.3. Статические детерминированные модели оптимизации запасов без дефицита и с дефицитом

43.3.1. Простейшая модель оптимального размера заказа

Предположим, что:

1) темп спроса на товар известен и постоянен;

2) получение заказа мгновенно;

3) закупочная цена не зависит от размера заказа;

4) дефицит не допускается.

Исходные данные: темп спроса, издержки заказа, издержки хранения.

Результат: оптимальный размер заказа, время между заказами, количество заказов за фиксированный период времени, совокупные издержки.

Размер заказа является постоянным. Заказ выполняется мгновенно. Уровень запасов убывает с постоянной интенсивностью, пока не достигает нулевого значения. В этот момент времени делается и мгновенно выполняется заказ и уровень запаса восстанавливается до максимального значения. При этом ***оптимальным решением задачи*** будет такой размер заказа, при котором минимизируются общие издержки за период, равные сумме издержек хранения и издержек заказа. Динамика изменения количества продукта s на складе показана на рис. 43.1.

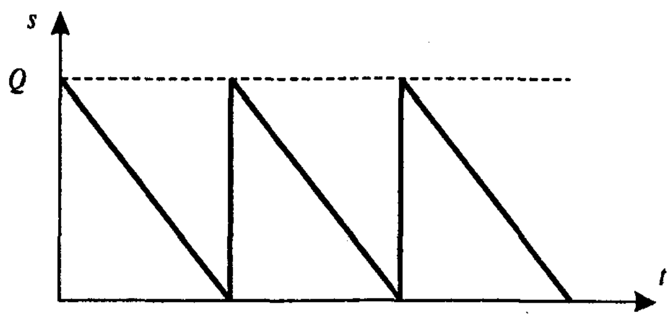


Рис. 43.1. **Динамика изменения количества продукта на складе**

Пусть Q – размер заказа;

T – продолжительность периода планирования;

D, d – величина спроса за период планирования и в единицу времени соответственно;

K – издержки одного заказа;

H, h – удельные издержки хранения за период и в единицу времени соответственно.

Тогда:

$C_1 = \dfrac{D}{Q} \cdot K$ – издержки за период планирования;

$C_2 = \dfrac{Q}{2} \cdot H$ – издержки хранения за период планирования;

$C = C_1 + C_2 = \dfrac{D}{Q} \cdot K + \dfrac{Q}{2} \cdot H$ – совокупные издержки.

Кривые издержек заказа C_1, издержек хранения C_2 и совокупных издержек C показаны на рис. 43.2.

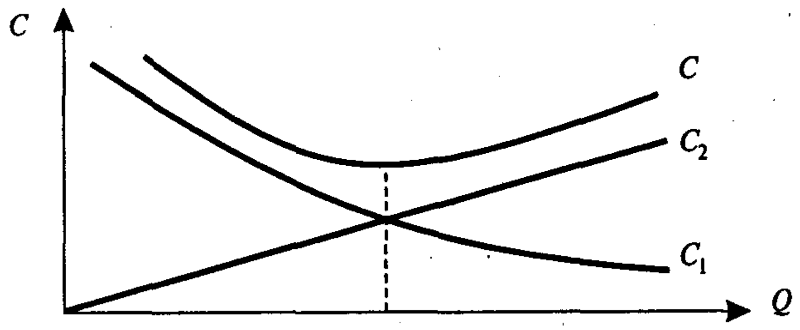


Рис. 43.2. **Кривые издержек**

Определив минимум функции совокупных издержек, получаем:

$Q^* = \sqrt{\dfrac{2dK}{h}} = \sqrt{\dfrac{2DK}{H}}$ – оптимальный размер заказа;

$N = \dfrac{D}{Q^*}$ – оптимальное число заказов за период;

$t = \dfrac{Q^*}{d} = \dfrac{T}{N}$ – время цикла (оптимальное время между заказами).

При этом оптимальный размер заказа не зависит от цены продукта.

Пример 43.3. Торговый агент компании занимается продажей последней модели марки автомобиля Volvo. Годовой спрос на эту модель оценивается в 4 000 единиц. Цена каждого автомобиля равна 90 тыс. усл. ед., а годовые издержки хранения составляют 10 % от цены самого автомобиля. Анализ показал, что средние издержки заказа составляют 25 тыс. усл. ед. на заказ. Время выполнения заказа – 8 дней. Ежедневный спрос на автомобили равен 20.

Чему равен оптимальный размер заказа? Каковы совокупные издержки? Каково оптимальное количество заказов в год? Каково оптимальное время между двумя заказами, если предположить, что количество рабочих дней в году равно 200?

Решение.

Исходные данные: величина спроса D = 4 000 единиц; издержки заказа K = 25 тыс. усл. ед.; издержки хранения h= 9/200 тыс. усл. ед.; цена за единицу p = 90 тыс. усл. ед.; время выполнения заказа L= 8 дней; ежедневный спрос d = 20 единиц; число рабочих дней T= 200.

Тогда, используя простейшую модель оптимального размера заказа, получаем: оптимальный размер заказа Q= 149 единиц; совокупные издержки C = 1 341 тыс. усл. ед.; оптимальное число заказов за год N = 26,83; число дней между двумя заказами t = 7,45.

43.3.2. Модель оптимального размера заказа с фиксированным временем его выполнения

Предположим, что:

1) темп спроса на товар известен и постоянен;

2) время выполнения заказа известно и постоянно;

3) закупочная цена не зависит от размера заказа;

4) дефицит не допускается.

Исходные данные: темп спроса, издержки заказа, издержки хранения, время выполнения заказа.

Результат: оптимальный размер заказа, время между заказами, точка восстановления запаса, количество заказов за фиксированный период времени, совокупные издержки.

Размер заказа является постоянным. Время выполнения заказа постоянно. Уровень запасов убывает с постоянной интенсивностью, пока не достигает точки восстановления R.

В этот момент делается заказ, который выполняется за время L. К моменту поступления заказа размер запаса на складе равен нулю.

Оптимальным решением задачи будет такой размер заказа Q^*, при котором минимизируются общие издержки за период, равные сумме издержек хранения и издержек заказа. Динамика изменения количества продукта s на складе показана на рис. 43.3.

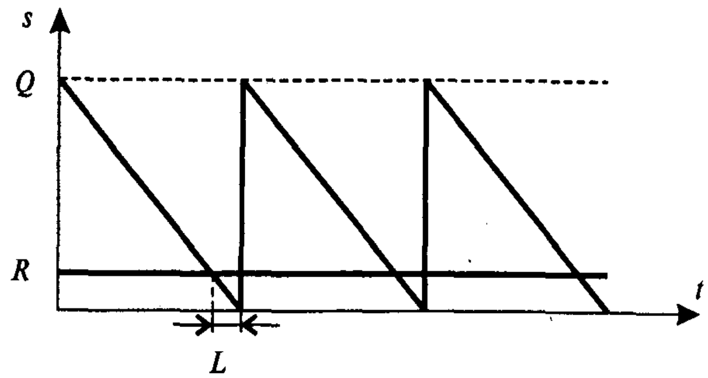


Рис. 43.3. **Динамика изменения количества продукта на складе**

Пусть Q – размер заказа;

T – продолжительность периода планирования;

D, d – величина спроса за период планирования и в единицу времени соответственно;

K – издержки одного заказа;

H, h – удельные издержки хранения за период и в единицу времени соответственно;

L – время выполнения заказа.

Тогда:

$C_1 = \frac{D}{Q} \cdot K$ – издержки за период планирования;

$C_2 = \frac{Q}{2} \cdot H$ – издержки хранения за период планирования;

$C = C_1 + C_2 = \frac{D}{Q} \cdot K + \frac{Q}{2} \cdot H$ – совокупные издержки;

$Q^* = \sqrt{\frac{2dK}{h}} = \sqrt{\frac{2DK}{H}}$ – оптимальный размер заказа;

$R = dL$ – точка восстановления запаса;

$N = \dfrac{D}{Q^*}$ – оптимальное число заказов за период;

$t = \dfrac{Q^*}{d} = \dfrac{T}{N}$ – время цикла (оптимальное время между заказами).

Кривые издержек заказа C_1, издержек хранения C_2 и совокупных издержек C показаны на рис. 43.2.

Пример 43.4 (п*оставка товара с фиксированным интервалом времени).* Магазин закупает продукцию на одной из фабрик. Годовой спрос на эту продукцию составляет 800 шт. Издержки заказа равны 450 грн, издержки хранения – 310 грн за одну упаковку (20 шт.) в год. Магазин заключил договор на поставку с фиксированным интервалом времени.

Количество рабочих дней в году – 300. Время поставки товара – 6 дней. Стоимость одной единицы продукции – 235 грн.

Чему равно оптимальное число заказов в течение года? Чему равна точка восстановления запаса? Каковы минимальные совокупные издержки?

Решение.

Оптимальный размер заказа:

$$Q^* = \sqrt{\frac{2DK}{H}} = \sqrt{\frac{2 \cdot 800 \cdot 450}{310/20}} = 215{,}5 \text{ шт.}$$

Число заказов в течение года: $N = \dfrac{D}{Q^*} = \dfrac{800}{215{,}5} = 3{,}7 \approx 4$.

Поскольку среднесуточный спрос равен 800/300 = 2,7 шт., точка восстановления запаса составит 2,7 · 6 = 16,2 шт.

Минимальные издержки заказа и хранения:

$$C = \frac{D}{Q} \cdot K + \frac{Q}{2} \cdot H = 4 \cdot 450 + 107{,}5 \cdot 15{,}5 = 3466{,}25 \text{ грн.}$$

43.3.3. Модель оптимального размера заказа с производством

Предположим, что:

1) темп спроса на товар известен и постоянен;

2) темп производства товара известен и постоянен;

3) время выполнения заказа известно и постоянно;

4) закупочная цена не зависит от размера заказа;

5) дефицит не допускается.

Исходные данные: темп спроса, темп производства, издержки заказа, издержки хранения, время выполнения заказа.

Результат: оптимальный размер заказа, время между заказами, точка восстановления запаса.

Фирма производит продукт самостоятельно, хранит его на складе и расходует с постоянным темпом. Если темп производства выше темпа спроса, то излишки продукта накапливаются на складе. Когда количество продукта на складе достигает максимального значения, производство прекращается и продукт расходуется со склада с постоянным темпом. Когда запас на складе достигает точки восстановления, производство возобновляется.

При этом ***оптимальным решением задачи*** будет такой размер заказа Q^*, при котором минимизируются общие издержки за период, равные сумме издержек хранения и издержек на возобновление (запуск) производства.

Динамика изменения количества продукта s на складе показана на рис. 43.4, где $tg\alpha = p - d, tg\beta = d$.

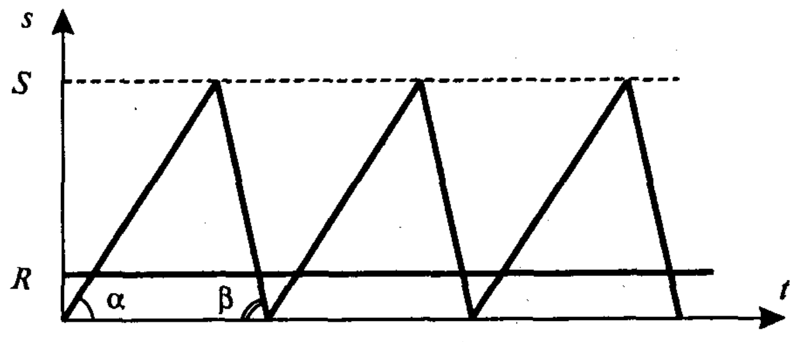


Рис. 43.4. **Динамика изменения количества продукта на складе**

Пусть Q – размер заказа;

p – темп производства;

T – продолжительность периода планирования;

D, d – величина спроса за период планирования и в единицу времени соответственно;

K – фиксированные издержки на запуск производства;

H, h – удельные издержки хранения за период и в единицу времени соответственно;

L – время, необходимое для запуска производства.

Тогда:

$C_1 = \frac{D}{Q} \cdot K$ – издержки на запуск производства;

$C_2 = \frac{Q}{2} \cdot H\left(1 - \frac{d}{p}\right)$ – издержки хранения;

$Q^* = \sqrt{\frac{2dK}{h\left(1 - \frac{d}{p}\right)}} = \sqrt{\frac{2DK}{H\left(1 - \frac{d}{p}\right)}}$ – оптимальный размер заказа;

$S^* = Q^*\left(1 - \frac{d}{p}\right)$ – максимальный уровень запасов;

$R = dL$ – точка восстановления запаса;

$N = \frac{D}{Q^*}$ – оптимальное число заказов за период;

$t = \frac{Q^*}{d} = \frac{T}{N}$ – время цикла (оптимальное время между заказами).

В этой модели оптимальный размер заказа также не зависит от цены продукта.

Пример 43.5 (*производство деталей).* На первом станке производятся детали в количестве 10 000 единиц в год. Эти детали используются для производства продукции на втором станке производительностью 2 400 единиц в год. Оставшиеся детали образуют запас. Издержки хранения составляют 0,5 грн за одну деталь в год. Стоимость производственного цикла на первом станке равна 800 грн. Определить оптимальный размер партии на первом станке.

Решение.

Оптимальный размер партии:

$$Q^* = \sqrt{\frac{2DK}{H\left(1 - \frac{d}{p}\right)}} = \sqrt{\frac{2 \cdot 2400 \cdot 800}{0{,}5\left(1 - \frac{2400}{10000}\right)}} = 3178{,}9 \text{ шт.}$$

43.3.4. Модель оптимального размера заказа с дефицитом

Предположим, что:

1) темп спроса на товар известен и постоянен;

2) время выполнения заказа известно и постоянно;

3) закупочная цена не зависит от размера заказа.

Исходные данные: темп спроса, издержки заказа, издержки хранения, издержки дефицита.

Результат: оптимальный размер заказа, время между заказами, точка восстановления запаса, совокупные издержки.

Размер заказа является постоянным. Уровень запасов убывает с постоянной интенсивностью. Допускается дефицит продукта.

После получения заказа фирма компенсирует дефицит и восстанавливает запас продукта на складе. Заказ делается тогда, когда дефицит продукта на складе достигает оптимального размера.

Оптимальным решением задачи будет такой размер заказа Q^*, при котором минимизируются общие издержки за период, равные сумме издержек хранения, издержек заказа и издержек дефицита. Динамика изменения количества продукта s на складе показана на рис. 43.5.

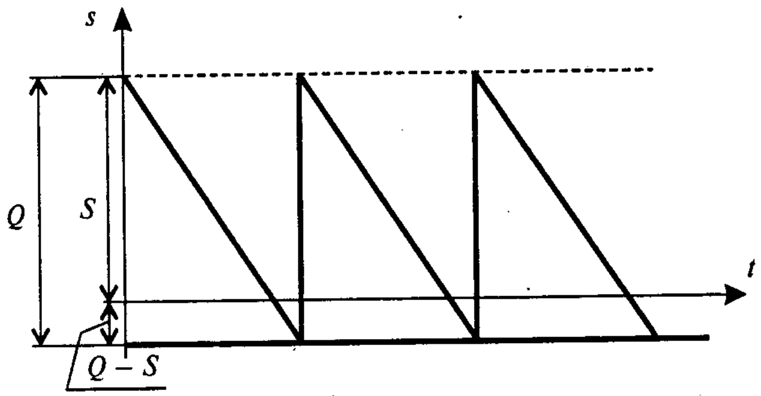


Рис. 43.5. **Динамика изменения количества продукта на складе**

Пусть Q – размер заказа;

T – продолжительность периода планирования;

D, d – величина спроса за период планирования и в единицу времени соответственно;

K – издержки одного заказа;

H, h – удельные издержки хранения за период и в единицу времени соответственно;

B, b – упущенная прибыль, возникающая вследствие дефицита одной единицы продукта, за период и в единицу времени соответственно;

S – максимальный запас продукции;

L – время выполнения заказа.

Тогда:

$C_1 = \frac{D}{Q} \cdot K$ – издержки заказа за период планирования;

$C_2 = \frac{S^2}{2Q} \cdot H$ – издержки хранения за период планирования;

$C_3 = \frac{(Q-S)^2}{2Q} \cdot B$ – издержки дефицита за период планирования;

$C = \frac{D}{Q} \cdot K + \frac{S^2}{2Q} \cdot H + \frac{(Q-S)^2}{2Q} \cdot B$ – совокупные издержки;

$Q^* = \sqrt{\frac{2dK}{h} \cdot \frac{b+h}{b}} = \sqrt{\frac{2DK}{H} \cdot \frac{B+H}{B}}$ – оптимальный размер заказа;

$S^* = \sqrt{\frac{2dK}{h} \cdot \frac{b}{b+h}} = \sqrt{\frac{2DK}{H} \cdot \frac{B}{B+H}}$ – максимальный размер запаса;

$Q^* - S^*$ – максимальный дефицит;

$R = dL$ – точка восстановления запаса.

Пример 43.6 (*планирование дефицита).* Рассмотрим вариант планирования дефицита примера 20.3. Допустим, по оценке менеджера, упущенная прибыль, связанная с отсутствием товара и утратой доверия клиентов, составляет 20 грн в год за одну единицу товара при условии, что издержки заказа и хранения остаются без изменения. Определить оптимальный размер заказа при плановом дефиците. Нужно ли менеджеру вводить систему с плановым дефицитом?

Решение.

Оптимальный размер заказа:

$$Q^* = \sqrt{\frac{2DK}{H} \cdot \frac{B+H}{B}} = 215{,}5 \cdot 1{,}3 = 280{,}15 \text{ шт.}$$

Максимальный размер запаса за один цикл:

$$S^* = \sqrt{\frac{2DK}{H} \cdot \frac{B}{B+H}} = 215{,}5 \cdot 0{,}75 = 161{,}8 \text{ шт.}$$

Совокупные издержки:

$$C = \frac{D}{Q} \cdot K + \frac{S^2}{2Q} \cdot H + \frac{(Q-S)^2}{2Q} \cdot B = 1800 + 724{,}2 + 499{,}97 = 3024{,}17 \text{ грн.}$$

Получено, что совокупные издержки при плановом дефиците меньше издержек без дефицита на 442,08 грн. Следовательно, целесообразно ввести систему с плановым дефицитом.

43.3.5. Модель оптимального размера заказа с количественными скидками

Предположим, что:

1) темп спроса на товар известен и постоянен;

2) время выполнения заказа известно и постоянно.

Исходные данные: темп спроса, издержки заказа, издержки хранения, цена товара, количественные скидки в случае закупки крупных партий товара.

Результат: оптимальный размер заказа, время между заказами, точка восстановления запаса, количество заказов за фиксированный период времени, совокупные издержки.

Пусть Q – размер заказа;

T – продолжительность периода планирования;

D, d – величина спроса за период планирования и за единицу времени соответственно;

K – издержки одного заказа;

H, h – удельные издержки хранения за период и в единицу времени соответственно.

Предположим, что известны числа $c_i, a_i, i = \overline{1,n},$ где c_i – цена продукта при размере заказа Q в интервале $a_{i-1} \le Q \le a_i$. Будем считать, что $a_0 = 0$ и $a_n = +\infty$. Тогда:

$\dfrac{D}{Q} \cdot K$ – издержки заказа за период планирования;

$\frac{Q}{2} \cdot H$ – издержки хранения за период планирования;

$c_i D$ – издержки на закупку товара;

$C_i = \frac{D}{Q} \cdot K + \frac{Q}{2} \cdot H + c_i D$ – совокупные издержки.

Оптимальный размер заказа определяется в результате решения n задач. Каждая из этих задач сводится к определению такого размера заказа $Q_i, i = \overline{1,n}$, при котором функция совокупных (общих) издержек $C_i = \frac{D}{Q} \cdot K + \frac{Q}{2} \cdot H + c_i D$ достигает минимума при ограничениях $a_{i-1} \le Q \le a_i$. Решение исходной задачи определяется из условия:

$$Q^* = \min_i \min_{Q_i} \left(C_i (Q_i) \right).$$

На рис. 43.6 изображены функции совокупных издержек для трех значений цен продукта.

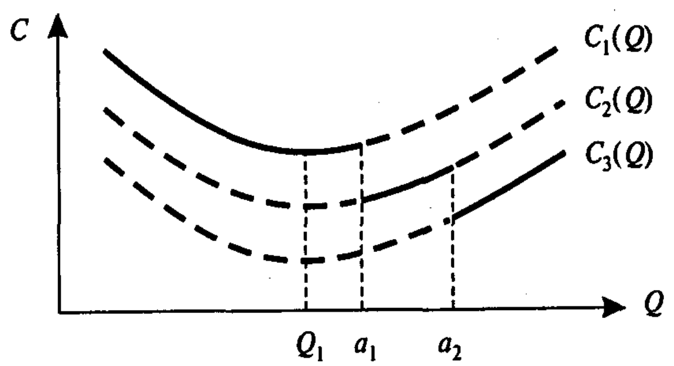


Рис. 43.6. **Функции совокупных издержек**

Значение цены c_1 определено на интервале $0 \le Q \le a_1$, цены c_2 – на интервале $a_1 \le Q \le a_2$, цены c_3 – на интервале $a_2 \le Q \le +\infty$. Соответственно, функция общих издержек $C_1(Q)$ определена при значении цены c_1на интервале $0 \le Q \le a_1$, функция $C_2(Q)$ – при значении цены c_2 на интервале $a_1 \le Q \le a_2$, функция $C_3(Q)$ – при значении цены c_3 на интервале $a_2 \le Q \le +\infty$. Минимальное значение функции $C_1(Q)$ в области ее допустимых значений достигается в точке Q_1, функции $C_2(Q)$ – в точке

a_1, функции $C_3(Q)$ – в точке a_2. Оптимальный размер заказа следует выбирать из величин Q_1, a_1 и a_2 по формуле:

$$Q^* = \min\left(C_1(Q_1), C_2(a_1), C_3(a_2)\right).$$

Пример 43.7 (*продажи со скидками).* Магазин продает комплекты детских конструкторов. Издержки заказа составляют 49 грн. Годовой спрос на конструкторы равен 5 000. В зависимости от размера заказа фирма предлагает скидки:

Вариант скидки	1	2	3
Размер заказа, шт.	0 – 1 000	1 000 – 2 000	Более 2 000
Размер скидки, %	0	4	5
Цена со скидкой	5,00	4,80	4,75

Годовые издержки хранения в процентном отношении к цене составляют 20 %. Найти размер заказа, минимизирующий общие издержки.

Решение.

Рассчитаем Q^* для каждого вида скидок: Q_1^* = 700, Q_2^* = 714, Q_3^* = 718. Так как Q_1^* находится в интервале между 0 и 1 000, то его необходимо взять равным 700. Оптимальный объем со скидкой Q_2^* меньше количества, необходимого для получения скидки, следовательно, его необходимо принять равным 1 000 единиц. Аналогично Q_3^* берем равным 2 000 единиц. Получим: Q_1^* = 700, Q_2^* = 1 000, Q_3^* = 2 000. Далее необходимо рассчитать общие издержки для каждого размера заказа и вида скидок, а затем выбрать наименьшее значение. Расчеты приведены в следующей таблице:

Вариант скидки	1	2	3
Цена со скидкой, грн	5,00	4,80	4,75
Размер заказа, шт.	700	1 000	2 000
Стоимость товара за год, грн	25 000	24 000	23 750
Годовые издержки заказа, грн	350,0	245,0	122,5
Годовые издержки хранения, грн	350	480	950
Общие годовые издержки, грн	25 700,0	24 725,0	24 822,5

Выбираем тот размер заказа, который минимизирует общие годовые издержки. Из таблицы видно, что заказ в размере 1 000 комплектов конструкторов будет минимизировать совокупные издержки.

43.4. Стохастические модели управления запасами

Мы отказываемся от предположения о постоянстве и детерминированности величины спроса на товар и предполагаем известным распределение величины спроса.

Пусть S – размер запаса на начало периода планирования;

D– величина спроса за период планирования (целое число);

H– удельные издержки хранения за период;

B– удельные издержки дефицита за период;

$p(D)$ – вероятность того, что величина спроса за период планирования составит D.

Функция распределения величины спроса:

$$F(x) = P(D < x) = \sum_{D=0}^{x-1} p(D).$$

В случае, когда величина спроса за период планирования превышает размер запаса $(D > S)$, возникает дефицит и соответствующие издержки дефицита.

Если запас больше, чем величина спроса $(S > D)$, то возникают издержки хранения.

Математическое ожидание величины издержек хранения за период планирования $C_1(S)$ для размера начального запаса S можно оценить следующим образом:

$$M(C_1(S)) = H \sum_{D=0}^{S} (S - D) p(D).$$

Математическое ожидание величины издержек дефицита за период планирования $C_2(S)$ для размера начального запаса S можно оценить следующим образом:

$$M(C_2(S)) = B \cdot \sum_{D=S+1}^{\infty} (D - S) p(D).$$

Математическое ожидание совокупных издержек $C(S)$ в этом случае имеет вид:

$$M(C(S)) = M(C_1(S)) + M(C_2(S)).$$

В стохастической модели *оптимальным* является такой размер начального запаса S^*, при котором математическое ожидание совокупных издержек $C(S^*)$ имеет минимальное значение, то есть такой размер запаса S^*, который удовлетворяет условию:

$$F(S^*) < \frac{B}{H+B} < F(S^*+1).$$

Если $F(S^*) = \dfrac{B}{H+B}$, то $C(S^*) = C(S^*+1)$ и оптимальными являются как размер запаса S^*, так и размер запаса S^*+1.

Пример 43.8 (*создание запаса продукции при дискретном спросе*). Небольшой салон специализируется на продаже телефонов стоимостью 2 000 грн. Затраты на хранение единицы продукции составляют 500 грн. Изучение спроса, проведенное в течение месяца, дало следующее распределение числа покупаемых телефонов:

Спрос, шт.	3	4	5	6	7
Вероятность	0,1	0,2	0,3	0,3	0,1

Найти оптимальный размер запаса.

Решение.

Суммарные затраты минимальны при размере запаса S^*, удовлетворяющем неравенству:

$$F(S^*) < \frac{B}{H+B} < F(S^*+1),$$

где $\dfrac{B}{H+B} = \rho$ – плотность убытков;

$F(S) = P(D < S)$ – функция распределения величины спроса.

Вычислим плотность убытков: $\rho = \dfrac{B}{H + B} = \dfrac{2000}{2500} = 0{,}8.$

Найдем значения функции распределения величины спроса:

Запас, шт.	3	4	5	6	7	Более 7
Спрос, шт.	3	4	5	6	7	Более 7
$F(x)$	0,0	0,1	0,3	0,6	0,9	1

Оптимальный размер запаса продукции удовлетворяет неравенству $F(6) < 0{,}8 < F(7)$.

Следовательно, размер запаса в 6 единиц будет оптимальным.

Вопросы для самодиагностики

1. Описать модель распределения инвестиционных ресурсов.
2. Привести классификацию моделей управления запасами.
3. Чему равен оптимальный размер заказа в детерминированной модели управления запасами?
4. Что необходимо знать для определения оптимального размера заказа в модели с производством?
5. Что необходимо знать для определения оптимального размера заказа в модели с дефицитом?
6. К чему приводит уменьшение размера заказа в модели управления запасами?
7. Что необходимо знать для определения оптимального размера заказа в модели с ценовыми скидками?
8. Какая модель называется стохастической?

Упражнения

43.1. Торговая компания приобретает в течение года 1 500 телевизоров для розничной продажи в своем магазине. Издержки хранения каждого телевизора равны 45 грн в год. Издержки заказа – 150 грн. Количество рабочих дней в году равно 300, время выполнения заказа – 6 дней.

Каков оптимальный размер заказа? Чему равны годовые издержки заказа? Чему равна точка восстановления запаса?

43.2. Менеджер по продажам организовывает продажу 400 гелевых матрацев в год, причем издержки хранения равны 1 тыс. грн за матрац в день, а издержки заказа – 40 тыс. грн. Количество рабочих дней равно 250, время выполнения заказа – 6 дней.

Каков оптимальный размер заказа? Чему равна точка восстановления запаса? Каков оптимальный размер заказа, если издержки хранения равны 1,5 тыс. грн?

43.3. В среднем частное малое производственное предприятие может производить 150 ножей в день. Дневной спрос на ножи примерно равен 40 шт. Фиксированные издержки производства составляют 100 грн, издержки хранения – 8 грн за нож в год. В году 250 рабочих дней.

Каков оптимальный размер производственного заказа? Чему равны издержки хранения? Чему равны совокупные издержки за год?

43.4. Годовой заказ на миксер для магазина бытовых товаров – 3 000 единиц, или 10 единиц в день. Издержки заказа равны 25 грн, издержки хранения – 0,4 грн в день. Так как миксер очень популярен, то в случае отсутствия товара покупатели обычно согласны подождать, пока не поступит следующая партия товара. Однако издержки вследствие дефицита равны 0,75 грн за миксер в день.

Сколько миксеров будет заказывать магазин? Каков максимальный дефицит? Чему равны совокупные издержки?

43.5. Магазин «Все для дома» закупает линолеум размером 2 х 3 м^2 в компании «Комфорт». В зависимости от размера заказа компания предлагает следующие скидки:

Размер заказа	9 кусков или менее	10 – 50 кусков	50 кусков и более
Цена 1 куска, тыс. грн	18	17,5	17,25

Издержки заказа равны 45 тыс. грн. Годовые издержки хранения составляют 50 % от закупочной цены, годовой спрос на линолеум равен 100 кускам.

Определить оптимальный размер заказа.

43.6. Мебельный салон продает в год около 1 000 спальных гарнитуров по цене 50 тыс. грн. Размещение одного заказа на поставку гарнитуров обходится в 40 тыс. грн.

Годовая стоимость хранения гарнитура составляет 25 % его цены. Салон может получить у поставщика скидку в 3 %, если размер заказа составит не менее 200 гарнитуров.

Следует ли салону воспользоваться этой скидкой?

43.7. Система управления запасами некоторого товара подчиняется основной модели. Каждый год с постоянной интенсивностью спрос составляет 15 000 ед. товара, издержки на организацию поставки составляют 10 грн на партию, цена единицы товара – 30 грн, а издержки на ее хранение – 7,5 грн в год.

Найти оптимальный размер партии, число поставок, продолжительность цикла.

44. Задачи массового обслуживания

44.1. Система массового обслуживания и характеристика ее элементов

Важную часть теории управления производством составляют основы знаний об очередях, иногда называемые теорией очередей, или теорией массового обслуживания. Модели очередей (как и линейное программирование и модели управления запасами) используются и в сфере управления материальным производством, и в сфере обслуживания. Анализ очередей в терминах длины очереди, среднего времени ожидания, среднего времени обслуживания и других факторов помогает лучше понять принципы организации системы обслуживания.

Основной экономический принцип совершенствования систем массового обслуживания (СМО) состоит в оценке общих ожидаемых затрат, включающих затраты на обслуживание и потери, которые несет система в результате ожидания клиента.

В СМО различают три основных этапа, которые проходит каждая заявка: поступление заявки на вход в систему; прохождение очереди; процесс обслуживания, после которого заявка покидает систему.

На каждом этапе используются определенные характеристики:

характеристики входа: число заявок на входе (размер популяции); режим поступления заявок в систему обслуживания; поведение клиентов;

характеристики очереди: длина; правило обслуживания;

характеристики процесса обслуживания: конфигурация системы обслуживания (число каналов и число фаз обслуживания); режим обслуживания.

Число потенциально возможных заявок (размер популяции) может считаться либо бесконечным (неограниченная популяция), либо конечным (ограниченная популяция).

Если число заявок, поступивших на вход системы с момента начала процесса обслуживания до любого заданного момента времени, является лишь малой частью потенциально возможного числа клиентов, популяция на входе рассматривается как *неограниченная.* Примеры неограниченных популяций: автомобили, проходящие через пропускные пункты на скоростных дорогах, покупатели в супермаркете и т. п.

Если количество заявок, которые могут поступить в систему, сравнимо с числом заявок, уже находящихся в системе массового обслуживания, популяция считается *ограниченной.* Пример ограниченной популяции: компьютеры, принадлежащие конкретному предприятию и поступающие на обслуживание в ремонтную мастерскую.

Заявки могут поступать в систему обслуживания в соответствии с определенным *режимом (*графиком). Появления клиентов считаются *случайными,* если они независимы друг от друга и точно непредсказуемы. Часто в задачах массового обслуживания число появлений в единицу времени может быть оценено с помощью пуассоновского распределения вероятностей.

Большинство моделей очередей основывается на предположении, что *поведение клиентов* является стандартным, то есть каждая поступающая в систему заявка встает в очередь, дожидается обслуживания и не покидает систему до тех пор, пока ее не обслужат. Другими словами, клиент (человек или машина), вставший в очередь, ждет до тех пор, пока он не будет обслужен, не покидает очередь и не переходит из одной очереди в другую. На практике клиенты могут покинуть очередь потому, что она оказалась слишком длинной. Может возникнуть и другая ситуация: клиенты дожидаются своей очереди, но по каким-то причинам уходят необслуженными.

Длина очереди может быть ограничена либо не ограничена. Длина очереди (очередь) *ограничена,* если она по каким-либо причинам (например, из-за физических ограничений) не может увеличиваться до бесконечности. Если очередь достигает своего максимального размера, то следующая заявка в систему не допускается и происходит отказ. Длина очереди *не ограничена,* если в очереди может находиться любое число заявок. Например, очередь автомобилей на бензозаправке.

Большинство реальных систем использует *правило обслуживания* «первым пришел – первым ушел» *(FIFO* – first in, first out). В некоторых случаях, например в приемной больницы, в дополнение к этому правилу могут устанавливаться различные *приоритеты.* Пациент с инфарктом в критическом состоянии, по-видимому, будет иметь приоритет в обслуживании по сравнению с пациентом, сломавшим палец. Порядок запуска компьютерных программ – другой пример установления приоритетов в обслуживании.

Конфигурация системы обслуживания определяется *числом каналов обслуживания* и *числом фаз.* Обычно количество каналов можно определить как число клиентов, обслуживание которых может быть начато одновременно, например: число мастеров в парикмахерской. Примеры *одноканальной* системы обслуживания: банк, в котором открыто единственное окошко для обслуживания клиентов, или ресторан, обслуживающий клиентов в автомобилях. Если же в банке открыто несколько окошек для обслуживания, клиент ожидает в общей очереди и подходит к первому освободившемуся окну, то мы имеем дело с *многоканальной* однофазовой системой обслуживания. Большинство банков, также, как почтовые отделения и авиакассы, являются многоканальными системами обслуживания.

Другая характеристика – *число фаз* (или последовательных этапов) *обслуживания* одного клиента.

Однофазовыми являются такие системы, в которых клиент обслуживается в одном пункте (на одном рабочем месте), затем покидает систему. Ресторан для обслуживания автомобилей, в котором официант получает деньги и приносит заказ в автомобиль, является примером однофазовой системы.

Если же в ресторане нужно сделать заказ в одном месте, оплатить его в другом и получить пищу в третьем, то мы имеем дело с *многофазовой* (три фазы) системой обслуживания.

На рис. 44.1 приведены системы обслуживания различной конфигурации.

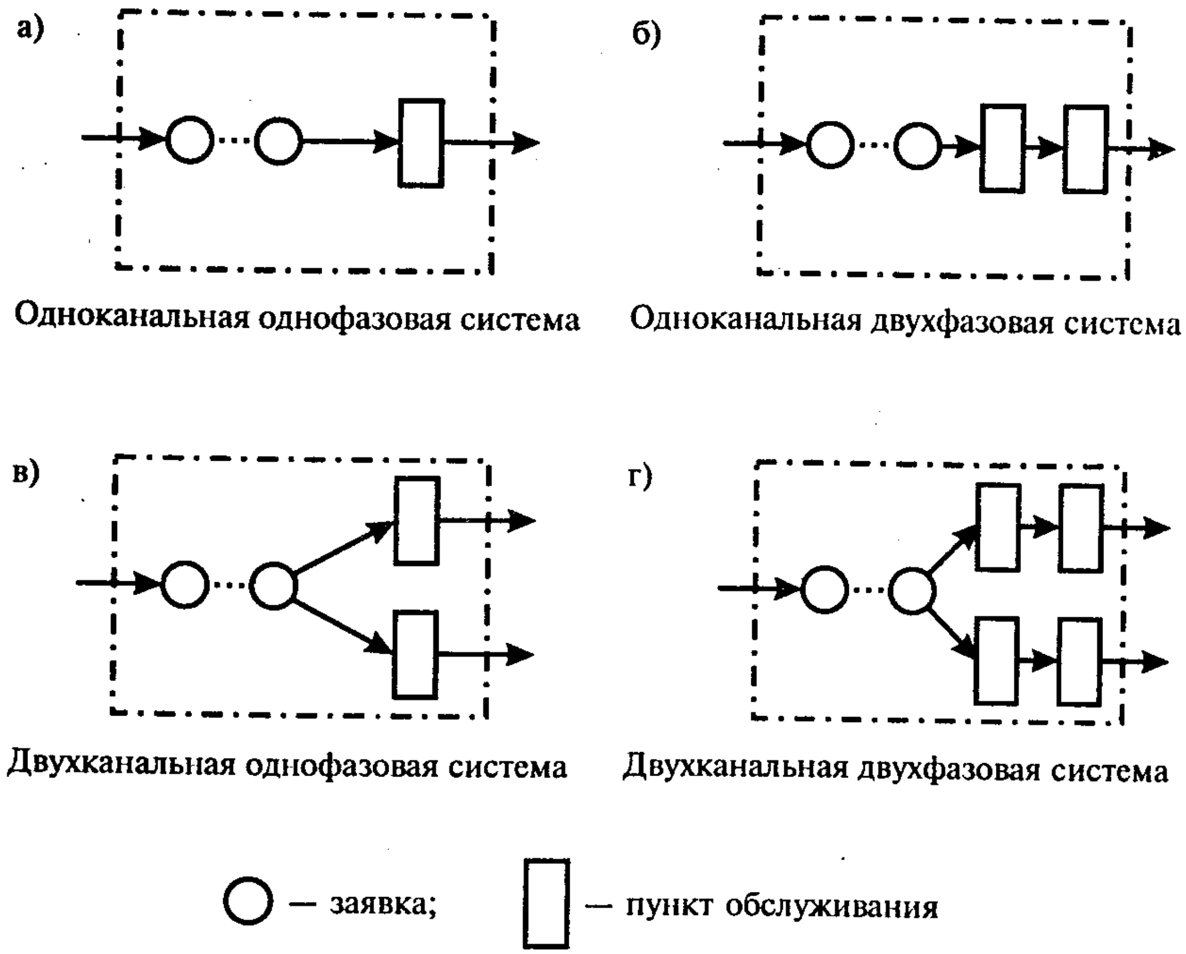


Рис. 44.1. **Конфигурации СМО**

Как и режим поступления заявок, *режим обслуживания* может характеризоваться либо постоянным, либо случайным временем обслуживания. При *постоянном* времени на обслуживание любого клиента затрачивается одинаковое время. Такая ситуация может наблюдаться на автоматической мойке автомобилей. Однако более часто встречаются ситуации, когда время обслуживания имеет *случайное* распределение.

Во многих случаях можно предположить, что время обслуживания подчиняется экспоненциальному распределению.

При анализе СМО используются технические и экономические характеристики. Наиболее часто используются следующие *технические характеристики:* среднее время, которое клиент проводит в очереди; средняя длина очереди; среднее время, которое клиент проводит в системе обслуживания (время ожидания плюс время обслуживания); среднее число клиентов в системе обслуживания; вероятность того, что система обслуживания окажется незанятой; вероятность определенного числа клиентов в системе.

Среди *экономических характеристик* наибольший интерес представляют следующие: издержки ожидания в очереди; издержки ожидания в системе; издержки обслуживания.

44.2. Характеристика простого потока требований

На практике наиболее распространенным является ***простейший поток*** требований (заявок), обладающий свойствами стационарности, ординарности и отсутствия последействия.

Стационарность характеризуется тем, что вероятность поступления определенного количества требований (заявок) в течение некоторого промежутка времени зависит только от длины этого промежутка и не зависит от начала отсчета.

Ординарность потока определяется невозможностью одновременного появления двух или более заявок.

Отсутствие последействия характеризуется тем, что поступление заявки не зависит от того, когда и сколько заявок поступило до этого момента.

В этом случае вероятность того, что число заявок, поступивших на обслуживание за промежуток времени t, равно k, определяется по закону Пуассона

$$P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!},$$

где e = 2,7182 – основание натурального логарифма;

λ – *интенсивность потока заявок,* то есть среднее число заявок в единицу времени:

$$\lambda = \frac{1}{\tau},$$

где τ – среднее значение интервала времени между двумя соседними заявками.

Для такого потока заявок время между двумя соседними заявками распределено экспоненциально с плотностью вероятности

$$f(t) = \lambda e^{-\lambda t}.$$

Случайное время ожидания в очереди начала обслуживания считают распределенным экспоненциально:

$$f(t) = \nu e^{-\nu t},$$

где ν – *интенсивность движения очереди,* то есть среднее число заявок, приходящих на обслуживание в единицу времени:

$$\nu = \frac{1}{\bar{t}_{оч}},$$

где $\bar{t}_{оч}$ – среднее значение времени ожидания в очереди.

Выходящий поток заявок связан с потоком обслуживания в канале, где длительность обслуживания $\bar{t}_{обсл}$ является случайной величиной и подчиняется показательному закону распределения с плотностью

$$f(t_{обсл}) = \mu e^{-\mu t},$$

где μ – *интенсивность потока обслуживания,* то есть среднее число заявок, обслуживаемых в единицу времени:

$$\mu = \frac{1}{\bar{t}_{обсл}},$$

где $\bar{t}_{обсл}$ – среднее время обслуживания.

Важной характеристикой СМО, объединяющей λ и μ, является *интенсивность нагрузки* $\rho = \dfrac{\lambda}{\mu}$.

44.3. Расчет параметров систем массового обслуживания

В зависимости от сочетания приведенных выше характеристик могут рассматриваться различные модели СМО.

Модель одноканальной системы массового обслуживания M/M/1 с пуассоновским входным потоком заявок и экспоненциальным временем обслуживания

Для систем этого типа выполняются следующие условия:

1. Заявки обслуживаются по принципу «первым пришел – первым обслужен» *(FIFO),* причем каждый клиент ожидает своей очереди до конца независимо от длины очереди.

2. Появления заявок являются независимыми событиями, однако среднее число заявок, поступающих в единицу времени, неизменно.

3. Процесс поступления заявок описывается пуассоновским распределением, причем заявки поступают из неограниченного множества.

4. Время обслуживания описывается экспоненциальным распределением вероятностей.

5. Темп обслуживания выше темпа поступления заявок.

Пусть λ – число заявок в единицу времени; μ – число клиентов, обслуживаемых в единицу времени; n – число заявок в системе.

Тогда:

$L_s = \frac{\lambda}{\mu - \lambda}$ – среднее число клиентов в системе;

$W_s = \frac{1}{\mu - \lambda}$ – среднее время обслуживания одного клиента в системе (время ожидания плюс время обслуживания);

$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$ – среднее число клиентов в очереди;

$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$ – среднее время ожидания клиента в очереди;

$r = \frac{\lambda}{\mu}$ – характеристика загруженности системы (доля времени, в течение которого система занята обслуживанием);

$P_0 = 1 - \frac{\lambda}{\mu}$ – вероятность отсутствия заявок в системе;

$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$ – вероятность того, что в системе находится более чем k заявок.

Пример 44.1. Механик автосервиса, может заменить масло в среднем в трех автомобилях в течение часа (то есть в среднем на одном автомобиле за 20 мин). Время обслуживания подчиняется экспоненциальному закону. Клиенты, нуждающиеся в этой услуге, приезжают в среднем по два в час, в соответствии с пуассоновским распределением. Клиенты обслуживаются в порядке прибытия и их число не ограничено.

Рассчитать основные характеристики системы обслуживания.

Решение.

На основе исходных данных получаем:

λ = 2 машины в час – количество машин, поступающих в течение часа;

μ = 3 машины в час – количество машин, обслуживаемых в течение часа.

Тогда:

$L_s = \frac{\lambda}{\mu - \lambda} = \frac{2}{3-2} = 2$ машины – среднее количество машин, находящихся в системе;

$W_s = \frac{1}{\mu - \lambda} = \frac{1}{3-2} = 1$ час – среднее время ожидания в системе;

$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{2^2}{3(3-2)} = 1{,}333$ машины – среднее количество машин, ожидающих в очереди;

$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{2}{3(3-2)} = 40$ мин. – среднее время ожидания в очереди;

$r = \frac{\lambda}{\mu} = \frac{2}{3} = 0{,}666$ – доля времени, в течение которого механик занят;

$P_0 = 1 - \frac{\lambda}{\mu} = 1 - 0{,}666 = 0{,}333$ – вероятность того, что в системе нет ни одного клиента.

Вероятности того, что в системе находится более чем k машин:

при k = 1 $P_{n>1} = \left(\frac{2}{3}\right)^2 = 0{,}444$, то есть существует 44,4 % шансов на то, что в системе находится более одной машины, и т. д.

Многоканальная система обслуживания M/M/S
с неограниченной очередью

Предполагается, что клиенты ожидают в общей очереди и обращаются в первый освободившийся канал обслуживания. В многоканальной системе поток заявок подчиняется *пуассоновскому* закону, а время обслуживания – *экспоненциальному.* Приходящий первым обслуживается первым, и все каналы обслуживания работают в одинаковом темпе.

Для расчета параметров многоканальной системы обслуживания удобно использовать соответствующее программное обеспечение.

Для многоканальной системы с неограниченной очередью должно выполняться условие $\frac{r}{n} < 1$, где r – параметр загрузки системы (среднее число занятых каналов); n – минимальное количество каналов, при котором очередь не будет расти до бесконечности. В противном случае предельные вероятности существовать не могут.

Формулы для описания системы M/M/S:

$P_0 = \left(1 + \frac{r}{1!} + \frac{r^2}{2!} + \ldots + \frac{r^n}{n!} + \frac{r^{n+1}}{n!(n-r)}\right)^{-1}$ – вероятность того, что система свободна;

$P_n = \frac{r^n}{n!} P_0$ – вероятность того, что в системе находится n заявок;

$P_q = \frac{r^{n+1}}{n!(n-r)} P_0$ – вероятность того, что заявка окажется в очереди;

$r = \frac{\lambda}{\mu}$ – среднее число занятых каналов;

$L_q = \frac{r^{n+1}}{n \cdot n! \left(1 - \frac{r}{n}\right)^2} P_0$ – среднее число заявок в очереди;

$L_s = L_q + r$ – среднее число заявок в системе;

$W_q = \frac{1}{\lambda} L_q$ – среднее время нахождения заявки в очереди;

$$W_s = \frac{1}{\lambda} L_s = W_q + \frac{r}{\lambda}$$ – среднее время нахождения заявки в системе.

Пример 44.2. Сберкасса имеет трех контролеров-кассиров (n = 3) для обслуживания вкладчиков. Поток вкладчиков поступает в сберкассу с интенсивностью λ = 30 чел./ч. Средняя продолжительность обслуживания контролером-кассиром одного вкладчика $\bar{t}_{обсл} = 3$ мин.

Определить характеристики сберкассы как объекта СМО.

Решение.

Интенсивность потока обслуживания $\mu = 1/\bar{t}_{обсл} =$ 1/3 = 0,333, интенсивность нагрузки r = 1,5.

1. Вероятность простоя контролеров-кассиров в течение рабочего дня:

$$P_0 = \left(1 + \frac{1{,}5}{1!} + \frac{1{,}5^2}{2!} + \frac{1{,}5^3}{3!} + \frac{1{,}5^4}{3!\,(3-1{,}5)}\right)^{-1} = 0{,}21.$$

2. Вероятность застать всех контролеров-кассиров занятыми:

$$P_n = \frac{1{,}5^3}{3!} \cdot 0{,}21 = 0{,}118.$$

3. Вероятность очереди: $P_q = \dfrac{1{,}5^4}{3!\,(3-1{,}5)} \cdot 0{,}21 = 0{,}118.$

4. Среднее число заявок в очереди:

$$L_q = \frac{1{,}5^4}{3 \cdot 3!\left(1 - \frac{1{,}5}{3}\right)^2} \cdot 0{,}21 = 0{,}236.$$

5. Среднее время ожидания заявки в очереди:

$$W_q = \frac{1}{0{,}5} \cdot 0{,}236 = 0{,}472 \text{ мин.}$$

6. Среднее время пребывания заявки в СМО:

$$W_s = 0{,}472 + 3 = 3{,}472 \text{ мин.}$$

Модель с постоянным временем обслуживания M/D/1

Некоторые системы имеют *постоянное,* а не экспоненциально распределенное время обслуживания. В таких системах клиенты обслуживаются в течение фиксированного периода времени, как, например, на автоматической мойке автомобилей.

Для модели с постоянным темпом обслуживания значения величин L_q и W_q вдвое меньше, чем соответствующие значения в модели, имеющей переменный темп обслуживания.

Формулы, описывающие модель:

$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}$ – средняя длина очереди;

$W_q = \frac{\lambda}{2\mu(\mu - \lambda)}$ – среднее время ожидания в очереди;

$L_s = L_q + \frac{\lambda}{\mu}$ – среднее число клиентов в системе;

$W_s = W_q + \frac{1}{\mu}$ – среднее время ожидания в системе.

Пример 44.3 (*утилизация отходов).* Компания собирает и утилизирует алюминиевые отходы и стеклянные бутылки. Водители автомобилей, доставляющих сырье для вторичной переработки, ожидают в очереди на разгрузку в среднем 15 мин. Время простоя водителя и автомобиля оценивается в 6 тыс. грн в час.

Новая автоматическая линия может обслуживать контейнеровозы с постоянным темпом 12 машин в час (5 мин. на одну машину). Время прибытия контейнеровозов подчиняется пуассоновскому закону с параметром λ = 8 автомобилей в час. Если новая линия будет использоваться, то амортизационные затраты составят 0,3 тыс. грн на один контейнеровоз. Следует ли использовать новую линию?

Решение.

Затраты на простой одного автомобиля в очереди за одну поездку в системе без автоматической линии составляют

$$W_q \cdot 6 = \frac{1}{4} \cdot 6 = 1{,}5 \text{ тыс. грн.}$$

В системе с автоматической линией время ожидания в очереди при λ = 8 автомобилей в час и μ = 12 автомобилей в час будет равно:

$$W_q = \frac{\lambda}{2\mu(\mu-\lambda)} = \frac{8}{2\cdot 12\cdot (12-8)} = \frac{1}{12} \text{ ч.}$$

Затраты на простой автомобиля в очереди в этом случае составят:

$$W_q \cdot 6 = \frac{1}{12}\cdot 6 = 0{,}5 \text{ тыс. грн.}$$

Сокращение времени простоя привело к сокращению затрат на простой одного автомобиля за одну поездку на сумму в 1,5 – 0,5 = 1 тыс. грн. При условии, что затраты по эксплуатации линии на один контейнеровоз составляют 0,3 тыс. грн, общие затраты составят 0,5 + 0,3 = 0,8 тыс. грн. Система с линией дает экономию в 1,5 – 0,8 = 0,7 тыс. грн.

Таким образом, новую линию использовать следует.

Модель с ограниченной популяцией

Если число потенциальных клиентов системы обслуживания *ограничено,* мы имеем дело со специальной моделью. Такая задача может возникнуть, например, если речь идет об обслуживании оборудования фабрики, имеющей пять станков.

Особенность этой модели по сравнению с тремя рассмотренными ранее в том, что существует *взаимозависимость* между длиной очереди и темпом поступления заявок.

Модель с ограниченной очередью

Модель отличается от предыдущих тем, что число мест в очереди *ограничено.* В этом случае заявка, прибывшая в систему, когда все каналы и места в очереди заняты, покидает систему необслуженной, то есть получает отказ.

Основной характеристикой качества системы является отказ заявке в обслуживании.

Ограничения на длину очереди могут быть из-за:

1) ограничения сверху времени пребывания заявки в очереди;

2) ограничения сверху длины очереди;

3) ограничения общего времени пребывания заявки в системе.

Формулы, описывающие модель:

$$P_0=\left(1+\frac{r}{1!}+\frac{r^2}{2!}+\ldots+\frac{r^n}{n!}+\frac{r^{n+1}}{n!(n-r)}\left(1-\left(\frac{r}{n}\right)^m\right)\right)^{-1}$$ – вероятность простоя каналов обслуживания, когда нет заявок;

$$P_{отк}=\frac{r^{n+m}}{n!\,n^m}\cdot P_0$$ – вероятность отказа в обслуживании;

$$P_{обсл}=1-P_{отк}$$ – вероятность обслуживания;

$$A=P_{обсл}\cdot\lambda$$ – абсолютная пропускная способность;

$$\bar{n}=\frac{A}{\mu}$$ – среднее число занятых каналов;

$$\bar{L}_q=\frac{r^{n+1}}{n\cdot n!\left(1-\frac{r}{n}\right)^2}P_0\cdot\left(1-\left(\frac{r}{n}\right)^m\left(m+1-m\frac{r}{n}\right)\right)$$ – среднее число заявок в очереди;

$$\bar{W}_q=\frac{1}{\lambda}\bar{L}_q$$ – среднее время ожидания обслуживания;

$$\bar{L}_s=\bar{L}_q+\bar{n}$$ – среднее число заявок в системе;

$$t=\frac{\bar{L}_s}{\lambda}$$ – среднее время пребывания в системе.

Пример 44.4. Магазин получает ранние овощи из пригородных теплиц. Автомобили с грузом прибывают в разное время с интенсивностью λ = 6 машин в день. Подсобные помещения и оборудование для подготовки овощей к продаже позволяют обрабатывать и хранить товар, привезенный двумя автомашинами (m = 2). В магазине работают три фасовщика (n = 3), каждый из которых в среднем может обрабатывать товар с одной машины в течение $\bar{t}_{обс}$ = 4 ч. Продолжительность рабочего дня при сменной работе составляет 12 ч.

Определить, какова должна быть емкость подсобных помещений, чтобы вероятность полной обработки товаров была $P^*_{обсл}$ ≥ 0,97.

Решение.

Определим интенсивность загрузки фасовщиков:

$$r = \frac{\lambda}{\mu} = \frac{6}{3} = 2, \quad \mu = \frac{1}{\bar{t}_{обс}} = \frac{1}{12/4} = 3 \text{ авт./дн.}$$

1. Найдем вероятность простоя фасовщиков при отсутствии машин (заявок):

$$P_0 = \left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3!\,(3-2)}\left(1 - \frac{2}{3}\right)^2\right)^{-1} = 0{,}128.$$

2. Вероятность отказа в обслуживании:

$$P_{отк} = \frac{2^5}{3!\,3^2} \cdot 0{,}128 = 0{,}075$$

3. Вероятность обслуживания: $P_{обсл} = 1 - 0{,}075 = 0{,}925.$

Так как $P_{обсл}$ = 0,925 < $P^*_{обсл}$ = 0,97, произведем аналогичные вычисления для m = 3, получим $P_0 = 0{,}122, \quad P_{отк} = 0{,}048, \quad P_{обсл} = 0{,}952.$

Так как $P_{обсл}$ = 0,952 < $P^*_{обсл}$ = 0,97, примем m = 4.

Для этого случая $P_0 = 0{,}12, \quad P_{отк} = 0{,}028, \quad P_{обсл} = 0{,}972.$

Таким образом, 0,972 > 0,97, емкость подсобных помещений необходимо увеличить до m = 4.

Для достижения заданной вероятности обслуживания можно увеличивать число фасовщиков, проводя последовательно вычисления СМО для n = 4, 5 и т. д. Задачу можно решить, увеличивая емкость подсобных помещений, число фасовщиков, уменьшая время обработки товаров.

Найдем остальные параметры СМО для рассчитанного случая при $P_0 = 0{,}12, \quad P_{отк} = 0{,}028, \quad P_{обсл} = 0{,}972.$

4. Абсолютная пропускная способность:

$$A = 0{,}972 \cdot 6 = 5{,}832 \text{ авт./дн.}$$

5. Среднее число занятых обслуживанием каналов (фасовщиков):

$$\bar{n} = \frac{5{,}832}{3} = 1{,}944.$$

6. Среднее число заявок в очереди:

$$\bar{L}_q = \frac{2^4}{3 \cdot 3!\left(1 - \frac{2}{3}\right)^2} \cdot 0{,}12 \cdot \left(1 - \left(\frac{2}{3}\right)^4 \cdot \left(4 + 1 - 4 \cdot \frac{2}{3}\right)\right) = 0{,}548.$$

7. Среднее время ожидания обслуживания:

$$\overline{W}_q = \frac{1}{6} \cdot 0{,}548 = 0{,}09 \text{ дн.}$$

8. Среднее число машин в магазине: $\bar{L}_s = 0{,}548 + 1{,}944 = 2{,}492$ авт.

9. Среднее время пребывания машины в магазине:

$$t = \frac{\bar{L}_s}{\lambda} = \frac{2{,}492}{6} = 0{,}415 \text{ дн.}$$

Вопросы для самодиагностики

1. Какие основные параметры определяют конфигурацию системы массового обслуживания?

2. Какое обычно распределение вероятностей используется в теории массового обслуживания для описания простейшего потока заявок, поступающих на вход системы?

3. Каким в теории массового обслуживания предполагается количество заявок в популяции?

4. Одна работница обслуживает тридцать станков, обеспечивая их запуск после остановки. Как можно охарактеризовать модель такой системы массового обслуживания?

5. Какое обычно распределение вероятностей используется в теории массового обслуживания для описания времени, затрачиваемого на обслуживание заявок?

6. Ремонт вышедших из строя компьютеров на фирме осуществляют три специалиста, работающие одновременно и независимо друг от

друга. Как можно охарактеризовать модель такой системы массового обслуживания?

Упражнения

44.1. На стоянке автомобилей возле магазина имеются 3 места, каждое из которых отводится под один автомобиль. Автомобили прибывают на стоянку с интенсивностью 20 автомобилей в час. Продолжительность пребывания автомобилей на стоянке составляет в среднем 15 мин. Стоянка на проезжей части не разрешается.

Определить среднее количество мест, не занятых автомобилями, и вероятность того, что прибывший автомобиль не найдет на стоянке свободного места.

44.2. АТС предприятия обеспечивает не более 5 переговоров одновременно. Средняя продолжительность разговоров составляет 1 мин. На станцию поступает в среднем 10 вызовов в сек.

Определить характеристики АТС как объекта СМО.

44.3. В грузовой речной порт поступает в среднем 6 сухогрузов в сутки. В порту имеются 3 крана, каждый из которых обслуживает 1 сухогруз в среднем за 8 ч. Краны работают круглосуточно.

Определить характеристики работы порта как объекта СМО и в случае необходимости дать рекомендации по улучшению его работы.

44.4. На автозаправочной станции установлены 2 колонки для выдачи бензина. Около станции находится площадка на 2 автомашины для ожидания заправки. На станцию прибывает в среднем одна машина в 3 мин. Среднее время обслуживания одной машины составляет 2 мин.

Определить характеристики работы автозаправочной станции как объекта СМО.

44.5. АТС поселка обеспечивает не более 5 переговоров одновременно. Время переговоров в среднем составляет около 3 мин. Вызовы на станцию поступают в среднем через 2 мин.

Определить вероятность того, что заявка получит отказ, среднее число занятых каналов, абсолютную пропускную способность АТС.

44.6. На автозаправочной станции (АЗС) имеются 3 колонки. Площадка при станции, на которой машины ожидают заправку, может вместить не более одной машины, и если она занята, то очередная машина, прибывшая к станции, в очередь не становится, а проезжает на соседнюю станцию. В среднем машины прибывают на станцию каждые 2 мин. Процесс заправки одной машины продолжается в среднем 2,5 мин.

Определить вероятность отказа, абсолютную пропускную способность АЗС, среднее число машин, ожидающих заправку, среднее время ожидания машины в очереди, среднее время пребывания машины на АЗС (включая обслуживание).

45. Задачи упорядочивания и координации. Сетевое планирование

45.1. Характеристика задач упорядочивания и координации. Постановка задачи оптимизации последовательности обработки изделий

Задачи упорядочивания и *координации* возникают при необходимости определения оптимального (рационального) управления отдельными операциями или комплексами операций.

В отличие от задач массового обслуживания в этих задачах определяются оптимальные последовательность, порядок выполнения работ, сроки начала и окончания каждой операции, объемы выделяемых ресурсов.

Оптимальными (рациональными) считаются решения, обеспечивающие наибольшую эффективность системы в целом.

Критерии, используемые в задачах упорядочивания:

максимальное использование производственных мощностей, то есть минимизация суммарного времени простоя оборудования;

минимизация суммарного времени обработки всех требований и заявок.

Классифицируют задачи упорядочивания:

по характеру обслуживания – детерминированные, стохастические;

по характеру учета времени – динамические, статистические.

Классическим примером задачи упорядочивания, находящим применение в различных отраслях промышленности, является задача планирования производственной линии – оптимизации последовательности обработки изделий.

Постановка задачи: имеется несколько изделий, каждое из которых надо обработать на двух машинах последовательно (сначала на первой, потом на второй). Известно время обработки каждого изделия на каждой машине (табл. 45.1).

Таблица 45.1

Исходные данные задачи

$N_{изд}$	1	2	3	…	n
t_{1j}	t_{11}	t_{12}	t_{13}	...	t_{1n}
t_{2j}	t_{21}	t_{22}	t_{23}	...	t_{2n}

Здесь t_{1j} – время обработки j-го изделия на первой машине;

t_{2j} – время обработки j-го изделия на второй машине.

Формализованное описание задачи:

время перехода детали от первой машины к другой незначительно и его можно не учитывать;

каждое изделие должно быть обработано в определенном порядке: вначале на первой машине, потом – на второй;

каждое обслуживание должно быть завершено прежде, чем начнется следующее, прерываться нельзя.

Пусть t_{2j}^0 – время простоя второй машины. Тогда суммарное время обработки изделий на второй машине составит: $Z = \sum_{j=1}^{n} t_{2j} + \sum_{j=1}^{n} t_{2j}^0$.

Так как $\sum_{j=1}^{n} t_{2j}$ известно, то необходимо минимизировать $\sum_{j=1}^{n} t_{2j}^0$.

45.2. Комбинаторные задачи упорядочивания

Задачи упорядочивания являются частными случаями *комбинаторных задач*.

Термин ***комбинаторная оптимизация*** означает, что существует конечное число допустимых вариантов, перебрав которые можно найти оптимальное решение. Экстремальные комбинаторные задачи, называемые также *задачами выбора*, состоят в отыскании среди конечного множества альтернатив одной, которой отвечает экстремальное значение принятой целевой функции.

Классическим примером комбинаторной задачи является *задача о коммивояжере*. 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 или с минимальными затратами на всю поездку.

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 или стоимость переезда из города i в город j (чтобы избежать бессмысленных значений $x_{ii}=1$, предполагается, что c_{ii} равны нулю).

Тогда математическая модель имеет вид:

$$F(X)=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}\rightarrow min,$$

$$\sum_{j=1}^{n}x_{ij}=1,\quad i=\overline{1,n},\quad \sum_{i=1}^{n}x_{ij}=1,\quad j=\overline{1,n}.$$

К приведенным ограничениям необходимо добавить условия на недопустимость подциклов, то есть повторного посещения городов (за исключением исходного).

Другой пример комбинаторных задач – это общая *задача календарного планирования,* которая 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.

Большинство комбинаторных и целочисленных типов задач, таких, как задача с неделимостями, задача коммивояжера, задача календарного планирования, принадлежат 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), то есть 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.

45.3. Содержание и сферы использования сетевых методов планирования и управления

Система *методов сетевого планирования и управления* (СПУ) – система методов планирования и управления разработкой крупных народнохозяйственных комплексов, научными исследованиями, конструкторской и технологической подготовкой производства, новых видов изделий, строительством и реконструкцией, капитальным ремонтом основных фондов путем применения сетевых графиков.

Система СПУ позволяет:

формировать календарный план реализации некоторого комплекса работ;

выявлять и мобилизовать резервы времени, трудовые, материальные и денежные ресурсы;

осуществлять управление комплексом работ по принципу «ведущего звена» с прогнозированием и предупреждением возможных срывов в ходе работ;

повышать эффективность управления в целом при четком распределении ответственности между руководителем разных уровней и исполнителями работ.

Современное сетевое планирование начинается с разбиения программы работ на операции. Определяются оценки продолжительности операций, и строится *сетевая модель* (график). Построение сетевой модели позволяет проанализировать все операции и внести улучшения в структуру модели до начала ее реализации. Строится календарный график, определяющий начало и окончание каждой операции, а также взаимосвязи с другими операциями графика. Календарный график выявляет критические операции, которым надо уделять особое внимание, чтобы закончить все работы в директивный срок. Что касается некритических операций, то календарный план позволяет определить резервы времени, которые можно выгодно использовать при задержке выполнения работ или эффективном применении как трудовых, так и финансовых ресурсов.

45.4. Элементы сетевого графика, методика его построения

Сетевая модель – графическое изображение плана выполнения комплекса работ, состоящего из нитей (работ) и узлов (событий), которые отражают логическую взаимосвязь всех операций.

В основе сетевого моделирования лежит изображение планируемого комплекса работ в виде графа.

Граф – схема, состоящая из заданных точек (вершин), соединенных системой линий. Отрезки, соединяющие вершины, называются *ребрами (дугами) графа*. *Ориентированным* называется такой граф, на котором стрелкой указаны направления всех его ребер (дуг), что позволяет определить, какая из двух его граничных вершин является начальной, а какая – конечной. Исследование таких сетей проводится методами теории графов.

Теория графов оперирует понятием пути, объединяющим последовательность взаимосвязанных ребер. Контур означает такой путь, у которого начальная вершина совпадает с конечной. *Сетевой график* – это ориентированный граф без контуров.

В сетевом моделировании имеются два основных элемента – работа и событие.

Работа – это активный процесс, требующий затрат ресурсов, либо пассивный (ожидание), приводящий к достижению намеченного результата.

Фиктивная работа – это связь между результатами работ (событиями), не требующая затрат времени и ресурсов.

Событие – это результат (промежуточный или конечный) выполнения одной или нескольких предшествующих работ.

Путь – это любая непрерывная последовательность (цепь) работ и событий.

Критический путь – это путь, не имеющий резервов и включающий самые напряженные работы комплекса. Работы, расположенные на критическом пути, называют *критическими*. Все остальные работы являются *некритическими* (ненапряженными) и обладают резервами времени, которые позволяют передвигать сроки их выполнения, не влияя на общую продолжительность выполнения всего комплекса работ.

При построении сетевых моделей необходимо соблюдать следующие правила:

1. Сеть изображается слева направо, и каждое событие с большим порядковым номером изображается правее предыдущего. Общее направление стрелок, изображающих работы, также в основном должно быть расположено слева направо, при этом каждая работа должна выходить из события с меньшим номером и входить в событие с большим номером.

2. Два соседних события могут объединяться лишь одной работой. Для изображения параллельных работ вводятся промежуточное событие и фиктивная работа (рис. 45.1).

3. В сети не должно быть тупиков, то есть промежуточных событий, из которых не выходит ни одна работа (рис. 45.2).

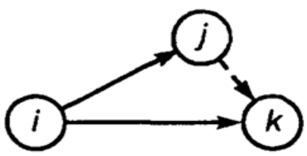


Рис. 45.1. **Параллельная работа**

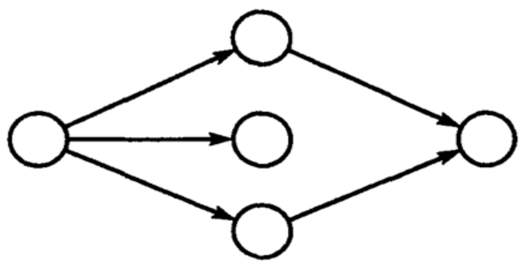


Рис. 45.2. **Пример тупика**

4. В сети не должно быть промежуточных событий, которым не предшествует хотя бы одна работа (рис. 45.3).

5. В сети не должно быть замкнутых контуров, состоящих из взаимосвязанных работ, создающих замкнутую цепь (рис. 45.4).

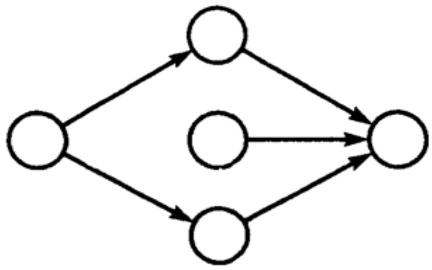


Рис. 45.3. **Ошибка сети**

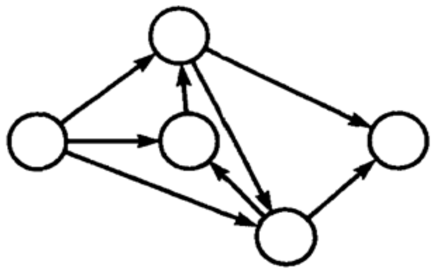


Рис. 45.4. **Замкнутая цепь**

Для правильной нумерации событий поступают следующим образом: нумерация событий начинается с исходного события, которому дается номер 1. Из исходного события 1 вычеркивают все исходящие из него работы, на оставшейся сети вновь находят событие, в которое не входит ни одна работа. Этому событию дается номер 2. Затем вычеркивают работы, выходящие из события 2, и вновь находят на оставшейся части сети событие, в которое не входит ни одна работа, ему присваивается номер 3, и так продолжается до завершающего события.

Пример нумерации сетевого графика показан на рис. 45.5.

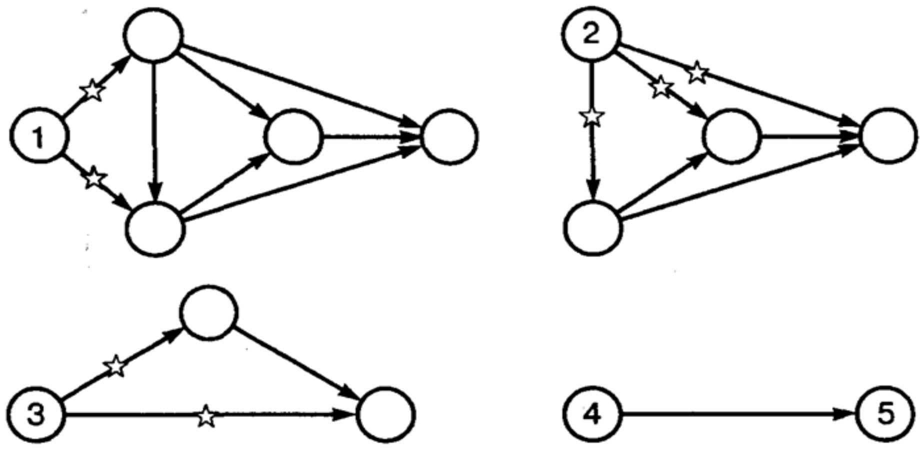


Рис. 45.5. **Нумерация сетевого графика**

Продолжительность выполнения работ устанавливается на основании действующих нормативов или по экспертным оценкам специалистов.

В первом случае временные оценки являются детерминированными (однозначными), во втором – стохастическими (вероятностными).

Пример 45.1. Построить сетевую модель программы опроса общественного мнения, которая включает разработку (A; 1 день) и распечатку анкет (B; 0,5 дня), прием на работу (C; 2 дня) и обучение (D; 2 дня) персонала, выбор опрашиваемых лиц (E; 2 дня), рассылку им анкет (F; 1 день) и анализ полученных данных (G; 5 дней).

Решение.

Из условия задачи известно содержание работ, но явно не указаны взаимосвязи между работами. Поэтому необходимо проанализировать смысл каждой конкретной работы и выяснить, какие из остальных работ должны ей непосредственно предшествовать.

Исходной работой, начинающей сетевой график, в данном случае является «прием на работу» (C), поскольку все остальные работы должны выполняться уже принятыми на работу сотрудниками. Перед выполнением всех работ по опросу общественного мнения сотрудников необходимо обучить персонал (D). Перед тем как разослать анкеты (F), их надо разработать (A), распечатать (B) и выбрать опрашиваемых лиц (E), причем работу с анкетами и выбор лиц можно выполнять одновременно. Завершающей работой проекта является анализ полученных данных (G), который нельзя выполнить без предварительной рассылки анкет (F).

В результате этих рассуждений построим сетевую модель и пронумеруем события модели (рис. 45.6).

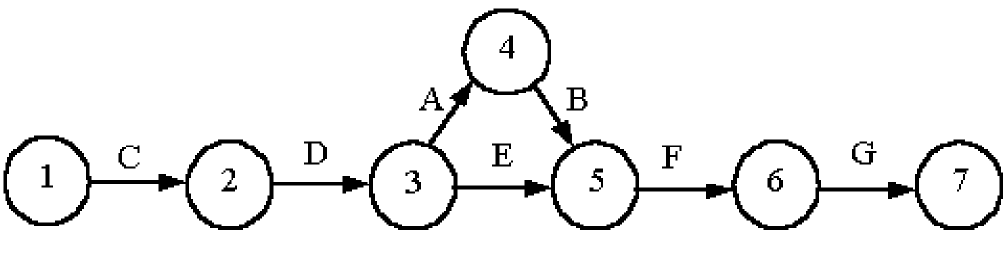


Рис. 45.6. **Сетевая модель программы опроса общественного мнения**

Пример 45.2. Построить сетевую модель, включающую работы A, B, C, ..., L, которая отображает следующее упорядочение работ:

1) A, B и C – исходные операции проекта;
2) A и B предшествуют D;
3) B предшествует E, F и H;
4) F и C предшествует G;
5) E и H предшествуют I и J;

6) C, D, F и J предшествуют K;

7) K предшествует L.

Решение.

В пункте 1) условия явно указано, что A, B и C являются исходными работами, поэтому изобразим их тремя стрелками, выходящими из исходного события 1.

Пункт 2) условия означает, что стрелки работ A и B должны закончиться в одном событии, из которого выйдет стрелка работы D. Но поскольку стрелки работ A и B также начинаются в одном событии, то имеет место параллельность работ, которая недопустима правилами построения сетевых моделей (рис. 45.7). Для ее устранения введем дополнительное событие 2, в которое войдет работа B, после чего соединим события 2 и 3, в которые входят работы A и B пунктирной стрелкой фиктивной работы. В данном случае фиктивная работа (2,3) не соответствует никакой реальной работе, а лишь отображает логическую связь между работами B и D.

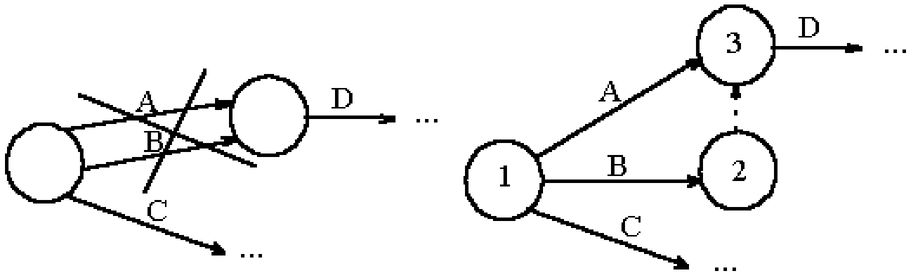


Рис. 45.7. **Устранение параллельности работ A и B**

Дальнейшее построение рассмотрим с помощью рис. 45.8.

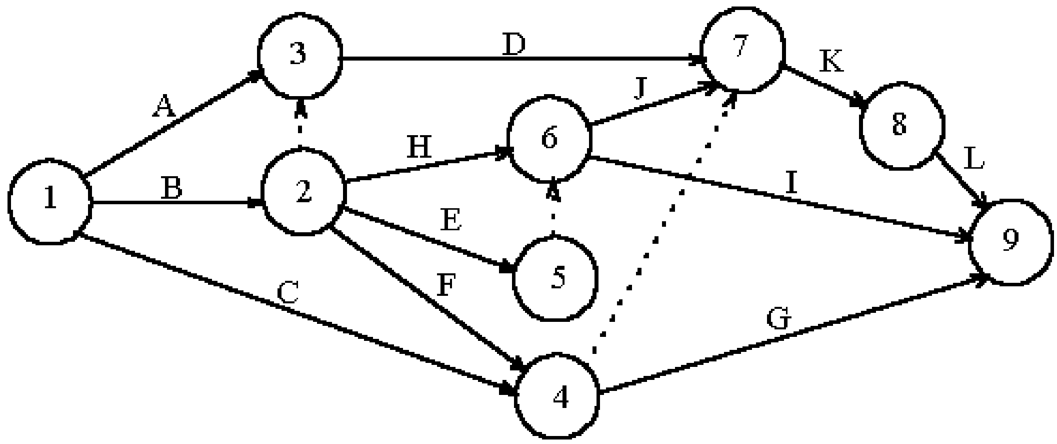


Рис. 45.8. **Сетевая модель задачи**

Согласно пункту 3) условия задачи из события 2 выходят три стрелки работ E, F и H.

Согласно пункту 4) условия задачи стрелки работ C и F должны войти в общее событие, из которого выйдет стрелка работы G.

Проблема с параллельностью работ E и H (пункт 5) условия задачи) решается путем введения дополнительного события 5 и фиктивной работы (5,6).

Для отображения в сетевой модели пункта 6) условия задачи введем стрелки работ D и J в событие 7, а связь работ F и C с работой K отобразим с помощью фиктивной работы (4,7). Стрелки работ F и C нельзя было напрямую вводить в событие 7, потому что после них должна следовать работа G, которая с работами D и J никак не связана.

Стрелка работы L выходит из события 8, то есть после окончания работы K в соответствии с пунктом 7) условия задачи.

Поскольку в условии не указано, что работы L, I и G предшествуют каким-либо другим работам, то эти работы являются завершающими и их стрелки войдут в завершающее событие 9. Нумерацию событий проводят после построения сетевого графика, следя за тем, чтобы номер начального события каждой работы был меньше номера ее конечного события.

45.5. Расчеты основных параметров сетевого графика

Календарное планирование предусматривает определение моментов начала и окончания каждой работы и других временных характеристик сетевого графика. Это позволяет проанализировать сетевую модель, выявить критические работы, непосредственно определяющие срок выполнения проекта, провести оптимизацию использования ресурсов (временных, финансовых, исполнителей).

Основным временным параметром сетевого графика является ***продолжительность критического пути***.

Расчет критического пути включает два этапа.

Первый называется *прямым проходом*. Вычисления начинают с исходного события и продолжают до тех пор, пока не будет достигнуто завершающее событие. Для каждого события определяется одно число, представляющее ранний срок его наступления.

На втором этапе, называемом *обратным проходом*, вычисления начинают с завершающего события и продолжают, пока не будет достигнуто исходное событие. Для каждого события вычисляется поздний срок его наступления.

Расчет сетевой модели начинают с временных параметров событий, которые вписывают непосредственно в вершины сетевого графика (рис. 45.9):

$T_p(i)$ – ранний срок наступления события i, минимально необходимый для выполнения всех работ, которые предшествуют событию i;

$T_n(i)$ – поздний срок наступления события i, превышение которого вызовет задержку наступления завершающего события сети;

$R(i)=T_n(i)-T_p(i)$ – резерв времени события i, то есть время, на которое может быть отсрочено наступление события i без нарушения сроков завершения проекта в целом.

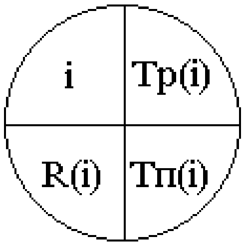


Рис. 45.9. **Отображение временных параметров событий на сетевом графике**

Ранние сроки свершения событий $T_p(i)$ рассчитываются от исходного ($И$) к завершающему ($З$) событию следующим образом:

1) для исходного события $И$: $T_p(И)=0$;

2) для всех остальных событий i:

$$T_p(i)=\max_{\forall(k,i)}\left[T_p(k)+t(k,i)\right],$$

где максимум берется по всем работам (k,i), входящим в событие i; $t(k,i)$ – длительность работы (k,i) (рис. 45.10).

Поздние сроки свершения событий $T_n(i)$ рассчитываются от завершающего к исходному событию:

1) для завершающего события $З$: $T_п(З) = T_р(З)$;

2) для всех остальных событий:

$$T_п(i) = \min_{\forall(i,j)} \left[T_п(j) - t(i,j)\right],$$

где минимум берется по всем работам (i,j), выходящим из события i; $t(i,j)$ – длительность работы (i,j) (рис. 45.11).

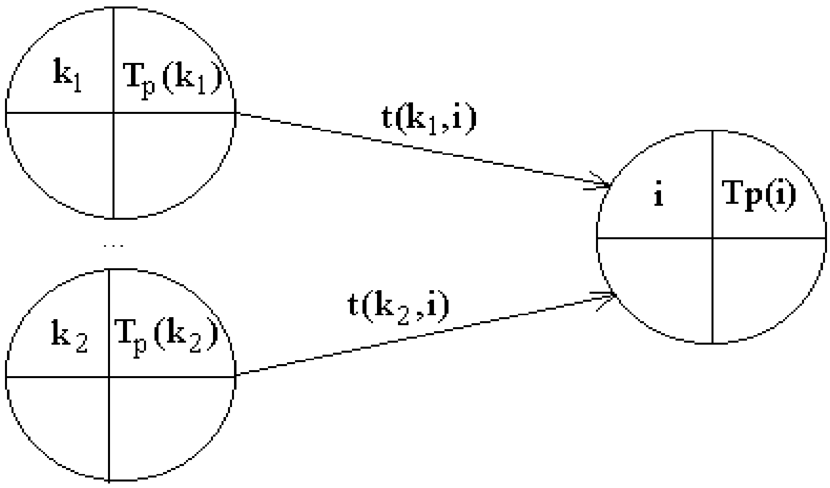


Рис. 45.10. **Расчет раннего срока $T_р(i)$ свершения события i**

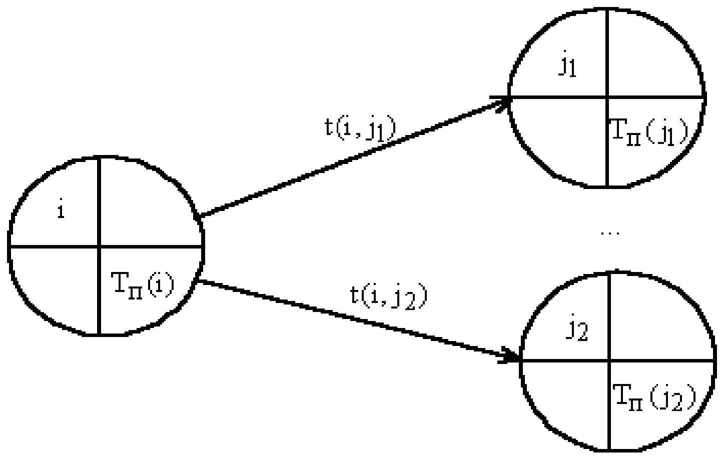


Рис. 45.11. **Расчет позднего срока $T_п(i)$ свершения события i**

Временные параметры работ определяются на основе ранних и поздних сроков событий:

$T_{рн}(i,j) = T_р(i)$ – ранний срок начала работы;

$T_{ро}(i,j) = T_р(i) + t(i,j)$ – ранний срок окончания работы;

$T_{по}(i,j) = T_п(j)$ – поздний срок окончания работы;

$T_{пн}(i,j) = T_п(j) - t(i,j)$ – поздний срок начала работы;

$R_п(i,j) = T_п(j) - T_р(i) - t(i,j)$ – полный резерв времени выполнения работы показывает максимальное время, на которое можно увеличить длительность работы (i,j) или отсрочить ее начало, чтобы не нарушился срок завершения проекта в целом;

$R_с(i,j) = T_р(j) - T_р(i) - t(i,j)$ – свободный резерв времени выполнения работы показывает максимальное время, на которое можно увеличить продолжительность работы (i,j) или отсрочить ее начало, не меняя ранних сроков начала последующих работ.

Путь – это последовательность работ в сетевом графике (в частном случае это одна работа), в которой конечное событие одной работы совпадает с начальным событием следующей за ней работы.

Полный путь – это путь от исходного до завершающего события.

Критический путь – максимальный по продолжительности полный путь. Работы, лежащие на критическом пути, называют **критическими**. Критические работы имеют нулевые свободные и полные резервы.

Подкритический путь – полный путь, ближайший по длительности к критическому пути.

Для проведения анализа временных параметров сетевой модели используют **график привязки**, который отображает взаимосвязь выполняемых работ во времени. По вертикальной оси графика привязки откладываются коды работ, по горизонтальной оси – отрезки, соответствующие длительностям работ (раннее начало и раннее окончание работ). График привязки можно построить на основе данных о продолжительности работ. При этом необходимо помнить, что работа (i,j) может выполняться только после того, как будут выполнены все предшествующие ей работы (k,i).

Пример 45.3. Компания разрабатывает строительный проект. Исходные данные по основным операциям проекта представлены в табл. 45.2.

Построить сетевую модель проекта, определить критические пути модели. Проанализировать, как влияет на ход выполнения проекта задержка работы D на 4 недели.

Таблица 45.2

Исходные данные

Название	Непосредственно предшествующие операции	Длительность, недели
A	–	4
B	–	6
C	A,B	7
D	B	3
E	C	4
F	D	5
G	E,F	3

Решение.

Построим сетевую модель и рассчитаем временные параметры событий. При поиске критических путей на сетевом графике будем использовать следующие условия его критичности: *необходимое условие* – нулевые резервы событий, лежащих на критическом пути; *достаточное условие* – нулевые полные резервы времени выполнения работ, лежащих на критическом пути.

Согласно необходимому условию два полных пути сетевой модели (рис. 45.12) $L_1 = 1,2,3,4,6,7$ и $L_2 = 1,3,4,6,7$ могут быть критическими.

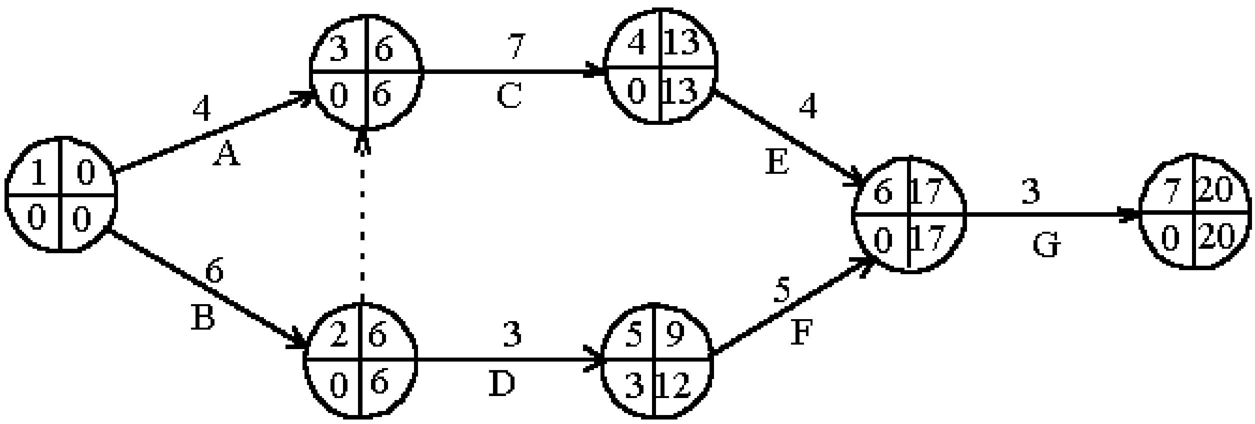


Рис. 45.12. **Сетевой график**

Проверим достаточное условие критичности для работ (1,2) и (1,3):

$$R_n(1,2) = T_n(2) - T_p(1) - t(1,2) = 6 - 0 - 6 = 0;$$

$$R_n(1,3) = T_n(3) - T_p(1) - t(1,3) = 6 - 0 - 4 = 2.$$

Путь L_2, начинающийся с работы (1,3) не является критическим, так как минимум одна из его работ (1,3) не является критической. Работа (1,3) имеет ненулевой полный резерв, а значит может быть задержана с выполнением, что недопустимо для критических работ.

Таким образом, сетевая модель имеет единственный критический путь $L_{kp} = 1,2,3,4,6,7$ длительностью $T_{kp} = 20$ недель. За выполнением работ этого пути необходим особый контроль, так как любое увеличение их длительности нарушит срок выполнения проекта в целом.

Работа D или (2,5) не является критической, ее полный резерв равен 3-м неделям. Это означает, что при задержке работы в пределах 3-х недель срок выполнения проекта не будет нарушен. Поэтому, если согласно условию работа D задержится на 4 недели, то весь проект закончится на 1 неделю позже.

45.6. Методы оптимизации сетевого графика по критерию времени

При поиске критических путей следует помнить, что *признаком критической работы* являются нулевые значения резервов времени. Это означает, что каждая последующая критическая работа будет начинаться *строго в момент окончания* предыдущей критической работы. Вследствие этого сдвиг любой из работ критического пути обязательно приведет к увеличению первоначальной длительности проекта (T_{kp}). Кроме того, следует учесть, что критический путь является *полным*, то есть соединяет исходное и завершающее события сети. Поэтому на графике привязки первая из работ критического пути всегда начинается в исходном событии сети с нулевого (начального) момента времени, а последняя из работ критического пути всегда завершается позже всех остальных работ сети в завершающем событии.

Из вышеприведенных соображений следует способ определения критического пути на графике привязки (все найденные работы выписываются последовательно справа налево):

1) найти на графике привязки и выписать работу (i,j), которая заканчивается позже всех остальных.

Это будет последняя работа критического пути (ее конечное событие будет иметь номер завершающего события сети);

2) из всех работ сети (k,i), конечное событие которых i совпадает с начальным событием i работы (i,j), найденной в п. 1), выбрать и выписать ту, которая на графике вплотную примыкает к работе (i,j);

3) из всех работ сети (l,k), конечное событие которых k совпадает с начальным событием k работы (k,i), найденной в п. 2), выбрать и выписать ту, которая на графике вплотную примыкает к работе (k,i);

4) продолжать п. 3) до тех пор, пока не будет найдена исходная работа сети, то есть начинающаяся в нулевой момент времени (ее начальное событие будет иметь номер исходного события сети, например 1).

Следует заметить, что если в сетевой модели несколько критических путей, то, выполняя вышеописанные действия, можно обнаружить несколько работ, удовлетворяющих сформулированным требованиям. В таком случае необходимо продолжать поиск по каждой из таких работ в отдельности.

Пример 45.4. По данным о кодах и длительностях работ в днях (табл. 45.3) построить график привязки сетевой модели, определить критические пути и их длительность. Определить свободные и полные резервы каждой работы, отметить на графике привязки свободные резервы работ.

Таблица 45.3

Исходные данные

(i,j)	1,2	1,3	1,4	1,5	2,3	3,6	3,7	4,5	4,6	5,7	6,7
$t(i,j)$, дни	3	3	2	10	2	5	9	10	6	1	4

Решение.

1. Построим график привязки (рис. 45.13).

2. Начнем поиск критических путей (справа налево) с работ, завершающих проект. На графике привязки (рис. 45.13) две работы (6,7) и (3,7), которые заканчиваются позже остальных в завершающем событии № 7.

Записываем работы, определенные как критические справа налево:

$$L_{kp1} = \ldots(6,7); \tag{45.1}$$

$$L_{kp2} = \ldots(3,7).$$

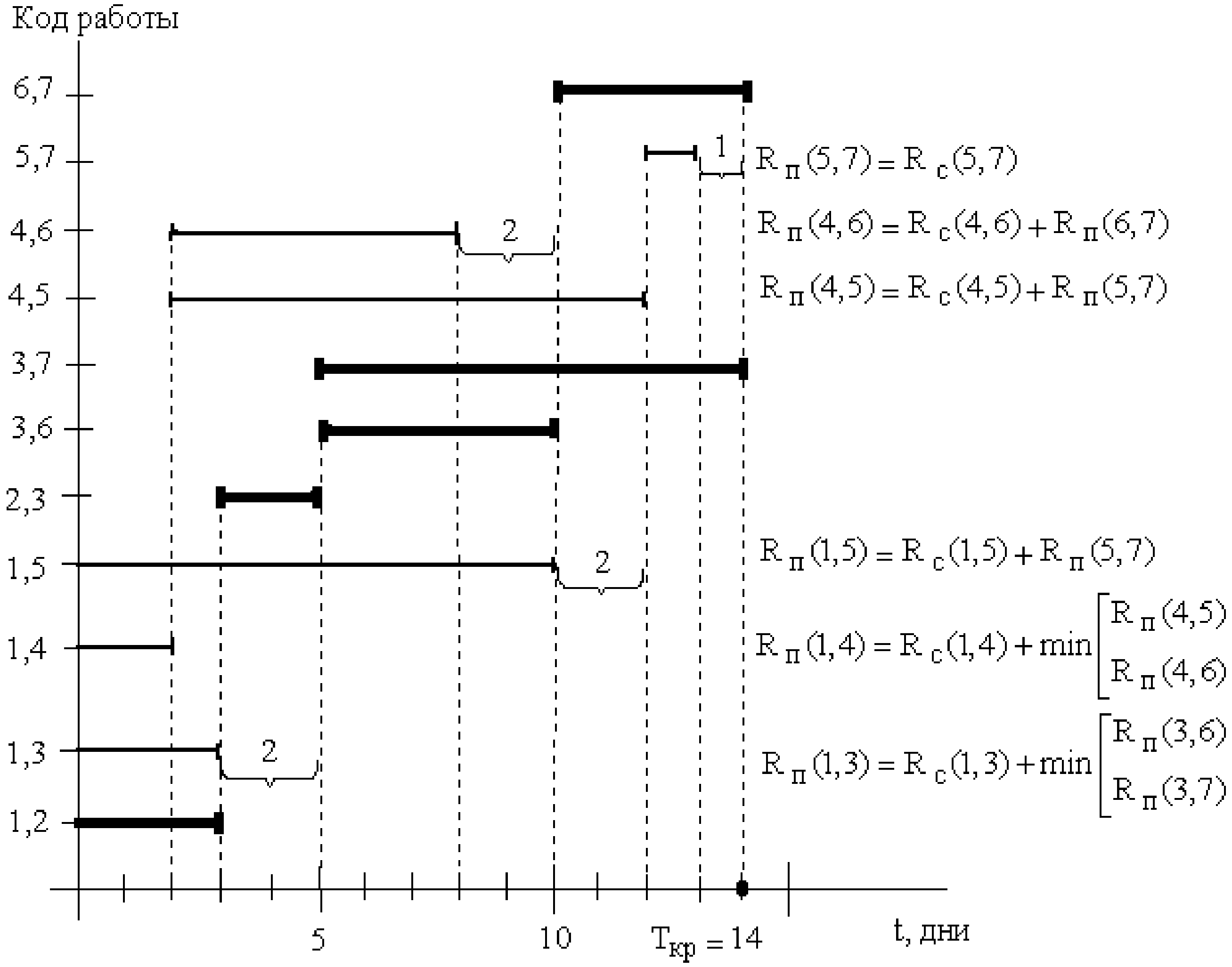


Рис. 45.13. **График привязки**

3. Найдем критическую работу из L_{kp1}, предшествующую (6,7). Код этой работы должен оканчиваться на 6. Таких работ две – (4,6) и (3,6). Но только одна из них, работа (3,6), по времени своего окончания вплотную «примыкает» на графике к началу работы (6,7). Допишем слева найденную критическую работу (3,6) к выражению (45.1):

$$L_{kp1} = \ldots(3,6)(6,7). \tag{45.2}$$

4. Найдем критическую работу из L_{kp1}, предшествующую (3,6). Код этой работы должен оканчиваться на 3. Таких работ две – (2,3) и (1,3).

Но только одна из них, работа (2,3), по времени своего окончания вплотную «примыкает» на графике к началу работы (3,6). Допишем слева найденную критическую работу (2,3) к выражению (45.2):

$$L_{kp1} = \ldots(2,3);(3,6);(6,7). \tag{45.3}$$

5. Найдем критическую работу из L_{kp1}, предшествующую (2,3). Код этой работы должен оканчиваться на 2. Работа (1,2) по времени своего окончания вплотную «примыкает» на графике к началу работы (2,3). С этой работы начинается критический путь L_{kp1}:

$$L_{kp1} = (1,2);(2,3);(3,6);(6,7).$$

6. Аналогичный поиск работ критического пути L_{kp2} приводит к результату $L_{kp2} = (1,2);(2,3);(3,7).$

В другой форме записи $L_{kp1} = 1,2,3,6,7$ и $L_{kp2} = 1,2,3,7$.

7. Для наглядности выделим на графике привязки критические работы жирной линией.

Определим резервы работ.

1. Для всех найденных критических работ впишем в табл. 45.4 *нулевые* значения свободного и полного резервов. Рассмотрим некритические работы, начиная с конца табл. 45.4.

Таблица 45.4

Резервы работ

i,j	$t(i,j)$	$R_c(i,j)$	$R_n(i,j)$	Критичность
1,2	3	**0**	**0**	Критическая
1,3	3	2	2	–
1,4	2	0	1	–
1,5	10	2	3	–
2,3	2	**0**	**0**	Критическая
3,6	5	**0**	**0**	Критическая
3,7	9	**0**	**0**	Критическая
4,5	10	0	1	–
4,6	6	2	2	–
5,7	1	1	1	–
6,7	4	**0**	**0**	Критическая

2. Работа (5,7), согласно графику привязки (см. рис. 45.13) заканчивается в 13-й день, а завершающее событие 7 сети, в которое она входит, наступает лишь в 14-й день. То есть, если работа (5,7) задержится на 1 день, то это не повлияет на срок выполнения проекта ($T_{kp}=14$ дней). Поскольку (5,7) завершающая работа сети, то ее полный и свободный резервы равны $R_n(5,7)=R_c(5,7)=1$.

3. Работа (4,6) заканчивается в 8-й день, в то время как последующая работа (6,7) начинается в 10-й день. То есть работа (4,6) может задержаться на 2 дня и это никак не повлияет на время начала последующей работы (6,7), то есть $R_c(4,6)=2$.

Используем правило: *полный резерв любой работы складывается из собственного свободного резерва и минимального из полных резервов непосредственно следующих работ.*

За работой (4,6) следует только критическая работа (6,7) с нулевым полным резервом. Поэтому $R_n(4,6)=R_c(4,6)+R_n(6,7)=2+0=2$.

4. Работа (4,5) заканчивается в 12-й день, в этот же день начинается следующая работа (5,7), то есть любая задержка выполнения работы (4,5) приведет к задержке начала работы (5,7). Это означает, что работа (4,5) не имеет свободного резерва $R_c(4,5)=0$. Но если сдвинуть во времени работу (4,5) на 1 день, то работа (5,7) также сдвинется на 1 день и это не нарушит срок выполнения проекта, так как у работы (5,7) есть временной резерв. Таким образом, согласно правилу:

$$R_n(4,5)=R_c(4,5)+R_n(5,7)=0+1=1.$$

5. Работа (1,5) заканчивается в 10-й день, в то время как последующая работа (5,7) начинается в 12-й день. То есть работа (1,5) может задержаться на 2 дня и это никак не повлияет на время начала последующей работы (5,7), то есть $R_c(1,5)=2$. Кроме того, поскольку последующая работа (5,7) имеет резерв в 1 день, то, в общем, работу (1,5) можно сдвинуть на 3 дня и это не нарушит сроков проекта (см. рис. 45.13), то есть $R_n(1,5)=R_c(1,5)+R_n(5,7)=2+1=3$.

6. Работа (1,4) заканчивается во 2-й день, и в этот же день начинаются следующие работы (4,5) и (4,6). То есть работа (1,4) не имеет свободного резерва времени $R_c(1,4)=0$.

Поскольку после работы (1,4) следуют две работы с различными полными резервами, то

$$R_n(1,4) = R_c(1,4) + min\{R_n(4,5);R_n(4,6)\} = 0 + min\{1;2\} = 0 + 1 = 1.$$

7. Работа (1,3) заканчивается в 3-й день, а следующие за ней работы (3,6) и (3,7) начинаются в 5-й день, то есть $R_c(1,3) = 2$. Поскольку обе последующие работы критические, то полный и свободный резерв работы (1,3) совпадают:

$$R_n(1,3) = R_c(1,3) + min\{R_n(3,6);R_n(3,7)\} = 2 + min\{0;0\} = 2 + 0 = 2.$$

8. Ненулевые свободные резервы работ обозначены на графике привязки фигурными скобками (см. рис. 45.13).

Вопросы для самодиагностики

1. Охарактеризовать задачи упорядочивания и координации.
2. Пояснить понятие комбинаторная оптимизация.
3. Описать методику построения сетевого графика.
4. Какие элементы и временные параметры сетевого графика?
5. Чему равно наиболее раннее время наступления события?
6. Чему равно наиболее позднее время наступления события?
7. Как сократить время выполнения проекта?
8. Чему равен полный резерв времени выполнения работы?

Упражнения

45.1. Рассмотреть следующий проект (табл. 45.5).

Найти критический путь.

За какое минимальное время может быть выполнен проект?

Сколько работ находится на критическом пути?

На сколько недель можно отложить выполнение работы *D* без отсрочки завершения проекта в целом?

На сколько недель можно отложить выполнение работы С без отсрочки завершения проекта в целом?

Таблица 45.5

Исходные данные

Работа	Непосредственно предшествующие работы	Время выполнения, недели
A	–	5
B	–	3
C	A	7
D	A	6
E	B	7
F	D, E	3
G	D, E	10
H	C, F	8

45.2. Университет рассматривает предложение о строительстве нового учебного корпуса. Работы, которые следует выполнить перед началом строительства, представлены в табл. 45.6.

Таблица 45.6

Исходные данные

Работа	Содержание работы	Непосредственно предшествующие работы	Время выполнения, недели
A	Определить место строительства	–	6
B	Разработать первоначальный проект	–	8
C	Получить разрешение на строительство	A, B	12
D	Выбрать архитектурную фирму	C	4
E	Разработать смету затрат на строительство	C	6
F	Разработать проект строительства	D, E	15
G	Обеспечить финансирование проекта	E	12
H	Нанять подрядчика	F, G	8

Найти критический путь. Сколько работ находится на критическом пути? (Фиктивные работы не учитываются.) Через какое минимальное время после принятия решения о реализации проекта можно начать работу по строительству учебного корпуса? На сколько недель можно отложить выбор архитектурной фирмы? Чему равно наиболее позднее время завершения работы по обеспечению финансирования?

45.3. Проект внедрения компьютерной программы состоит из восьми работ (табл. 45.7).

Таблица 45.7

Исходные данные

Работа	Непосредственно предшествующие работы	Время выполнения, недели
A	–	3
B	–	6
C	A	2
D	B, C	5
E	D	4
F	E	3
G	B, C	9
H	F, G	3

Найти критический путь. Сколько времени потребуется для выполнения проекта? Сколько работ на критическом пути? Чему равно наиболее раннее время начала работы *C*?

На сколько дней можно отложить выполнение работы *C* без отсрочки завершения проекта в целом?

Чему равно наиболее позднее время окончания работы *F?* На сколько дней можно отложить выполнение работы *F* без отсрочки завершения проекта?

46. Задачи и модели замены

46.1. Сущность и классификация задач замены

Одной из важных экономических проблем является определение оптимальной стратегии в замене старых станков, агрегатов, машин на новые. Старение оборудования включает его физический и моральный износ, в результате чего растут производственные затраты по выпуску продукции на старом оборудовании, увеличиваются затраты на его ремонт и обслуживание, снижаются производительность и ликвидная стоимость.

Наступает время, когда старое оборудование выгоднее продать, заменить новым, чем эксплуатировать ценой больших затрат; причем его можно заменить новым оборудованием того же вида или новым, более совершенным.

Оптимальная стратегия замены оборудования состоит в определении оптимальных сроков замены. Критерием оптимальности при этом может служить прибыль от эксплуатации оборудования, которую следует оптимизировать, или суммарные затраты на эксплуатацию в течение рассматриваемого промежутка времени, подлежащие минимизации.

Процессы замены распадаются на два основных класса в зависимости от вида износа оборудования: либо оно устаревает и далее заменяется, либо оно полностью выходит из строя.

Классификация задач замены оборудования:

по характеру замены оборудования: задачи замены оборудования длительного пользования; задачи замены оборудования с целью предупреждения отказа; задачи замены оборудования выбора оптимального плана предварительного ремонта с целью уменьшения отказа;

по характеру затрат на оборудование: дискретные; непрерывные;

по выходу из строя оборудования: детерминированные; случайные;

по времени учета затрат на оборудование: без приведения затрат к текущему моменту; с приведением затрат к текущему моменту.

46.2. Постановка задачи замены оборудования длительного пользования

Производственное оборудование длительного пользования с течением времени изнашивается и подлежит предупредительно-восстановительному ремонту. Оборудование также устаревает (морально и физически) и подлежит полной замене. При этом *постановка задачи* такова: определить такие сроки ремонта и замены оборудования, при которых минимизируется сумма затрат на ремонт и замену оборудования при его старении. Иногда в оборудовании выходят из строя отдельные элементы (например, микросхемы) – в данном случае требуется определить такие сроки профилактического ремонта по замене вышедших из строя деталей, чтобы потери на данный элемент были минимальными. Здесь также имеет место профилактический контроль по обнаружению неисправностей. Требуется определить такие сроки контроля, при которых минимизируется сумма затрат на проведение контроля, а также минимизируется сумма потерь от простоя оборудования вследствие выхода из строя отдельных элементов.

46.3. Динамическая модель замены оборудования

Задачу замены оборудования можно представить в виде динамической модели.

Введем обозначения:

$r(t)$ – стоимость продукции, производимой за один год на единице оборудования возраста t лет;

$u(t)$ – ежегодные затраты на обслуживание оборудования возраста t лет;

$s(t)$ – остаточная стоимость оборудования возраста t лет;

P – покупная цена оборудования.

Рассмотрим период N лет, в пределах которого требуется определить оптимальный цикл замены оборудования.

Обозначим через $f_N(t)$ максимальный доход, получаемый от оборудования возраста t лет за оставшиеся N лет цикла его использования при условии оптимальной стратегии.

Возраст оборудования отсчитывается в направлении течения процесса. Так, $t=0$ соответствует случаю использования нового оборудования. Временные же шаги процесса нумеруются в обратном направлении по отношению к ходу процесса. Так, $N=1$ относится к одной временной стадии, остающейся до завершения процесса, а $N=N$ – к началу процесса (рис. 46.1).

Рис. 46.1. **Цикл замены оборудования**

На каждом этапе N-шагового процесса должно быть принято решение о сохранении или замене оборудования. Выбранный вариант должен обеспечивать получение максимальной прибыли.

Шаги процесса

Функциональные уравнения, основанные на принципе оптимальности Беллмана, имеют вид:

$$f_N(t) = \max\begin{cases} r(t) - u(t) + f_{N-1}(t+1) \to \textit{сохранение}, \\ s(t) - P + r(0) - f_{N-1}(1) \to \textit{замена}, \end{cases} \qquad (46.1)$$

$$f_1(t) = \max\begin{cases} r(t) - u(t) \to \textit{сохранение}, \\ s(t) - P + r(0) - u(0) \to \textit{замена}. \end{cases} \qquad (46.2)$$

Уравнение (46.1) описывает N-шаговый процесс, а (46.2) – одношаговый. Оба уравнения состоят из двух частей: верхняя строка определяет доход, получаемый при сохранении оборудования; нижняя – доход, получаемый при замене оборудования и продолжении процесса работы на новом оборудовании.

В уравнении (46.1) функция $r(t) - u(t)$ есть разность между стоимостью произведенной продукции и эксплуатационными издержками на N-м этапе процесса.

Функция $f_{N-1}(t+1)$ характеризует суммарную прибыль от $(N-1)$ оставшихся шагов для оборудования, возраст которого в начале осуществления этих этапов составляет $(t+1)$ лет.

Нижняя строка (46.1) характеризуется следующим образом: функция $s(t) - P$ представляет чистые издержки по замене оборудования, возраст которого t лет.

Функция $r(0)$ выражает доход, получаемый от нового оборудования возраста 0 лет. Предполагается, что переход от работы на оборудовании возраста t лет к работе на новом оборудовании совершается мгновенно, то есть период замены старого оборудования и переход на работу на новом оборудовании укладываются в один и тот же этап.

Последняя функция $f_{N-1}(1)$ в (46.1) представляет собой доход от оставшихся $N-1$ этапов, до начала осуществления которых возраст оборудования составляет один год.

Аналогичная интерпретация может быть дана уравнению для одношагового процесса. Здесь нет слагаемого вида $f_0(t+1)$, так как N принимает значение 1, 2,..., N.

Равенство $f_0(t)=0$ следует из определения функции $f_N(t)$.

Уравнения (46.1) и (46.2) являются рекуррентными соотношениями, которые позволяют определить величину $f_N(t)$ в зависимости от $f_{N-1}(t+1)$.

Структура этих уравнений показывает, что при переходе от одного шага процесса к следующему возраст оборудования увеличивается с t до $(t+1)$ лет, а число оставшихся шагов уменьшается с N до $(N-1)$.

Расчет начинают с использования уравнения (46.1). Уравнения (46.1) и (46.2) позволяют оценить варианты замены и сохранения оборудования, с тем чтобы принять тот из них, который предполагает больший доход. Эти соотношения дают возможность не только выбрать линию поведения при решении вопроса о сохранении или замене оборудования, но и определить прибыль, получаемую при принятии каждого из этих решений.

Пример 46.1. Определить оптимальный цикл замены оборудования при следующих исходных данных: P = 10, $S(t)$ = 0, $f(t) = r(t) - u(t)$, представленных в табл. 46.1.

Таблица 46.1

Исходные данные

N	0	1	2	3	4	5	6	7	8	9	10	11	12
$f_N(t)$	10	9	8	7	6	5	4	3	2	1	0	0	0

Решение.

Уравнения (46.1) и (46.2) запишем в следующем виде:

$$f_N(t) = \max\begin{cases} f(t) + f_{N-1}(t+1), \\ -P + f(0) + f_{N-1}(1), \end{cases}$$

$$f_1(t) = max\begin{cases} f(t), \\ -P + f(0). \end{cases}$$

Для N = 1:

$$f_1(0) = max\begin{cases} f(0), \\ -P + f(0) \end{cases} = max\begin{cases} 10 \\ -10+10 \end{cases} = 10,$$

$$f_1(1) = max\begin{cases} f(1), \\ -P + f(0) \end{cases} = max\begin{cases} 9 \\ -10+10 \end{cases} = 9,$$

……………………………………………………………

$$f_1(12) = max\begin{cases} f(12), \\ -P + f(0) \end{cases} = max\begin{cases} 0 \\ -10+10 \end{cases} = 0.$$

Для N = 2:

$$f_2(0) = \max\begin{cases} f(0) + f_1(1), \\ -P + f(0) + f_1(1) \end{cases} = \max\begin{cases} 10+9 \\ -10+10+9 \end{cases} = 19,$$

$$f_2(1) = \max\begin{cases} f(1) + f_1(2), \\ -P + f(0) + f_1(1) \end{cases} = \max\begin{cases} 9+8 \\ -10+10+9 \end{cases} = 17,$$

………………………………………………………………………………

Вычисления продолжаем до тех пор, пока не будет выполнено условие $f_1(t) > f_2(t)$, то есть в данный момент оборудование необходимо заменить, так как величина прибыли, получаемая в результате замены оборудования, больше, чем в случае использования старого.

Результаты расчетов помещаем в таблицу, момент замены отмечаем звездочкой, после чего дальнейшие вычисления по строчке прекращаем (табл. 46.2).

Можно не решать каждый раз уравнение (46.3), а вычисления проводить в таблице.

Например, вычислим $f_4(t)$:

$$f_4(0) = f_1(0) + f_3(1) = 10 + 24 = 34 > f_3(0) = 24,$$

$$f_4(1) = f_1(1) + f_3(2) = 9 + 21 = 30 > f_3(1),$$

$$f_4(2) = f_1(2) + f_3(3) = 8 + 18 = 26 > f_3(2),$$

$$f_4(3) = f_1(3) + f_3(4) = 7 + 17 = 24 = f_3(\ ),$$

$$f_4(4) = f_1(4) + f_3(5) = 6 + 17 = 23 < f_3(\ ).$$

Дальнейшие расчеты для $f_4(\ )$ прекращаем, так как

$$f_4(4) = 23 < f_3(\ ) = 24.$$

Таблица 46.2

Результаты расчетов

$F_N(t)$	0	1	2	3	4	5	6	7	8	9	10	11	12
	N	$N-1$	11	10	9	8	7	6	5	4	3	2	1
$f_1(t)$	10	9	8	7	6	5	4	3	2	1	0	0	0
$f_2(t)$	19	17	15	13	11	9	9*						
$f_3(t)$	27	24	21	18	17	17*							
$f_4(t)$	34	30	26	24	24*								
$f_5(t)$	40	35	32	31	30	30*							
$f_6(t)$	45	41	39	37	36	35	35*						
$f_7(t)$	51	48	45	43	41	41*							
$f_8(t)$	58	54	51	48	48*								
$f_9(t)$	64	60	56	55	54	54*							
$f_{10}(t)$	70	65	63	61	60	60*							
$f_{11}(t)$	75	72	69	67	66	65	65*						
$f_{12}(t)$	82	78	75	73	72	72*							

По результатам вычислений и по линии, разграничивающей области решений сохранения и замены оборудования, находим оптимальный цикл замены оборудования. Для данной задачи он составляет 4 года.

Вопросы для самодиагностики

1. В чем сущность задачи замены?
2. Привести классификацию задач замены.

3. Сформулировать постановку задачи замены оборудования.

4. Описать динамическое программирование как метод решения задачи замены оборудования.

Упражнения

46.1. К началу рассматриваемого периода на предприятии установлено новое оборудование. Зависимость производительности этого оборудования от времени его работы, а также затраты на содержание и ремонт при различном времени его использования приведены в табл. 46.3.

Таблица 46.3

Исходные данные

Наименование	Время, в течение которого используется оборудование, годы					
	0	1	2	3	4	5
Годовой выпуск продукции, млн. грн	80	75	65	60	60	55
Ежегодные затраты на содержание и ремонт оборудования, млн. грн	20	25	30	35	45	55

Известно, что затраты, связанные с приобретением и установкой нового оборудования, идентичного установленному, составляют 40 млн грн, а заменяемое оборудование списывается.

Составить такой план замены оборудования в течение пяти лет, при котором общий доход за данный период времени максимален.

46.2. К началу анализируемого периода на предприятии установлено новое оборудование.

Определить оптимальный цикл замены оборудования при следующих исходных данных: закупочная цена оборудования (P) составляет 12 ден.ед.; остаточная стоимость оборудования $s(t)=0$; $f_N(t)=r(t)-u(t)$ – максимальный доход, получаемый от оборудования возраста t лет за оставшиеся N лет цикла использования оборудования при условии оптимальной стратегии, где $r(t)$ – стоимость продукции, выпускаемой за год на единице оборудования возраста t лет, $u(t)$ –

ежегодные затраты на обслуживание оборудования возраста t лет;
N = 8 лет.

Зависимость $f_N()$ от N задана в табл. 46.4.

Таблица 46.4

Исходные данные

N	0	1	2	3	4	5	6	7	8
$f_N()$	12	11	10	8	6	4	2	0	0

47. Задачи с условиями неопределенности и конфликта

47.1. Характеристика задач стохастического программирования

Принятию управленческих решений, как правило, предшествует анализ возможных альтернатив по определенному критерию эффективности, что позволяет определить оптимальное решение, которое является основой разработки управленческого решения. Если информация относительно возможных альтернатив является полной и метод определения оптимального решения известен, то такие задачи являются *детерминированными*. То есть математическая модель задачи рассматривается в условиях определенности. К таким задачам, например, принадлежат задачи линейного программирования.

Большинство реальных экономических задач не содержат полной информации относительно возможных альтернатив или эти альтернативы могут реализовываться или не реализовываться, но вероятности этих событий неизвестны. Следовательно, возникает необходимость принимать решение в условиях *неопределенности*. Если перечень возможных вариантов мероприятий является полным и критерий, по которому определяется эффективность этих вариантов, является известным, однако реализация определенного варианта может происходить или не происходить с вероятностью, которая предварительно известна, то существует *стохастическая неопределенность*. Такие условия, когда известны все возможны альтернативы, а также известны вероятности их реализации (закон распределения вероятностей для каждой из альтернатив), определяются как *условия риска*.

Принятие решения в условиях риска приводит к тому, что реализация альтернативы, которая считается оптимальной, является случайным событием. Это круг задач *стохастического программирования*.

Таким образом, ***стохастическое программирование*** (stochastic programming) – раздел математического программирования, совокупность методов решения оптимизационных задач вероятностного характера. Это означает, что либо параметры ограничений (условий) задачи, либо параметры целевой функции, либо и те и другие являются случайными величинами (содержат случайные компоненты).

Основным принципом решения таких задач является переход от условий неопределенности к условиям риска или условиям определенности.

47.2. Характеристика задач теории игр

В экономике иногда приходится сталкиваться с ситуацией, когда при наличии многих участников эффективность решения одного из них зависит от того, какие решения приняли другие участники. Например, доход предприятия от продажи изделия зависит не только от установленной на него цены, но и от количества купленных покупателем изделий. Или при выборе ассортимента товаров, выпускаемых предприятием, нужно учитывать, какой ассортимент товаров выпускают другие предприятия. Такие задачи решаются методами ***теории игр***.

Теория игр, как самостоятельная наука, возникла в 1944 г. Она соответствует жизненным ситуациям, в которых сталкиваются интересы двух или нескольких участников. Интересы каждого участника различны и каждый участник достигает своей цели различным путем. Такие ситуации называются *конфликтными* и являются предметом изучения теории игр.

Формализация содержательного анализа конфликта представляет собой его математическую модель, которая называется ***игрой***. Участники игры называются *игроками*. Игроками могут быть как отдельные участники или предприятия, так и коллективы, имеющие общие интересы. Теория игр изучает оптимальное поведение игроков в играх.

Для конфликта характерно то, что ни один из его участников заранее не знает решений, принятых другими участниками, то есть вынуждены действовать в условиях риска и неопределенности.

Неопределенность игры может зависеть не только от сознательных действий участников, но и от действий стихийных сил (например природы). В зависимости от источника неопределенности различают игры: комбинаторные, азартные, стратегические. Примером комбинаторной игры является шахматная игра, когда особенности правил игры вызывают такое разнообразие в ее развитии, что результат заранее предсказать невозможно. Примером азартной игры является игра в монету, в кости, карты и др., в которых источником неопределенности выступает случайный фактор. Для стратегической игры источником неопределенности является отсутствие информации о действиях противника, то есть о его стратегии.

Игра бывает парной и множественной. В парной игре сталкиваются интересы двух противников, в множественной – трех и более. Наиболее распространенными являются парные игры.

Пусть A и B – два участника игры (два противника). Предположим, что интересы участников поддаются некоторому количественному описанию, то есть результат игры – выигрыш – определяется некоторым числом. План, по которому игрок совершает выбор ситуации, при любой фактической информации, называется ***стратегией игрока***. *Оптимальной* называется стратегия, которая при многократном повторении игры обеспечивает данному игроку максимально возможный выигрыш. Простейшей стратегической игрой является игра с нулевой суммой, то есть сумма выигрыша равна нулю. Это значит, что выигрыш первого игрока (v) равен проигрышу второго игрока (u), то есть $u+v=0$.

Для определенности, предположим, что игрок A выигрывает, игрок B проигрывает. Игрок A выбирает свои стратегии первым, его стратегии $A_1, A_2, ..., A_m$. Игрок B выбирает одну из своих стратегий $B_1, B_2, ..., B_n$. Тогда выигрыш первого игрока и проигрыш второго игрока есть функция стратегий A_i, B_j, то есть $v = f(A_i, B_j)$. Обозначим $f(A_i, B_j) = a_{ij}$. Тогда можно построить матрицу, которая называется ***платежной*** $\Pi_{(m\times n)} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix}_{(m\times n)}$, где a_{ij} – выигрыш первого игрока, если он выбрал i-ю стратегию и при этом второй игрок выбрал j-ю стратегию.

Для определения своей наилучшей стратегии игрок A стремится максимизировать свой минимальный выигрыш. Он просматривает строки и в каждой строке выбирает $\min$ элемент α_i и из этих элементов выбирает $\max$, то есть его позиция *максиминная*

$$\alpha = \max_j \min_i a_{ij},$$

где α – гарантированный выигрыш игрока A.

α называется ***нижней ценой*** (или границей) ***игры***.

Игрок B стремится минимизировать свой максимальный проигрыш, то есть он в каждом столбце выбирает $\max$ элемент β_j и среди них выбирает $\min$, то есть его позиция *минимаксная*

$$\beta = \min_i \max_j a_{ij}.$$

β называется ***верхней ценой игры***.

Если второй игрок будет придерживаться своей минимаксной стратегии, то гарантировано, что он во всяком случае проиграет не больше β.

Для матричной игры выполняется условие $\alpha \le \beta$. Если $\alpha = \beta$, то такая игра называется *игрой с седловой точкой*, а пара оптимальных стратегий ($A_{iопт}, B_{jопт}$) – ***седловой точкой*** матрицы.

В этом случае элемент $a_{ij} = v$ называется *ценой игры*. Игра называется определенной и любые отклонения игроков от своих стратегий невозможны.

Если игра имеет седловую точку, то говорят, что она решается в *«чистых» стратегиях*. *Чистой стратегией* называется выбор игроком конкретной стратегии.

Если платежная матрица не имеет седловой точки, то есть $\alpha < \beta$, то поиск решения игры приводит к применению сложной стратегии, которая составляется из случайного применения двух и больше стратегий с определенными вероятностями. Такая сложная стратегия называется *смешанной*.

В игре, матрица которой имеет размерность $m \times n$, стратегии первого игрока задаются наборами вероятностей $X = (x_1, x_2, ..., x_m)$, с которыми игрок применяет свои чистые стратегии.

Эти наборы можно рассматривать как m-мерные векторы, для координат которых

$$\sum_{i=1}^{m} x_i = 1, \quad x_i \geq 0, \quad i = \overline{1,m}.$$

Аналогично для второго игрока наборы вероятностей определяют n-мерные векторы $Y = (y_1, y_2, ..., y_n)$, для координат которых

$$\sum_{j=1}^{n} y_j = 1, \; y_j \geq 0, \; j = \overline{1,n}.$$

Выигрыш первого игрока при использовании смешанных стратегий определяют как математическое ожидание выигрыша, он равен:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j.$$

В основной теореме теории игр утверждается, что каждая конечная игра имеет, по крайней мере, одно решение, возможно, в области смешанных стратегий.

Применение оптимальной стратегии позволяет получить выигрыш, который равен цене игры: $\alpha \leq \nu \leq \beta.$

Применение первым игроком оптимальной стратегии должно обеспечить ему при любых действиях второго игрока выигрыш не меньше цены игры. Поэтому выполняются соотношения: $\sum_{i=1}^{m} a_{ij} x_{ionm} \geq \nu.$

Аналогично для второго игрока оптимальная стратегия должна обеспечить при любых стратегиях первого игрока проигрыш, который не превышает цену игры, то есть справедливы соотношения: $\sum_{j=1}^{n} a_{ij} y_{jonm} \leq \nu.$

Если платежная матрица не содержит седловой точки, то задача определения смешанной стратегии тем сложнее, чем больше размерность матрицы. Поэтому матрицы большой размерности целесообразно упростить, уменьшив их размерность путем вычеркивания дублирующих (одинаковых) и сознательно невыгодных стратегий, а именно *доминантных* строк (элементы которых меньше элементов любой параллельной строки) и *доминирующих* столбцов (элементы которых больше элементов любого параллельного столбца).

Теория игр находится в тесной связи с линейным программированием, потому что каждая игра двух участников может быть представлена как задача линейного программирования и решена симплексным методом, и наоборот, каждая задача линейного программирования может быть представлена как игра двух участников.

Графический метод решения матричной игры

Графический метод применим к играм, в которых хотя бы один игрок имеет только две стратегии.

Рассмотрим игру $(2 \times n)$ (табл. 47.1).

Таблица 47.1

Матричная игра $(2 \times n)$

Вероятности		Второй игрок				Стратегии первого игрока
		y_1	y_2	…	y_n	
Первый игрок	x_1	a_{11}	a_{12}	…	a_{1n}	A_1
	x_2	a_{21}	a_{22}	…	a_{2n}	A_2
Стратегии второго игрока		B_1	B_2	…	B_n	

Предполагаем, что игра не имеет седловой точки.

Обозначим: x_1 – вероятность применения первым игроком 1-й стратегии, x_2 – вероятность применения первым игроком 2-й стратегии ($x_2 = 1 - x_1$), y_1 – вероятность применения вторым игроком 1-й стратегии, y_2 – вероятность применения вторым игроком 2-й стратегии и т. д., y_n – вероятность применения вторым игроком n-й стратегии.

Ожидаемый выигрыш первого игрока при применении вторым игроком 1-й стратегии равен математическому ожиданию выигрыша:

$$M(y) = a_{11}x_1 + a_{21}x_2.$$

Аналогично найдем ожидаемые выигрыши первого игрока при использовании вторым игроком $2,3,...,n$-й стратегий. Полученные результаты запишем в табл. 47.2.

Таблица 47.2

Результаты решения матричной игры $[2\times n]$

Чистые стратегии второго игрока	Ожидаемые выигрыши первого игрока
1	$a_{11}x_1+a_{21}x_2$
2	$a_{12}x_1+a_{22}x_2$
…	…
n	$a_{1n}x_1+a_{2n}x_2$

Для решения задачи используем номограммы на параллельных шкалах. Проведем две оси A_1 и A_2 на расстоянии 1 ед. перпендикулярно оси l и нанесем на них выигрыши игрока A в зависимости от стратегий игрока B (рис. 47.1).

На шкалах A_1 и A_2 отмечены значения проигрыша игрока B – результаты реализации стратегий $B_1,B_2,...,B_n$.

Так как позиция игрока A максиминная, то находим нижнюю границу игры и на ней $\max$ точку.

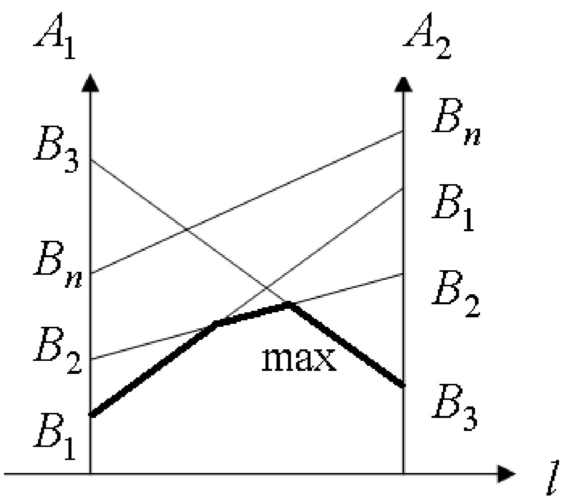


Рис. 47.1. **Графическое решение матричной игры** $[2\times n]$

На графике рис. 47.1 это пересечение стратегий B_2,B_3, то есть $y_1=y_4=...=y_n=0$ и задача сводится к простейшей парной игре с платежной матрицей $\Pi=\begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}$.

Решим матричную игру для игрока A:

$$\begin{cases} a_{12}x_1 + a_{22}x_2 = \nu, \\ a_{13}x_1 + a_{23}x_2 = \nu, \\ x_1 + x_2 = 1. \end{cases}$$

Так как события – реализация стратегий A_1 и A_2 – образуют полную группу, то $x_1 + x_2 = 1$.

Решим матричную игру для игрока B:

$$\begin{cases} a_{12}y_2 + a_{13}y_3 = \nu, \\ a_{22}y_2 + a_{23}y_3 = \nu, \\ y_2 + y_3 = 1. \end{cases}$$

Решив системы трех линейных уравнений с тремя неизвестными, получим оптимальные планы для игроков A и B.

Рассмотрим игру $(m \times 2)$ (табл. 47.3).

Таблица 47.3

Матричная игра $(m \times 2)$

Вероятности		Второй игрок		Стратегии первого игрока
		y_1	y_2	
Первый игрок	x_1	a_{11}	a_{12}	A_1
	x_2	a_{21}	a_{22}	A_2
	…	…	…	…
	x_m	a_{m1}	a_{m2}	A_m
Стратегии второго игрока		B_1	B_2	

Если нет седловой точки, решаем ее графическим методом.

Построим номограмму (рис. 47.2). Поскольку позиция игрока B минимаксная, то находим верхнюю границу игры и на ней $\min$ точку. На графике рис. 47.2 $\min$ точка на пересечении стратегий A_1, A_2, то есть $x_3 = x_4 = \ldots = x_m = 0$ и задача сводится к простейшей парной игре с платежной матрицей $\varPi = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

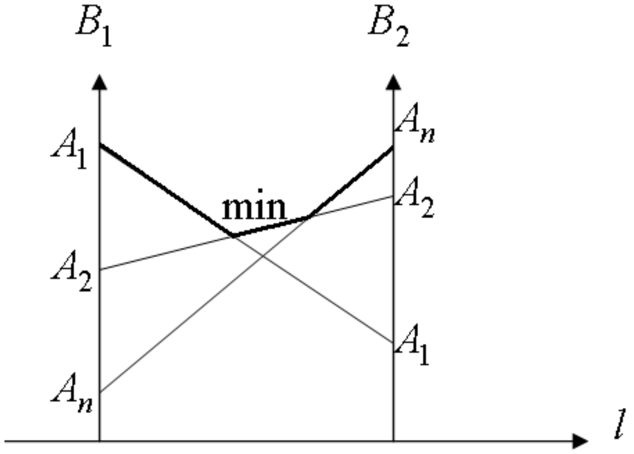


Рис. 47.2. **Графическое решение матричной игры** $(n \times 2)$

Тогда для игрока A:

$$\begin{cases} a_{11}x_1 + a_{21}x_2 = \nu, \\ a_{12}x_1 + a_{22}x_2 = \nu, \\ x_1 + x_2 = 1. \end{cases}$$

Для игрока B:

$$\begin{cases} a_{11}y_1 + a_{12}y_2 = \nu, \\ a_{21}y_1 + a_{22}y_2 = \nu, \\ y_1 + y_2 = 1. \end{cases}$$

Решив полученные системы трех линейных уравнений с тремя неизвестными, получим оптимальные планы для игроков A и B.

Пример 47.1. Решить матричную игру, заданную платежной матрицей:

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

Решение.

Определяем нижнюю и верхнюю цену игры:

$$\alpha = \max(3;2;1) = 3,$$
$$\beta = \min(7;3;6;9) = 3.$$

$\alpha = \beta = 3$, то есть игра с седловой точкой, цена игры равна 3.

$X_{опт} = (1;0;0)$, $Y_{опт} = (0;1;0;0)$.

Пример 47.2. Решить матричную игру, заданную платежной матрицей:

$$П=\begin{pmatrix} 7 & 6 & 5 & 4 & 2 \\ 5 & 4 & 3 & 2 & 3 \\ 5 & 6 & 6 & 3 & 5 \\ 2 & 3 & 3 & 4 & 4 \end{pmatrix}_{4\times 5}.$$

Решение.

Пусть A_1, A_2, A_3, A_4 – стратегии игрока A; B_1, B_2, B_3, B_4, B_5 – стратегии игрока B.

Обозначим: x_1 – вероятность применения первым игроком 1-й стратегии; x_2, x_3, x_4 – вероятности применения первым игроком 2, 3, 4-й стратегий соответственно (так как события A_1, A_2, A_3, A_4 образуют полную группу, то $x_1 + x_2 + x_3 + x_4 = 1$); y_1 – вероятность применения вторым игроком 1-й стратегии; y_2, y_3, y_4, y_5 – вероятности применения вторым игроком 2, 3, 4, 5-й стратегий соответственно ($y_1 + y_2 + y_3 + y_4 + y_5 = 1$).

Размерность платежной матрицы сокращается путем вычеркивания доминантных строк и доминирующих столбцов, то есть A_2, A_4 и B_1, B_2, B_3. Поэтому $x_2 = x_4 = y_1 = y_2 = y_3 = 0$ и матрица принимает вид: $\begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix}_{2\times 2}$.

Тогда для игрока A:

$$\begin{cases} 4x_1 + 3x_4 = \nu, \\ 2x_1 + 5x_4 = \nu, \\ x_1 + x_4 = 1. \end{cases}$$

Для игрока B:

$$\begin{cases} 4y_4 + 2y_5 = \nu, \\ 3y_4 + 5y_5 = \nu, \\ y_4 + y_5 = 1. \end{cases}$$

Решив эти системы, получим:

$$x_1 = x_4 = 1/2,\ \nu = 7/2; \quad y_4 = 3/4,\ y_5 = 1/4,\ \nu = 7/2.$$

Оптимальная стратегия первого игрока: $X_{опт} = (1/2, 0, 1/2, 0),$ при этом цена игры $v = 7/2.$

Оптимальная стратегия второго игрока: $Y_{опт} = (0, 0, 0, \frac{3}{4}, \frac{1}{4})$, при этом цена игры $v = 7/2.$

47.3. Сведение матричной игры к задаче линейного программирования

Пусть задана платежная матрица

$$\underset{(m\times n)}{\Pi} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}_{(m\times n)}.$$

Игрок A выбирает свои стратегии первым, его стратегии $A_1, A_2, ..., A_m$. Игрок B выбирает одну из своих стратегий $B_1, B_2, ..., B_n$.

Стратегии игрока A задаются набором вероятностей $X = (x_1, x_2, ..., x_m)$, который можно рассматривать как m-мерный вектор с координатами $\sum_{i=1}^{m} x_i = 1, \quad x_i \ge 0, \quad i = \overline{1, m}.$ Аналогично для второго игрока набор вероятностей определяет n-мерный вектор $Y = (y_1, y_2, ..., y_n)$, для координат которого $\sum_{j=1}^{n} y_j = 1,\ y_j \ge 0,\ j = \overline{1, n}.$

Матрица размерности $(m\times n)$, седловой точки нет и преобразования невозможны.

Пусть $a_{ij} \ge 0$ (если есть $a_{ij} < 0$, то можно ко всем элементам матрицы добавить одно и то же число).

Составим математическую модель для игроков A и B.

Поскольку игрок A стремится максимизировать свой выигрыш, то система ограничений будет иметь неравенства типа $\geq$:

$$\begin{cases} \sum_{i=1}^{m} a_{ij} x_i \geq \nu, \\ x_i \geq 0,\ \sum_{i=1}^{m} x_i = 1,\ j = \overline{1,n}. \end{cases}$$

Так как игрок B стремится минимизировать свой проигрыш, то в системе ограничений неравенства будут типа $\leq$:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} y_j \leq \nu, \\ y_j \geq 0,\ \sum_{j=1}^{n} y_j = 1,\ i = \overline{1,m}. \end{cases}$$

Преобразуем системы, разделим все неравенства и уравнения на ν и сделаем замену: $\dfrac{x_i}{\nu} = t_i,\ \dfrac{y_j}{\nu} = u_j$.

Тогда получим:

$$Z = \sum_{i=1}^{m} t_i = \frac{1}{\nu} \ (\min),$$

$$\begin{cases} \sum_{i=1}^{m} a_{ij} t_i \geq 1,\ j = \overline{1,n}, \\ t_i \geq 0,\ i = \overline{1,m}. \end{cases}$$

$$f = \sum_{j=1}^{n} u_j = \frac{1}{\nu} \ (\max),$$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} u_j \leq 1,\ i = \overline{1,m}, \\ u_j \geq 0,\ j = \overline{1,n}. \end{cases}$$

Получили две взаимнодвойственные задачи. Достаточно одну из них (лучше вторую) решить симплекс-методом, решение другой получим или в симплекс-таблице, или по теоремам двойственности.

Пример 47.3. Решить матричную игру, заданную платежной матрицей:

$$\Pi = \begin{pmatrix} 4 & 3 & 4 & 2 \\ 3 & 4 & 6 & 5 \\ 2 & 5 & 1 & 3 \end{pmatrix}.$$

Решение.

Определяем нижнюю и верхнюю цены игры: $\alpha = \max\{2;3;1\} = 3$, $\beta = \min\{4;5;6;5\} = 4$. $\alpha \neq \beta$, то есть игра без седловой точки, в смешанных стратегиях.

Игрок A выбирает свои стратегии A_1, A_2, A_3, вероятности которых x_1, x_2, x_3. Игрок B выбирает одну из своих стратегий B_1, B_2, B_3, B_4 с вероятностями y_1, y_2, y_3, y_4.

Составим математическую модель для игроков A и B.

A:

$$x_1 + x_2 + x_3 = 1,$$
$$\begin{cases} 4x_1 + 3x_2 + 2x_3 \geq \nu, \\ 3x_1 + 4x_2 + 5x_3 \geq \nu, \\ 4x_1 + 6x_2 + x_3 \geq \nu, \\ 2x_1 + 5x_2 + 3x_3 \geq \nu, \end{cases}$$
$$x_i \geq 0,\ i = \overline{1,3}.$$

B:

$$y_1 + y_2 + y_3 + y_4 = 1,$$
$$\begin{cases} 4y_1 + 3y_2 + 4y_3 + 2y_4 \leq \nu, \\ 3y_1 + 4y_2 + 6y_3 + 5y_4 \leq \nu, \\ 2y_1 + 5y_2 + y_3 + 3y_4 \leq \nu, \end{cases}$$
$$y_j \geq 0,\ j = \overline{1,4}.$$

Разделим все неравенства и уравнения на ν и сделаем замену: $\dfrac{x_i}{\nu} = t_i,\ \dfrac{y_j}{\nu} = u_j$.

Тогда

A:

$$Z = t_1 + t_2 + t_3 \ (\to \min)$$
$$\begin{cases} 4t_1 + 3t_2 + 2t_3 \geq 1, \\ 3t_1 + 4t_2 + 5t_3 \geq 1, \\ 4t_1 + 6t_2 + t_3 \geq 1, \\ 2t_1 + 5t_2 + 3t_3 \geq 1, \end{cases}$$
$$t_i \geq 0,\ i = \overline{1,3}.$$

B:

$$f = u_1 + u_2 + u_3 + u_4 \ (\to \max)$$
$$\begin{cases} 4u_1 + 3u_2 + 4u_3 + 2u_4 \leq 1, \\ 3u_1 + 4u_2 + 6u_3 + 5u_4 \leq 1, \\ 2u_1 + 5u_2 + u_3 + 3u_4 \leq 1, \end{cases}$$
$$u_j \geq 0,\ j = \overline{1,4}.$$

Решим вторую задачу симплекс-методом, получим

$$U_{опт} = \left(\frac{3}{14};0;0;\frac{1}{14}\right),\ f_{\max} = \frac{2}{7} \Rightarrow \nu = \frac{7}{2},\ Y_{опт} = \left(\frac{3}{4};0;0;\frac{1}{4}\right)\ (y_j = \nu \cdot u_j).$$

В строке z_j напротив исходного единичного базиса получим решение двойственной задачи: $T_{опт} = \left(\frac{1}{7}; \frac{1}{7}; 0\right)$, $X_{опт} = \left(\frac{1}{2}; \frac{1}{2}; 0\right)$ $\left(x_i = \nu \cdot t_i\right)$.

Вопросы для самодиагностики

1. Как моделируются рискованные ситуации в экономике на базе теории игр? Что такое матричная игра и как она интерпретируется в экономике?

2. Сформулировать математическую модель парной матричной игры с нулевой суммой. Что такое платежная матрица?

3. Как определяется матричная игра в чистых и смешанных стратегиях? Что называется доминантной и доминирующей стратегиями?

4. Как геометрически интерпретируется матричная игра?

5. Какое условие решения матричной игры $2 \times n$ в смешанных стратегиях? Какое условие решения матричной игры $m \times 2$ в смешанных стратегиях?

6. Как решить парную матричную игру симплексным методом?

Упражнения

47.1. Упростить платежную матрицу и найти решение матричной игры:

а) $A = \begin{pmatrix} 3 & 1 & 4 & 5 \\ 4 & 3 & 5 & 7 \\ 2 & 1 & 3 & 3 \\ 5 & 5 & 4 & 3 \end{pmatrix}$; б) $A = \begin{pmatrix} 2 & 3 & 1 & 4 \\ 4 & 3 & 4 & 5 \\ 3 & 2 & 1 & 4 \\ 5 & 4 & 1 & 6 \end{pmatrix}$; в) $A = \begin{pmatrix} 4 & 1 & 3 & 2 \\ 4 & 2 & 4 & 3 \\ 5 & 2 & 4 & 4 \\ 1 & 4 & 7 & 5 \end{pmatrix}$;

г) $A = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 2 & 3 & 3 \\ 0 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 \end{pmatrix}$; д) $A = \begin{pmatrix} 1 & 4 & 2 & 0 \\ 0 & 3 & 1 & 1 \\ 2 & 5 & 2 & 1 \\ 3 & 4 & 1 & 2 \end{pmatrix}$; е) $A = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 2 & 3 & 3 & 4 \\ 0 & 2 & 1 & 3 \\ 5 & 2 & 3 & 5 \end{pmatrix}$.

47.2. Лицо, принимающее решение, проводит анализ внедрения двух бизнес-проектов компании. Прибыль компании в зависимости от принятия соответствующей стратегии конкурента компании задана матрицей

$$P=\begin{pmatrix} 8 & 2 & 8 & 9 & 1 \\ 10 & 7 & 3 & 1 & 5 \end{pmatrix}.$$

Решить заданную матричную игру графически и аналитически.

47.3. Для отопления помещения необходимо закупить топливо. Затраты топлива и цена на него зависят от состояния погоды в зимний период, который представлен матрицей

$$P=\begin{pmatrix} 2 & 4 & 6 & 8 & 3 \\ 10 & 0 & 5 & 7 & 9 \end{pmatrix}.$$

Решить заданную матричную игру графически и аналитически.

47.4. Предприятие изготавливает три вида продукции и получает прибыль, которая зависит от спроса на продукцию (повышенный спрос, нормальный, умеренный, низкий). Элементы приведенной матрицы характеризуют прибыль отдельного вида продукции в зависимости от состояния спроса.

$$A=\begin{pmatrix} 7 & 6 & 7 & 5 \\ 6 & 7 & 9 & 8 \\ 5 & 8 & 4 & 6 \end{pmatrix}.$$

Решить заданную матричную игру.

48. Многокритериальные задачи в менеджменте

48.1. Характеристика многокритериальных оптимизационных задач

На практике часто требуется найти экстремальные значения нескольких экономических показателей. В этом случае математическая модель имеет несколько целевых функций, причем некоторые из них исследуются на максимум, а другие – на минимум. Поэтому ставится задача отыскания такого компромиссного (субоптимального) решения модели, в котором значения всех рассматриваемых экономических показателей были бы приближены к экстремальным.

Поиск компромиссного решения относится к *многокритериальным задачам оценки оптимальности*.

Концепции, на которых строятся многокритериальные модели, являются важными в теории управления. Во многих примерах экономических задач осуществляется поиск нескольких целей (критериев), среди которых достаточно сложно выделить одну. Например, в задаче планирования с долгосрочными целями есть: максимизация прибыли, максимизация доли рынка на конец планового периода, максимизация основного капитала на конец планового периода. Такие цели не поддаются непосредственному сравнению или комбинированию. Безусловно, это конфликтующие цели, послабление требований к одной цели ведет к возможности получить лучшие результаты по другой.

Обозначим i-й частный критерий через $F_i(\overline{X})$, где $\overline{X}$ – допустимое решение, а область допустимых решений – через Q.

Если учесть, что изменением знака функции всегда можно свести задачу минимизации к задаче максимизации, то задачу многокритериальной оптимизации можно сформулировать следующим образом:

$$F(\overline{X}) = \{F_1(\overline{X}), F_2(\overline{X}), ..., F_m(\overline{X})\} \to \max, \qquad (48.1)$$

$$\overline{X} \in Q. \qquad (48.2)$$

Вектор $\overline{X}^* \in Q$ называется ***эффективным*** (*оптимальным по Парето*) ***решением*** многокритериальной задачи, если не существует такого вектора $\overline{X} \in Q$, что

$$F_i(\overline{X}) \geq F_i(\overline{X}^*), i = \overline{1,m}, \qquad (48.3)$$

причем хотя бы для одного значения i имеет место строгое неравенство.

Множество допустимых решений, для которых невозможно одновременно улучшить все частные показатели эффективности (то есть улучшить хотя бы один из них, не ухудшая остальных), принято называть ***областью Парето***, или *областью компромиссов*, а принадлежащие ей решения – ***эффективными***, или *оптимальными по Парето*.

48.2. Методы многокритериальной оптимизации управленческих решений

В настоящее время разработано несколько подходов к многоцелевому моделированию (многокритериального принятия решения). Среди них использование *теории полезности*, поиск *Парето-оптимума* с помощью многокритериального линейного программирования, *аналитический иерархический процесс* (разработан Т. Саати) и *целевое программирование* (предложенное А. Чансом и В. Купером).

Целевое программирование можно рассматривать как *эвристический подход* для многоцелевых моделей на основе концепции линейного программирования. В общем случае целевое программирование используется для линейных моделей. Иногда его рассматривают как попытку математической интерпретации понятия удовлетворения. Этот термин был введен Х. Саймоном, чтобы охарактеризовать ситуацию поиска не оптимального, а «достаточно хорошего» решения. То есть, когда необходимо привести несколько целей одновременно хотя бы к минимально удовлетворительному уровню.

Многокритериальные задачи математически недостаточно разработаны и для практической деятельности решаются следующими способами:

1. Производится ранжирование показателей, т. е. расположение их в порядке значимости, важности. Затем приступают к поиску решения, оптимального по наиболее важному из них. Задавшись допустимой величиной изменения первого критерия, ищут решение по второму критерию, наилучшему в полученной области, и т. д. Порядок значимости и допустимые диапазоны выбирают, исходя из сущности проблемы.
2. Построение единого (интегрального) показателя эффективности посредством суммирования произведений имеющихся показателей на «весовые» коэффициенты (коэффициенты важности показателей).
3. Превращение всех целевых функций, кроме одной, в ограничения.

Рассмотрим математическую модель экономической задачи, в которой две целевые функции и система линейных ограничений.

Найдем компромиссное решение по двум показателям, один из которых требует отыскания максимума, а другой – минимума:

$$F_1 = \sum_{j=1}^{n} c_j x_j \to \max, \quad F_2 = \sum_{j=1}^{n} d_j x_j \to \min,$$

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} x_j \le b_i, \; x_j \ge 0, \; j = \overline{1,n}, \; i = \overline{1,m},$$

где F_1, F_2 – значения целевых функций (экономические показатели), для упрощения записи опущены обозначения аргумента;

a_{ij}, c_j, d_j, b_i – коэффициенты;

x_j – переменные.

Решим задачу по каждому показателю в отдельности и найдем оптимальные значения $F_{1\max}, F_{2\min}$.

Проделав преобразования над целевыми функциями, получим математическую модель для отыскания компромиссного решения задачи с двумя целевыми функциями:

$$W = x_{n+1} \to min,$$

$$\begin{cases} \sum_{j=1}^{n} c_j x_j + F_{1\max} x_{n+1} \ge F_{1\max}, \\ \sum_{j=1}^{n} d_j x_j - F_{2\min} x_{n+1} \le F_{2\min}, \\ \sum_{j=1}^{n} a_{ij} x_j \le b_i, \end{cases}$$

$$x_j \ge 0, \quad j = \overline{1,n}, \quad i = \overline{1,m},$$

где W – целевая функция;

x_{n+1} – наибольшее относительное значение экономических показателей.

Математическая модель будет аналогичной в случае получения компромиссных решений задач, имеющих три целевые функции и более.

Пример 48.1. Фирма выпускает два вида изделий по цене 2 ден. ед. и 3 ден. ед. соответственно. По результатам маркетинговых исследований спрос на изделия второго вида не менее 1 тыс. ед. в год. Для производства изделий используются материалы *A* и *B,* запасы которых на фирме составляют 18 и 15 т соответственно. Для изготовления 1 тыс. изделий норма расхода материала *A* для изделий 1-го вида составляет 3 т, а для изделий 2-го вида – 5 т.

Для изготовления 1 тыс. изделий материала *B* расходуется: для изделий 1-го вида – 5 т, для изделий 2-го вида – 3 т. Себестоимость изделий 1-го вида – 1 ден. ед., а 2-го вида – 2 ден. ед.

Найти оптимальное решение по производству изделий 1-го и 2-го видов, чтобы прибыль и количество выпускаемых изделий были максимальными, себестоимость минимальной.

Решение.

Обозначим: x_1 – количество изделий 1-го вида, тыс. ед.; x_2 – количество изделий 2-го вида, тыс. ед.

Математическая модель задачи будет иметь вид:

$$F_1 = x_1 + x_2 \to \max, F_2 = 2x_1 + 3x_2 \to \max, F_3 = x_1 + 2x_2 \to \min,$$

$$\begin{cases} 3x_1 + 5x_2 \le 18, \\ 5x_1 + 3x_2 \le 15, \\ x_2 \ge 1, \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0.$$

Решим задачу по каждой целевой функции в отдельности.

Получим: $X_{1опт} = (1{,}31; 2{,}81),\ X_{2опт} = (1{,}31; 2{,}81),\ X_{3опт} = (0; 1),$

$$F_{1\max} = 4{,}125,\ F_{2\max} = 11{,}063,\ F_{3\min} = 2{,}0.$$

Математическая модель задачи отыскания компромиссного решения:

$$W = x_3 \to \min,$$

$$\begin{cases} x_1 + x_2 + 4{,}125x_3 \ge 4{,}125, \\ 2x_1 + 3x_2 + 11{,}063x_3 \ge 11{,}063, \\ x_1 + 2x_2 - 2x_3 \le 2, \\ 3x_1 + 5x_2 \le 18, \\ 5x_1 + 3x_2 \le 15, \\ x_2 \ge 1,\ x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0. \end{cases}$$

Решая задачу на компьютере (например, в среде MS Excel с помощью опции «Поиск решения»), получим $X_{компр} = (1{,}07;1)$.

Таким образом, фирме целесообразно выпускать 1,07 тыс. изделий 1-го вида и 1 тыс. изделий 2-го вида.

Вопросы для самодиагностики

1. Что называется многокритериальной задачей?
2. Привести примеры многокритериальных задач в экономике.
3. Какие способы решения многокритериальных задач?

Упражнения

Составить математическую модель отыскания компромиссного решения и найти его:

48.1. $F_1 = x_1 + 2x_2 \to \max,\ F_2 = 4x_1 + x_2 \to \min$

$$\begin{cases} x_1 + 2x_2 \le 6, \\ x_1 \ge 1, \\ x_1 - x_2 \le 4, \\ x_1 + 3x_2 \ge -3,\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

48.2. $F_1 = 2x_1 + x_2 \to \max,\ F_2 = 2x_1 + x_2 \to \min$

$$\begin{cases} 3x_1 + 4x_2 \ge 12, \\ -x_2 \ge -3, \\ -2x_1 + 4x_2 \ge -8, \\ -x_1 \ge -7,\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

48.3. $F_1 = x_1 + 3x_2 \to \max,\ F_2 = 2x_1 + x_2 \to \min$

$$\begin{cases} 4x_1 + 6x_2 \le 24, \\ x_2 \ge 1, \\ 6x_1 + 3x_2 \le 18, \\ x_1 \le 4,\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Предметный указатель

Л

М

Н

Т

У

Предметный указатель

О

П

Т

У

Предметный указатель

Предметный указатель

Т

У