

УДК 51(075.8)  
ББК 22.1я73  
М997

## ***Высшая математика и ее приложения к биологии***

Рецензенты:

доктор физико-математических наук, профессор *Ю. С. Хохлов* (зав. кафедрой математической статистики и эконометрики Тверского государственного университета);

кандидат физико-математических наук, доцент *А. Б. Плачев* (Московский государственный институт радиотехники, электроники и автоматики (технический университет))

**Мятлев В. Д.**

М997 Теория вероятностей и математическая статистика. Математические модели : учебное пособие для студ. высш. учеб. заведений / В. Д. Мятлев, Л. А. Панченко, Г. Ю. Ризниченко, А. Т. Терехин. — М. : Издательский центр «Академия», 2009. — 320 с. — (Университетский учебник. Высшая математика и ее приложения к биологии).

ISBN 978-5-7695-4704-1

В учебном пособии изложены преимущественно вопросы анализа биологических данных и в качестве основы использованы методы теории вероятностей и математической статистики; представлено моделирование динамики биологических процессов, основанное на применении математического аппарата дифференциальных уравнений.

Для студентов биологических специальностей высших учебных заведений.

УДК 51(075.8)  
ББК 22.1я73

*Оригинал-макет данного издания является собственностью Издательского центра «Академия», и его воспроизведение любым способом без согласия правообладателя запрещается*

ISBN 978-5-7695-4704-1

© Коллектив авторов, 2009  
© Образовательно-издательский центр «Академия», 2009  
© Оформление. Издательский центр «Академия», 2009

# ПРЕДИСЛОВИЕ

Учебное пособие по математике для биологических специальностей вузов основано на материале курса «Математические методы в биологии», преподаваемого с середины 80-х годов XX в. на биологическом факультете МГУ им. М. В. Ломоносова. В книге подробно рассматриваются разделы математики, находящие наибольшее применение в биологии (теория вероятностей и математическая статистика, теория дифференциальных уравнений), с акцентом на их практическое использование в биологических исследованиях.

Книга состоит из двух частей (в учебной программе биологического факультета МГУ эти части соответствуют третьему и четвертому семестрам общего математического обучения). Задача первой части — обучение методам статистического анализа полевых и экспериментальных биологических данных на основе методов теории вероятностей и математической статистики. Необходимо отметить, что теория вероятностей, имеющая самостоятельную прикладную ценность, в данном контексте рассматривается главным образом в качестве методологической и аппаратной основы математической статистики. Задача второй части — обучение построению и анализу математических моделей, описывающих динамику биологических процессов, с использованием дифференциальных и дискретных рекуррентных уравнений.

Первая часть книги написана В. Д. Мятлевым, Л. А. Панченко и А. Т. Терехиным, вторая — Г. Ю. Ризниченко.

# ВВЕДЕНИЕ

Цель учебного пособия — обучить студентов применению математических методов в биологических исследованиях, или, иначе, построению и анализу математических моделей биологических объектов, которые часто называют *биологическими системами*, чтобы подчеркнуть их сложность.

*Модель* — это аналог *моделируемого объекта*, более доступный для исследования. Модели могут быть как *натурными* (например, уменьшенная модель самолета или автомобиля, испытываемая в аэродинамической трубе, или животное, используемое в качестве модели человека при определении токсичности нового лекарственного средства), так и *абстрактными*. Не замечая этого, мы постоянно пользуемся абстрактными *вербальными* (словесными) моделями и, следует признать, что они удовлетворяют подавляющую часть наших потребностей не только в повседневной, но также в научной и производственной деятельности. *Математические модели*, основанные на использовании чисел, символьных обозначений, графиков, строгих определений и правил вывода, также относят к абстрактному моделированию. По сравнению с вербальными моделями они более удобны для описания количественных, геометрических и структурных аспектов моделируемого объекта (системы) и менее уязвимы в отношении скрытых логических пропусков и противоречий.

Математические модели можно классифицировать по разным основаниям. В частности, их можно разделить на *вероятностные* (*вероятностно-статистические, статистические, стохастические*) модели — в таких моделях принципиально учитывается неполная предсказуемость результата при заданных известных условиях, и *детерминистические модели*, в которых исходные условия предопределяют получение однозначного конечного результата. В первой части книги рассматриваются вероятностные, а во второй — преимущественно детерминистические модели.

Можно также разделить математические модели на **функциональные** и **дифференциальные**. Функциональные модели описывают связь одних характеристик моделируемого объекта с другими в виде функции или алгебраического уравнения (неявной функции). В частности, эти модели могут описывать динамику поведения объекта, если аргументом функции служит время и/или возраст. Дифференциальные модели строятся на основе описания локальных во времени и/или в пространстве связей компонентов моделируемой системы. Они специально предназначены для моделирования динамики, которая, однако, в отличие от функциональных моделей, в дифференциальных моделях задается неявно (локально), а ее явная функциональная зависимость от времени находится путем аналитического или численного (компьютерного) интегрирования модели. Дифференциальные модели рассматриваются во второй части данной книги, функциональные — в первой.

Функциональный характер вероятностно-статистических моделей наиболее явно проявляется в **регрессионной модели**, выражающей так называемую зависимую переменную  $y$  в виде функции одной или нескольких независимых переменных  $x$  с добавлением случайной ошибки  $\varepsilon$ :

$$y = f(x) + \varepsilon.$$

Однако даже простая задача оценивания распределения вероятностей значений некоторой наблюдаемой величины  $y$  может рассматриваться как анализ вырожденной функциональной модели вида

$$y = \varepsilon,$$

в котором основное внимание сосредоточено на случайной ошибке  $\varepsilon$ .

Дифференциальные модели, называемые также **динамическими системами** и используемые для моделирования биологических систем, чаще всего задаются с помощью **систем дифференциальных уравнений**. Заметим, что в последнем предложении слово «**система**» было использовано в трех разных смыслах, обозначающих соответственно модель, моделируемый объект и средство моделирования (совокупность уравнений). Методологически более правильно было бы применить в этих трех случаях разные термины, однако следует учитывать сложившуюся практику и ориентироваться на контекст.

# ЧАСТЬ I

## СТАТИСТИЧЕСКИЙ АНАЛИЗ БИОЛОГИЧЕСКИХ ДАННЫХ

---

### ГЛАВА 1

#### ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Задача *теории вероятностей* состоит в построении и анализе вероятностных математических моделей реальных явлений, учитывающих недетерминированные, случайные аспекты этих явлений.

В биологии ввиду сложности изучаемых ею объектов и процессов особенно часто приходится пользоваться вероятностными моделями. Например, уже в такой простой задаче, как сравнение размеров самцов и самок некоторого вида животных, мы встречаемся с той трудностью, что отдельные особи, как самцы, так и самки, имеют разные размеры, причем некоторые самки крупнее некоторых самцов, а некоторые самцы мельче некоторых самок.

Выход состоит в рассмотрении данной ситуации с вероятностной точки зрения, согласно которой размеры отдельных особей представляют собой случайно выбранные значения из двух потенциально бесконечных множеств (*генеральных совокупностей*) самцов и самок данного вида.

Вероятностное описание генеральной совокупности считается заданным, если заданы доли встречаемости элементов всех возможных размеров в этой генеральной совокупности.

Далее будет показано, что удобно задавать распределение долей разных размеров с помощью некоторой функции, называемой *функцией распределения вероятностей*. При таком подходе задача сравнения размеров самцов и самок сводится к задаче сравнения соответствующих функций распределения.

Данная глава посвящена рассмотрению основных понятий теории вероятностей и математических свойств вероятностных моделей.

Для проверки адекватности вероятностных моделей и их подгонки к реально наблюдаемым данным служат **методы математической статистики**, которые будут рассмотрены в гл. 2.

## 1.1. Случайные события и их вероятности

Исходным понятием теории вероятностей является понятие **случайного испытания** (эквивалентные термины: **опыт, эксперимент, наблюдение**), т. е. действия, приводящего к некоторому результату, который, вообще говоря, нельзя однозначно предсказать, зная **комплекс условий проведения испытания**. Например, нельзя заранее наверняка угадать, гербом или решкой упадет подброшенная вверх монета. Нельзя однозначно предсказать пол будущего цыпленка по внешнему виду яйца.

Предполагается, что случайное испытание можно повторять неограниченное число раз, сохраняя комплекс условий, но получая, вообще говоря, разные результаты.

### 1.1.1. Пространство элементарных событий

Множество интересующих нас возможных результатов, которые могут произойти при проведении испытания с заданным комплексом условий, называют **пространством (множеством) элементарных событий** (иногда используют термин «**пространство элементарных исходов**»). Пространство элементарных событий  $\Omega = \{\omega\}$  должно быть **полной системой событий**, а именно, в результате испытания обязательно должно произойти одно и только одно из элементарных событий (далее дано точное определение полной системы событий).

При бросании монеты пространство элементарных событий состоит из двух элементарных событий:  $\Omega = \{\omega_1, \omega_2\}$  — «герб» и «решка». При определении пола вылупившегося из яйца цыпленка пространство элементарных событий также состоит из двух элементарных событий:  $\Omega = \{\omega_1, \omega_2\}$  — «самец» и «самка».

При бросании кубика с пронумерованными от 1 до 6 гранями (игральная кость) пространство элементарных событий состоит из шести элементарных событий:  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ , где  $\omega_i$  соответствует номеру грани, которая оказывается после бросания наверху.

При определении массы новорожденного элементарными событиями будут все возможные результаты взвешивания. Количество элементарных событий в данном случае зависит от возможного диапазона изменения массы и точности взвешивания. Часто в подобных ситуациях удобно расширить множество элементарных событий, отказавшись от указания диапазона изменения измеряемой величины и допустив неограниченную точность измерения. В этом случае множество элементарных событий будет состоять из всех действительных чисел от  $-\infty$  до  $+\infty$ . Конечно, при этом среди элементарных событий будут такие, которые заведомо не могут появиться в результате испытания, например  $-2$  г,  $10^{100}$  г и т. п., однако включение в  $\Omega$  лишних невозможных элементарных событий не противоречит определению пространства элементарных событий.

Пространство элементарных событий может порождаться не одной, а двумя и более измеряемыми характеристиками. Например, если одновременно фиксируются масса и рост новорожденного, то пространство элементарных событий состоит из пар действительных чисел, т. е. геометрически представляет собой плоскость.

### 1.1.2. Операции над событиями

Элементарные события можно объединять в более сложные события. Например, объединение элементарных событий  $\{\omega_1, \omega_3, \omega_5\}$  при бросании кубика образует событие «выпадение нечетного номера». Вообще, любое подмножество пространства элементарных событий называется *случайным (составным) событием*, или просто *событием*. Событие  $\Omega$ , включающее все элементарные события, называется *достоверным*, или *полным*. Считается, что событие происходит, когда происходит одно из составляющих его элементарных событий. Например, событие  $\{\omega_1, \omega_3, \omega_5\}$  происходит, когда происходит или  $\omega_1$ , или  $\omega_3$ , или  $\omega_5$ . Полное событие происходит всегда, поскольку, по определению, в результате испытания обязательно должно произойти одно из элементарных событий. Полезно ввести понятие *невозможного*, или *пустого*, события  $\emptyset$ , т. е. события, не содержащего ни одного элементарного события. Невозможное событие не происходит никогда.

Поскольку все события являются множествами элементарных событий, их можно комбинировать, используя теоретико-множественные операции.

**Объединением**, или **суммой**, событий  $A$  и  $B$  называется событие (обозначается  $A \cup B$  или  $A + B$ ), состоящее из всех элементарных событий, входящих хотя бы в одно из событий  $A$  или  $B$ . Событие  $A \cup B$  происходит, когда происходит хотя бы одно из событий  $A$  или  $B$ .

**Пересечением**, или **произведением**, событий  $A$  и  $B$  называется событие (обозначается  $A \cap B$  или  $AB$ ), состоящее из всех элементарных событий, входящих одновременно и в  $A$ , и в  $B$ . Событие  $A \cap B$  происходит, когда происходят одновременно  $A$  и  $B$ .

**Дополнением**, или **отрицанием**, события  $A$  называется событие (обозначается  $\bar{A}$ ), включающее все элементарные события, не входящие в  $A$ . Событие  $\bar{A}$  происходит, когда не происходит  $A$ .

События  $A$  и  $B$  называются **непересекающимися**, или **несовместными**, если их пересечение есть невозможное событие, т. е.  $AB = \emptyset$ .

Система событий  $A_1, \dots, A_n$  называется **полной системой событий**, если  $A_1 + \dots + A_n = \Omega$  и  $A_i A_j = \emptyset$  для всех  $i \neq j$ .

Примером полной системы будет система всех элементарных событий  $\Omega = \{\omega\}$ . Любая пара событий  $A$  и  $\bar{A}$  также представляет собой полную систему.

### 1.1.3. Вероятности событий

Понятие вероятности имеет два аспекта: содержательный и формально-математический.

С содержательной точки зрения вероятность некоторого события определяется относительными шансами появления этого события по сравнению с другими событиями.

Экспериментально эти шансы реализуются при многократном повторении событий. **Частота появления события** в длинном ряду испытаний является естественной оценкой вероятности события.

Точный смысл изложенному дает формальное аксиоматическое определение вероятности.

Для заданного пространства элементарных событий  $\Omega$  **вероятностью события**  $A$  называется функция  $P(A)$ , удовлетворяющая следующим аксиомам Колмогорова [А. Н. Колмогоров, 1974]:

- 1)  $P(A) \geq 0$  для любого  $A$ ;
- 2)  $P(\Omega) = 1$ ;



3)  $P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots$ , если  $A_i A_j = \emptyset$  для любых  $i \neq j$ .

Из приведенных аксиом непосредственно вытекают следующие свойства вероятности:

- 1)  $0 \leq P(A) \leq 1$  для любого  $A$ ;
- 2)  $P(\emptyset) = 0$ ;
- 3)  $P(A + B) = P(A) + P(B) - P(AB)$ .

**Замечание.** Ранее случайное событие было определено как любое подмножество множества  $\Omega$ . Однако, если, например,  $\Omega$  является множеством действительных чисел, то среди множества всех подмножеств этого множества можно найти подмножества  $A$  настолько необычные, можно сказать «уродливые» (так называемые **неизмеримые подмножества**), что с ними нельзя связать какую-либо функцию  $P(A)$  — см., например, [В. Н. Тутубалин, 1992]. Следовательно, не любое подмножество из  $\Omega$  может быть событием с точки зрения аксиоматического определения вероятности — оно должно быть **измеримым**. С другой стороны, согласно аксиоме 3 счетное объединение событий также должно быть событием. Эти два требования удовлетворяются, если события принадлежат некоторой  **$\sigma$ -алгебре** измеримых подмножеств множества  $\Omega$  (т. е. они измеримы и их дополнения и конечные или счетные объединения или пересечения принадлежат той же  $\sigma$ -алгебре). В случае евклидова пространства такой  $\sigma$ -алгеброй может быть  **$\sigma$ -алгебра борелевских множеств**, образованная любыми параллелепипедами (в одномерном случае — отрезками), а также их дополнениями и конечными или счетными объединениями и пересечениями.

Естественность аксиоматического определения вытекает из доказываемой на его основе теоремы Бернулли, связывающей аксиоматическое определение вероятности с ее частотной интерпретацией.

**Теорема Бернулли.** Вероятность отклонения частоты появления события в  $n$  независимых испытаниях (понятие независимости будет введено позднее) от вероятности этого события  $P(A)$  на величину, большую, чем заданное число  $\varepsilon$ , стремится к 0 при  $n$ , стремящемся к бесконечности («частота стремится к вероятности по вероятности»):

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{m}{n} - P(A) \right| > \varepsilon \right\} = 0,$$

где  $m$  — число появлений события  $A$  в  $n$  испытаниях.

Эта теорема (она будет доказана далее при рассмотрении неравенства Чебышева, см. 1.2.5) утверждает, как нам и хотелось бы, что при увеличении числа испытаний частота стре-

мится к вероятности, однако не в обычном детерминированном смысле, а в вероятностном. При этом не исключается возможность того, что даже при очень большом числе испытаний может получиться значение частоты  $m/n$ , сильно отличающееся от вероятности  $P(A)$ , однако вероятность этого события очень мала. Например, не исключено, что при 100 бросаниях правильной монеты мы получим 100 «гербов», т. е. частоту  $m/n = 100/100 = 1$ , которая сильно отличается от вероятности  $P(A) = 0,5$ , однако вероятность такого большого отклонения частоты от вероятности  $P(A)$  равна ничтожно малой величине  $1/2^{100}$ .

Другим доводом в пользу выбранной системы аксиом будет наличие конкретных вероятностных моделей, в которых эта система аксиом естественна. Рассмотрим две такие модели.

#### 1.1.4. Модель равновероятных элементарных событий

В модели равновероятных элементарных событий (называемой также *классической моделью вероятности*) число элементарных событий конечно и они считаются равновероятными. Примерами могут служить испытания с бросанием монеты (два равновероятных события), кубика (шесть равновероятных событий) или результаты скрещивания двух гетерозиготных особей  $Aa$  (четыре равновероятных события:  $AA$ ,  $Aa$ ,  $aA$ ,  $aa$ ).

Вероятность события в классической модели определяется как отношение числа элементарных событий, составляющих это событие, к числу всех элементарных событий. Например, вероятность получения потомка с доминантным признаком в результате скрещивания двух гетерозиготных особей равна  $3/4$ , поскольку это событие включает (часто говорят — «ему благоприятствуют») три элементарных события ( $AA$ ,  $Aa$  и  $aA$ ) из четырех возможных (опыт Менделя).

Очевидно, что свойства вероятности, постулируемые в аксиоматическом определении, здесь непосредственно следуют из определения вероятности как отношения числа «благоприятствующих» элементарных событий к общему числу возможных равновероятных событий.

Решение вероятностных задач с использованием классической модели часто облегчается применением комбинаторных формул. Эти формулы определяют число элементарных событий в опытах, состоящих в случайном выборе элементов из  $n$  различных элементов. При этом в постановке каждого такого опыта

строго указано, каким способом производится выбор и что понимается под различными выборами.

Существуют две принципиально различные схемы выбора. В первой схеме выбор осуществляется без возвращения элементов, т. е. каждый отобранный элемент исключается из исходного множества элементов. Назовем эту схему **выбором без возвращения**. Во второй схеме выбор осуществляется поэлементно с обязательным возвращением отобранного элемента на каждом шаге и тщательным перемешиванием исходного множества перед следующим выбором. Назовем эту схему **выбором с возвращением**. После того как выбор осуществлен, отобранные элементы могут быть либо упорядочены в порядке выбора из исходного множества (в этом случае будем говорить, что «**порядок выбора элементов важен**»), либо нет (в этом случае будем говорить, что «**порядок выбора элементов не важен**»).

Следовательно, для определения числа способов выбора  $m$  элементов из исходного множества  $n$  элементов надо рассмотреть четыре различные постановки задачи.

1. Выбор без возвращения. Порядок выбора важен. В этом случае для выбора первого элемента имеется  $n$  возможностей, второго —  $n - 1$ , третьего —  $n - 2$  и так далее до  $m$ -го, для которого остается  $n - m + 1$  возможностей. Это дает  $n(n-1)(n-2) \dots \times (n - m + 1)$  вариантов, называемых **размещениями** из  $n$  элементов по  $m$  элементов. Умножив и разделив полученное выражение на  $(n - m)! = (n - m)(n - m - 1) \dots \cdot 2 \cdot 1$ , получаем следующую формулу для числа размещений:  $A_n^m = n! / (n - m)!$  В частности, для числа размещений из  $n$  элементов по  $n$  элементов, называемых также **перестановками** из  $n$  элементов, получаем  $P_n = A_n^n = n!$

2. Выбор без возвращения. Порядок выбора не важен. Число таких комбинаций, называемых **сочетаниями** из  $n$  элементов по  $m$  элементов, получается делением числа размещений из  $n$  элементов по  $m$  элементов на число перестановок из  $m$  элементов, равное  $m! = m(m-1) \dots \cdot 2 \cdot 1$ , что дает следующую формулу для числа сочетаний:  $C_n^m = n! / [m!(n - m)!]$ .

3. Выбор с возвращением. Порядок выбора важен. В этом случае для выбора каждого из  $m$  элементов имеется  $n$  возможностей, что дает  $\bar{A}_n^m = n^m$  комбинаций, называемых **размещениями с повторениями** из  $n$  элементов по  $m$  элементов (заметим, что  $m$  может быть как меньше, так и больше  $n$ ).

4. Выбор с возвращением. Порядок выбора не важен. В этом случае число комбинаций, называемых **сочетаниями с по-**

**вторениями** из  $n$  элементов по  $m$  элементов ( $m$  может быть как меньше, так и больше  $n$ ), определяется формулой:  $\bar{C}_n^m = C_{n+m-1}^m$ . Действительно, сочетания с повторениями можно получить как перестановки из  $n + m - 1$  объектов, состоящих из  $m$  фишек и  $n - 1$  разделителей. В этих перестановках фишки, оказавшиеся перед первым разделителем, считаются элементами, принадлежащими к первому из  $n$  типов элементов, оказавшиеся после первого разделителя — принадлежащими ко второму типу, после второго — к третьему и т. д. Поскольку внутренний порядок фишек и разделителей не влияет на результат, числе всех перестановок  $(n + m - 1)!$  надо разделить на число перестановок фишек  $m!$  и число перестановок разделителей  $(n - 1)!$ , что дает приведенную выше формулу.

### Задачи для самостоятельного решения

1. Сколькими способами можно разместить 12 мышей, занумерованных от 1 до 12, в четырех клетках  $A, B, C, D$  по три мыши в каждой?

2. В распоряжении агрохимика имеется шесть различных типов минеральных удобрений. Ему необходимо провести эксперименты по изучению совместного влияния любой тройки минеральных удобрений. Сколько всего экспериментов ему придется провести, если: а) порядок внесения удобрений несущественен? б) существенен?

3. Для лечения некоторой хронической болезни применяются пять лекарств  $a, b, c, d, e$ . Врач хочет провести сравнительное исследование трех из этих пяти лекарств. Три исследуемых лекарства врач отбирает из данных пяти случайным образом. Чему равна вероятность того, что: а) лекарство  $a$  будет исследовано? б) будут исследованы лекарства  $a$  и  $b$ ? в) будет исследовано по крайней мере одно из лекарств  $a$  и  $b$ ?

4. Классифицируются  $n$  особей по  $r$  признакам,  $n \geq r$ . Найдите вероятность того, что никакие две особи не принадлежат к одному и тому же классу. Все возможные распределения особей по классам равновероятны.

### 1.1.5. Геометрическая модель

Рассмотрим некоторую область  $\Omega$  на плоскости площадью  $S_\Omega$ . Определим вероятность произвольного участка  $A$  площадью  $S_A$  в этой области как долю его площади в площади всей области:

$$P(A) = S_A/S_\Omega.$$

С помощью такой модели можно, например, описать эксперимент, состоящий в регистрации попадания космической частицы на фотопластинку.

Определение вероятности в соответствии с приведенной формулой полностью согласуется с аксиоматическим определением, поскольку из геометрических свойств площадей непосредственно следует, что  $P(A) \geq 0$ ,  $P(\Omega) = 1$ , и  $P(A + B) = P(A) + P(B)$ , если  $AB = \emptyset$ .

### 1.1.6. Условная вероятность и независимость

*Условной вероятностью*  $P(A|B)$  события  $A$  при условии, что произошло событие  $B$ , называют отношение

$$P(A|B) = \frac{P(AB)}{P(B)}$$

вероятности пересечения событий  $A$  и  $B$  к вероятности события  $B$  (предполагается, что  $P(B) \neq 0$ ).

Из определения условной вероятности непосредственно следует **правило умножения**:

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

— вероятность произведения событий равна произведению условной вероятности на вероятность условия.

Рассматривая условные вероятности при условии  $B$ , мы фактически переходим к новому пространству элементарных событий, которое является частью первоначального и включает только элементарные события, соответствующие событию  $B$ . Доля таких элементов в исходном пространстве элементарных событий равна  $P(B)$ . Событие  $A|B$  можно интерпретировать как извлечение элементарного события, соответствующего событию  $A$ , из совокупности элементарных событий, соответствующих событию  $B$ . Доля таких элементарных событий в исходном пространстве элементарных событий равна  $P(AB)$ , а доля их в  $B$  равна  $P(AB)/P(B)$ .

**Пример 1.1.** Найдем вероятность того, что потомок, полученный в результате скрещивания гетерозиготных родителей, гомозиготен, если он имеет доминантный признак.

Вероятность наличия у потомка доминантного признака равна, как нам уже известно,  $3/4$ . Вероятность гомозиготности при наличии доминантного аллеля, т. е. вероятность пересечения

событий  $\{AA, aa\}$  и  $\{AA, Aa, aA\}$ , равна вероятности события  $\{AA\}$ , т. е.  $1/4$ . Таким образом, искомая условная вероятность равна  $1/4 : 3/4 = 1/3$ .

С понятием условной вероятности тесно связано понятие независимости событий, хотя можно определить понятие независимости и безотносительно к условной вероятности.

Два события называются *независимыми*, если вероятность их пересечения равна произведению их вероятностей:

$$P(AB) = P(A)P(B).$$

Из этого определения сразу следует, что условная вероятность события  $A$  при условии, что произошло независимое событие  $B$ , равна безусловной вероятности события  $A$ :

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

и также, что условная вероятность события  $B$  при условии, что произошло независимое событие  $A$ , равна безусловной вероятности события  $B$ :

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(B)P(A)}{P(A)} = P(B).$$

Более того, любое из условий  $P(A|B) = P(A)$  или  $P(B|A) = P(B)$  влечет выполнение соотношения

$$P(AB) = P(A)P(B),$$

определяющего независимость.

**Пример 1.2.** Рассмотрим два события:  $A$  — попадание частицы в левую половину фотопластинки,  $B$  — попадание частицы в нижнюю половину фотопластинки. События  $A$  и  $B$  имеют вероятности  $1/2$ . Событие  $AB$ , соответствующее попаданию частицы в пересечение  $A$  и  $B$ , т. е. в левый нижний угол фотопластинки, имеет вероятность  $1/4$ . Поскольку  $1/4 = 1/2 \cdot 1/2$ , т. е.  $P(AB) = P(A)P(B)$ , события  $A$  и  $B$  независимы.

**Пример 1.3.** Зависимы ли события «наличие доминантного признака»  $\{AA, Aa, aA\}$  и гомозиготность  $\{AA, aa\}$ ? Вероятность первого события равна  $3/4$ , вероятность второго —  $1/2$ . Вероятность их пересечения равна  $1/4$ . Поскольку  $1/4 \neq 3/4 \cdot 1/2$ , события не являются независимыми.

К этому же выводу можно прийти иначе. В примере 1.1 было показано, что вероятность гомозиготности при наличии доми-

нантного аллеля равна  $1/3$ . Вообще же вероятность гомозиготного потомка равна  $1/2$ . Таким образом, условная вероятность не равна безусловной и, следовательно, события зависимы.

**Пример 1.4.** Рассмотрим случайное испытание, состоящее из двух последовательных бросаний монеты. Результатом этого события будут четыре равновероятных элементарных события: герб — герб, герб — решка, решка — герб, решка — решка (или  $ГГ$ ,  $ГР$ ,  $РГ$ ,  $РР$ ). Зависимы ли события «выпадение герба в первом бросании»  $A = \{ГГ, ГР\}$  и «выпадение герба во втором бросании»  $B = \{ГГ, РГ\}$ ? Имеем  $P(A) = 1/2$ ,  $P(B) = 1/2$ ,  $P(AB) = P(ГГ) = 1/4$ . Поскольку  $P(AB) = P(A)P(B)$ , то события независимы.

### Задачи для самостоятельного решения

1. Два охотника стреляют в волка, причем каждый производит по одному выстрелу. Для первого охотника вероятность попадания в цель равна 0,7, для второго — 0,8. Какова вероятность хотя бы одного попадания? Как изменится результат, если охотники произведут по два выстрела?

2. Докажите, что события  $A$  и  $\bar{B}$  ( $\bar{A}$  и  $\bar{B}$ ) независимы, если независимы  $A$  и  $B$ .

3. Докажите, что если два события  $A$  и  $B$  с положительными вероятностями несовместны, то они зависимы.

4. Задача «осторожного фальшивомонетчика». Дворцовый чеканщик кладет в каждый ящик вместимостью в 100 монет одну фальшивую. Король подозревает чеканщика и подвергает проверке монеты, взятые наудачу по одной в каждом из 100 ящичков. Какова вероятность того, что чеканщик не будет разоблачен?

5. Известно, что при трех бросаниях игральной кости цифра 6 выпала хотя бы один раз. Какова вероятность того, что она выпала два раза?

6. Предположим, что в семье с тремя детьми все возможные распределения детей по полу равновероятны. Событие  $A$  — «в семье имеются дети обоих полов» и событие  $B$  — «в семье имеется не более одной девочки». 1) События  $A$  и  $B$  независимы? 2)  $A$  для семьи, имеющей двоих детей, события  $A$  и  $B$  независимы?

#### 1.1.7. Формула полной вероятности и формула Байеса

Пусть имеется полная система событий  $A_1 + \dots + A_n = \Omega$ ,  $A_i A_j = \emptyset$  для  $i \neq j$ . Тогда любое событие  $B$  можно представить в виде суммы непересекающихся событий  $B = BA_1 + \dots + BA_n$  и, следовательно,

$$P(B) = P(BA_1) + \dots + p(BA_n).$$

Вспомнив правило умножения, можно записать

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n) = \\ &= \sum_{i=1}^n P(B|A_i)P(A_i) \end{aligned} \quad (1.1)$$

— *формула полной вероятности.*

**Пример 1.5.** Пусть имеются три трудноразличимые при диагнозе болезни  $A_1, A_2, A_3$ , встречающиеся с частотой 50, 40 и 10 %, и есть метод лечения  $B$ , приводящий к успеху соответственно в 70, 75 и 90 % случаев. Какова вероятность излечения для пациента, страдающего одной (неизвестно, какой именно) из болезней  $A_1, A_2$  или  $A_3$ ?

Имеем  $P(A_1) = 0,5; P(A_2) = 0,4; P(A_3) = 0,1; P(B|A_1) = 0,7; P(B|A_2) = 0,75; P(B|A_3) = 0,9$ . Получаем:  $P(B) = 0,5 \cdot 0,7 + 0,4 \times 0,75 + 0,1 \cdot 0,9 = 0,74$ .

Из соотношений, полученных в соответствии с правилом умножения  $P(A_i B) = P(B|A_i)P(A_i)$  и  $P(A_i B) = P(A_i|B)P(B)$ , находим

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}.$$

Заменяв  $P(B)$  в знаменателе по формуле полной вероятности, получаем **формулу Байеса**, или **формулу вероятностей гипотез**:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}. \quad (1.2)$$

Вероятности  $P(A_i), i = 1, \dots, n$ , в данном контексте носят название **априорных вероятностей** гипотез  $A_i$ , а вероятности  $P(A_i|B), i = 1, \dots, n$ , называются **апостериорными вероятностями гипотез**  $A_i$ . Формула Байеса, таким образом, позволяет, исходя из результатов эксперимента, корректировать имеющиеся знания о вероятностях интересующих нас событий.

**Пример 1.6.** Рассмотрим данные примера 1.5, изменив вопрос. Предположим, что лечение по методу  $B$  оказалось успешным. Какие выводы можно сделать на основании этого факта от-



носителем диагноза? Формула Байеса дает следующие результаты:

$$P(A_1|B) = 0,5 \cdot 0,7/0,74 = 0,47;$$

$$P(A_2|B) = 0,4 \cdot 0,75/0,74 = 0,4;$$

$$P(A_3|B) = 0,1 \cdot 0,9/0,74 = 0,13.$$

Итак, вероятность диагноза  $A_1$  несколько снизилась, а вероятность диагноза  $A_3$  повысилась, что естественно, поскольку лечение  $B$  более эффективно при болезни  $A_3$ .

### Задачи для самостоятельного решения

1. Предположим, что 5 % всех мужчин и 0,25 % всех женщин дальтоники. Наугад выбранное лицо оказалось дальтоником. Какова вероятность того, что это мужчина? (Считать, что мужчин и женщин одинаковое число.)

2. При рентгеновском обследовании вероятность обнаружить заболевание туберкулезом у больного туберкулезом равна 0,9. Вероятность принять здорового человека за больного равна 0,01. Пусть доля больных туберкулезом по отношению ко всему населению равна 0,001. Найдите вероятность того, что человек здоров, если он был признан больным при обследовании.

3. Среди 25 экзаменационных билетов 5 «хороших». Два студента по очереди берут по одному билету. Найдите вероятности следующих событий:  $A$  — первый студент взял «хороший» билет;  $B$  — второй студент взял «хороший» билет;  $C$  — оба студента взяли «хорошие» билеты.

4. Ваши друзья могут с равной вероятностью играть в одну из двух игр. В одной игре используется одна игральная кость, в другой — две игральные кости. Счет в любой игре равен количеству очков, выпавших на одной кости, или на обеих костях вместе. Вы слышите, что в какой-то игре у них выпало два очка. Какова вероятность того, что они играют в игру с одной костью?

5. В первой урне 8 белых и 2 черных шара, во второй урне 16 черных шаров. Из каждой урны наудачу извлекли по одному шару, а затем из этих шаров наудачу взяли один шар. Найдите вероятность того, что он белый.

6. В родительской популяции генотипы  $AA$ ,  $Aa$ ,  $aa$  встречаются с частотами  $p^2$ ,  $2pq$ ,  $q^2$ , соответственно  $p + q = 1$ . Найдите распределение генотипов потомков (предполагается случайное скрещивание).

7. Найдите три вероятности того, что потомок имеет генотип  $AA$ , если одна из родительских особей (мужская или женская) имеет генотип: (1)  $AA$ , (2)  $Aa$ , (3)  $aa$ . Другая родительская особь имеет один из генотипов  $AA$ ,  $Aa$ ,  $aa$ . В родительской популяции генотипы  $AA$ ,  $Aa$ ,  $aa$

встречаются с частотами  $p^2, 2pq, q^2$ , соответственно, и  $p + q = 1$  (предполагается случайное скрещивание).

8. В родительской популяции генотипы  $AA, Aa, aa$  встречаются с частотами  $p^2, 2pq, q^2$ , соответственно, и  $p + q = 1$  (предполагается случайное скрещивание). Известно, что их потомок имеет генотип  $AA$ . Найдите вероятность того, что его брат имеет генотип  $Aa$ .

### 1.1.8. Формула Бернулли

При проведении  $n$  независимых испытаний, в каждом из которых вероятность появления события (часто называемого «успехом») равна  $p$  ( $0 < p < 1$ ), вероятность наступления этого события ровно  $k$  раз описывается **формулой Бернулли**

$$P_n(k) = C_n^k p^k q^{n-k}, \quad (1.3)$$

где  $q = 1 - p$ .

Доказательство формулы Бернулли следует из того, что событие « $n$  испытаний закончились  $k$  «успехами» и  $n - k$  «неуспехами» состоит из стольких составляющих несовместных событий, сколько существует способов выбора  $k$  предметов из  $n$  предметов. Следовательно, количество этих составляющих событий равно  $C_n^k$ , а вероятность каждого из них равна  $p^k q^{n-k}$  (так как проводится  $n$  независимых испытаний).

Из формулы (1.3) следует, что вероятность наступления события не менее  $k_1$  раз и не более  $k_2$  раз ( $k_1 \leq k_2$ ) равна

$$P_n(k_1, k_2) = \sum_{i=k_1}^{k_2} P_n(i).$$

#### Задачи для самостоятельного решения

1. Что вероятнее произойдет при бросании правильной монеты — герб выпадет один раз при двух бросаниях или герб выпадет три раза при шести бросаниях?

2. Вероятность хотя бы одного появления события  $A$  в четырех независимых опытах равна 0,5. Какова вероятность появления события  $A$  в одном опыте, если в каждом опыте эта вероятность одинакова?

3. Всхожесть семян некоторого сорта растений оценивается с вероятностью 0,8. Какая вероятность больше: из десяти наугад выбранных семян взойдет хотя бы одно или из двенадцати семян взойдут хотя бы два?

4. В некоторой популяции насекомых 30 % особей инфицированы. Что вероятнее — найти хотя бы одну инфицированную особь из 10 наугад выбранных насекомых или более одной из 12?

5. При одном цикле осмотра радиолокационной станции, следящей за космическим объектом, объект будет обнаружен с вероятностью 0,5. Сколько надо провести циклов осмотра, чтобы вероятность обнаружения объекта была больше 0,99?

## 1.2. Случайные величины и их распределения

*Одномерной случайной величиной*  $\xi$ , или просто *случайной величиной*, называют любую числовую функцию, определенную на пространстве элементарных событий  $\Omega$ .

Так как одномерная случайная величина  $\xi$  представляет собой числовую функцию на пространстве элементарных событий, то любая числовая функция  $\varphi(\xi)$  от случайной величины в соответствии с определением также является случайной величиной.

*Пример 1.7.* Рассмотрим пространство элементарных событий, которое получается в результате независимых бросаний двух монет. В этом примере пространство элементарных событий состоит из четырех элементарных событий, которым соответствует вероятность 0,25. Определим на этом пространстве случайную величину  $\xi$ , равную числу гербов, появившихся при бросании двух монет. Очевидно, что значения случайной величины  $\xi$  равны 0, 1, 2 и случайная величина принимает эти значения с вероятностями 0,25; 0,5; 0,25 соответственно.

*Пример 1.8.* Пусть имеется очень большое количество кроликов  $N$ , среди которых доля альбиносов равна 0,05. Предположим, что из общего количества  $N$  кроликов наугад выбрано 40. Рассмотрим пространство элементарных событий  $\Omega$  для представления результатов данного эксперимента. В этом примере пространство элементарных событий состоит из  $C_N^{40}$  равновероятных элементарных событий. Определим теперь на этом пространстве случайную величину  $\xi$ , равную числу альбиносов среди 40 выбранных кроликов. Множество возможных значений случайной величины  $\xi$  есть числа от 0 до 40 включительно. Из формулы Бернулли следует, что вероятность того, что случайная величина  $\xi$  примет значение  $i$ , где  $i = 0, 1, 2, \dots, 40$ , равна

$$P(\xi = i) = P_{40}(i) = C_{40}^i (0,05)^i (0,95)^{40-i}.$$

### 1.2.1. Функция распределения случайной величины

**Функцией распределения вероятностей**, или просто **функцией распределения** (иногда применяют термин **кумулятивная функция распределения**), случайной величины  $\xi$  называется функция  $F(x)$ , равная для любого значения  $x$  вероятности события  $\xi \leq x$ :

$$F(x) = P(\xi \leq x). \quad (1.4)$$

В некоторых источниках литературы нестрогое неравенство в формуле (1.4) заменяют строгим. Часто также применяют другое обозначение функции распределения вероятностей случайной величины  $F_{\xi}(x)$ .

Из определения (1.4) легко вывести следующие свойства функции распределения:

- 1)  $0 \leq F(x) \leq 1$ ;
- 2)  $F(x)$  — неубывающая функция;
- 3)  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;
- 4)  $\lim_{x \rightarrow +\infty} F(x) = 1$ ;
- 5)  $P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1)$ .

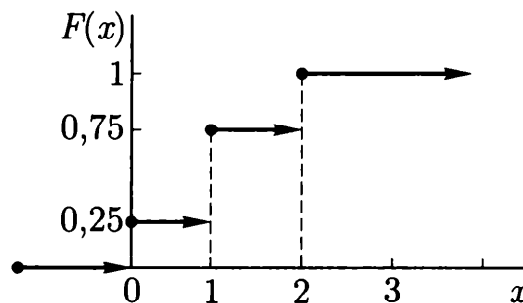


Рис. 1.1. Функция распределения  $F(x)$  случайной величины (см. пример 1.7)

На рис. 1.1 приведен график функции распределения вероятностей случайной величины из примера 1.7.

### 1.2.2. Дискретные случайные величины

Случайные величины бывают двух типов: дискретные — принимающие конечное или счетное число значений, и непрерывные — принимающие любое значение на некотором непрерывном промежутке действительной числовой оси. В этом подразделе рассмотрим дискретные случайные величины.

**Дискретной случайной величиной**  $\xi$  называется случайная величина, принимающая конечное или счетное множество значений  $x_0, x_1, x_2, \dots$

Обозначим множество всех возможных значений, которые принимает дискретная случайная величина  $\xi$ , через  $x_0, x_1, x_2, \dots$ , а вероятности, с которыми  $\xi$  принимает эти значения, —  $p_0, p_1, p_2, \dots$ . Тогда  $\sum_i p_i = 1$ . Распределение дискретной случайной величины  $\xi$  будет полностью описано, если указать для любого  $i$  вероятность  $p_i$  того, что  $\xi$  принимает значение  $x_i$ , т. е.  $p_i = P(\xi = x_i)$ . Функция распределения  $F(x)$  дискретной случайной величины  $\xi$  при этом оказывается равной

$$F(x) = P(\xi \leq x) = \sum_{x_i \leq x} p_i.$$

Таким образом,  $F(x)$  — ступенчатая функция, равная постоянной на любом интервале, не содержащем точек  $x_i$ , и имеющая в каждой точке  $x_i$  скачок вверх на величину  $p_i$ .

Таким образом, чтобы задать дискретную случайную величину  $\xi$ , достаточно описать множество всех возможных значений случайной величины  $x_0, x_1, x_2, \dots$ , а также указать числа  $p_i$ , такие, что

$$1) p_i \geq 0, \quad i = 0, 1, 2, \dots$$

$$2) \sum_i p_i = 1.$$

Наиболее распространенными формами представления дискретных случайных величин являются табличная

$x_0$	$x_1$	$x_2$	$\dots$
$p_0$	$p_1$	$p_2$	$\dots$

и графическая (рис. 1.2 — 1.5), отображающие зависимость вероятности  $p_i = P(\xi = x_i)$  от значения случайной величины  $x_i$ . Функция, выражающая эту зависимость, называется **распределением вероятностей дискретной случайной величины**.

Наиболее известными примерами дискретных случайных величин являются: случайная величина, распределенная по дискретному равномерному закону; биномиально распределенная случайная величина; случайная величина, распределенная по закону Бернулли; случайная величина, распределенная по закону Пуассона.

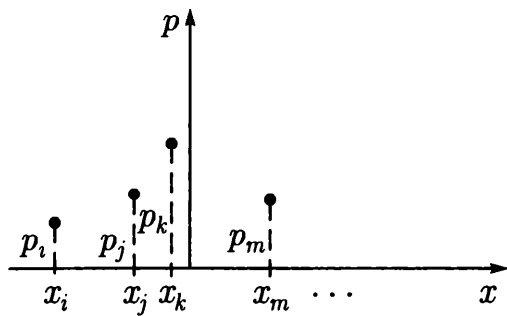


Рис. 1.2. Распределение вероятностей дискретной случайной величины

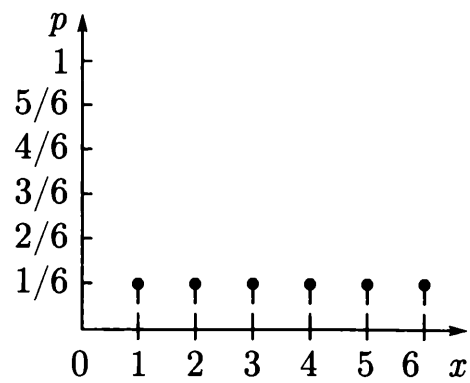


Рис. 1.3. Распределение вероятностей дискретного равномерного распределения ( $n = 6$ )

Случайная величина, принимающая  $n$  ( $n \geq 1$ ) значений  $x_1, x_2, \dots, x_n$  с вероятностями  $p_i = 1/n$ , называется **случайной величиной, распределенной по дискретному равномерному закону**. На рис. 1.3 рассматриваемая случайная величина (для  $n = 6$ ) представлена в графической форме. Случайная величина, распределенная по дискретному равномерному закону, является моделью событий с равновероятными исходами (см. пример с бросанием игральной кости).

Случайная величина, принимающая два значения: 0 и 1 с вероятностями  $q = 1 - p$  и  $p$ , соответственно ( $0 < p < 1$ ), называется распределенной по **закону Бернулли** с параметром  $p$ . Случайная величина, распределенная по закону Бернулли, — удачная модель для описания многих конкретных испытаний, имеющих два исхода (наиболее известный пример — бросание правильной монеты — здесь  $p = q = 1/2$ ), в том числе и в биологии, на-

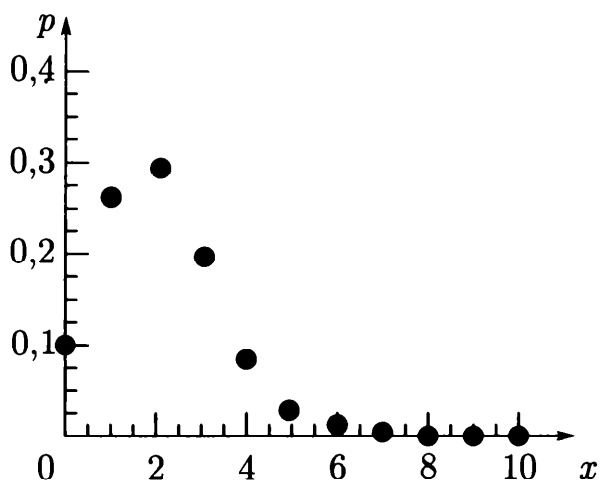


Рис. 1.4. Распределение вероятностей биномиально распределенной случайной величины ( $n = 10$ ;  $p = 0,2$ )

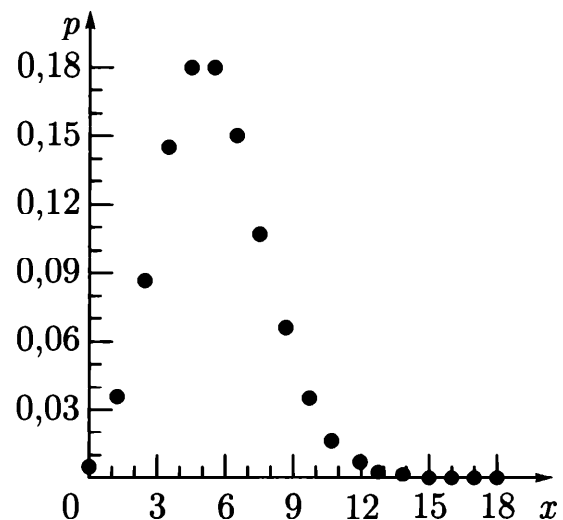


Рис. 1.5. Распределение вероятностей пуассоновской случайной величины ( $\lambda = 5$ )

пример, присутствие или отсутствие некоторого признака, пол родившегося детеныша и т. д.

Случайная величина  $\xi$  — число успехов в  $n$  испытаниях (см. формулу (1.3)), принимающая  $n + 1$  значение  $0, 1, 2, \dots, n$  с вероятностями

$$p_i = P(\xi = i) = C_n^i p^i q^{n-i},$$

где  $i = 0, 1, 2, \dots, n$ ,  $q = 1 - p$ ,  $0 < p < 1$ , называется **биномиально распределенной случайной величиной**, а  $n$  и  $p$  — параметрами распределения. На рис. 1.4 биномиальная случайная величина представлена в графической форме. Пример использования биномиальной случайной величины дан в примере 1.8.

Заметим, что случайная величина, распределенная по закону Бернулли, является частным случаем биномиальной случайной величины для  $n = 1$ .

Случайная величина  $\xi$ , принимающая счетное множество значений  $0, 1, 2, \dots$  с вероятностями

$$p_i = P(\xi = i) = \frac{\lambda^i e^{-\lambda}}{i!},$$

где  $i = 0, 1, 2, \dots$ ,  $\lambda > 0$ , называется **случайной величиной, распределенной по закону Пуассона**.

Величина  $\lambda$  называется **параметром распределения Пуассона**.

Случайная величина, распределенная по закону Пуассона, может быть получена как предел биномиальной случайной величины при  $n \rightarrow \infty, p \rightarrow 0, np = \text{const} = \lambda$ .

На рис. 1.5 случайная величина, распределенная по закону Пуассона, представлена в графической форме. Случайная величина, распределенная по закону Пуассона, служит моделью числа появления некоторого события за единицу времени, численности бактерий в единице объема, численности животных на единице площади и т. п.

Распределение Пуассона иногда называют «распределением вероятностей редких событий», поскольку оно хорошо описывает ситуацию случайно и независимо друг от друга появляющихся событий в течение заданного периода времени (регистрация радиоактивных частиц в счетчике Гейгера, телефонные звонки, появление посетителей в малопосещаемом магазине и т. п.). При этом существенна именно независимость событий, а их «редкость» требуется лишь для того, чтобы можно было пренебречь вероятностью одновременного появления двух событий. Если параметр  $\lambda$  относится к единице времени, то периоду

времени длительностью  $t$  будет соответствовать пуассоновское распределение с параметром  $\lambda t$ . Соответственно вероятность того, что в течение периода  $t$  не произойдет ни одного события, равна

$$p_0 = P(\xi = 0) = e^{-\lambda t}.$$

Например, если появление события влечет гибель организма, то  $p_0 = e^{-\lambda t}$  можно интерпретировать как вероятность того, что организм доживет до возраста  $t$  («*вероятность дожития*»). Параметр  $\lambda$  в этом случае называют *интенсивностью смертности* (*смертностью*).

Из приведенной формулы видно, что чем больше  $\lambda$ , тем меньше вероятность дожить до заданного возраста  $t$ , и, конечно, чем больше заданный возраст, тем меньше вероятность до него дожить (классический пример — время жизни стакана в столовой).

Из других часто используемых дискретных распределений отметим без определения отрицательное биномиальное распределение и гипергеометрическое распределение.

### Задачи для самостоятельного решения

1. Нарисуйте график функции распределения случайной величины, распределенной по закону Бернулли.

2. Нарисуйте график функции распределения случайной величины, распределенной по дискретному равномерному закону ( $n = 6$ ).

3. Вероятность хотя бы одного появления события  $A$  при четырех независимых опытах равна 0,61. Какова вероятность появления события  $A$  при одном опыте, если при каждом опыте эта вероятность одинакова?

4. Имеется пять независимых проб воздуха единичного объема. Число бактерий в каждой пробе распределено по закону Пуассона с параметром  $\lambda$ , равным 2. Найдите вероятность того, что, по крайней мере, в одной пробе имеется не менее трех бактерий.

### 1.2.3. Непрерывные случайные величины

*Непрерывной случайной величиной* называется случайная величина, для которой функция распределения  $F(x)$  непрерывна, и существует функция  $f(x)$ , называемая *функцией плотности вероятностей* (*функцией плотности*), такая, что

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (1.5)$$



Формула (1.5) с учетом интерпретации интеграла как площади проиллюстрирована графически на рис. 1.6.

Справедливы следующие свойства функции плотности:

1)  $f(x) \geq 0$ ;

2) 
$$\int_{-\infty}^{+\infty} f(x) dx = 1$$
;

3) 
$$P(x_1 < \xi \leq x_2) = \int_{x_1}^{x_2} f(x) dx.$$

Свойство 3 функции плотности проиллюстрировано на рис. 1.7. Функция плотности служит наглядной формой представления непрерывных случайных величин, являющихся моделью для описания многих экспериментов, связанных с измерением количественных признаков (например, массы животного, длины какой-либо части тела, температуры, влажности, урожайности и т. д.). С помощью функции плотности легко сравнить вероятности того, что случайная величина примет значение из одного или другого интервала. Знание функции плотности случайной величины  $\xi$  дает полную информацию о непрерывной случайной величине  $\xi$ , т. е. знание о ее распределении.

Простейший пример непрерывной случайной величины — равномерно распределенная случайная величина.

Случайная величина  $\xi$  называется **равномерно распределенной на отрезке**  $[a, b]$  ( $a$  и  $b$  — параметры распределения), если ее функция плотности имеет вид

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leq x \leq b; \\ 0 & \text{— в противном случае.} \end{cases}$$

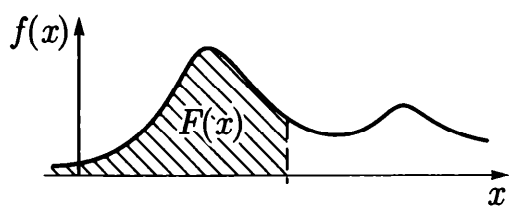


Рис. 1.6. Связь между функцией распределения  $F(x)$  и функцией плотности  $f(x)$

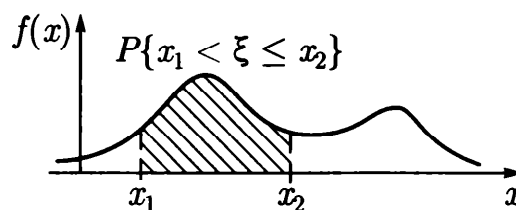


Рис. 1.7. Определение вероятности попадания случайной величины  $\xi$  в интервал от  $x_1$  до  $x_2$  по функции плотности

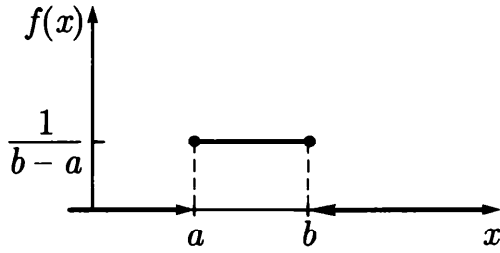


Рис. 1.8. Функция плотности равномерно распределенной на отрезке  $[a, b]$  случайной величины

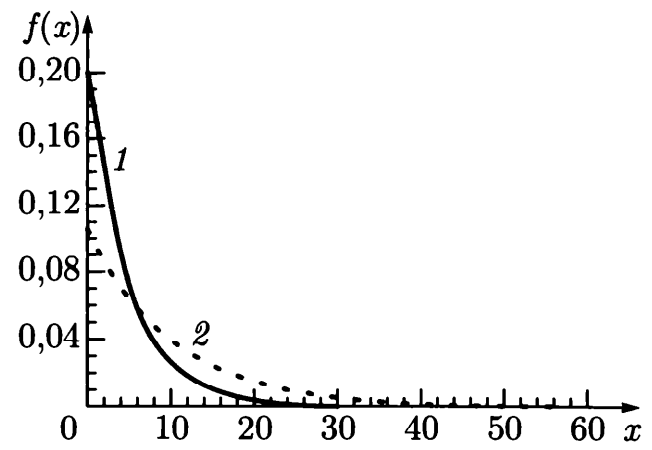


Рис. 1.9. Функция плотности экспоненциального распределения:

$$1 - Q = 0,2; 2 - Q = 0,1$$

Нетрудно проверить, что данная функция удовлетворяет свойствам 1—3. График функции плотности равномерно распределенной на отрезке  $[a, b]$  случайной величины приведен на рис. 1.8. Вероятность того, что равномерно распределенная случайная величина  $\xi$  примет значение из отрезка определенной длины  $[x_1, x_2]$ , целиком лежащего между  $a$  и  $b$ , не зависит от расположения этого отрезка и равна  $(x_2 - x_1)/(b - a)$ . Вероятность принять значения на отрезке  $[x_1, x_2]$ , расположенного вне отрезка  $[a, b]$ , равна нулю.

Непрерывная случайная величина  $\xi$  называется **распределенной по экспоненциальному закону** с параметром  $Q$ , если функция плотности  $f(x)$  случайной величины  $\xi$  имеет вид

$$f(x) = \begin{cases} Qe^{-Qx}, & \text{если } x \geq 0; \\ 0 & \text{— в противном случае,} \end{cases}$$

где  $Q > 0$ .

График функции плотности экспоненциально распределенной случайной величины представлен на рис. 1.9. Функция распределения экспоненциального распределения равна 0 при  $x < 0$ , а если  $x \geq 0$ , то она равна

$$F(x) = \int_0^x Qe^{-Qs} ds = 1 - e^{-Qx}.$$

Например, вероятность дожития до определенного возраста имеет экспоненциальное распределение (см. обсуждение пуассоновского распределения).

Из других часто используемых непрерывных распределений отметим треугольное распределение, распределение Эрланга, распределение Вейбулла, гамма-распределение, бета-распределение.

Современные пакеты прикладных программ позволяют получать графики функции распределения, распределения вероятностей дискретной случайной величины и плотности непрерывной случайной величины.

### Задачи для самостоятельного решения

1. Нарисуйте график функции распределения случайной величины, распределенной по равномерному закону на отрезке  $[0, 1]$ .
2. Нарисуйте график функции распределения случайной величины, распределенной по экспоненциальному закону с параметром  $Q = 2$ .

#### 1.2.4. Нормальное и связанные с ним распределения

Построение статистических моделей чаще всего основывается на предположении о **нормальности** используемых распределений.

Непрерывная случайная величина  $\xi$  называется **нормально распределенной случайной величиной**, т. е. **распределенной по нормальному**, или **гауссовому, закону** с параметрами  $\mu$  и  $\sigma^2$  (записывается  $\xi \sim N(\mu, \sigma^2)$ ), если функция плотности  $f(x)$  случайной величины  $\xi$  имеет вид

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

где  $-\infty < x < +\infty$ ,  $-\infty < \mu < +\infty$ ,  $\sigma > 0$ .

График функции плотности нормально распределенной случайной величины представлен на рис. 1.10 в двух вариантах:  $\mu = 5$ ,  $\sigma = 1$  и  $\mu = 10$ ,  $\sigma = 3$ . Заметим, что максимальное значение  $f(x)$  достигается при  $x = \mu$  ( $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$ ) и график  $f(x)$  симметричен относительно прямой, проходящей через точку  $\mu$  параллельно оси ординат.

При практическом использовании нормального распределения оказывается полезным следующее его свойство: площадь графика под функцией плотности  $f(x)$  нормального распределения над интервалом  $(\mu - k\sigma, \mu + k\sigma)$  не зависит от  $\mu$  и  $\sigma$  и равна приблизительно 0,68 ( $k = 1$ ); 0,95 ( $k = 1,96$ ); 0,99 ( $k = 2,58$ ),

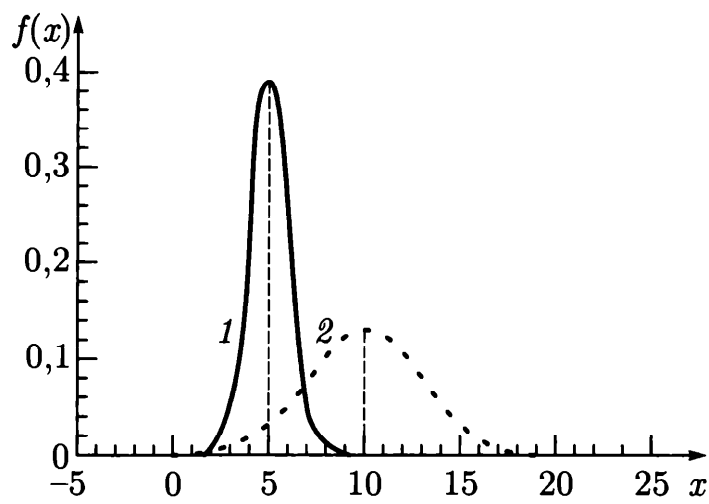


Рис. 1.10. Функция плотности нормального распределения:

$$1 - \mu = 5, \sigma = 1; 2 - \mu = 10, \sigma = 3$$

т. е. вероятности нормально распределенной случайной величины принять значение в интервале  $(\mu - k\sigma, \mu + k\sigma)$  равны

$$\begin{aligned} P(\mu - \sigma \leq \xi \leq \mu + \sigma) &\approx 0,68; \\ P(\mu - 1,96\sigma \leq \xi \leq \mu + 1,96\sigma) &\approx 0,95; \\ P(\mu - 2,58\sigma \leq \xi \leq \mu + 2,58\sigma) &\approx 0,99. \end{aligned}$$

Если  $\xi \sim N(\mu, \sigma^2)$ , то можно доказать, что случайная величина  $a + b\xi \sim N(a + b\mu, b^2\sigma^2)$ , где  $a$  и  $b$  — константы. В частности, случайная величина

$$\frac{\xi - \mu}{\sigma} \sim N(0, 1).$$

Это распределение называется **стандартным нормальным распределением**. Функция плотности стандартного нормального распределения имеет вид

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

Многие наблюдаемые в реальных экспериментах случайные величины подчиняются приблизительно нормальному закону распределения. По этой причине значительная часть классической статистической теории предполагает нормальность распределения рассматриваемой случайной величины.

Теоретическое основание для этого предположения дает **центральная предельная теорема**, согласно которой распределение суммы независимых случайных величин (точное определение независимости случайных величин будет дано в 1.3),

ни одна из которых не доминирует, при увеличении числа слагаемых сходится к нормальному распределению (более точную формулировку теоремы можно найти, например, в книге [В. Н. Тутубалин, 1992]). В частности, условия центральной предельной теоремы выполняются, если слагаемые одинаково распределены. Например, распределение биномиальной случайной величины, которую можно рассматривать как сумму  $n$  независимых случайных величин, распределенных по Бернулли, при увеличении  $n$  будет приближаться к нормальному распределению с параметрами  $np$  и  $np(1 - p)$  (*теорема Муавра — Лапласа*).

В приложениях статистики часто используют связанные с нормальным распределением: распределение  $\chi^2$  (хи-квадрат), распределение Стьюдента (часто применяют обозначение  $t$ -распределение) и  $F$ -распределение. Плотности этих распределений выражаются через стандартные математические функции в виде некоторых (довольно громозких) формул, которые и служат определением распределений. Здесь приведем «конструктивное» определение, раскрывающее возможности использования этих распределений в математической статистике.

Предположим, что каждая из  $n$  независимых случайных величин  $\xi_1, \xi_2, \dots, \xi_n$  (определение независимости случайных величин см. в 1.3) распределена нормально с параметрами 0 и 1 ( $\xi_i \sim N(0, 1)$ ). Тогда случайная величина

$$\eta = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2$$

имеет *распределение  $\chi^2$  (хи-квадрат)* с  $n$  степенями свободы (записывается  $\eta \sim \chi_n^2$ ).

График функции плотности распределения  $\chi^2$  изображен на рис. 1.11 в двух вариантах:  $n = 5$  и  $n = 10$ .

Предположим теперь, что каждая из  $n + 1$  независимых случайных величин  $\xi_0, \xi_1, \xi_2, \dots, \xi_n$  распределена нормально с параметрами 0 и  $\sigma^2$  ( $\xi_i \sim N(0, \sigma^2)$ ). Тогда случайная величина

$$\eta = \frac{\xi_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n \xi_i^2}}$$

имеет *распределение Стьюдента* с  $n$  степенями свободы (записывается  $\eta \sim t_n$ ).

График функции плотности  $t$ -распределения приведен на рис. 1.12 в двух вариантах:  $n = 2$  и  $n = 10$ . Заметим, что мак-

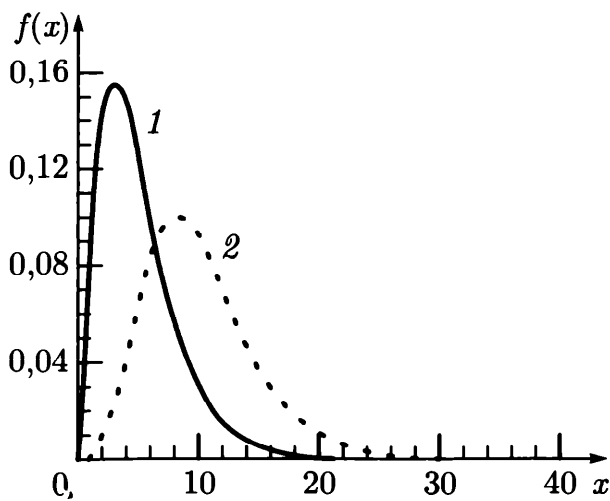


Рис. 1.11. Функция плотности  $\chi^2$ -распределения:

1 —  $n = 5$ ; 2 —  $n = 10$

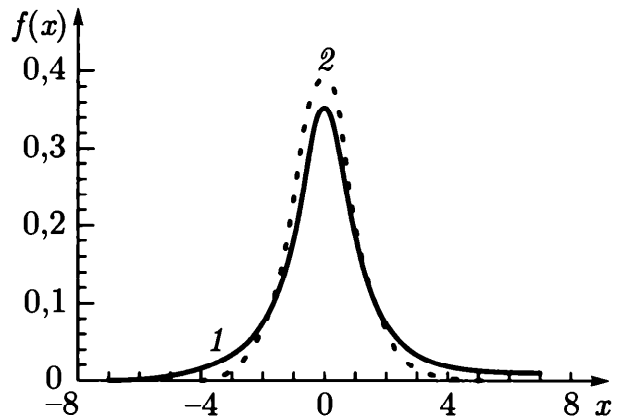


Рис. 1.12. Функция плотности  $t$ -распределения:

1 —  $n = 2$ ; 2 —  $n = 10$

симметричное значение плотности распределения достигается при  $x = 0$  и график симметричен относительно оси ординат.

Предположим теперь, что каждая из  $n + m$  независимых случайных величин  $\xi_1, \xi_2, \dots, \xi_n, \xi_{n+1}, \dots, \xi_{n+m}$  распределена нормально с параметрами 0 и  $\sigma^2$ , т. е.  $\xi_i \sim N(0, \sigma^2)$ . Тогда случайная величина

$$\eta = \frac{\frac{1}{n} \sum_{i=1}^n \xi_i^2}{\frac{1}{m} \sum_{i=n+1}^{n+m} \xi_i^2}$$

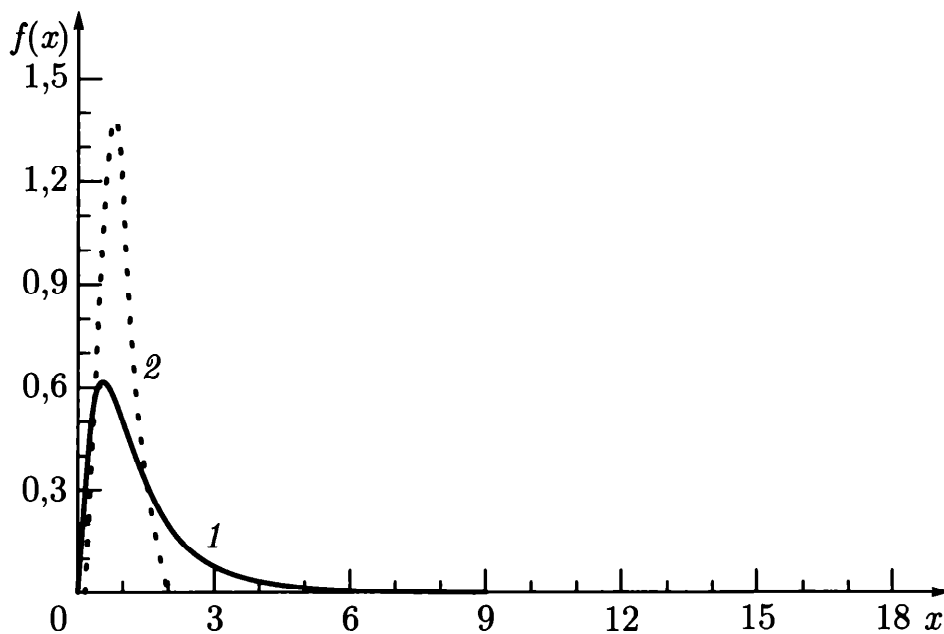


Рис. 1.13. Функция плотности  $F$ -распределения:

1 —  $n = 5, m = 5$ ; 2 —  $n = 50, m = 50$

имеет *F-распределение* с  $n$  и  $m$  степенями свободы (записывается  $\eta \sim F_{n,m}$ ).

График функции плотности *F-распределения* представлен на рис. 1.13 в двух вариантах:  $n = 5, m = 5$  и  $n = 50, m = 50$ .

Иногда возникает ситуация, когда рассматриваемую случайную величину удастся с помощью несложных изменений преобразовать в нормально распределенную случайную величину. Например, в биологических исследованиях часто возникает логнормальное распределение.

Непрерывная случайная величина  $\xi$  называется распределенной по *логнормальному закону* с параметрами  $\mu$  и  $\sigma^2$ , если случайная величина  $\eta = \ln \xi$  распределена нормально ( $\eta \sim N(\mu, \sigma^2)$ ).

### 1.2.5. Характеристики распределений

Полная информация о случайной величине дается ее распределением вероятностей — дискретным распределением в дискретном случае, функцией плотности — в непрерывном, либо функцией распределения в любом из этих двух случаев.

Для решения многих задач достаточно знать лишь некоторые числовые характеристики, называемые характеристиками распределения случайной величины, важнейшими среди которых являются: математическое ожидание, дисперсия, среднее квадратичное отклонение, коэффициент вариации, моменты, центральные моменты, коэффициент асимметрии, коэффициент эксцесса, медиана, мода, первая квартиль, третья квартиль, интерквартильный размах, квантили.

*Математическим ожиданием* случайной величины  $\xi$  (применяют также термины «среднее», «генеральное среднее») называется число  $M(\xi)$  (другие обозначения:  $E(\xi), m_\xi, m, \mu$ ), равное

$$M(\xi) = \sum_i x_i p_i$$

— для дискретной случайной величины и

$$M(\xi) = \int_{-\infty}^{+\infty} x f(x) dx$$

— для непрерывной случайной величины.

Математическое ожидание  $M(\xi)$  является характеристикой положения центра распределения, или мерой центральной тен-

## Основные характеристики некоторых теоретических распределений

Распределение	Формула для вычисления	
	математического ожидания	дисперсии
Дискретное равномерное (множество возможных значений: $x_1, x_2, \dots, x_n$ )	$\frac{1}{n} \sum_{i=1}^n x_i$	$\frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2$
Бернулли с параметром $p$	$p$	$p(1-p)$
Биномиальное с параметрами $n$ и $p$	$np$	$np(1-p)$
Пуассона с параметром $\lambda$	$\lambda$	$\lambda$
Равномерное на отрезке $[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Нормальное с параметрами $\mu$ и $\sigma^2$	$\mu$	$\sigma^2$
$\chi^2$ с $n$ степенями свободы	$n$	$2n$
$t$ -распределение с $n$ степенями свободы	$0$	$\frac{n}{n-2}$
$F$ -распределение с $n$ и $m$ степенями свободы	$\frac{m}{m-2}$	$\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$
Экспоненциальное с параметром $Q$	$1/Q$	$1/Q^2$
Логнормальное с параметрами $\mu, \sigma^2$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

денции, точнее, средним (взвешенным по вероятностям) значений случайной величины.

В табл. 1.1 приведены значения математических ожиданий для некоторых распределений. Читателю советуем хотя бы для двух распределений, одного дискретного и одного непрерывного вычислить математические ожидания самостоятельно.

### Свойства математического ожидания

1. Если  $\varphi(\xi)$  — некоторая функция от случайной величины, то при довольно общих ограничениях ее математическое ожидание вычисляется по формулам:



$$M(\varphi(\xi)) = \sum_i \varphi(x_i) p_i$$

— для дискретной случайной величины и

$$M(\varphi(\xi)) = \int_{-\infty}^{+\infty} \varphi(x) f(x) dx$$

— для непрерывной случайной величины.

2.  $M(a\xi + b) = aM(\xi) + b$ , где  $a$  и  $b$  — константы.

Действительно, используя свойство 1, определение случайной величины и свойства суммы, получаем для дискретной случайной величины

$$M(a\xi + b) = \sum_{i=1}^k (ax_i + b)p_i = a \sum_{i=1}^k x_i p_i + b \sum_{i=1}^k p_i = aM(\xi) + b.$$

(Для непрерывной случайной величины доказательство аналогично.) В частности,  $M(a\xi) = aM(\xi)$  и  $M(b) = b$ , т. е. константу можно выносить за знак математического ожидания и математическое ожидание константы равно этой константе.

3. Для любых двух случайных величин  $\xi$  и  $\eta$

$$M(\xi + \eta) = M(\xi) + M(\eta).$$

Приведем здесь и в следующем пункте доказательство только для дискретной случайной величины. Обозначим через  $p_{ij} = P(\xi = x_i, \eta = y_j)$ ,  $p_{i.} = P(\xi = x_i)$ ,  $p_{.j} = P(\eta = y_j)$ , тогда

$$\sum_{j=1}^l p_{ij} = p_{i.}, \quad \sum_{i=1}^k p_{ij} = p_{.j}.$$

$$\begin{aligned} M(\xi + \eta) &= \sum_{i=1}^k \sum_{j=1}^l (x_i + y_j) p_{ij} = \sum_{i=1}^k x_i \sum_{j=1}^l p_{ij} + \sum_{j=1}^l y_j \sum_{i=1}^k p_{ij} = \\ &= \sum_{i=1}^k x_i p_{i.} + \sum_{j=1}^l y_j p_{.j} = M(\xi) + M(\eta). \end{aligned}$$

4. Если  $\xi$  и  $\eta$  — независимые случайные величины, то

$$M(\xi\eta) = M(\xi)M(\eta).$$

Действительно,

$$\begin{aligned}
M(\xi\eta) &= \sum_{i=1}^k \sum_{j=1}^l x_i y_j p_{ij} = \sum_{i=1}^k \sum_{j=1}^l x_i y_j p_{i \cdot} p_{\cdot j} = \\
&= \sum_{i=1}^k x_i p_{i \cdot} \sum_{j=1}^l y_j p_{\cdot j} = M(\xi)M(\eta)
\end{aligned}$$

(для дискретных независимых случайных величин, по определению,  $p_{ij} = p_{i \cdot} p_{\cdot j}$ ).

Для описания многих практически важных свойств случайной величины необходимо знание не только ее математического ожидания, но и отклонения возможных ее значений от среднего значения. Например, у случайных величин, принимающих значения  $\pm 1$  или  $\pm 10$  с вероятностями  $1/2$ , математические ожидания совпадают и равны нулю, но отклонения их значений от нуля разные.

Для измерения разброса значений случайной величины около среднего значения часто используют такие характеристики, как дисперсия, среднее квадратичное отклонение, коэффициент вариации.

**Дисперсией случайной величины** называется число  $D(\xi)$  (другие обозначения:  $\sigma^2(\xi)$ ,  $\sigma_{\xi}^2$ ), равное

$$D(\xi) = \sum_i (x_i - M(\xi))^2 p_i,$$

если  $\xi$  — дискретная случайная величина, и

$$D(\xi) = \int_{-\infty}^{+\infty} (x - M(\xi))^2 f(x) dx,$$

если  $\xi$  — непрерывная случайная величина.

Дисперсия характеризует средний квадрат отклонения случайной величины от своего математического ожидания.

В табл. 1.1 приведены формулы для вычисления дисперсий некоторых распределений. Читателю советуем, хотя бы для двух распределений — одного дискретного и одного непрерывного — вычислить дисперсии самостоятельно.

### **Свойства дисперсии**

1.  $D(\xi) = M[\xi - M(\xi)]^2 = M(\xi^2) - [M(\xi)]^2$ .

Действительно, пользуясь свойствами математического ожидания, получаем

$$M[\xi - M(\xi)]^2 = M(\xi^2) - 2[M(\xi)]^2 + [M(\xi)]^2 = M(\xi^2) - [M(\xi)]^2.$$

2.  $D(\xi) \geq 0$ , поскольку по определению дисперсия является суммой неотрицательных слагаемых или интегралом от неотрицательной функции.

3.  $D(a\xi + b) = a^2D(\xi)$ , где  $a$  и  $b$  — константы.

Действительно,

$$\begin{aligned} D(a\xi + b) &= M[a\xi + b - M(a\xi + b)]^2 = M[a\xi + b - aM(\xi) - b]^2 = \\ &= M[a\xi - aM(\xi)]^2 = a^2M[\xi - M(\xi)]^2 = a^2D(\xi). \end{aligned}$$

В частности,  $D(a\xi) = a^2D(\xi)$  и  $D(b) = 0$ , т. е. константу можно выносить за знак дисперсии в квадрате и дисперсия константы равна нулю.

4. Если  $\xi$  и  $\eta$  — независимые случайные величины, то

$$D(\xi \pm \eta) = D(\xi) + D(\eta).$$

Действительно,

$$\begin{aligned} D(\xi \pm \eta) &= M(\xi \pm \eta)^2 - [M(\xi \pm \eta)]^2 = \\ &= M(\xi)^2 \pm 2M(\xi\eta) + M(\eta)^2 - [M(\xi)]^2 \mp 2M(\xi\eta) - [M(\eta)]^2 = \\ &= M(\xi)^2 - [M(\xi)]^2 + M(\eta)^2 - [M(\eta)]^2 = D(\xi) + D(\eta). \end{aligned}$$

**Неравенство Чебышева** позволяет оценивать вероятность отклонения случайной величины  $\xi$  от своего математического ожидания при известной  $D(\xi)$ :

$$P(|\xi - M(\xi)| \geq \varepsilon) \leq D(\xi)/\varepsilon^2. \quad (1.6)$$

Приведем вывод формулы (1.6) для дискретной случайной величины. Пусть случайная величина  $\xi$  имеет конечную дисперсию  $D(\xi) = \sum_i [x_i - M(\xi)]^2 p_i$  и пусть  $\varepsilon$  — любое положительное число. Если в этой сумме выбросить все слагаемые, для которых  $|x_i - M(\xi)| < \varepsilon$ , а в оставшихся слагаемых, для которых  $|x_i - M(\xi)| \geq \varepsilon$ , заменить выражения  $[x_i - M(\xi)]^2$  их минимальным значением  $\varepsilon^2$ , то получим

$$D(\xi) \geq \varepsilon^2 \sum_{|x_i - M(\xi)| \geq \varepsilon} p_i.$$

Но

$$P(|\xi - M(\xi)| \geq \varepsilon) = \sum_{|x_i - M(\xi)| \geq \varepsilon} p_i,$$

откуда следует доказываемое неравенство.

С помощью неравенства Чебышева, в частности, легко доказывается сформулированная ранее теорема Бернулли. Действительно, пусть  $\xi = m/n$ , тогда  $M(\xi) = M(m/n) = np/n = p$  и  $D(\xi) = D(m/n) = npq/n^2 = pq/n \rightarrow 0$  при  $n \rightarrow \infty$ . Заменяя в неравенстве Чебышева  $M(\xi)$  на  $p$  и  $D(\xi)$  на  $pq/n$ , получим доказательство теоремы Бернулли.

**Средним квадратичным отклонением** случайной величины (используются также термины «среднеквадратическое отклонение» и «стандартное отклонение») называется корень квадратный из дисперсии

$$\sigma(\xi) = \sqrt{D(\xi)}.$$

Среднее квадратичное отклонение, следовательно, является, как и дисперсия, мерой рассеяния распределения, но измеряется, в отличие от дисперсии, в тех же единицах, которые используют для измерения значений случайной величины.

В дальнейшем часто будет использовано преобразование **стандартизации** случайной величины  $\xi$ :

$$\eta = \frac{\xi - M(\xi)}{\sigma(\xi)}.$$

Докажите (с использованием свойств математического ожидания и дисперсии), что  $M(\eta) = 0$ , а  $D(\eta) = 1$ .

**Коэффициентом вариации** случайной величины  $\xi$  называется число  $V(\xi)$ , равное, если  $M(\xi) > 0$ ,

$$V(\xi) = \frac{\sigma(\xi)}{M(\xi)},$$

Таким образом, коэффициент вариации является, как и дисперсия, и среднее квадратичное отклонение, мерой рассеяния распределения, но служит для измерения среднего квадратичного отклонения в долях математического ожидания.

**Моментом порядка  $\nu$**  (иногда применяют термин «**начальный момент порядка  $\nu$** ») случайной величины называется число  $\mu_\nu$ , равное

$$\mu_\nu = M(\xi^\nu).$$

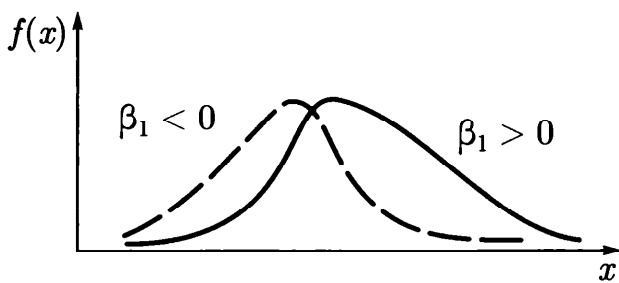


Рис. 1.14. Пример распределений с положительной ( $\beta_1 > 0$ ) и отрицательной ( $\beta_1 < 0$ ) асимметрией

**Центральным моментом порядка  $\nu$**  случайной величины  $\xi$  называется число  $\mu_{\nu,0}$ , равное

$$\mu_{\nu,0} = M[\xi - M(\xi)]^\nu.$$

Таким образом, математическое ожидание случайной величины является моментом первого порядка, а дисперсия — центральным моментом второго порядка. Термин «момент» заимствован из теоретической механики. Известно, что первый момент  $\mu_{1,0}$ , т. е.  $M(\xi)$ , — это абсцисса центра тяжести массы распределения, а второй центральный момент  $\mu_{2,0}$ , т. е.  $D(\xi)$ , — момент инерции массы относительно перпендикулярной оси, проходящей через центр тяжести  $x = M(\xi)$ .

Кроме этих моментов наиболее часто используются третий и четвертый центральные моменты.

**Коэффициентом асимметрии** случайной величины  $\xi$  называется величина

$$\beta_1 = \frac{\mu_{3,0}}{[D(\xi)]^{3/2}};$$

**коэффициентом эксцесса** случайной величины  $\xi$  — величина

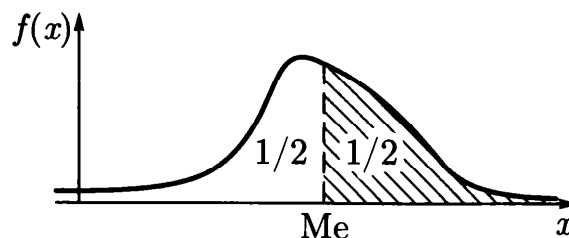
$$\beta_2 = \frac{\mu_{4,0}}{[D(\xi)]^2} - 3.$$

Если плотность распределения случайной величины симметрична, то коэффициент асимметрии  $\beta_1 = 0$ . На рис. 1.14 приведены графики функций плотности в двух случаях:  $\beta_1 > 0$ ,  $\beta_1 < 0$ . Для нормального распределения  $\beta_2 = 0$ . Если распределение сосредоточено вокруг среднего теснее, чем нормальное, то  $\beta_2 < 0$ , в противном случае —  $\beta_2 > 0$ .

Как отмечалось, математическое ожидание является характеристикой положения центра распределения. Другими характеристиками положения являются медиана и мода.

**Медианой** случайной величины  $\xi$  называется число  $Q_2$ , такое, что  $P(\xi \leq Q_2) \geq 1/2$  и  $P(\xi \geq Q_2) \geq 1/2$ . Другое распространенное обозначение  $Me$ .

Рис. 1.15. Медиана распределения



Если  $\xi$  — непрерывная случайная величина, то определение медианы можно проиллюстрировать графиком функции плотности (рис. 1.15).

Если распределение случайной величины симметрично, как, например, в случае нормального распределения, то медиана совпадает с математическим ожиданием. Однако для несимметричных распределений математическое ожидание и медиана, вообще говоря, не совпадают.

**Модой** ( $Mo$ ) непрерывной случайной величины называется такое значение  $x$ , в котором  $f(x)$  достигает своего локального максимума.

Мода есть «центр сгущения» случайной величины в смысле наиболее часто встречающихся значений. Распределение с одной модой называется унимодальным, с несколькими модами — мультимодальным. Для симметричного унимодального распределения мода совпадает с математическим ожиданием, а следовательно, и с медианой.

Для несимметричных распределений математическое ожидание, медиана и мода, вообще говоря, не совпадают. На рис. 1.16 для несимметричного унимодального распределения показаны все три характеристики положения распределения.

Можно ввести еще две характеристики распределения, определяемые аналогично медиане, — первую и третью квартили.

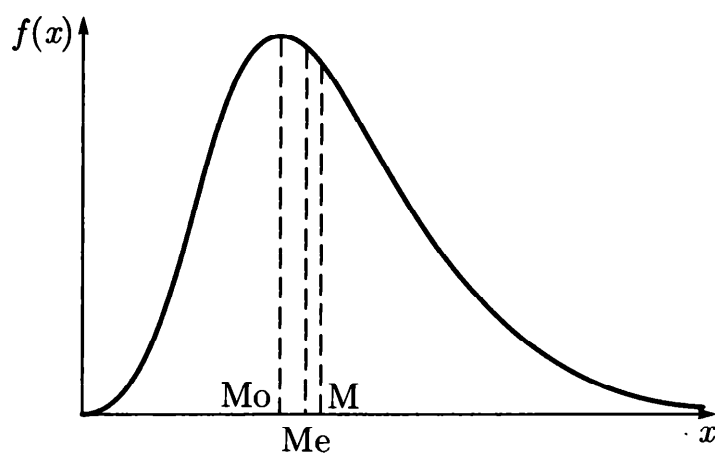
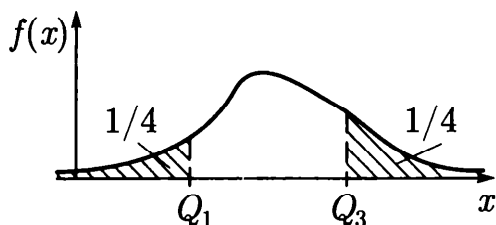


Рис. 1.16. Мода ( $Mo$ ), медиана ( $Me$ ) и математическое ожидание ( $M$ ) случайной величины  $\xi$



**Первой квартилью** (часто используется термин «**нижняя квартиль**») распределения случайной величины  $\xi$  называется число  $Q_1$ , такое, что

Рис. 1.17. Первая и третья квартили распределения

$$P(\xi \leq Q_1) = 1/4,$$

**третьей квартилью** (**верхней квартилью**) распределения случайной величины  $\xi$  называется число  $Q_3$ , такое, что

$$P(\xi > Q_3) = 1/4.$$

Если  $\xi$  — непрерывная случайная величина, то определение квартили удобно иллюстрировать графически (рис. 1.17).

Вероятность того, что случайная величина примет значение в промежутке  $[Q_1, Q_3]$ , равна  $P(Q_1 < \xi \leq Q_3) = 1/2$ .

Разность  $Q_3 - Q_1$  называется **интерквартильным размахом** и может служить, аналогично среднеквадратическому отклонению, мерой рассеяния значений случайной величины.

**Квантилью порядка  $p$**  распределения  $F(x)$  называется число  $C_p$ , такое, что  $F(C_p) = p$ .

Для непрерывной случайной величины это определение иллюстрирует рис. 1.18.

Из определения квантили следует, что медиана — это квантиль порядка 0,5, первая квартиль — квантиль порядка 0,25, а третья квартиль — квантиль порядка 0,75.

Для некоторых наиболее распространенных в математической статистике распределений созданы таблицы квантилей.

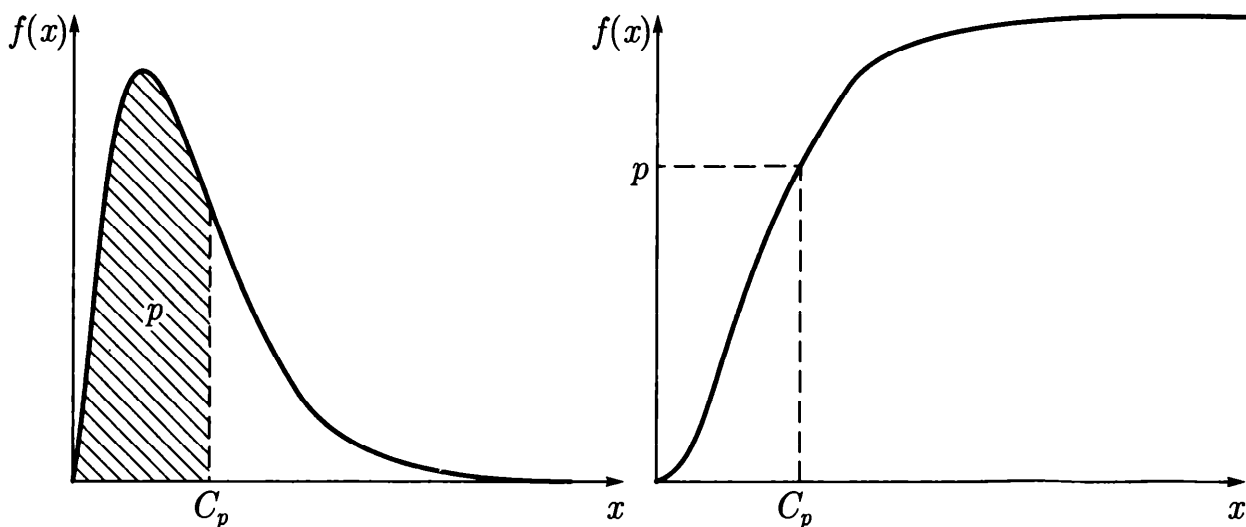


Рис. 1.18. Графическая иллюстрация определения квантили

В пакетах прикладных программ по статистической обработке вместо термина «квантиль» используется термин «*процентиль*», когда порядок квантили выражается в процентах.

### Задачи для самостоятельного решения

1. Вычислите математическое ожидание и дисперсию для распределения Бернулли с параметром  $p$ .
2. Вычислите математическое ожидание и дисперсию для биномиального распределения с параметрами  $n$  и  $p$ .
3. Вычислите математическое ожидание для пуассоновского распределения.
4. Станок может сломаться при изготовлении детали с вероятностью  $p$ . Число изготовленных деталей до остановки — случайная величина. Найдите закон ее распределения и математическое ожидание.
5. Докажите, что дисперсия числа появлений успеха при одном испытании не превосходит  $0,25$ .
6. В некоторой партии зерна  $\frac{1}{3}$  зерен невсхожие. Наугад выбираем 4 зерна. Случайная величина — число всхожих среди них зерен. Найдите закон ее распределения, математическое ожидание и дисперсию.
7. Каково должно быть математическое ожидание числа бактерий в единичной пробе, чтобы вероятность того, что в пробе имеется хотя бы одна бактерия, была не меньше  $\frac{1}{2}$ ?
8. Имеется пять проб воздуха единичного объема. Математическое ожидание числа бактерий в каждой пробе равно 2. Найдите вероятность того, что, по крайней мере, в одной пробе имеется не менее двух бактерий.

### 1.3. Многомерные случайные величины

На каждом пространстве элементарных событий  $\Omega$  можно определить несколько случайных величин.

**Пример 1.9.** Рассмотрим, как и в примере 1.7, пространство элементарных событий, которое получается в результате независимых бросаний двух монет. Пространство элементарных событий состоит из четырех элементарных событий  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  — «герб» + «герб», «герб» + «решка», «решка» + «герб», «решка» + «решка». Определим на этом пространстве случайную величину  $\xi_1$ , равную числу выпавших гербов, и случайную величину  $\xi_2$ , равную числу выпавших решек. Очевидно, что  $\xi_1(\omega_1) = 2$ ,  $\xi_1(\omega_2) = \xi_1(\omega_3) = 1$ ,  $\xi_1(\omega_4) = 0$  и случайная величина  $\xi_1$  принимает эти значения с вероятностями  $0,25$ ;  $0,5$ ;  $0,25$  соответственно. А  $\xi_2(\omega_1) = 0$ ,  $\xi_2(\omega_2) = \xi_2(\omega_3) = 1$ ,  $\xi_2(\omega_4) = 2$  и  $\xi_2$



принимает эти значения с вероятностями 0,25; 0,5; 0,25. Вектор  $\xi = \{\xi_1, \xi_2\}$  является двумерной случайной величиной.

**Пример 1.10.** Двумерной случайной величиной является вектор  $\vec{\xi} = \{\xi_1, \xi_2\}$ , где  $\xi_1$  и  $\xi_2$  — широта и долгота места падения метеорита на Землю. А четырехмерной случайной величиной является вектор  $\vec{\xi} = \{\xi_1, \xi_2, \xi_3, \xi_4\}$ , где  $\xi_1, \xi_2, \xi_3, \xi_4$  — длина и ширина чашелистика и лепестка цветка ириса.

Совокупность  $m$  функций, определенных на одном и том же множестве элементарных событий, называется ***m*-мерной случайной величиной**  $\xi = \{\xi_1, \dots, \xi_m\}$ . Многомерная случайная величина полностью определяется ее **функцией распределения** вероятностей

$$F(x_1, \dots, x_m) = P(\xi_1 \leq x_1, \dots, \xi_m \leq x_m),$$

удовлетворяющей следующим условиям:

- 1)  $0 \leq F(x_1, \dots, x_m) \leq 1$ ;
- 2)  $F(x_1, \dots, x_m)$  не убывает по каждому аргументу;
- 3)  $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_i, \dots, x_m) = 0$  для любого  $i$ ;
- 4)  $\lim_{x_i \rightarrow +\infty, i=1,2,\dots,m} F(x_1, \dots, x_m) = 1$ ;
- 5)  $\lim_{x_j \rightarrow \infty, j \neq i} F(x_1, \dots, x_i, \dots, x_m) = F_i(x_i)$ , где  $F_i(x_i)$  — функция

распределения одномерной случайной величины  $\xi_i$ .

Одномерные случайные величины  $\xi_1, \dots, \xi_m$  называются **независимыми**, если их совместная функция распределения равна произведению одномерных функций распределения

$$F(x_1, x_2, \dots, x_m) = F_1(x_1)F_2(x_2) \dots F_m(x_m).$$

### 1.3.1. Дискретные многомерные случайные величины

Многомерная случайная величина называется **дискретной**, если составляющие ее случайные величины являются дискретными. Многомерная дискретная случайная величина полностью определяется набором значений вероятностей

$$p_{i_1, \dots, i_m} = p(\xi_1 = x_{i_1}, \dots, \xi_m = x_{i_m}),$$

заданных для любой комбинации значений  $x_{i_1}, \dots, x_{i_m}$  случайных величин  $\xi_1, \dots, \xi_m$ . Функция распределения в этом случае выражается через вероятности  $p_{i_1 \dots i_m}$  следующей формулой:

$$F(x_1, \dots, x_m) = \sum_{x_{i_1} \leq x_1 \dots x_{i_m} \leq x_m} p_{i_1 \dots i_m}.$$

**Пример 1.11.** Двумерная случайная величина  $(\xi, \eta)$  принимает значения  $(x_i, y_j), i = 1, \dots, k; j = 1, \dots, l$ . Вероятности всех возможных пар значений  $(x_i, y_j)$  можно представить в виде таблицы:

$\xi$	$\eta$					Итог
	$y_1$		$y_j$		$y_l$	
$x_1$	$p_{11}$		$p_{1j}$		$p_{1l}$	$p_{1.}$
$x_i$	$p_{i1}$		$p_{ij}$		$p_{il}$	$p_{i.}$
$x_k$	$p_{k1}$		$p_{kj}$		$p_{kl}$	$p_{k.}$
Итог	$p_{.1}$		$p_{.j}$		$p_{.l}$	

В итоговых строке и столбце записаны суммы по столбцам и строкам

$$p_{.j} = \sum_{i=1}^k p_{ij}; p_{i.} = \sum_{j=1}^l p_{ij}.$$

Итоговый столбец определяет **одномерное (маргинальное) распределение случайной величины  $\xi$** , а итоговая строка — одномерное (маргинальное) распределение случайной величины  $\eta$ . Если разделить все вероятности  $j$ -го столбца на итоговую вероятность  $p_{.j}$ , то получим условные вероятности значений  $\xi$  при условии  $\eta = y_j$

$$p_{1j}/p_{.j}, \dots, p_{ij}/p_{.j}, \dots, p_{kj}/p_{.j},$$

определяющие **условное распределение случайной величины  $\xi$**  при фиксированном значении другой случайной величины  $\eta$ , равном  $y_j$ . Аналогично определяется условное распределение  $\eta$  при заданном значении  $\xi$ .

Если  $p_{ij} = p_{.j}p_{i.}$  для любых  $i, j (i \neq j)$ , то случайные величины  $\xi$  и  $\eta$  являются независимыми.

### 1.3.2. Непрерывные многомерные случайные величины

Многомерная случайная величина называется **непрерывной**, если непрерывна ее функция распределения  $F(x_1, \dots, x_m)$  и существует функция плотности  $f(x_1, \dots, x_m)$ , такая, что

$$F(x_1, \dots, x_m) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_m} f(x_1, \dots, x_m) dx_1, \dots, dx_m.$$

Рассмотрим непрерывную двумерную случайную величину  $(\xi, \eta)$  с плотностью  $f(x, y)$  и функцией распределения  $F(x, y)$ . **Одномерная (маргинальная) плотность**  $\xi$  получается путем интегрирования по  $y$  двумерной плотности

$$f_{\xi}(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

Аналогично определяется маргинальная плотность  $\eta$

$$f_{\eta}(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

**Условная плотность случайной величины**  $\xi$  при заданном значении случайной величины  $\eta = y$  задается формулой

$$f(x|y) = \frac{f(x, y)}{\int_{-\infty}^{+\infty} f(x, y) dx} = \frac{f(x, y)}{f_{\eta}(y)},$$

откуда получаем непрерывный аналог формулы умножения вероятности («совместная плотность равна произведению условной плотности на плотность условия»):

$$f(x, y) = f(x|y)f_{\eta}(y).$$

Аналогично условное распределение  $\eta$  задается формулой

$$f(y|x) = \frac{f(x, y)}{\int_{-\infty}^{+\infty} f(x, y) dy} = \frac{f(x, y)}{f_{\xi}(x)},$$

откуда

$$f(x, y) = f(y|x)f_{\xi}(x).$$

Непрерывные случайные величины будут независимыми, если

$$f(x, y) = f_{\xi}(x)f_{\eta}(y).$$

Характеристикой связи случайных величин  $\xi$  и  $\eta$  является коэффициент корреляции.

**Коэффициентом корреляции** двух случайных величин  $\xi$  и  $\eta$  называется отношение

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi)D(\eta)}},$$

где  $\text{cov}(\xi, \eta) = M[(\xi - M(\xi))(\eta - M(\eta))]$  — **ковариация**  $\xi$  и  $\eta$ . Если  $\rho(\xi, \eta) = 0$ , то случайные величины  $\xi$  и  $\eta$  называются некоррелированными.

### **Свойства коэффициента корреляции**

1. Для любых случайных величин  $\xi$  и  $\eta$   $|\rho(\xi, \eta)| \leq 1$ .
2. Если  $\xi$  и  $\eta$  — независимые случайные величины, то  $\rho(\xi, \eta) = 0$  (обратное, вообще говоря, неверно).
3.  $\rho(\xi, \eta) = +1$  или  $-1$  тогда и только тогда, когда случайные величины  $\xi$  и  $\eta$  связаны линейной зависимостью  $\xi = a\eta + b$ .

**Двумерным нормальным распределением** называют распределение вероятностей двумерной непрерывной случайной величины  $(\xi, \eta)$ , если функция плотности имеет вид

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right] \right\},$$

где  $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$  — некоторые параметры,  $\sigma_1 > 0, \sigma_2 > 0, |\rho| \leq 1$ . Можно доказать, что маргинальные случайные величины  $\xi$  и  $\eta$ , соответствующие этому двумерному распределению, — нормально распределенные случайные величины, причем  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  — их математические ожидания и дисперсии, а  $\rho$  — коэффициент корреляции.

Определение нормального распределения можно обобщить и на случай более двух измерений. Здесь не будем приводить формулу плотности вероятности **многомерного нормального распределения** ввиду ее громоздкости, а отметим лишь, что в число его параметров входят наряду с математическими ожиданиями и дисперсиями одномерных случайных величин также и все их попарные корреляции.

Если составляющие двумерной нормально распределенной случайной величины  $\xi$  и  $\eta$  некоррелированы, то  $\rho(\xi, \eta) = 0$ . Тогда функция плотности  $\xi$  и  $\eta$  равна

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} = f_\xi(x) f_\eta(y).$$

Следовательно, если составляющие двумерной нормально распределенной случайной величины некоррелированы, то они и независимы. Таким образом, для составляющих двумерного нормального распределения понятия независимости и некоррелированности равносильны.

Пусть имеется  $n$  случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ . Рассмотрим новую случайную величину  $\eta = \varphi(\xi_1, \xi_2, \dots, \xi_n)$ , где  $\varphi$  — некоторая функция. Тогда функция распределения случайной величины  $\eta$  задается (в случае, когда  $\xi_i$  — непрерывные случайные величины) формулой

$$F(x) = P(\eta \leq x) = \int \dots \int_{D_x} f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n,$$

где  $f(x_1, x_2, \dots, x_n)$  — функция плотности  $n$ -мерной случайной величины  $(\xi_1, \dots, \xi_n)$ ;  $D_x$  — область  $n$ -мерного пространства, содержащая все точки  $x_1, x_2, \dots, x_n$ , такие, что  $\varphi(x_1, x_2, \dots, x_n) \leq x$ .

Математическое ожидание случайной величины  $\eta$  равно

$$M(\eta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \varphi(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Предположим теперь, что  $\eta = a_0 + a_1\xi_1 + \dots + a_n\xi_n$ . Тогда можно доказать, воспользовавшись свойствами математического ожидания и дисперсии, что

$$M(\eta) = a_0 + a_1M(\xi_1) + \dots + a_nM(\xi_n)$$

и

$$D(\eta) = \sum_{i=1}^n a_i^2 D(\xi_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{cov}(\xi_i, \xi_j).$$

Пусть теперь  $\xi_1, \xi_2, \dots, \xi_n$  независимы и одинаково распределены. Обозначим  $M(\xi_i) = \mu$ ;  $D(\xi_i) = \sigma^2$ ;  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ , тогда

$$M(\bar{\xi}) = \mu; D(\bar{\xi}) = \frac{\sigma^2}{n}.$$

Если дополнительно предположить, что все случайные величины  $\xi_1, \xi_2, \dots, \xi_n$  нормально распределены, т. е.  $\xi_i \sim N(\mu, \sigma^2)$ ,  $i = 1, 2, \dots, n$ , то можно доказать, что случайная величина  $\eta = a_0 + a_1\xi_1 + \dots + a_n\xi_n$  также нормально распределена. Отсюда, в частности, следует, что  $\bar{\xi} \sim N(\mu, \sigma^2/n)$ .

## ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Задача *математической статистики*, в строгом понимании этого термина, состоит в разработке и применении методов описания реальных явлений вероятностными моделями, исходя из данных, полученных в результате наблюдений за этими явлениями.

В более широком смысле математическая статистика понимается как совокупность методов планирования экспериментов и обработки данных, полученных в результате экспериментов или наблюдений, причем эти методы могут не основываться на вероятностных моделях. При таком широком понимании вместо термина «математическая статистика» часто используют термин «*анализ данных*».

Исторически вначале сформировались методы обработки данных, не связанные тесно с теорией вероятности, так называемая *дескриптивная*, или *описательная, статистика*. С начала XX в. начали интенсивно развиваться методы анализа данных, основанные на вероятностных моделях, прежде всего, методы *статистического оценивания* и *статистической проверки гипотез*, о которых в основном и будет идти речь в данной главе.

Развитие вычислительной техники привело к появлению новых методов анализа. Некоторые из них разработаны на основе подходов, отличных от теоретико-вероятностного, — геометрические, оптимизационные и др. Вероятностное обоснование этих методов либо отсутствует, либо недостаточно развито, что затрудняет количественную оценку степени достоверности получаемых выводов аналитическими средствами классической математической статистики. Однако в последние годы, в связи с быстрым ростом производительности ЭВМ, получили распространение процедуры так называемого *случайного моделирования* (*пермутационные методы, бут-стрэп*), позволяющие оценить статистические свойства получаемых решений без использования аналитических методов.

Начнем рассмотрение методов математической статистики с ее исходного понятия — случайной выборки.

## 2.1. Случайная выборка и ее описание

Понятие *случайной выборки* тесно связано с понятиями случайного испытания и случайной величины, которые были рассмотрены в гл. 1. Случайная выборка представляет собой совокупность наблюдений  $x_1, x_2, \dots, x_n$  случайной величины  $\xi$ , полученных в  $n$  независимых случайных испытаниях. Число полученных наблюдений  $n$  называется *объемом выборки*. Образно можно представить процесс получения случайной выборки как извлечение наудачу значений из гипотетической бесконечной *генеральной совокупности*, где разные значения содержатся в пропорциях, соответствующих распределению случайной величины, и тщательно перемешаны между собой. Случайная величина может быть не только одномерной, но и многомерной, например, когда у случайно выбранного животного измеряется несколько характеристик: размер, масса и т. д. В этом случае каждая из компонент выборки будет вектором.

Конкретная случайная выборка — это просто набор значений случайной величины  $\xi$  (во избежание чрезмерного усложнения обозначений будем считать ее одномерной). Однако при оценке информации, которую несет эта выборка, ее необходимо рассматривать как одну из бесконечного числа потенциально возможных выборок объема  $n$ , т. е. как векторную  $n$ -мерную случайную величину

$$\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n),$$

компонентами которой являются  $n$  элементов выборки. Из условий получения выборки следует, что случайные величины  $\xi_1, \xi_2, \dots, \xi_n$  имеют одинаковые функции распределения  $F(x_1), F(x_2), \dots, F(x_n)$ , совпадающие с функцией распределения  $F(x)$  исходной случайной величины  $\xi$ . Кроме того, случайные величины  $\xi_1, \xi_2, \dots, \xi_n$  по определению случайной выборки независимы, поэтому их совместная функция распределения равна произведению одномерных функций распределения

$$G(x_1, x_2, \dots, x_n) = F(x_1)F(x_2) \dots F(x_n).$$

Как правило, информация, содержащаяся в выборке, интересует нас не столько сама по себе, сколько как информация о всей генеральной совокупности. Для этого выборка должна



быть правильно организованной и представительной. Существует специальный раздел математической статистики — *планирование выборочных обследований*. Здесь рассмотрим только один способ получения выборки — *простой случайный выбор*. Суть такого выбора состоит в извлечении наудачу из тщательно перемешанной генеральной совокупности  $n$  значений. На практике дело обстоит сложнее. Предположим, что был проведен отлов  $n$  взрослых животных определенного вида и измерена их масса. Какую генеральную совокупность представляет эта выборка — всех животных данного вида, живущих в разных местах, прошлых и будущих, или популяцию, обитающую в данной местности, или популяцию, обитающую в данной местности в данный год? Конечно, чем уже понимать генеральную совокупность, описываемую данной выборкой, тем ближе истина, но, возможно, тем меньший интерес она будет представлять.

Итак, имеется случайная выборка  $x_1, x_2, \dots, x_n$  значений случайной величины  $\xi$  с неизвестным распределением  $F(x)$ . Как разумно распорядиться этими значениями, чтобы получить представление о распределении  $F(x)$ , т. е. о генеральной совокупности, из которой извлечена эта выборка?

Можно использовать следующий эвристический принцип: будем считать, что исследуемая генеральная совокупность близка к гипотетической генеральной совокупности, состоящей только из значений  $x_1, x_2, \dots, x_n$ , содержащихся в ней в равной пропорции (значения  $x_1, x_2, \dots, x_n$  предполагаются упорядоченными по возрастанию, т. е. представлены в виде *вариационного ряда*), т. е. случайная величина  $\xi$  близка к случайной величине  $\hat{\xi}$ , принимающей  $n$  значений  $x_1, x_2, \dots, x_n$  с вероятностями  $1/n$  (это, действительно, максимум информации о значениях случайной величины и их вероятностях, которую можно извлечь из выборки). Распределение случайной величины  $\hat{\xi}$  называется эмпирическим распределением случайной величины  $\xi$ , а ее функция распределения  $\hat{F}(x)$  — *эмпирической функцией распределения*. Очевидно, что каждой выборке соответствует своя эмпирическая функция распределения, т. е. можно сказать, что  $\hat{F}(x)$  — случайная функция.  $\hat{F}(x)$  представляет собой ступенчатую функцию, возрастающую от 0 до 1 скачками высотой  $1/n$  в точках  $x_1, x_2, \dots, x_n$  (если некоторое значение повторяется  $k$  раз, то ему будет соответствовать один скачок величиной  $k/n$ ). Можно определить эмпирическую функцию формулой  $\hat{F}(x) = n_x/n$ , где  $n_x$  — число значений выборки, не превосходящих  $x$ .

**Пример 2.1.** Пусть случайная величина  $\xi$  — длина лепестка случайно выбранного цветка ириса разноцветного (*Iris versicolor*). Следующий ряд чисел содержит значения длин (в см) 50 случайно выбранных лепестков:

4,6 4,5 4,8 4,6 5,1 3,9 4,5 4,7 4,7 4,5  
 4,0 3,8 4,5 4,0 4,4 4,4 4,5 4,1 4,5 3,5  
 4,2 3,3 4,2 4,2 4,6 3,9 4,5 3,5 3,7 3,9  
 4,3 4,2 4,0 4,7 4,4 4,1 4,9 4,7 4,3 3,0  
 4,1 4,7 3,6 4,9 4,0 4,0 4,4 4,8 5,0 3,3

Представим эти значения в виде вариационного ряда:

3,0 3,3 3,3 3,5 3,5 3,6 3,7 3,8 3,9 3,9  
 3,9 4,0 4,0 4,0 4,0 4,0 4,1 4,1 4,1 4,2  
 4,2 4,2 4,2 4,3 4,3 4,4 4,4 4,4 4,4 4,5  
 4,5 4,5 4,5 4,5 4,5 4,5 4,6 4,6 4,6 4,7  
 4,7 4,7 4,7 4,7 4,8 4,8 4,9 4,9 5,0 5,1

На рис. 2.1 представлено построенное на основе этого вариационного ряда эмпирическое распределение случайной величины  $\xi$  для этой выборки, т. е. распределение случайной величины  $\hat{\xi}$ , а на рис. 2.2 — график эмпирической функции распределения  $\hat{F}(x)$ .

Поскольку эмпирическая функция распределения  $\hat{F}(x)$  является оценкой для  $F(x)$  (можно доказать, что при  $n \rightarrow \infty$  вероятность того, что максимальное расхождение между  $\hat{F}(x)$  и  $F(x)$

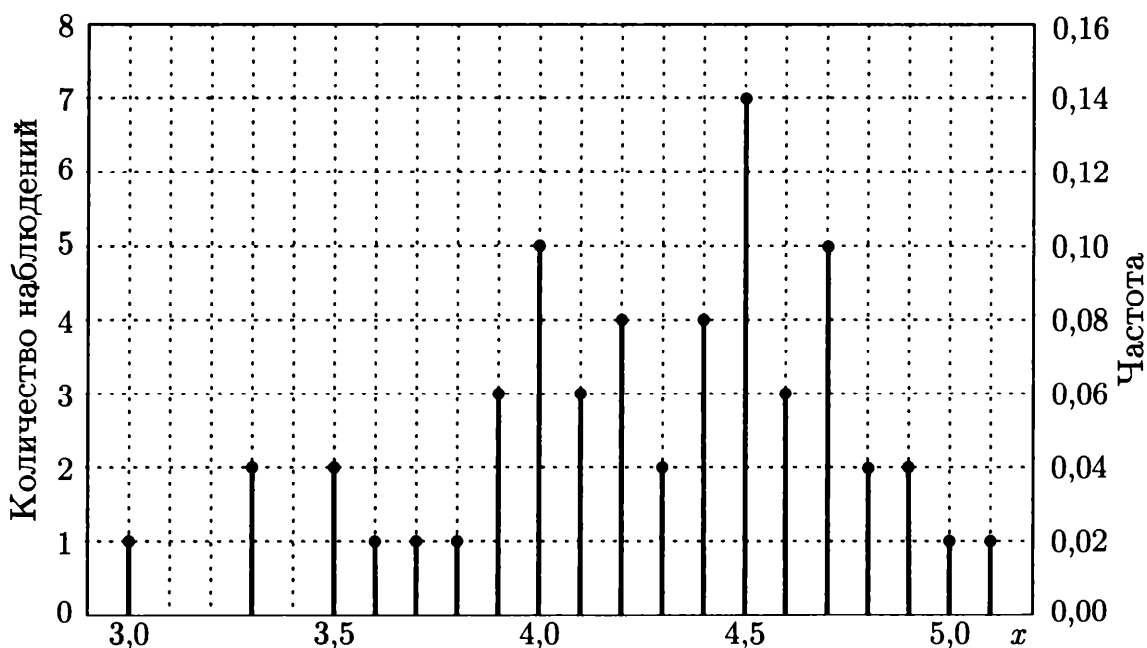


Рис. 2.1. График эмпирического распределения случайной величины

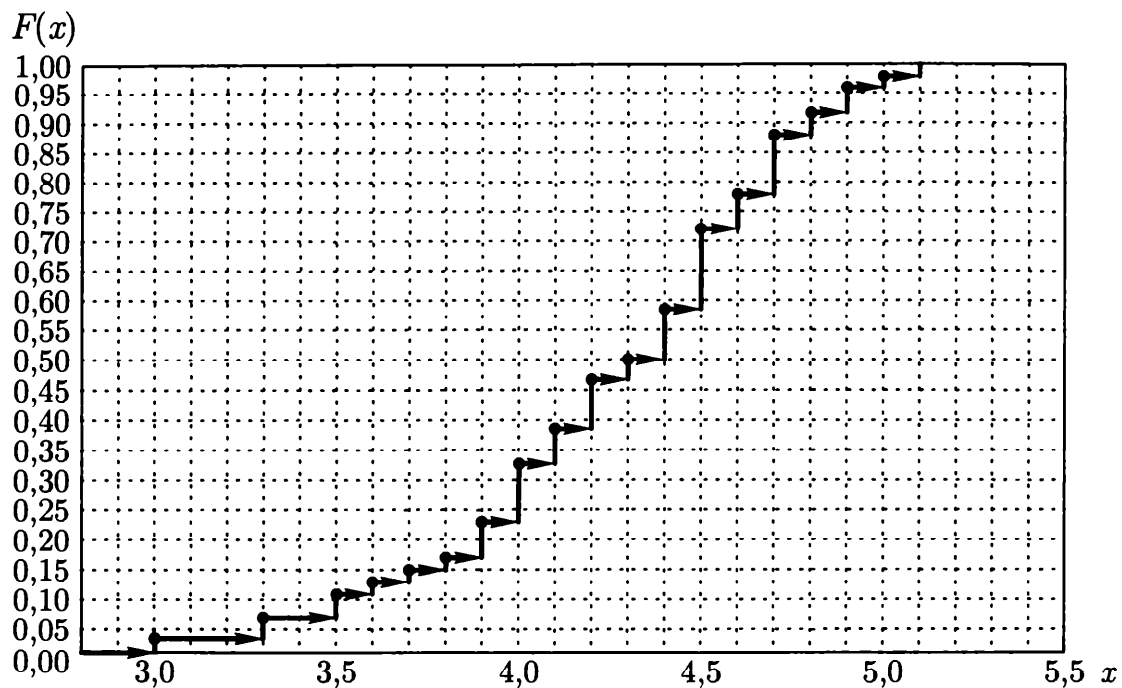


Рис. 2.2. График эмпирической функции распределения

не превзойдет заданного малого числа  $\xi$ , стремится к единице), можно взять характеристики  $\hat{F}(x)$  в качестве оценок характеристик генерального распределения.

В табл. 2.1 приведены полученные таким образом формулы для некоторых выборочных характеристик.

Таблица 2.1

**Формулы выборочных характеристик**

Характеристика	Формула
Выборочный момент порядка $k$	$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n x_i^k$
Выборочный центральный момент порядка $k$	$\hat{\mu}_{k,0} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$
Выборочное среднее — первый нецентральный момент	$\bar{x} = \hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n x_i$
Выборочная дисперсия — второй центральный момент	$s_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Выборочный коэффициент асимметрии	$\hat{\beta}_1 = \frac{\hat{\mu}_{3,0}}{\hat{\mu}_{2,0}^{3/2}}$
Выборочный коэффициент эксцесса	$\hat{\beta}_2 = \frac{\hat{\mu}_{4,0}}{\hat{\mu}_{2,0}^2}$

Зная эмпирическую функцию распределения, можно также найти эмпирические квантили, квартили и интерквартильную широту точно так же, как в случае обычной (теоретической) функции распределения. А именно, **выборочная квантиль**  $\hat{C}_p$  порядка  $p$  получается как абсцисса точки пересечения прямой  $y = p$  с графиком эмпирической функции распределения  $\hat{F}(x)$  (если пересечением является не точка, а отрезок, то в качестве квантили можно взять абсциссу середины этого отрезка). Выборочные квантили  $\hat{C}_{0,25}$ ,  $\hat{C}_{0,50}$ ,  $\hat{C}_{0,75}$  дают соответственно **выборочную нижнюю квартиль**  $\hat{Q}_1$ , **выборочную медиану**  $\hat{Q}_2$  и **выборочную верхнюю квартиль**  $\hat{Q}_3$ , а разность  $(\hat{C}_{0,75} - \hat{C}_{0,25})$  — **выборочную интерквартильную широту**. Еще одна характеристика разброса значений случайной величины — **размах**  $R$  — определяется как разность между максимальным и минимальным значением в выборке.

Если наблюдается не одномерная, а двумерная случайная величина  $(\xi, \eta)$ , т. е. выборка состоит из  $n$  пар значений  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $\dots$ ,  $(x_n, y_n)$ , то можно вычислить **выборочную ковариацию** для  $\xi$  и  $\eta$

$$\text{cov}(\xi, \eta) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

**Выборочный коэффициент корреляции** определяется формулой

$$r_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sqrt{s_{\xi 0}^2 s_{\eta 0}^2}},$$

где

$$s_{\xi 0}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad s_{\eta 0}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

**Выборочную моду**  $x_{\text{mod}}$  целесообразно оценивать для дискретного и непрерывного генерального распределения различными способами. В дискретном случае оценкой моды будет значение, встретившееся в выборке наибольшее число раз. В непрерывном случае повторяющиеся значения редки или вообще от-

сутствуют, поэтому следует разбить диапазон изменения наблюдаемых значений точками  $a_0, a_1, \dots, a_k$  на  $k$  равных промежутков  $[a_0, a_1], (a_1, a_2], \dots, [a_{k-1}, a_k]$  длиной  $h = (a_k - a_0)/k$  и в качестве выборочной моды взять середину интервала, в который попало наибольшее число значений.

В табл. 2.2 даны значения выборочных характеристик, вычисленные для примера данных о длинах лепестков ириса.

Результаты разбиения диапазона изменения значений выборки на  $k$  интервалов равной длины  $h$  и последующего подсчета числа значений  $n_i$ , попавших в интервалы  $i = 1, 2, \dots, k$ , можно представить графически. Построив над каждым интервалом  $i$  столбик высотой  $n_i$ , получим *гистограмму*. Если при построении гистограммы оперировать не числом значений, попавших в интервалы, а их относительной частотой в выборке  $v_i = \frac{n_i}{nh}$  и откладывать столбики высотой  $v_i$ , то полученная гистограмма будет *выборочным аналогом плотности* непрерывного распределения ( $n_i$  и  $v_i$  часто называются, соответственно, абсолютными и относительными частотами). В частности, сумма пло-

Таблица 2.2

### Значения выборочных характеристик

Характеристика	Значение
Выборочное среднее	$\bar{x} = 4,26$
Выборочная дисперсия (несмещенная оценка, см. 2.2.1)	$s^2 = \frac{n}{n-1} s_0^2 = 0,2164$
Выборочный коэффициент асимметрии	$\hat{\beta}_1 = -0,61$
Выборочный коэффициент эксцесса	$\hat{\beta}_2 = 0,05$
Минимум	$x_{\min} = 3,0$
Максимум	$x_{\max} = 5,1$
Размах	$R = 2,1$
Нижняя квартиль	$\hat{Q}_1 = 4,00$
Верхняя квартиль	$\hat{Q}_3 = 4,60$
Интерквартильный размах	$\hat{Q}_3 - \hat{Q}_1 = 0,60$
Медиана	$\hat{Q}_2 = 4,35$
Мода	$x_{\text{mod}} = 4,50$

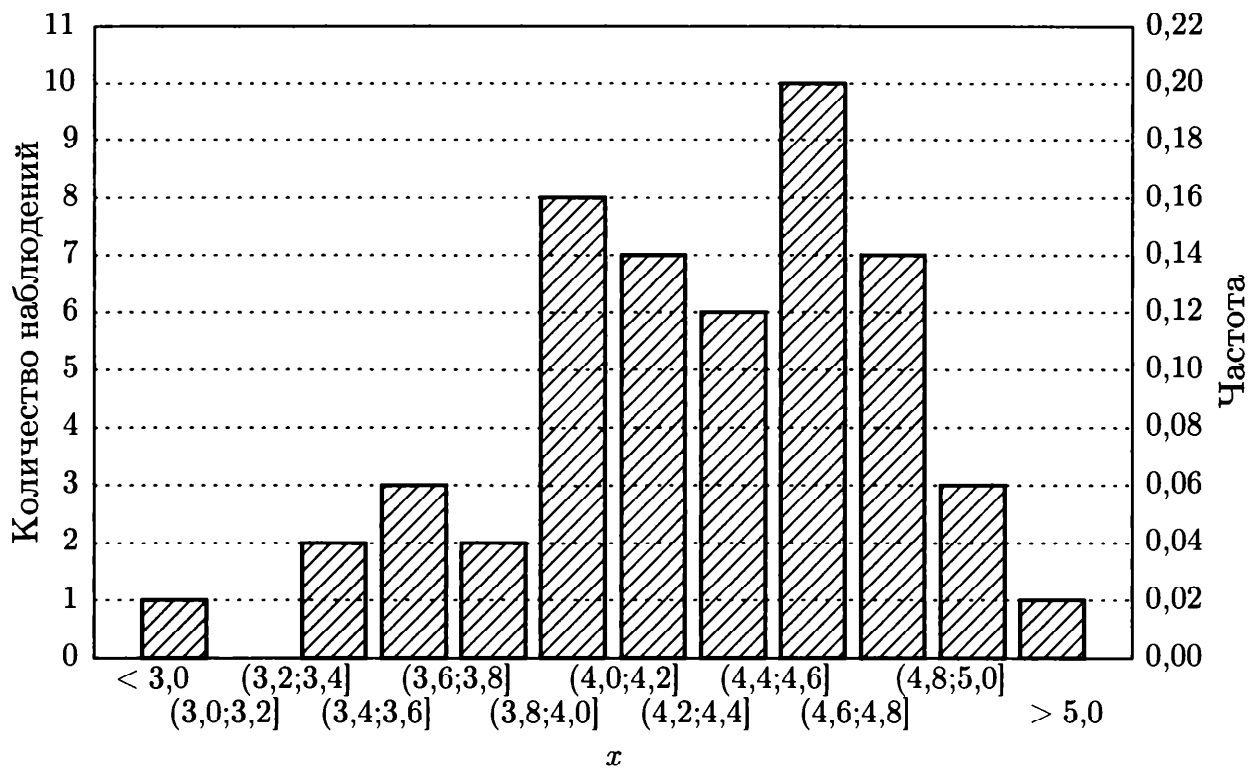


Рис. 2.3. Гистограмма

щадей всех столбиков гистограммы будет равна единице, как и для плотности распределения. Очевидно, что поскольку частоты  $\nu_i$  пропорциональны численностям  $n_i$ , гистограммы для **численностей** и **частот** различаются только масштабом шкалы по оси ординат. На рис. 2.3 приведен пример гистограммы для рассмотренной выше выборки, содержащей 50 значений длин лепестков ириса (заметим, что рис. 2.3 можно рассматривать как результат преобразования рис. 2.1 путем группировки значений на оси абсцисс).

При построении гистограммы возникает трудно формализуемая проблема выбора оптимальной длины  $h$  интервалов разбиения. Обычно число интервалов разбиения выбирают из расчета, чтобы в каждый интервал попало в среднем не менее десяти наблюдений (при очень малых объемах выборки это число уменьшают). Таким образом, при увеличении объема выборки  $n$  можно уменьшать длины интервалов разбиения и более детально характеризовать выборку, а следовательно, и порождающее ее генеральное распределение. По сравнению с эмпирической функцией распределения гистограмма более наглядна, однако при ее построении привносится элемент субъективизма. Гистограмму можно построить и на основе интервалов разбиения разной длины, но тогда процедура ее построения несколько усложняется и поэтому такие гистограммы используются довольно редко.

## 2.2. Статистическое оценивание

Допустим, имеется случайная выборка  $x_1, x_2, \dots, x_n$  значений некоторой случайной величины  $\xi$ , распределение которой может быть либо полностью неизвестным, либо частично известным, например, может быть известна его функциональная форма (вид функции распределения или плотности в непрерывном случае или формулы, определяющей вероятности отдельных значений, — в дискретном). Если вид распределения неизвестен, то нас могут интересовать различные его характеристики — математическое ожидание, мода, медиана, дисперсия, интерквартильная широта, моменты, асимметрия, эксцесс и т. д. Если же вид распределения известен, а неизвестны лишь значения определяющих его параметров, нас могут интересовать также (или даже прежде всего) значения этих параметров, например параметров  $\mu$  и  $\sigma^2$  — для нормального, параметра  $\lambda$  — для пуассоновского и параметра  $p$  — для биномиального распределений. Задача оценивания неизвестного параметра или характеристики  $\theta$  распределения случайной величины  $\xi$  состоит в вычислении на основе значений выборки  $x_1, x_2, \dots, x_n$  величины  $\hat{\theta}(x_1, x_2, \dots, x_n)$ , в каком-то смысле близкой к оцениваемому параметру или характеристике и называемой *точечной оценкой*  $\theta$ , поскольку за оценку неизвестного параметра  $\theta$  принимается конкретное значение (точка), вычисляемое по выборке [М. Кендалл, А. Стьюарт, 1973]. Точечные оценки будут рассмотрены в 2.2.1, интервальные — в 2.2.2.

### 2.2.1. Точечное статистическое оценивание

При рассмотрении выборочных характеристик случайной величины (математического ожидания, дисперсии и т. д.) как оценок характеристик неизвестных генеральных совокупностей фактически решалась задача оценивания. Однако такой подход был скорее интуитивным. Попробуем его формализовать, т. е. уточнить смысл близости *выборочной оценки*  $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$  к оцениваемому параметру  $\theta$ . Заметим, прежде всего, что  $\hat{\theta}$  является случайной величиной, значения которой меняются от одной выборки к другой. Для некоторых выборок значение  $\hat{\theta}$  может оказаться очень близким к оцениваемому параметру  $\theta$ , для других — далеким. Можно потребовать, чтобы хотя бы в среднем оценки для разных выборок группировались вокруг оцениваемого параметра, т. е. выполнялось условие  $M(\hat{\theta}) = \theta$ . Оценки,

удовлетворяющие этому условию, называются **несмещенными**. Отметим, что в зависимости от контекста  $\hat{\theta}(x_1, x_2, \dots, x_n)$  рассматривается либо как случайная величина (и только в этом случае можно говорить о математическом ожидании  $\hat{\theta}$  в определении несмещенности), либо как алгебраическая формула для вычисления значения оценки по выборке, либо просто как конкретное значение, полученное для конкретной выборки.

Можно доказать, пользуясь свойствами математического ожидания и определением выборки, что **выборочное математическое ожидание** (чаще используется термин «**выборочное среднее**»)  $\bar{x}$  является несмещенной оценкой истинного математического ожидания (генерального среднего)  $M(\xi)$ . Действительно, имеем

$$M\bar{x} = M\left\{\frac{1}{n}\sum_{i=1}^n x_i\right\} = \frac{1}{n}\sum_{i=1}^n M(x_i) = \frac{1}{n}\sum_{i=1}^n M(\xi) = M(\xi)$$

(мы воспользовались тем очевидным фактом, что распределение любой компоненты выборки  $x_i$  совпадает с распределением анализируемой случайной величины  $\xi$ ).

Однако, вопреки нашей интуиции, математическое ожидание выборочного второго центрального момента

$$\hat{\mu}_{2,0} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$$

не равно генеральной дисперсии. Для получения **несмещенной оценки дисперсии**  $s^2$  надо разделить сумму квадратов на  $n - 1$ , а не на  $n$ . Действительно, пользуясь свойствами математического ожидания, в частности тем, что математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий, получаем

$$\begin{aligned} M(s^2) &= M\left(\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2\right) = \\ &= \frac{1}{n-1}M\left(\sum_{i=1}^n ((x_i - M(\xi)) + (M(\xi) - \bar{x}))^2\right) = \\ &= \frac{1}{n-1}M\left(\sum_{i=1}^n (x_i - M(\xi))^2 + 2(M(\xi) - \bar{x}) \times \right. \end{aligned}$$



$$\begin{aligned}
& \times \sum_{i=1}^n (x_i - M(\xi)) + n(M(\xi) - \bar{x})^2 \Big) = \\
& = \frac{1}{n-1} M \left( \sum_{i=1}^n (x_i - M(\xi))^2 - 2n(M(\xi) - \bar{x})^2 + n(M(\xi) - \bar{x})^2 \right) = \\
& = \frac{1}{n-1} M \left( \sum_{i=1}^n (x_i - M(\xi))^2 - n(\bar{x} - M(\xi))^2 \right) = \\
& = \frac{1}{n-1} \left( \sum_{i=1}^n M((x_i - M(\xi))^2) - nM((\bar{x} - M(\xi))^2) \right) = \\
& = \frac{1}{n-1} (nD(\xi) - \frac{n}{n}D(\xi)) = D(\xi).
\end{aligned}$$

Таким образом, несмещенной оценкой для  $D(\xi)$  будет  $s^2$  — сумма квадратов отклонений от среднего, деленная на  $n - 1$ :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Следует отметить, что несмотря на то что выборочный центральный момент второго порядка не является несмещенной оценкой для дисперсии  $D(\xi)$ , его смещение (отклонение математического ожидания оценки от  $D(\xi)$ ), равное  $\frac{D(\xi)}{n}$ , стремится к нулю при  $n \rightarrow \infty$ . Оценки, удовлетворяющие этому свойству, называются **асимптотически несмещенными** и часто, когда не удастся найти несмещенную оценку для оцениваемого параметра, довольствуются асимптотически несмещенными оценками. Конечно, нас не очень интересует, что происходит со смещением при  $n \rightarrow \infty$ , когда рассматривается конкретная выборка фиксированного объема  $n$ . Скорее, мы предпочитаем асимптотически несмещенные оценки из-за того, что, как правило, их смещение относительно невелико при не очень малых объемах выборки. Например, смещение  $\hat{\mu}_{2,0}$  при  $n = 50$  равно 2 %.

Конечно, поскольку  $s^2$  — несмещенная оценка для  $D(\xi)$ , то ею и следует пользоваться. Однако нас чаще интересует не оценка дисперсии  $D(\xi)$  случайной величины  $\xi$ , а оценка ее среднеквадратического отклонения  $\sqrt{D(\xi)}$ . Естественно взять в качестве оценки среднеквадратического отклонения корень квадратный несмещенной оценки, т. е.  $s$ , однако необходимо учитывать, что  $s$  не будет несмещенной оценкой для  $\sqrt{D(\xi)}$ , а будет лишь асимптотически несмещенной.

Итак, на примере получения несмещенной оценки дисперсии мы убедились, что «естественный» способ получения оценок, состоящий в использовании характеристик выборочного распределения в качестве оценок характеристик генерального распределения, не всегда приводит к наилучшим результатам. Существует ряд регулярных приемов получения оценок: метод моментов, метод максимального правдоподобия (будут рассмотрены далее), но они не всегда приводят к наилучшим оценкам (например, с точки зрения несмещенности). Поэтому общий подход к задаче оценивания состоит в том, чтобы в качестве претендента на оценку характеристики или параметра (для краткости будем иногда использовать только термин «параметр»)  $\theta$  генерального распределения по случайной выборке  $x_1, x_2, \dots, x_n$  брать любые функции  $\hat{\theta}(x_1, x_2, \dots, x_n)$  от компонент выборки (вместо словосочетания «функция от выборки» применяют термин «*статистика*»), среди которых должна быть выбрана наилучшая.

У нас уже имеется один критерий для сравнения оценок — наличие или отсутствие несмещенности (хотя бы асимптотической). Однако этого недостаточно. Действительно, было показано, что выборочное среднее  $\bar{x}$  — несмещенная оценка для математического ожидания  $M(\xi)$ . Но можно предложить другие несмещенные оценки. Например, полусумма первого и второго значений выборки  $(x_1 + x_2)/2$ , как легко показать, будет также несмещенной оценкой для  $M(\xi)$ . Конечно, интуитивно мы чувствуем, что  $\bar{x}$  — более предпочтительная оценка, поскольку более полно учитывает информацию, содержащуюся в выборке. Однако необходим какой-то критерий, позволяющий чисто формально показать, что  $\bar{x}$  лучше оценок типа  $(x_1 + x_2)/2$ . Таким критерием может служить состоятельность оценки. Оценка  $\hat{\theta}(x_1, x_2, \dots, x_n)$  называется *состоятельной*, если при  $n \rightarrow \infty$  она сходится по вероятности к оцениваемому параметру  $\theta$ , т. е. если для любого положительного  $\varepsilon$  выполняется условие

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \varepsilon) = 0.$$

*Достаточным условием состоятельности* оценки  $\hat{\theta}$  является ее несмещенность и стремление дисперсии оценки  $\hat{\theta}$  к нулю при увеличении объема выборки, т. е. оценка будет состоятельной, если  $M(\hat{\theta}) = \theta$  и  $\lim_{n \rightarrow \infty} D(\hat{\theta}) = 0$ . Справедливость этого утверждения непосредственно следует из неравенства Чебышева, которое в данном случае имеет следующий вид:

$$P(|\hat{\theta} - M(\hat{\theta})| > \varepsilon) < \frac{D(\hat{\theta})}{\varepsilon^2}.$$

Пользуясь этим достаточным условием, покажем, что выборочное среднее  $\bar{x}$  является состоятельной оценкой математического ожидания. Поскольку несмещенность  $\bar{x}$  уже была доказана, осталось показать, что  $D(\bar{x}) \rightarrow 0$  при  $n \rightarrow \infty$ . Действительно, получаем, воспользовавшись свойствами дисперсии,

$$D(\bar{x}) = D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n D(x_i) = \frac{1}{n^2} n D(\xi) = \frac{1}{n} D(\xi) \xrightarrow{n \rightarrow \infty} 0.$$

Таким образом,  $\bar{x}$  — состоятельная оценка для математического ожидания  $M(\xi)$  случайной величины  $\xi$ . И наоборот, как легко видеть, полусумма  $(x_1 + x_2)/2$  первого и второго значений выборки, будучи несмещенной, не является состоятельной, поскольку ее дисперсия не стремится к нулю при неограниченном увеличении  $n$

$$D\left(\frac{x_1 + x_2}{2}\right) = \frac{1}{4}(D(x_1) + D(x_2)) = \frac{1}{2}D(\xi).$$

Аналогично можно доказать, что несмещенная оценка дисперсии  $s^2$  является состоятельной оценкой для генеральной дисперсии  $D(\xi)$ .

Еще одним желательным свойством оценки является ее эффективность. Несмещенная оценка  $\hat{\theta}(x_1, x_2, \dots, x_n)$  параметра  $\theta$  называется **эффективной оценкой**, если  $D(\hat{\theta}) \leq D(\theta^*)$  для любой другой несмещенной оценки  $\theta^*(x_1, x_2, \dots, x_n)$ . Дело в том, что две оценки, будучи обе несмещенными и состоятельными, могут различаться своими дисперсиями. Например, следующая оценка для математического ожидания

$$\tilde{x} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$$

будет, как легко проверить, несмещенной и состоятельной. Однако она не будет эффективной, так как ее дисперсия  $\frac{D(\xi)}{n-1}$  больше дисперсии  $\bar{x}$ , которая равна  $\frac{D(\xi)}{n}$ .

Ранее было отмечено, что любая функция от выборки может рассматриваться как потенциальная оценка параметра или характеристики наблюдения. Тем не менее, по крайней мере, для

получения начальных выражений для оценок полезно использовать регулярные методы. Полученные таким методом оценки следует проверить на несмещенность, состоятельность и эффективность и затем, если необходимо, подкорректировать (подобно тому, как это было сделано с первоначально смещенной оценкой дисперсии). Наиболее часто для получения оценок применяются методы моментов и максимального правдоподобия.

**Метод моментов**, предложенный К. Пирсоном (1894), а еще ранее в частных случаях применявшийся К. Гауссом, состоит в приравнивании некоторого числа выборочных моментов соответствующим моментам неизвестного теоретического распределения. Полученная система уравнений будет содержать в качестве неизвестных параметры теоретического распределения (через которые выражаются теоретические моменты) и известные выборочные значения, поэтому при решении этой системы относительно неизвестных параметров получаются выражения для них в виде функций от выборочных значений. Можно отметить, что ранее, приравнивая математическое ожидание первому выборочному моменту, а дисперсию — второму центральному выборочному моменту, был применен этот метод (во всяком случае, для нормального распределения, у которого параметры  $\mu$  и  $\sigma^2$  совпадают соответственно с первым моментом  $\mu_1$  и вторым центральным моментом  $\mu_{2,0}$ ).

**Метод максимального правдоподобия** был впервые предложен Р. Фишером (1912) и состоит в том, что в качестве оценок параметров распределения берутся такие значения неизвестных параметров  $\alpha_1, \alpha_2, \dots, \alpha_k$ , которым соответствует максимальное «правдоподобие»  $L$  (от англ. *likelihood*) появления полученной выборки  $x_1, x_2, \dots, x_n$ , т. е. максимальная вероятность (в дискретном случае)

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \alpha_1, \alpha_2, \dots, \alpha_k) &= \\ &= P(x_1, x_2, \dots, x_n; \alpha_1, \alpha_2, \dots, \alpha_k) = \\ &= P(x_1; \alpha_1, \alpha_2, \dots, \alpha_k) P(x_2; \alpha_1, \alpha_2, \dots, \alpha_k) \dots P(x_n; \alpha_1, \alpha_2, \dots, \alpha_k) \end{aligned}$$

или максимальная плотность вероятности (в непрерывном случае)

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \alpha_1, \alpha_2, \dots, \alpha_k) &= \\ &= f(x_1, x_2, \dots, x_n; \alpha_1, \alpha_2, \dots, \alpha_k) = \\ &= f(x_1; \alpha_1, \alpha_2, \dots, \alpha_k) f(x_2; \alpha_1, \alpha_2, \dots, \alpha_k) \dots f(x_n; \alpha_1, \alpha_2, \dots, \alpha_k). \end{aligned}$$

Поскольку *функция правдоподобия*  $L(x_1, x_2, \dots, x_n; \alpha_1, \alpha_2, \dots, \alpha_k)$  зависит как от неизвестных параметров теоретического распределения, так и от выборочных значений, то, решая задачу ее максимизации по  $\alpha_1, \alpha_2, \dots, \alpha_k$ , можно найти выражения неизвестных параметров  $\alpha_1, \alpha_2, \dots, \alpha_k$  через известные выборочные значения  $x_1, x_2, \dots, x_n$  [Г. Крамер, 1975]. Пример использования метода максимального правдоподобия для оценивания неизвестных параметров будет приведен в подразд. 3.3.5, посвященном логистической регрессии.

### 2.2.2. Интервальное статистическое оценивание

Часто помимо конкретного значения оценки необходимо знать свойства оценки, которые ассоциируются с ее точностью и надежностью. Этим требованиям отвечают так называемые интервальные оценки.

*Интервальная оценка* — это некоторый интервал  $(\theta_{\text{нижн}}, \theta_{\text{верхн}})$ , называемый *доверительным интервалом*, концы которого (*доверительные пределы*) зависят от выборочных значений и заданной *доверительной вероятности*  $\gamma$  и который с заданной вероятностью  $\gamma$  содержит теоретическое («истинное») значение  $\theta$  оцениваемого неизвестного параметра, т. е.

$$P(\theta_{\text{нижн}} < \theta < \theta_{\text{верхн}}) = \gamma$$

(дополнение  $\gamma$  до 1 будем обозначать  $\varepsilon$ , т. е.  $\gamma = 1 - \varepsilon$ ). Наиболее часто используются значения доверительной вероятности  $\gamma$ , равные 0,95 или 0,99 (95%-й и 99%-й доверительные интервалы). Если  $P(\theta < \theta_{\text{нижн}}) = P(\theta > \theta_{\text{верхн}}) = \varepsilon/2$ , то доверительный интервал называется *центральным*, в противном случае — *нецентральным*. Обычно применяют центральные доверительные интервалы, но иногда нецентральные интервалы оказываются более подходящими. Например, при оценке доли всхожих семян важнее более точно определить нижнюю границу доверительного интервала  $\theta_{\text{нижн}}$ . В этом случае следует взять  $\varepsilon_1$  по возможности малым и  $P(\theta < \theta_{\text{нижн}}) = \varepsilon_1$ ,  $P(\theta > \theta_{\text{верхн}}) = \varepsilon_2$ , где  $\varepsilon_1 + \varepsilon_2 = \varepsilon$ .

#### *Доверительный интервал для математического ожидания нормально распределенной случайной величины с известной дисперсией*

Проще всего понять логику интервального оценивания на примере построения доверительного интервала для математиче-

ского ожидания нормально распределенной случайной величины с известной дисперсией. Пусть  $\xi$  — нормально распределенная случайная величина с неизвестным математическим ожиданием  $M(\xi) = \mu$  и дисперсией  $D(\xi) = \sigma^2$ , т. е.  $\xi \sim N(\mu, \sigma^2)$ , и имеется выборка значений этой случайной величины  $x_1, x_2, \dots, x_n$  объема  $n$ . Требуется найти доверительный интервал для  $\mu$  с доверительной вероятностью  $\gamma$ .

Ранее было показано, что выборочное среднее (для любого распределения, в том числе и нормального) имеет математическое ожидание, равное математическому ожиданию исходной случайной величины, т. е.  $\mu$ , а дисперсия — дисперсии исходной случайной величины, деленной на  $n$ , т. е.  $\sigma^2/n$ . Следовательно, статистика

$$u = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \quad (2.1)$$

полученная путем стандартизации выборочного среднего  $\bar{x}$ , будет иметь нулевое математическое ожидание и единичную дисперсию. Поскольку, как известно, линейные комбинации нормально распределенных случайных величин также имеют нормальное распределение, а случайная величина  $u$  фактически является линейной комбинацией нормально распределенных случайных величин  $x_1, x_2, \dots, x_n$ , то  $u$  будет стандартно распределенной случайной величиной, т. е.  $u \sim N(0, 1)$ . Стандартное нормальное распределение — это конкретное, полностью заданное распределение, квантили которого можно найти в соответствующих таблицах (или вычислить путем численного интегрирования). В частности, можно найти симметричные относительно центра распределения границы, внутрь которых  $u$  попадает с заданной вероятностью  $\gamma$ :

$$P(u_{\epsilon/2} < u < u_{1-\epsilon/2}) = \gamma$$

или, с учетом симметрии нормального распределения,

$$P(-u_{1-\epsilon/2} < u < u_{1-\epsilon/2}) = \gamma \quad (2.2)$$

(через  $u_{\epsilon/2}$  и  $u_{1-\epsilon/2}$  обозначены квантили стандартного нормального распределения порядка  $\epsilon/2$  и  $1 - \epsilon/2$ ). В частности, справедливы следующие неравенства:

$$P(-1,96 < \mu < 1,96) = 0,95 \quad \text{и} \quad P(-2,58 < u < 2,58) = 0,99.$$

Подставив в (2.2) выражение для  $u$  из (2.1), получим

$$P(-u_{1-\varepsilon/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < u_{1-\varepsilon/2}) = \gamma \quad (2.3)$$

или, после преобразований,

$$P\left(\bar{x} - u_{1-\varepsilon/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_{1-\varepsilon/2} \frac{\sigma}{\sqrt{n}}\right) = \gamma.$$

Это означает, что интервал  $\left(\bar{x} - u_{1-\varepsilon/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + u_{1-\varepsilon/2} \frac{\sigma}{\sqrt{n}}\right)$  будет  $100\gamma\%$ -м доверительным интервалом для неизвестного математического ожидания  $\mu$  нормального распределения с известной дисперсией  $\sigma^2$ . В частности,  $95\%$ -м доверительным интервалом будет интервал  $\left(\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}\right)$ , а  $99\%$ -м —  $\left(\bar{x} - 2,58 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2,58 \frac{\sigma}{\sqrt{n}}\right)$ . Видим, что ширина доверительного интервала уменьшается при уменьшении  $\sigma$ , увеличении объема выборки и снижении доверительной вероятности.

### *Доверительный интервал для математического ожидания нормально распределенной случайной величины с неизвестной дисперсией*

В случае неизвестной дисперсии постановка задачи и ход рассуждений при построении доверительного интервала аналогичны случаю известной дисперсии. Разница состоит в том, что в выражении (2.1) неизвестное среднеквадратическое отклонение  $\sigma$  заменяется его выборочной оценкой  $s$ :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}. \quad (2.4)$$

Полученная таким образом статистика  $t$ , будучи довольно сложной функцией от нормально распределенных случайных величин  $x_1, x_2, \dots, x_n$ , уже не будет нормально распределенной. Можно доказать, что статистика  $t$  имеет  $t$ -распределение с  $n - 1$  степенями свободы. Отсюда следует, что справедливо равенство

$$P\left(-t_{n-1, 1-\varepsilon/2} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{n-1, 1-\varepsilon/2}\right) = \gamma,$$

аналогичное уравнению (2.3) и отличающееся от него заменой  $\sigma$  на  $s$  и квантилей нормального распределения на соответствующие

щие квантили  $t$ -распределения с  $n - 1$  степенями свободы. Соответственно  $100\gamma\%$ -й доверительный интервал для неизвестного математического ожидания  $\mu$  нормального распределения с неизвестной дисперсией  $\sigma^2$  будет иметь следующий вид

$$\left( \bar{x} - t_{n-1, 1-\varepsilon/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\varepsilon/2} \frac{s}{\sqrt{n}} \right), \text{ или } \left( \bar{x} \pm t_{n-1, 1-\varepsilon/2} \frac{s}{\sqrt{n}} \right).$$

Известно, что этот доверительный интервал и доверительный интервал, полученный ранее, являются *робастными*, т. е. нечувствительными к умеренным отклонениям от предположения о нормальности распределения. Во всяком случае, как отмечается в пособиях по математической статистике, при объеме выборки не менее 15 становится целесообразно использовать приведенные доверительные интервалы для математического ожидания и в случае умеренного отклонения от предположения о нормальности.

Заметим, что при  $n \rightarrow \infty$   $t$ -распределение приближается к нормальному, а его квантили — к квантилям нормального распределения.

Например, при  $n - 1 = 60$  квантиль  $t_{60; 0,975}$  равна 2,00, что не очень сильно отличается от аналогичного значения  $u_{0,975} = 1,96$  для нормального распределения (особенно на фоне выборочных флуктуаций  $\bar{x}$  и  $s$ ). Поэтому при числе наблюдений порядка нескольких десятков можно пользоваться нормальным приближением для  $t$ -распределения.

Однако при небольшом числе степеней свободы различие между квантилями  $t$ -распределения и нормального распределения довольно значительно. Например, для  $n - 1 = 1$  имеем  $t_{1; 0,975} = 12,71$ , для  $n - 1 = 2$  —  $t_{2; 0,975} = 4,30$ , для  $n - 1 = 5$  —  $t_{5; 0,975} = 2,57$ . Но уже при  $n - 1 = 9$  (т. е. для выборки из 10 наблюдений) получаем значение  $t_{9; 0,975} = 2,26$ , что не очень сильно отличается от 1,96.

Возвращаясь к примеру с длинами лепестков ириса и учитывая, что  $\bar{x} = 4,26$ ,  $s = 0,47$ ,  $t_{49; 0,975} = 2,01$  (при числе степеней свободы  $n - 1 = 49$ ), а также предполагая, что распределение длин лепестков нормально (далее будет рассмотрена процедура проверки этого предположения), получаем, что 95%-м доверительным интервалом для математического ожидания длины лепестка будет интервал (4,13; 4,39), т. е. можно утверждать, что с вероятностью 0,95 неизвестное  $\mu$  находится между 4,13 и 4,39 (точнее, найденный доверительный интервал с вероятностью 0,95 покрывает неизвестное значение  $\mu$ ).



**Доверительный интервал для неизвестной дисперсии нормально распределенной случайной величины (при неизвестном математическом ожидании)**

Для нахождения доверительного интервала для неизвестной дисперсии нормально распределенной случайной величины рассмотрим статистику

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}.$$

Можно показать, что эта статистика имеет  $\chi^2$ -распределение с  $n - 1$  степенями свободы. Следовательно, справедливо равенство

$$P\left(\chi_{n-1, \varepsilon/2}^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_{n-1, 1-\varepsilon/2}^2\right) = \gamma,$$

которое можно переписать в виде

$$P\left(\frac{(n-1)s^2}{\chi_{n-1, 1-\varepsilon/2}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{n-1, \varepsilon/2}^2}\right) = \gamma.$$

Таким образом,  $100\gamma\%$ -й доверительный интервал для неизвестной дисперсии  $\sigma^2$  нормального распределения с неизвестным математическим ожиданием  $\mu$  будет иметь следующий вид:

$$\left(\frac{(n-1)s^2}{\chi_{n-1, 1-\varepsilon/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, \varepsilon/2}^2}\right),$$

где  $\chi_{n-1, \varepsilon/2}^2$  и  $\chi_{n-1, 1-\varepsilon/2}^2$  — квантили  $\chi^2$ -распределения с  $n - 1$  степенями свободы.

В частности, для длины лепестков ириса, учитывая, что  $s^2 = 0,22$ ,  $n - 1 = 49$ ,  $\chi_{49; 0,025}^2 = 70,24$  и  $\chi_{49; 0,975}^2 = 31,55$ , получаем, что  $95\%$ -м доверительным интервалом для дисперсии (в предположении нормальности распределения) будет интервал  $(0,15; 0,34)$ . Полученный доверительный интервал несимметричен относительно значения  $s^2 = 0,22$ , так как несимметрично распределение  $\chi^2$  (однако он является центральным).

Заметим, что полученный доверительный интервал для дисперсии в отличие от доверительного интервала для математического ожидания чувствителен к отклонениям от исходного предположения о нормальности распределения, поэтому перед его применением следует проверить гипотезу о нормальности распределения.

## Доверительный интервал для неизвестного параметра $p$ биномиального распределения

Пусть проведено  $n$  независимых испытаний, в которых некоторое событие  $A$  произошло  $k$  раз. Требуется найти точечную и интервальную оценки неизвестной вероятности  $p$  появления этого события.

Эту задачу можно рассматривать в двух эквивалентных формулировках. В первой формулировке считается, что получено  $n$  наблюдений случайной величины  $\xi_1$ , принимающей с вероятностью  $p$  значение 1 в случае появления события  $A$  и с вероятностью  $1 - p$  значение 0 — в случае непоявления события  $A$  (*распределение Бернулли*).

Во второй формулировке считается, что имеется только одно наблюдение случайной величины  $\xi$  — числа появлений события  $A$  в одном сложном испытании (*биномиальное распределение*).

Поскольку математическое ожидание  $\xi_1$ , как известно, равно  $p$ , получение оценки для  $p$  равносильно получению оценки для математического ожидания  $\xi_1$ . Несмещенной, состоятельной и эффективной оценкой для математического ожидания является выборочное среднее, которое в данном случае совпадает с частотой  $\nu = k/n$  появления события  $A$  в выборке. Таким образом, выборочная частота является несмещенной, состоятельной и эффективной оценкой для неизвестной вероятности.

Во второй формулировке математическое ожидание  $\xi$  равно  $np$ , а выборочное среднее для одного наблюдения равно самому наблюдению, т. е.  $k$ . И поскольку оценкой для  $np$  служит  $k$  (среднее по одному наблюдению), то оценкой для  $p$  снова будет  $k/n$ .

Построение доверительных интервалов несколько проще обсуждать в терминах биномиального распределения. Можно построить как точные доверительные интервалы для  $p$ , так и приближенные. Математические выкладки нахождения точных доверительных интервалов довольно громоздки и здесь приведем лишь окончательные формулы для доверительных интервалов. Напротив, выражения для приближенных доверительных интервалов легко получаются на основе применения центральной предельной теоремы, однако они применимы лишь при достаточно большом  $n$  — ориентировочно при  $n > \frac{25}{\nu(\nu - 1)}$ .

**Точный доверительный интервал для параметра биномиального распределения** имеет вид

$$\left( \frac{k}{k + (n - k + 1)F_{2(n-k+1),2k,1-\varepsilon/2}}, \frac{(k + 1)F_{2(k+1),2(n-k),1-\varepsilon/2}}{n - k + (k + 1)F_{2(k+1),2(n-k),1-\varepsilon/2}} \right),$$

где  $n$  — число испытаний;  $k$  — число появлений события  $A$ ;  $F_{f,g,\alpha}$  — квантиль порядка  $\alpha$  распределения  $F$  с  $f, g$  степенями свободы.

Рассмотрим в качестве примера данные о длине лепестков ириса с точки зрения оценки вероятности появления в случайной выборке «длинных» лепестков, а именно лепестков длиной 4,5 см и более. В данном случае  $n = 50$  и  $k = 21$ . Следовательно, точечной оценкой для  $p$  будет значение  $v = k/n = 0,42$ , а 95%-м доверительным интервалом — интервал

$$\left( \frac{21}{21 + 30F_{60;42;0,975}}, \frac{22F_{42;58;0,975}}{29 + 22F_{42;58;0,975}} \right).$$

Учитывая, что  $F_{60;42;0,975} = 1,78$  и  $F_{42;58;0,975} = 1,74$ , получаем окончательно интервал  $(0,28; 0,57)$ .

**Приближенный доверительный интервал для параметра биномиального распределения.** Поскольку число появлений события  $A$  в  $n$  испытаниях равно сумме чисел появлений этого события в отдельных испытаниях (0 или 1), в соответствии с центральной предельной теоремой при больших  $n$  распределение биномиальной случайной величины будет близко к нормальному. Учитывая, что дисперсия биномиальной случайной величины равна  $np(1 - p)$ , получаем для ее математического ожидания приближенные доверительные пределы  $k \pm u_{1-\varepsilon/2} \sqrt{np(1 - p)}$ , а для параметра  $p$ , соответственно, —

$$v \pm u_{1-\varepsilon/2} \sqrt{\frac{p(1 - p)}{n}}.$$

Заменив параметр  $p$  его выборочной оценкой  $v$ , получим окончательно для приближенного доверительного интервала параметра  $p$  следующее выражение:

$$v \pm u_{1-\varepsilon/2} \sqrt{\frac{v(1 - v)}{n}}.$$

В частности, для примера с лепестками ириса, подставив  $v = 0,42$ ;  $u_{0,975} = 1,96$  и  $n = 50$ , получим приближенный доверительный интервал  $(0,28; 0,56)$ , не слишком отличающийся от точного доверительного интервала.

## *Доверительный интервал для неизвестного параметра $\lambda$ пуассоновского распределения*

Пусть число появлений некоторого события  $A$  является случайной величиной, имеющей пуассоновское распределение, и пусть в результате наблюдения событие  $A$  произошло  $k$  раз. Требуется найти точечную и интервальную оценку неизвестного параметра  $\lambda$  пуассоновского распределения. Поскольку  $\lambda$  является математическим ожиданием пуассоновской случайной величины, то несмещенной, состоятельной и эффективной оценкой для  $\lambda$  будет выборочное среднее, которое для рассматриваемой ситуации единственного наблюдения совпадает с  $k$ . Что касается доверительных пределов для параметра  $\lambda$ , то, как и в случае биномиального распределения, можно предложить точное и приближенное решения. Выражения для приближенных доверительных пределов также основаны на применении центральной предельной теоремы и возможны лишь при достаточно большом  $k$  — ориентировочно при  $k > 25$ .

*Точный доверительный интервал для параметра пуассоновского распределения* имеет вид

$$\left( \frac{1}{2} \chi_{2k, \varepsilon/2}^2, \frac{1}{2} \chi_{2k+2, 1-\varepsilon/2}^2 \right),$$

где  $k$  — число появлений события  $A$ ;  $\chi_{f, \alpha}^2$  — квантиль порядка  $\alpha$   $\chi^2$ -распределения с  $f$  степенями свободы [Л. Н. Большев, Н. В. Смирнов, 1983; К. А. Браунли, 1977].

Рассмотрим следующий пример. Пусть после фильтрования 1 мл воды на фильтровальной бумаге обнаружено 100 клеток фитопланктона. Предполагая, что число клеток в заданном объеме воды имеет пуассоновское распределение, найти 95%-е доверительные пределы для параметра  $\lambda$  этого распределения. Подставив  $k = 100$ ,  $\chi_{200; 0,025}^2 = 162,7$  и  $\chi_{202; 0,975}^2 = 243,4$  в приведенные выражения, получим 95%-й доверительный интервал: (81,3; 121,7).

*Приближенный доверительный интервал для параметра пуассоновского распределения.* Поскольку число появлений события  $A$  при наблюдении пуассоновской случайной величины  $\xi$  можно аппроксимировать биномиальным распределением, а последнее — нормальным, то при не слишком малых  $k$  в соответствии с центральной предельной теоремой распределение случайной величины  $\xi$  будет близко к нормальному. Учитывая, что дисперсия пуассоновской случайной величины равна  $\lambda$ ,

получаем для ее математического ожидания приближенные доверительные пределы  $k \pm u_{1-\varepsilon/2} \sqrt{\lambda}$ . Заменяв параметр  $\lambda$  его выборочной оценкой  $k$ , получим окончательно для приближенного доверительного интервала параметра  $\lambda$  следующее выражение:

$$k \pm u_{1-\varepsilon/2} \sqrt{k}.$$

В частности, для предыдущего примера, подставив  $k = 100$  и  $u_{0,975} = 1,96$ , получим приближенный доверительный интервал  $(80,4; 119,6)$ , не слишком отличающийся от точного доверительного интервала.

### **Приближенный доверительный интервал для неизвестного коэффициента корреляции двумерного нормального распределения**

Рассмотрим теперь вопрос построения доверительного интервала для коэффициента корреляции. Пусть  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  — случайная выборка объема  $n$  из двумерного нормального распределения. Пусть  $\rho$  — коэффициент корреляции случайных величин  $\xi$  и  $\eta$ , а  $r$  — выборочный коэффициент корреляции. Распределение коэффициента корреляции  $r$ , особенно при значениях  $\rho$ , близких к 0 или 1, может сильно отличаться от нормального. Однако распределение следующей функции от  $r$ , называемой **преобразованием Фишера**, довольно хорошо аппроксимируется нормальным распределением

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

со средним  $M(z) = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$  и дисперсией  $D(z) = \frac{1}{n-3}$ . Соответственно стандартизованная случайная величина будет иметь стандартное нормальное распределение

$$u = \sqrt{n-3} \left( z - \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \right) \sim N(0, 1)$$

и с вероятностью  $\gamma$  будет заключена в пределах  $\pm u_{1-\varepsilon/2}$ , т. е.

$$P \left( -u_{1-\varepsilon/2} < \sqrt{n-3} \left( z - \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \right) < u_{1-\varepsilon/2} \right) = \gamma.$$

Решив неравенство под знаком вероятности относительно неизвестного коэффициента корреляции  $\rho$ , получим окончательно

$$P \left( \frac{\exp \left\{ 2 \left( z - \sqrt{\frac{1}{n-3}} u_{1-\varepsilon/2} \right) \right\} - 1}{\exp \left\{ 2 \left( z - \sqrt{\frac{1}{n-3}} u_{1-\varepsilon/2} \right) \right\} + 1} < \rho < \frac{\exp \left\{ 2 \left( z + \sqrt{\frac{1}{n-3}} u_{1-\varepsilon/2} \right) \right\} - 1}{\exp \left\{ 2 \left( z + \sqrt{\frac{1}{n-3}} u_{1-\varepsilon/2} \right) \right\} + 1} \right) = \gamma.$$

Заметим, что полученный доверительный интервал для коэффициента корреляции чувствителен к отклонениям от исходного предположения о двумерной нормальности случайных величин  $\xi$  и  $\eta$ .

**Пример 2.2.** Пусть объем выборки  $n = 10$ , а вычисленное по выборке значение  $r = 0,6$ , тогда 95%-м доверительным интервалом для неизвестного коэффициента корреляции будет интервал  $(0,05; 0,88)$ .

## 2.3. Статистическая проверка гипотез

Пусть  $x_1, x_2, \dots, x_n$  — случайная выборка значений случайной величины  $\xi$ , имеющей некоторое полностью или частично неизвестное распределение  $F(x)$ . В предыдущем подразделе рассматривались методы получения оценок параметров или характеристик этого неизвестного распределения. Однако часто нас интересуют не столько конкретные количественные оценки, сколько правильность или ошибочность некоторых утверждений, относящихся к распределению наблюдаемой случайной величины. Например, является ли это распределение нормальным или нет? равно математическое ожидание заданному значению или нет? Если кроме выборки  $x_1, x_2, \dots, x_n$  имеется выборка  $y_1, y_2, \dots, y_n$  значений другой случайной величины  $\eta$ , то можно поставить вопрос о том, равны или нет математические ожидания случайных величин  $\xi$  и  $\eta$ . Если имеется выборка  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  двумерной случайной величины  $(\xi, \eta)$ , то может возникнуть вопрос о равенстве нулю коэффициента корреляции между  $\xi$  и  $\eta$ .

### 2.3.1. Логика проверки статистических гипотез

Решению задач проверки гипотез о генеральном распределении по выборке из этого распределения посвящен специальный раздел математической статистики — *проверка статисти-*

**ческиx гипотез.** Логика проверки гипотез в математической статистике (она напоминает логику доказательства от противного) состоит в следующем. Вначале предполагается, что проверяемая гипотеза (ее принято называть **нулевой гипотезой** и обозначать  $H_0$ ) верна. В этом предположении ищется распределение вероятностей некоторой функции  $g(x_1, x_2, \dots, x_n)$  от значений выборки, называемой **статистикой критерия** (правило проверки гипотезы принято называть **критерием**), и в области значений этой статистики выделяется некоторая область  $W$ , называемая **критической областью**, такая, что вероятность  $P(g \in W)$  попадания выборочного значения статистики  $g$  в эту область не превосходит заданного малого значения  $\alpha$ , называемого **уровнем значимости критерия** (обычно полагают  $\alpha$  равным 0,05 или 0,01). Если для данной конкретной выборки  $g$  попадает в критическую область  $W$ , то гипотеза  $H_0$  **отвергается** (говорят — «отвергается на уровне значимости  $\alpha$ »), поскольку вероятность этого события, если в действительности верна  $H_0$ , мала. Если  $g$  не попадает в критическую область  $W$ , то говорят, что «гипотеза  $H_0$  **не отвергается** на уровне значимости  $\alpha$ » или «полученные данные не дают оснований отвергнуть гипотезу  $H_0$  на уровне значимости  $\alpha$ ».

Очевидно, что можно разными способами задать статистику критерия  $g(x_1, x_2, \dots, x_n)$ , а для заданной статистики — разными способами выбрать критическую область  $W$ , удовлетворяющую условию  $P(g \in W) = \alpha$ . Поэтому следует выбирать  $g$  и  $W$  в некотором смысле наилучшими из возможных, а именно такими, чтобы полученный критерий был **наиболее мощным**.

Для определения понятия **мощности критерия** введем понятие **альтернативной гипотезы**  $H_1$ , т. е. гипотезы, которая выполняется, если не выполняется нулевая гипотеза  $H_0$ . Тогда в терминах правильности или ошибочности принятия  $H_0$  и  $H_1$  можно указать четыре потенциально возможных результата применения критерия к выборке (табл. 2.3).

Как видим, мощность критерия — это вероятность принятия при применении данного критерия альтернативной гипотезы  $H_1$  при условии, что она верна. Очевидно, что при фиксированной вероятности ошибки 1-го рода (ее задаем самостоятельно, и она не зависит от свойств критерия) критерий будет тем лучше, чем больше его мощность (т. е. чем меньше вероятность ошибки 2-го рода).

Проиллюстрируем основные понятия рассмотренной методологии на простом примере проверки гипотезы о равенстве ма-

## Возможные результаты применения критерия к выборке

Верна гипотеза	Принята гипотеза	
	$H_0$	$H_1$
$H_0$	$1 - \alpha$ — вероятность правильно принять $H_0$ , когда верна $H_0$	$\alpha$ — вероятность ошибочно принять $H_1$ , когда верна $H_0$ (вероятность ошибки 1-го рода, уровень значимости)
$H_1$	$\beta$ — вероятность ошибочно принять $H_0$ , когда верна $H_1$ (вероятность ошибки 2-го рода)	$1 - \beta$ — вероятность правильно принять $H_1$ , когда верна $H_1$ (мощность критерия)

тематического ожидания нормально распределенной случайной величины с известной дисперсией заданному числу (с точки зрения практического применения этот пример несколько искусственен, поскольку дисперсия наблюдаемой случайной величины обычно неизвестна).

### 2.3.2. Проверка гипотез о математических ожиданиях

Итак, пусть  $\xi \sim N(\mu, \sigma^2)$ ,  $H_0: \mu = \mu_0$  и  $H_1: \mu = \mu_1$  и пусть имеется выборка  $x_1, x_2, \dots, x_n$  значений случайной величины  $\xi$  объема  $n$ . Предположим, что  $H_0$  верна и в качестве статистики критерия выбираем стандартизованное выборочное среднее

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

При верной  $H_0$  статистика  $u$  имеет стандартное нормальное распределение  $u \sim N(0, 1)$ , представленное на рис. 2.4. На этом рисунке также указана критическая область уровня  $\alpha = 0,05$ , состоящая из двух бесконечных полуинтервалов  $(-\infty, -1,96]$  и  $[1,96, \infty)$ , вероятность попадания в каждый из которых статистики  $u$  равна  $\alpha/2 = 0,025$ .

Имеются формализованные подходы к выбору статистик критериев и построению критических областей, приводящие к наиболее мощным критериям, но здесь их не рассматриваем. Неформальное же правило состоит в том, чтобы выбирать в качестве статистики величину, характеризующую степень отклонения от нулевой гипотезы. Очевидно, разность  $\bar{x} - \mu_0$  удовлетворяет этому условию, а деление на константу  $\sigma/\sqrt{n}$ , сохраняя это



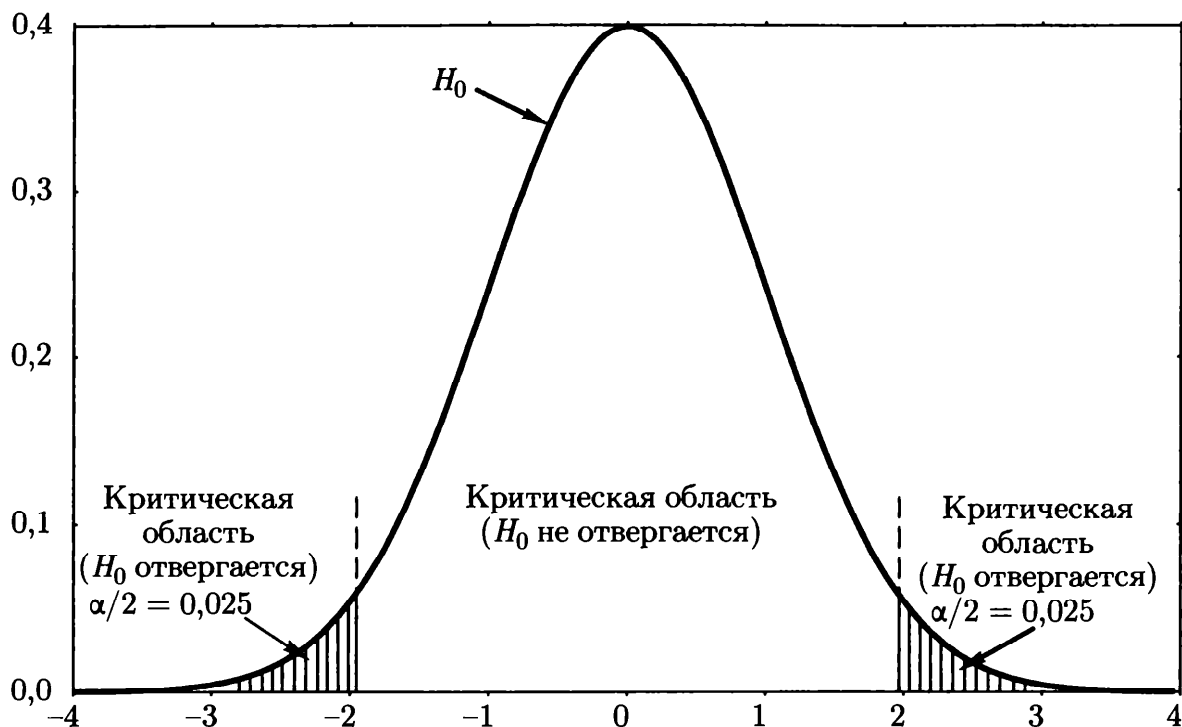


Рис. 2.4. Пример критической области

качество, приводит к величине  $u$  с полностью заданным распределением, что позволяет выбрать критическую область с требуемым уровнем значимости. Другое неформальное правило выбора критической области состоит в том, чтобы она включала значения статистики, соответствующие наибольшим отклонениям от нулевой гипотезы (на рис. 2.4 эта рекомендация соблюдена).

До сих пор мы рассматривали свойства критерия в предположении, что верна гипотеза  $H_0$ . А что происходит, когда верна альтернативная гипотеза  $H_1$ ? В этом случае распределение статистики критерия  $u$  изменится. Чтобы его найти, произведем преобразование

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu_0 - \mu_1 + \mu_1}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}},$$

из которого следует, что при гипотезе  $H_1$  распределение статистики  $u$  отличается от стандартного нормального сдвигом на величину  $\Delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$ , т. е.  $u \sim N(\Delta, 1)$  при выполнении  $H_1$ .

На рис. 2.5 взаимное расположение плотностей распределения статистики  $u$  при гипотезах  $H_0$  и  $H_1$  показано для случая  $\mu_0 = 0$ ,  $\mu_1 = 2$ ,  $\sigma^2 = 1$ ,  $\alpha = 0,05$  и  $n = 1$ . Вероятности ошибки 2-го рода  $\beta$  соответствует площадь под кривой функции плотности при  $H_1$  на промежутке от  $-1,96$  до  $1,96$ , где не отвергается гипотеза  $H_0$ , а следовательно, ошибочно не принимается гипотеза  $H_1$ . В данном случае вероятность ошибки 2-го рода,  $\beta = 0,48$ ,

довольно велика. Это произошло главным образом потому, что мал объем выборки — имеется всего одно наблюдение,  $n = 1$ . При увеличении  $n$  распределение, соответствующее альтернативной гипотезе  $H_1$ , будет сдвигаться вправо, поскольку величина  $\Delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$  будет увеличиваться, что приведет к уменьшению  $\beta$  (см. рис. 2.5). Очевидно также, что большей величине разности  $\mu_1 - \mu_0$  соответствует большая величина  $\Delta$  и, следовательно, меньшая вероятность ошибки 2-го рода. Вероятность ошибки 2-го рода уменьшается также при уменьшении дисперсии  $\sigma^2$  наблюдаемой случайной величины. Кроме того,  $\beta$  уменьшается при увеличении  $\alpha$ , однако не принято принимать  $\alpha$  больше 0,05. При уменьшении  $\alpha$  вероятность ошибки 2-го рода  $\beta$ , напротив, растет, поэтому не следует принимать  $\alpha$  слишком малым, если число наблюдений  $n$  мало, разность между  $\mu_1$  и  $\mu_0$  невелика, а дисперсия  $\sigma^2$  большая.

Содержательно, ошибка 1-го рода — это ошибка ложного обнаружения несуществующего отклонения от нулевой гипотезы (ложного обнаружения несуществующего эффекта). Ошибка 2-го рода — это ошибка ложного необнаружения существующего отклонения от нулевой гипотезы (ложного необнаружения существующего эффекта). Мощность критерия — его способность обнаружить имеющееся отклонение от нулевой гипотезы.

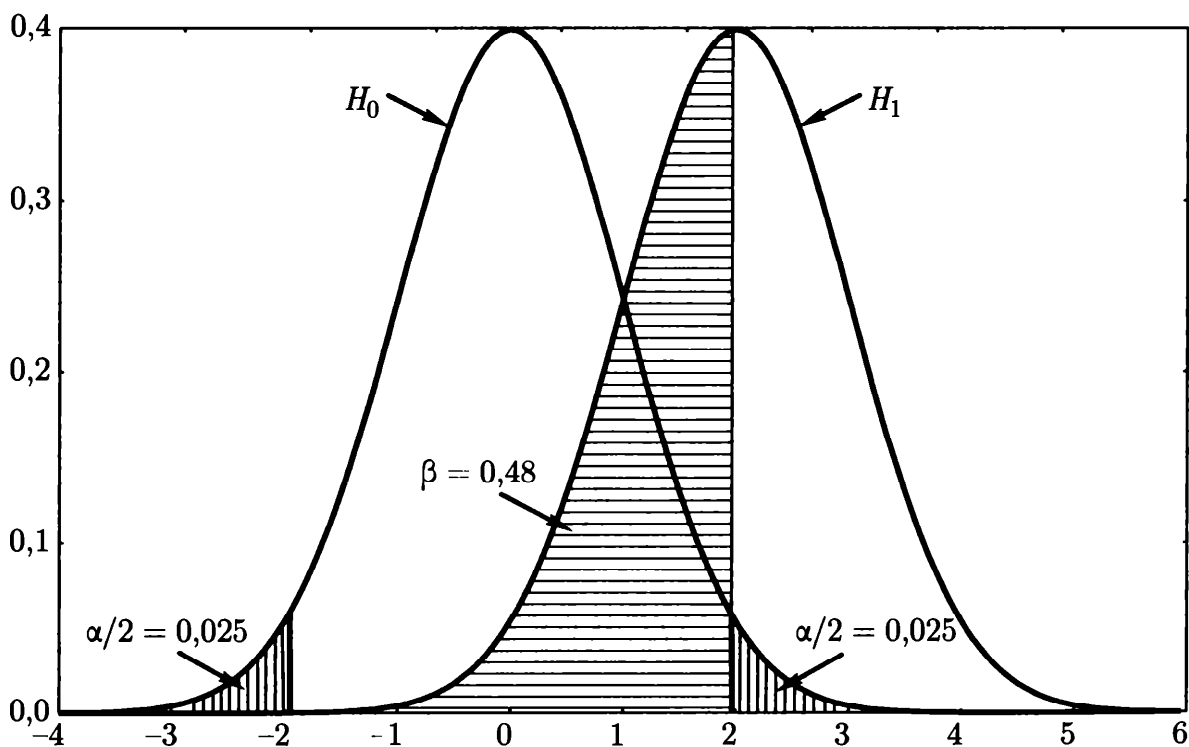


Рис. 2.5. Взаимосвязь между вероятностями ошибок 1-го и 2-го рода при двусторонней альтернативе

В приведенном примере предполагалось, что альтернативной гипотезе  $H_1$  соответствует вполне определенное распределение  $\xi \sim N(2, 1)$ , что позволило найти конкретное значение вероятности ошибки 2-го рода. Такого рода альтернативные гипотезы называются *простыми альтернативами*. Однако на практике чаще встречается ситуация, когда конкретной нулевой гипотезе противопоставляется целый спектр альтернатив. Например,  $H_1: M\xi \neq \mu_0$  или  $H_1: M\xi > \mu_0$ . Такого рода альтернативные гипотезы называются *сложными альтернативами*. В случае сложной альтернативной гипотезы нельзя определить величину вероятности ошибки 2-го рода. Например, в рассматриваемой ситуации она может быть значительной даже при очень большом числе наблюдений, если различие между  $\mu_1$  и  $\mu_0$  мало. Поэтому в ситуации, когда статистика критерия не попадает в критическую область, не утверждают категорично, что «нулевая гипотеза принимается», а формулируют вывод более осторожно: «нулевая гипотеза не отвергается». Тем самым подчеркивается, что хотя и не обнаружены отклонения от нулевой гипотезы, при верной  $H_1$  их можно ошибочно не обнаружить с вероятностью  $\beta$ , которая нам не известна и которая, возможно, довольно значительна. Если же  $\beta$  действительно велика, то утверждение «нулевая гипотеза принимается» не представляет большой ценности. Например, положив равной нулю вероятность ошибки 1-го рода, независимо от результатов наблюдений, всегда будем принимать гипотезу  $H_0$ , поскольку критическая область будет включать всю область определения статистики критерия. Однако при этом вероятность ошибки 2-го рода будет равна единице, т. е. если даже отклонение от нулевой гипотезы имеется, то оно с вероятностью «единица» не будет обнаружено.

Сложные альтернативы могут быть *двусторонними* ( $H_1: M(\xi) \neq \mu_0$ ) и *односторонними* ( $H_1: M(\xi) > \mu_0$  или  $H_1: M(\xi) < \mu_0$ ). Если имеется достоверная информация о направлении отклонения от нулевой гипотезы, то использование односторонней альтернативы предпочтительнее, поскольку повышает мощность критерия. Например, если известно, что отклонение математического ожидания  $\mu$  от гипотетического значения  $\mu_0$  может произойти только в большую сторону, то в качестве альтернативы следует взять гипотезу  $H_1: M(\xi) > \mu_0$ . Критическая область уровня  $\alpha$  в этом случае будет состоять не из двух бесконечных полуинтервалов  $(-\infty, u_{\alpha/2}]$  и  $[u_{1-\alpha/2}, \infty)$ , а из одного  $[u_{1-\alpha}, \infty)$ .

На рис. 2.6 ситуация с односторонней альтернативой представлена для случая  $\mu_0 = 0$ ,  $\mu_1 = 2$ ,  $\sigma^2 = 1$ ,  $\alpha = 0,05$  и  $n = 1$ .

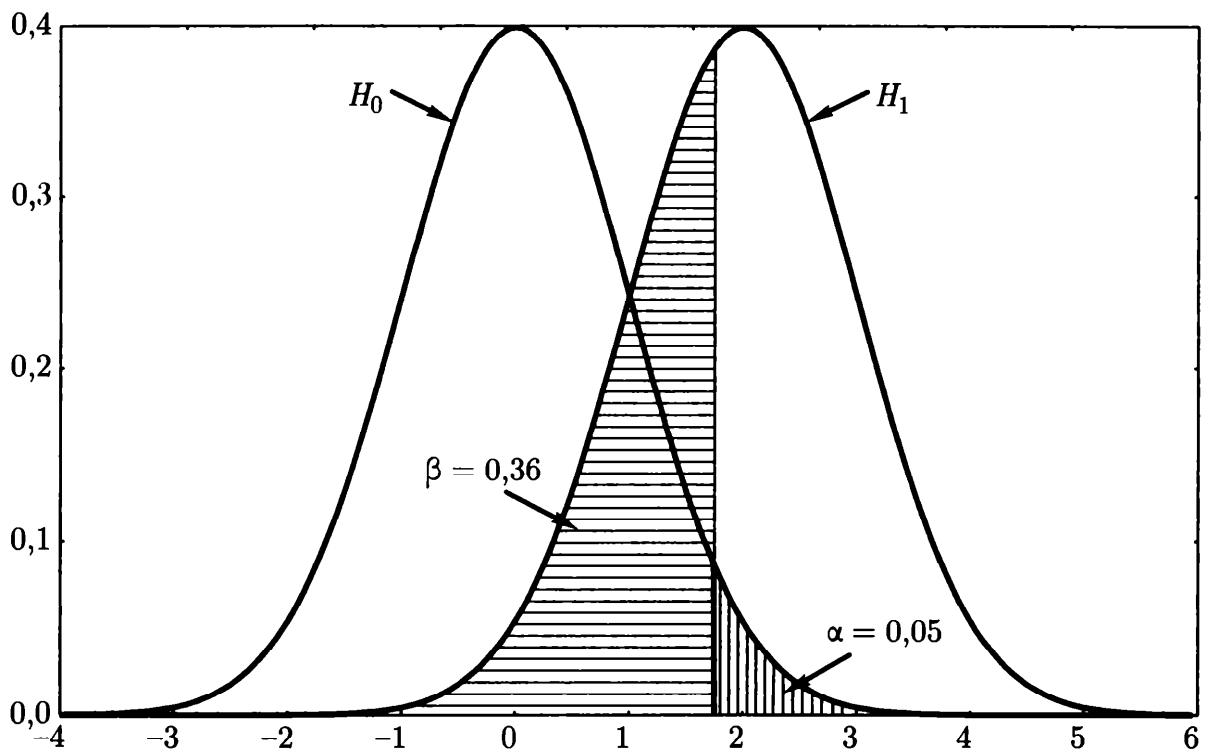


Рис. 2.6. Взаимосвязь между вероятностями ошибок 1-го и 2-го рода при односторонней альтернативе

Вероятности ошибки 2-го рода  $\beta$  при условии, что верна гипотеза  $H_1$ , соответствует площадь под кривой плотности статистики критерия  $g$ . На промежутке от  $-\infty$  до 1,64  $\beta = 0,36$ , что меньше величины  $\beta = 0,48$  для аналогичной двусторонней альтернативы, представленной на рис. 2.5.

**Пример 2.3.** Известно, что датчик генерирует случайные числа, нормально распределенные с дисперсией 1, но есть сомнения в том, что математическое ожидание равно 0. Требуется проверить гипотезу о равенстве математического ожидания нулю при двусторонней альтернативе по следующей случайной выборке объема  $n = 25$ :

0,830	0,177	-0,294	0,471	-0,044	0,635	2,209	-0,394	-0,404
1,257	1,137	-0,839	1,668	0,751	0,416	-0,922	1,473	-0,317
0,220	0,414	0,428	1,088	-1,130	-0,015	0,142		

Выборочное среднее равно  $\bar{x} = 0,358$ , следовательно, для статистики критерия получаем

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0,358 - 0}{1/\sqrt{25}} = 1,79.$$

Значение 1,79 не выходит за двусторонние 5%-е критические пределы  $\pm u_{0,975} = \pm 1,96$ , поэтому гипотеза не отвергается.

На самом деле математическое ожидание датчика было равным 0,25, т. е. отклонение ошибочно не было обнаружено — при

проверке гипотезы была сделана ошибка 2-го рода. Очевидно, мощность критерия при данном числе наблюдений  $n = 25$ , данной разности между гипотетическим и истинным математическими ожиданиями  $\mu_1 - \mu_0 = 0,25$  и данной дисперсии  $\sigma^2 = 1$  недостаточна. (В другом эксперименте с этим же датчиком была получена выборка значений объема  $n = 100$ . Выборочное среднее оказалось равным  $\bar{x} = 0,218$ , а выборочное значение статистики — равным 2,18, что дало основание отвергнуть нулевую гипотезу.)

**Замечание.** В современных пакетах статистических программ кроме вычисленного значения статистики критерия  $u_{\text{набл}}$  (в данном случае  $u_{\text{набл}} = 1,79$ ) выдается и так называемое *p-значение* (от англ. *p-value*, или просто *p*), равное в случае двусторонней альтернативы вероятности выхода за пределы  $\pm u_{\text{набл}}$ :

$$p = P(|u| \geq u_{\text{набл}})$$

(в случае односторонних альтернатив определение *p*-значения изменяется очевидным образом). Вместо сравнения  $u_{\text{набл}} = 1,79$  с границами критической области  $\pm u_{0,975} = \pm 1,96$  проще сравнить соответствующее *p*-значение, равное  $p = 0,073$ , с уровнем значимости  $\alpha = 0,05$ . Неравенство  $p > \alpha$  означает, что  $u_{\text{набл}}$  не попало в критическую область и, следовательно,  $H_0$  не должна быть отвергнута, и, наоборот, если  $p \leq \alpha$ , то  $H_0$  должна быть отвергнута. В рассмотренном примере  $0,073 > 0,05$ , следовательно,  $H_0$  не должна быть отвергнута.

### **Проверка гипотезы о равенстве заданному числу математического ожидания нормально распределенной случайной величины с неизвестной дисперсией (одновыборочный *t*-критерий)**

Аналогично случаю построения доверительного интервала для неизвестного математического ожидания нормально распределенной случайной величины, в случае неизвестной дисперсии в качестве статистики критерия проверки гипотезы о равенстве математического ожидания заданному числу возьмем ту же статистику, что и в случае с известной дисперсией, но с заменой неизвестного среднеквадратического отклонения  $\sigma$  его выборочной оценкой  $s$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

При верной  $H_0$  статистика  $t$  имеет *t*-распределение с  $n - 1$  степенями свободы. Соответственно критическая область для про-

верки гипотезы  $H_0: \mu = \mu_0$  против двусторонней альтернативы  $H_1: \mu \neq \mu_0$  будет состоять из двух бесконечных полуинтервалов  $(-\infty, t_{n-1, \alpha/2}]$  и  $[t_{n-1, 1-\alpha/2}, \infty)$  против односторонней альтернативы  $H_1: \mu > \mu_0$  — из одного полуинтервала  $[t_{n-1, 1-\alpha}, \infty)$  и против односторонней альтернативы  $H_1: \mu < \mu_0$  — также из одного полуинтервала  $(-\infty, t_{n-1, \alpha}]$ , где  $t_{\alpha/2}, t_{1-\alpha/2}, t_{1-\alpha}, t_{\alpha}$  обозначают квантили  $t$ -распределения с  $n - 1$  степенями свободы соответствующего уровня значимости (в силу симметричности  $t$ -распределения справедливы равенства  $t_{\alpha/2} = -t_{1-\alpha/2}$  и  $t_{\alpha} = -t_{1-\alpha}$ ).

**Пример 2.4.** Рассмотрим пример 2.3 в предположении, что дисперсия неизвестна. В этом случае необходимо вычислить оценку среднеквадратического отклонения, которая оказывается равной  $s = 0,830$ . Выборочное значение статистики критерия соответственно равно

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0,358 - 0}{0,830/\sqrt{25}} = 2,15.$$

Это значение необходимо сравнить с 5%-ми двусторонними критическими пределами, равными  $\pm t_{24; 0,025} = \pm 2,06$ . Выборочное значение статистики выходит за эти пределы, следовательно, гипотеза о равенстве математического ожидания нулю должна быть отвергнута на уровне значимости 5 %.

**Замечание.** Несмотря на то что применение  $t$ -критерия требует нормальности исходной случайной величины, его можно использовать и при умеренных отклонениях от нормальности и не слишком малых  $n$  ( $n \geq 15$ ).

### ***Проверка гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин (двухвыборочный $t$ -критерий)***

**Независимые выборки.** Предположим, что имеются случайные выборки  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_m$  значений двух независимых нормально распределенных случайных величин  $\xi \sim N(M(\xi), D(\xi))$  и  $\eta \sim N(M(\eta), D(\eta))$  и требуется проверить гипотезу  $H_0: M(\xi) = M(\eta)$  о равенстве математических ожиданий этих случайных величин.

1. Если известно, что дисперсии случайных величин  $\xi$  и  $\eta$  равны, т. е.  $D(\xi) = D(\eta) = \sigma^2$  (значение  $\sigma^2$  неизвестно), то можно получить следующую объединенную несмещенную оценку для  $\sigma^2$ :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2}{n + m - 2}.$$

В этом случае  $s^2/n$  и  $s^2/m$  будут несмещенными оценками для дисперсии выборочных средних  $\bar{x}$  и  $\bar{y}$ , а сумма  $s^2/n + s^2/m$  — несмещенной оценкой для дисперсии разности средних  $\bar{x} - \bar{y}$ . Соответственно статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n} + \frac{s^2}{m}}},$$

как можно показать, будет иметь  $t$ -распределение с  $n + m - 2$  степенями свободы. Критическая область уровня  $\alpha$  для проверки гипотезы  $H_0: M(\xi) = M(\eta)$  против двусторонней альтернативы  $H_1: M(\xi) \neq M(\eta)$  будет состоять из двух бесконечных полуинтервалов  $(-\infty, t_{n+m-2, \alpha/2}]$  и  $[t_{n+m-2, 1-\alpha/2}, \infty)$  против односторонней альтернативы  $H_1: M(\xi) > M(\eta)$  — из полуинтервала  $[t_{n+m-2, 1-\alpha}, \infty)$  и против альтернативы  $H_1: M(\xi) < M(\eta)$  — из полуинтервала  $(-\infty, t_{n+m-2, \alpha}]$ , где  $t_{n+m-2, \alpha/2}$ ,  $t_{n+m-2, 1-\alpha/2}$ ,  $t_{n+m-2, 1-\alpha}$ ,  $t_{n+m-2, \alpha}$  обозначают соответствующие квантили  $t$ -распределения с  $n + m - 2$  степенями свободы.

2. Если нет оснований считать, что дисперсии случайных величин  $\xi$  и  $\eta$  равны, то для каждой из дисперсий  $D(\xi)$  и  $D(\eta)$  вычисляется своя оценка

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad \text{и} \quad s_y^2 = \frac{\sum_{j=1}^m (y_j - \bar{y})^2}{m - 1}$$

и соответственно модифицируется статистика критерия

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}},$$

которая, как можно показать, имеет  $t$ -распределение с числом степеней свободы, равным целой части от  $1/k$ , где  $k$  выражается следующей формулой

$$k = \frac{\left(\frac{s_x^2/n}{s_x^2/n + s_y^2/m}\right)^2}{n - 1} + \frac{\left(\frac{s_y^2/m}{s_x^2/n + s_y^2/m}\right)^2}{m - 1}.$$

**Связанные выборки.** Пусть теперь  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_n$  — связанные случайные выборки из нормальных распределений  $\xi \sim N(M(\xi), D(\xi))$  и  $\eta \sim (M(\eta), D(\eta))$ . Например,  $x_i$  и  $y_i$  — результаты измерения давления до и после приема лекарства соответственно, или длина и ширина цветка ириса. Для проверки гипотезы  $H_0: M(\xi) = M(\eta)$  при двусторонней альтернативе  $H_1: M(\xi) \neq M(\eta)$  используется статистика

$$t = \frac{(\bar{x} - \bar{y})\sqrt{n}}{\sqrt{s_x^2 + s_y^2 - 2s_{xy}}}, \quad \text{где} \quad s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

При условии, что верна гипотеза  $H_0$ , статистика критерия имеет  $t$ -распределение с  $n - 1$  степенями свободы. Критическая область уровня  $\alpha$  состоит из двух интервалов  $(-\infty, -t_{n-1, 1-\alpha/2})$  и  $(t_{n-1, 1-\alpha/2}, \infty)$ , где  $t_{n-1, 1-\alpha/2}$  — квантиль  $t$ -распределения порядка  $1 - \alpha/2$  с числом степеней свободы  $n - 1$ . Так построенный критерий называется парным  $t$ -критерием или  $t$ -критерием для связанных выборок.

**Замечание.** Для связанных выборок  $t$ -критерий совпадает с одновыборочным  $t$ -критерием для выборки  $z_1, z_2, \dots, z_n$ , где  $z_i = x_i - y_i$ ,  $i = 1, \dots, n$ , для проверки нулевой гипотезы  $H_0: M(z) = 0$  против двусторонней альтернативы  $H_1: M(z) \neq 0$ .

### 2.3.3. Проверка гипотез о дисперсиях

#### *Проверка гипотезы о равенстве заданному числу дисперсии нормально распределенной случайной величины (одновыборочный $\chi^2$ -критерий)*

Для проверки гипотезы  $H_0: \sigma^2 = \sigma_0^2$  о равенстве дисперсии  $\sigma^2$  нормально распределенной случайной величины  $\xi$  заданному числу  $\sigma_0^2$  рекомендуется использовать статистику

$$\chi^2 = (n - 1) \frac{s^2}{\sigma_0^2}.$$

Можно показать, что эта статистика, при условии, что верна гипотеза  $H_0$ , распределена по закону  $\chi^2$  с  $n - 1$  степенями свободы. Критическая область уровня  $\alpha$  при двусторонней альтернативе  $H_1: \sigma^2 \neq \sigma_0^2$  состоит из двух промежутков:  $[0, \chi_{n-1, \alpha/2}^2]$  и  $[\chi_{n-1, 1-\alpha/2}^2, \infty)$ , где  $\chi_{n-1, \alpha/2}^2$  и  $\chi_{n-1, 1-\alpha/2}^2$  — квантили порядка  $\alpha/2$  и  $1 - \alpha/2$  распределения  $\chi^2$  с  $n - 1$  степенями свободы. Для



односторонней альтернативы  $H_1: \sigma^2 > \sigma_0^2$  критическая область имеет вид  $[\chi_{n-1, 1-\alpha}^2, \infty)$ , а для альтернативы  $H_1: \sigma^2 < \sigma_0^2$  — соответственно  $[0, \chi_{n-1, \alpha}^2]$ .

### **Проверка гипотезы о равенстве дисперсий двух независимых нормально распределенных случайных величин (двухвыборочный $F$ -критерий)**

Ранее было показано, что процедура проверки гипотезы о равенстве двух математических ожиданий двух нормально распределенных случайных величин упрощается, если их дисперсии одинаковы. Следующий критерий позволяет проверить нулевую гипотезу  $H_0: D(\xi) = D(\eta)$  о равенстве дисперсий двух нормально распределенных случайных величин. В качестве статистики критерия используется отношение несмещенных оценок дисперсий этих случайных величин

$$F = \frac{s_x^2}{s_y^2}.$$

При условии, что верна гипотеза  $H_0$ , можно доказать, что статистика критерия имеет  $F$ -распределение с  $n - 1$  и  $m - 1$  степенями свободы ( $n$  и  $m$  — объемы двух выборок). Соответственно критическая область уровня  $\alpha$  для проверки гипотезы  $H_0: D(\xi) = D(\eta)$  против двусторонней альтернативы  $H_1: D(\xi) \neq D(\eta)$  будет состоять из двух интервалов:  $(0, F_{n-1, m-1, \alpha/2}]$  и  $[F_{n-1, m-1, 1-\alpha/2}, \infty)$ , где  $F_{n-1, m-1, \alpha/2}$ ,  $F_{n-1, m-1, 1-\alpha/2}$  — квантили порядка  $\alpha/2$  и  $1 - \alpha/2$   $F$ -распределения с  $n - 1$  и  $m - 1$  степенями свободы. Для односторонней альтернативы  $H_1: D(\xi) > D(\eta)$  критическая область имеет вид  $[F_{n-1, m-1, 1-\alpha}, \infty)$ , а для альтернативы  $H_1: D(\xi) < D(\eta)$  — соответственно  $[0, F_{n-1, m-1, \alpha}]$ . Если в качестве статистики использовать отношение большей оценки дисперсии к меньшей, то в качестве критической области (для случая  $s_x^2 > s_y^2$ ) при двусторонней альтернативе следует использовать одностороннюю критическую область  $[F_{n-1, m-1, 1-\alpha/2}, \infty)$ . Это позволяет ограничиться таблицами  $F$ -распределения, содержащими значения функции распределения только для аргументов больших единицы.

**Замечание.** В отличие от  $t$ -критерия  $F$ -критерий чувствителен к отклонениям исходных случайных величин от нормальности. При значительных отклонениях от нормальности, особенно при небольшом числе наблюдений, его применять не следует.

### 2.3.4. Сравнение параметров двух биномиальных распределений

Пусть две независимые биномиально распределенные случайные величины  $\xi$  и  $\eta$  с параметрами  $n, p_\xi$  и  $m, p_\eta$ , соответственно, при проведении независимых испытаний приняли значения  $k$  и  $l$ . Требуется проверить гипотезу  $H_0: p_\xi = p_\eta$  о равенстве параметров  $p_\xi$  и  $p_\eta$ . Для этого можно использовать статистику

$$u = \frac{\left(v_\xi - \frac{1}{2n}\right) - \left(v_\eta + \frac{1}{2m}\right)}{\sqrt{v(1-v) \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

где  $v_\xi > v_\eta$  и  $v$  — выборочные частоты, вычисленные по первой, второй и объединенной выборкам:  $v_\xi = k/n$ ,  $v_\eta = l/m$  и  $v = (k+l)/(n+m)$ . Если верна гипотеза  $H_0$ , то для  $v_\xi, v_\eta$ , не очень близких к 0 или 1, и при достаточно больших  $n, m$  эта статистика имеет приближенно стандартное нормальное распределение [К. А. Браунли, 1977].

Критическая область уровня значимости  $\alpha$  для проверки гипотезы  $H_0: p_\xi = p_\eta$  против двусторонней альтернативы  $H_1: p_\xi \neq p_\eta$  будет состоять из двух бесконечных полуинтервалов  $(-\infty, u_{\alpha/2}]$  и  $[u_{1-\alpha/2}, \infty)$  против односторонней альтернативы  $H_1: p_\xi > p_\eta$  — из одного полуинтервала  $[u_{1-\alpha}, \infty)$  и против односторонней альтернативы  $H_1: p_\xi < p_\eta$  — также из одного полуинтервала  $(-\infty, u_\alpha]$ , где  $u_{\alpha/2}, u_{1-\alpha/2}, u_{1-\alpha}$  и  $u_\alpha$  обозначают квантили соответствующего порядка стандартного нормального распределения.

*Замечание.* Имеется также точный критерий Фишера для проверки этой гипотезы [К. А. Браунли, 1977; Н. В. Глотов и [др.], 1982].

### 2.3.5. Сравнение параметров двух пуассоновских распределений

Пусть две независимые случайные величины  $\xi$  и  $\eta$ , имеющие пуассоновское распределение с параметрами  $\lambda_\xi$  и  $\lambda_\eta$ , соответственно, при проведении испытаний приняли значения  $k$  и  $l$ . Требуется проверить гипотезу  $H_0: \lambda_\xi = \lambda_\eta$  о равенстве параметров  $\lambda_\xi$  и  $\lambda_\eta$  распределений этих случайных величин. Для этого можно использовать статистику

распределение которой при выполнении  $H_0$  и при  $k + l > 5$  довольно точно приближается стандартным нормальным распределением. Соответственно, как и в подразд. 2.3.4, критическая область уровня значимости  $\alpha$  для проверки гипотезы  $H_0: p_\xi = p_\eta$  против двусторонней альтернативы  $H_1: p_\xi \neq p_\eta$  будет состоять из двух бесконечных полуинтервалов  $(-\infty, u_{\alpha/2}]$  и  $[u_{1-\alpha/2}, \infty)$  против односторонней альтернативы  $H_1: p_\xi > p_\eta$  — из одного полуинтервала  $[u_{1-\alpha}, \infty)$  и против односторонней альтернативы  $H_1: p_\xi < p_\eta$  — также из одного полуинтервала  $(-\infty, u_\alpha]$ .

### 2.3.6. Проверка гипотезы о равенстве заданному числу коэффициента корреляции

Пусть  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  — случайная выборка пар значений двумерной случайной величины  $(\xi, \eta)$ , имеющей двумерное нормальное распределение. Требуется проверить гипотезу  $H_0: \rho = \rho_0$  о равенстве коэффициента корреляции  $\rho$  этого двумерного распределения заданному числу  $\rho_0$ . Для проверки этой гипотезы можно использовать статистику

$$u = \frac{\sqrt{n-3}}{2} \left( \ln \frac{1+r}{1-r} - \ln \frac{1+\rho_0}{1-\rho_0} \right), \quad (2.5)$$

распределение которой при выполнении  $H_0$  и при достаточно большом  $n$  довольно точно приближается стандартным нормальным распределением. Соответственно, как и в двух предыдущих подразделах, критическая область уровня значимости  $\alpha$  для проверки гипотезы  $H_0: \rho = \rho_0$  против двусторонней альтернативы  $H_1: \rho \neq \rho_0$  будет состоять из двух бесконечных полуинтервалов  $(-\infty, u_{\alpha/2}]$  и  $[u_{1-\alpha/2}, \infty)$  против односторонней альтернативы  $H_1: \rho > \rho_0$  — из одного полуинтервала  $[u_{1-\alpha}, \infty)$  и против односторонней альтернативы  $H_1: \rho < \rho_0$  — также из одного полуинтервала  $(-\infty, u_\alpha]$ .

Обычно проверяется гипотеза о равенстве коэффициента корреляции нулю, что в случае двумерного нормального распределения, как ранее отмечалось, эквивалентно проверке гипотезы о независимости  $\xi$  и  $\eta$ . В этом случае выражение (2.5) упрощается

$$u = \frac{\sqrt{n-3}}{2} \left( \ln \frac{1+r}{1-r} \right). \quad (2.6)$$

**Пример 2.5.** Пусть объем выборки  $n = 10$ , вычисленное по выборке значение  $r = 0,6$ , и требуется проверить гипотезу  $H_0: \rho = 0$  против альтернативы  $H_1: \rho \neq 0$ .

Выборочное значение статистики  $u$ , вычисленное по формуле (2.6), равно 1,83. Поскольку оно не выходит за двусторонние 5%-е критические пределы стандартного нормального распределения  $\pm 1,96$ , нет оснований отвергнуть нулевую гипотезу об отсутствии корреляции. Если бы у нас были основания предполагать, что корреляционная зависимость в случае ее наличия может быть только положительной, то следовало бы использовать для проверки  $H_0$  одностороннюю критическую область, которая для  $\alpha = 0,05$  представляет собой бесконечный полуинтервал  $[1,64; \infty)$ . Значение 1,83 попадает в эту критическую область и, следовательно, гипотеза об отсутствии корреляции должна была быть отвергнута. Заметим, что число наблюдений в данном примере недостаточно велико для уверенного использования данного приближенного критерия. Если к этому добавить тот факт, что выборочное значение статистики критерия находится вблизи границы критической области, то следует заключить, что по имеющимся данным нельзя сделать надежного вывода ни о наличии, ни об отсутствии корреляции.

Отметим, что если бы значение  $r = 0,6$  было получено для  $n = 50$ , то выборочное значение статистики  $u$  было бы равно 4,75, и гипотеза однозначно должна была быть отвергнута не только на уровне значимости 5 %, но и 1 % (и даже более высоким, так как вероятность того, что стандартно распределенная случайная величина примет значение, большее 4,75, равна 0,000001).

### 2.3.7. Критерии согласия

Все рассмотренные ранее критерии принято относить к группе так называемых *параметрических критериев*. Применение этих критериев требует знания вида распределения наблюдаемых случайных величин (нормальное, биномиальное, пуассоновское, двумерное нормальное и др.), и проверяемая гипотеза касается параметров данных распределений. Прежде чем применять параметрические методы, необходимо убедиться в том, что мы действительно имеем дело с распределением требуемого вида.

Предположение о виде распределения случайной величины — статистическая гипотеза, которую можно проверить с помощью

экспериментальных данных. Критерии для проверки согласия между распределением выборочных значений и заданным теоретическим распределением называются **критериями согласия**.

Пусть имеется выборка  $x_1, x_2, \dots, x_n$  значений случайной величины  $\xi$  с неизвестной функцией распределения  $F(x)$ . Требуется проверить гипотезу  $H_0: F(x) = F_0(x)$  о том, что случайная величина имеет некоторое заданное распределение  $F_0(x)$  против альтернативной гипотезы  $H_1: F(x) \neq F_0(x)$ . Распределение  $F_0(x)$  может быть либо задано полностью (простая нулевая гипотеза), либо с точностью до значений параметров (сложная нулевая гипотеза). Во втором случае фактически проверяется принадлежность распределения к заданному виду, например проверяется гипотеза о нормальности. Часто это делается в целях обоснования применения для обработки полученных данных методов, требующих принадлежности распределения к заданному типу.

Рассмотрим два критерия согласия.

**Критерий согласия  $\chi^2$ .** Относительно распределения  $F(x)$  не делается никаких предположений, оно может быть как непрерывным, так и дискретным. Для проверки простой нулевой гипотезы статистика критерия  $\chi^2$  вычисляется следующим образом. Область изменения значений выборки разбивается на  $k$  интервалов с таким расчетом, чтобы число наблюдений  $n_i$ , попавших в  $i$ -й интервал,  $i = 1, \dots, k$ , в большинстве интервалов было не менее 10. Для каждого из интервалов вычисляется также вероятность  $p_i$  попадания в этот интервал случайной величины при условии выполнения гипотезы  $H_0$ . Статистика  $\chi^2$  равна нормированной сумме квадратов отклонений числа наблюдений от гипотетической частоты  $np_i$ :

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}.$$

Для легкости запоминания эту формулу можно рассматривать как сумму квадратов пуассоновских случайных величин  $n_i$ , стандартизованных путем вычитания из них гипотетических средних  $np_i$  и деления разности на их среднеквадратические отклонения  $(np_i)^{1/2}$ .

Если верна гипотеза  $H_0$ , то при достаточно большом  $n$  (не менее 50) распределение данной статистики хорошо приближается  $\chi^2$ -распределением с  $k - 1$  степенями свободы (одна степень свободы вычитается даже при полностью заданном  $F_0(x)$ , поскольку

ку значения  $n_i$  связаны соотношением  $n_1 + n_2 + \dots + n_k = n$ ). Следовательно, критическое множество уровня значимости  $\alpha$  состоит из одного полуинтервала  $[\chi_{k-1,1-\alpha}^2, +\infty)$ , где  $\chi_{k-1,1-\alpha}^2$  — квантиль  $\chi^2$ -распределения с числом степеней свободы  $k - 1$  порядка  $1 - \alpha$ .

Необходимость в проверке простых гипотез возникает относительно редко. Наиболее интересным для практики является случай, когда  $F_0(x)$  известна с точностью до значений  $r$  параметров,  $F_0(x) = F_0(x, \theta_1, \dots, \theta_r)$ , где  $\theta_1, \dots, \theta_r$  — неизвестные параметры. Тогда статистика критерия имеет вид

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i},$$

где  $\hat{p}_i = p_i(\hat{\theta}_1, \dots, \hat{\theta}_r)$  — оценки вероятностей  $p_i$ ,  $i = 1, \dots, k$ ;  $\hat{\theta}_1, \dots, \hat{\theta}_r$  — оценки параметров  $\theta_1, \dots, \theta_r$ , определяемые через наблюдаемые частоты  $n_1, \dots, n_k$ .

Если нулевая гипотеза  $H_0$  верна, то статистика критерия  $\chi^2$  при  $n \rightarrow \infty$  распределена асимптотически как  $\chi^2$  с числом степеней свободы  $k - r - 1$ . Следовательно, критическое множество уровня значимости  $\alpha$  состоит из полуинтервала  $[\chi_{k-r-1,1-\alpha}^2, +\infty)$ , где  $\chi_{k-r-1,1-\alpha}^2$  — квантиль  $\chi^2$ -распределения с числом степеней свободы  $k - r - 1$  порядка  $1 - \alpha$ .

Часто оценки неизвестных параметров  $\hat{\theta}_1, \dots, \hat{\theta}_r$  определяются не по наблюдаемым частотам  $n_i$ , а по всей выборке. Например, при проверке нормальности ожидаемую частоту в  $i$ -м интервале  $n\hat{p}_i$  находят, используя выборочное среднее  $\bar{x}$  и выборочную дисперсию  $s^2$ , определенные по всей выборке. В этом случае статистика критерия  $\chi^2$  при справедливости  $H_0$  не распределена асимптотически как  $\chi_{k-r-1}^2$ , ее функция распределения заключена между функциями распределений  $\chi_{k-1}^2$  и  $\chi_{k-r-1}^2$ . Различием между ними можно пренебречь при больших  $k$ . Но для малых  $k$  полезно убедиться, что выборочное значение статистики критерия  $\chi_{\text{выб}}^2 > \chi_{k-r-1,1-\alpha}^2$  и  $\chi_{\text{выб}}^2 > \chi_{k-1,1-\alpha}^2$ .

**Одновыборочные критерии Колмогорова и Смирнова.** Эти критерии применяют для проверки соответствия непрерывного распределения  $F(x)$  заданному  $F_0(x)$ . Статистика Колмогорова для проверки гипотезы  $H_0: F(x) = F_0(x)$  против альтернативы  $H_1: F(x) \neq F_0(x)$  может служить мерой близости эмпирической функции распределения  $\hat{F}(x)$  и теоретической  $F_0(x)$ :

$$D_n = \sup_x |\hat{F}(x) - F_0(x)|.$$

Статистика Смирнова  $D_n^+$  для проверки гипотезы  $H_0$  против альтернативы  $H_1^+ : F(x) > F_0(x)$  имеет вид

$$D_n^+ = \sup_x [\hat{F}(x) - F_0(x)].$$

Для случая простой нулевой гипотезы распределение статистик  $D_n$  и  $D_n^+$  при справедливости  $H_0$  не зависит от типа  $F_0(x)$ . Если верна нулевая гипотеза, предельным распределением статистики  $\sqrt{n}D_n$  при  $n \rightarrow \infty$  является распределение Колмогорова, а Н. В. Смирнов получил точное предельное распределение статистики  $D_n^+$ . Соответственно критическое множество уровня значимости  $\alpha$  для проверки гипотезы  $H_0$  против альтернативы  $H_1$  состоит из полуинтервала  $[D_{n,1-\alpha}, +\infty)$ , а против альтернативы  $H_1^+$  — из полуинтервала  $[D_{n,1-\alpha}^+, +\infty)$ , где  $D_{n,1-\alpha}$ ,  $D_{n,1-\alpha}^+$  — критические значения статистик  $D_n$  и  $D_n^+$ , соответственно, уровня значимости  $\alpha$ . При  $\alpha \leq 0,2$  имеем  $D_{n,1-\alpha} \approx D_{n,1-\alpha/2}^+$  с большой точностью (большей, чем 0,00005).

В случае сложной нулевой гипотезы, когда  $F_0(x)$  известна с точностью до параметров,  $F_0(x) = F_0(x, \theta_1, \dots, \theta_r)$ , где  $\theta_1, \dots, \theta_r$  — неизвестные параметры, статистика критерия для проверки гипотезы  $H_0$  против двусторонней альтернативы имеет вид:

$$\hat{D}_n = \sup_x \left| \hat{F}(x) - F_0(x, \hat{\theta}_1, \dots, \hat{\theta}_r) \right|,$$

где  $\hat{\theta}_1, \dots, \hat{\theta}_r$  — оценки неизвестных параметров.

При условии, что нулевая гипотеза  $H_0$  верна, распределение статистики  $\hat{D}_n$  (и  $\hat{D}_n^+$ ) уже зависит от конкретного вида распределения  $F_0(x, \theta_1, \dots, \theta_r)$ . Для некоторых типов распределений — нормального, показательного, логистического — Лиллиефорсом получены таблицы критических значений статистики  $\hat{D}_n$  при условии, что гипотеза  $H_0$  верна [Н. W. Lilliefors, 1967; R. R. Sokal, F. J. Rohlf, 1995]. Соответственно критическое множество уровня значимости  $\alpha$  для проверки гипотезы  $H_0$  против альтернативы  $H_1$  состоит из одного полуинтервала  $[\hat{D}_{n,1-\alpha}, +\infty)$ , где  $\hat{D}_{n,1-\alpha}$  — критическое значение статистики  $\hat{D}_n$  для заданных  $\alpha$ ,  $n$  и  $F_0(x)$ .

Статистика  $\hat{D}_n$  может быть преобразована к виду, практически не зависящему от  $n$ . Например, для нормального распределения получено следующее выражение для модифицированной формы статистики Колмогорова  $\hat{D}^*$  [Ю. Н. Тюрин, 1978]

$$\hat{D}^* = \hat{D}_n(\sqrt{n} - 0,01 + 0,85/\sqrt{n}).$$

Это дает возможность проводить проверку гипотезы практически при всех  $n$ , зная значения  $\hat{D}_{1-\alpha}^*$  для небольшого набора значений  $\alpha$ . В частности, для  $\alpha = 0,05$  имеем  $\hat{D}_{0,95}^* = 0,895$ .

При проверке гипотезы о нормальности распределения с неизвестными средним и дисперсией критерии Колмогорова и Смирнова являются более мощными, чем критерий  $\chi^2$ .

Заметим, что в англоязычной литературе критерии Колмогорова и Смирнова называют одновыборочным критерием Колмогорова — Смирнова, а статистики  $D_n$  и  $D_n^+$  — одновыборочными статистиками Колмогорова — Смирнова.

Среди других критериев согласия отметим критерий Шапиро — Уилка для проверки нормальности [Г. Хан, С. Шапиро, 1969].

**Пример 2.6.** Пусть получена следующая выборка 50 значений случайной величины  $\xi$  с неизвестным распределением:

45	89	93	40	91	60	2	59	87	78
57	39	50	0	35	91	67	62	25	93
19	98	55	78	34	45	86	31	15	95
50	52	35	66	0	44	93	36	29	44
17	85	17	63	34	43	100	75	84	9

Проверим гипотезу о том, что эта случайная величина имеет нормальное распределение. После разбиения области изменения выборочных значений на пять равных интервалов получаем следующие эмпирические и гипотетические частоты:

Интервал.....	$(-\infty, 20]$	$(20, 40]$	$(40, 60]$	$(60, 80]$	$(80, \alpha)$
Эмпирическая частота $n_i$ .....	8	10	12	7	13
Гипотетическая частота $np_i$ .....	6,1	9,7	13,4	11,6	9,2

Гипотетические частоты вычислялись для нормального распределения  $N(\mu = 53,3, \sigma^2 = 840,4)$  с параметрами, оцененными по выборке. Соответственно число степеней свободы статистики критерия равно  $5 - 1 - 2 = 2$ . Выборочное значение статистики  $\chi^2 = 4,2$ , а критическая граница  $\chi_{2;0,95}^2 = 6,0$ , т. е. критическое множество уровня значимости  $\alpha = 0,05$  состоит из полуинтервала  $[6, 0; +\infty)$ . Следовательно, нет оснований отвергнуть гипотезу о нормальности.

В действительности выборка была получена с помощью датчика случайных чисел, равномерно распределенных на отрезке  $[0, 100]$ , т. е. мы видим, что при данном числе наблюдений



(в общем-то, конечно, небольшом для проверки гипотезы о типе распределения) критерий  $\chi^2$  не обнаруживает отклонения от нормальности.

Выборочное значение статистики критерия Колмогорова — Смирнова равно  $D_5 = 0,11$ , что также не выходит за 5%-й предел этого критерия в предположении, что гипотетические средние равны выборочным. Однако в случае неизвестных параметров гипотетического нормального распределения лучше пользоваться модификацией критерия Колмогорова — Смирнова [H. W. Lilliefors, 1967]. Но  $\hat{D}_5 = 0,11 < D_{5;0,95}^* = 0,895$ , что не дает оснований отвергнуть гипотезу и по этому критерию.

**Пример 2.7.** Расчеты, аналогичные предыдущим, проведенные для выборки объема 150 значений случайной величины, равномерно распределенной на отрезке  $[0, 100]$ , дали значение  $\chi^2 = 7,6$ , что меньше  $\chi_{4;0,95}^2 = 9,5$ . Следовательно, нет оснований отвергнуть гипотезу о нормальности по критерию  $\chi^2$  и в этом случае. По критерию Колмогорова — Смирнова (Лиллиефорса) гипотеза отвергалась на уровне 1 %. Полученные результаты показывают, что критерий Колмогорова — Смирнова (Лиллиефорса) является более мощным, чем критерий  $\chi^2$  при проверке гипотезы о нормальности распределения с неизвестными средним и дисперсией.

**Пример 2.8.** Расчеты статистик критериев согласия для данных примера 2.1, содержащих 50 выборочных значений длины лепестка ириса разноцветного, приводят к значению статистики  $\chi^2$ , равному 2,1, и значению статистики  $\hat{D}_n$ , равному 0,117. В этом случае гипотеза о нормальности не отвергается ни критерием  $\chi^2$ , ни критерием Колмогорова — Смирнова — Лиллиефорса.

**Пример 2.9.** В некоторых классических экспериментах с селекцией гороха Г. Мендель наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Они приведены в табл. 2.4 вместе с теоретическими вероятностями, вычисленными в соответствии с теорией наследственности Менделя [М. Дж. Кендалл, А. Стьюарт, 1973].

В этом случае теоретическое распределение дискретно и известно полностью. Для проверки согласия экспериментальных данных теоретическому распределению используем критерий  $\chi^2$  для простой гипотезы. Значение статистики, вычисленное по вы-

## Результаты эксперимента Г. Менделя

Семена	Численность	
	эмпирическая	ожидаемая
Круглые и желтые	315	$556 \cdot \frac{9}{16} = 313$
Морщинистые и желтые	101	$556 \cdot \frac{3}{16} = 104$
Круглые и зеленые	108	$556 \cdot \frac{3}{16} = 104$
Морщинистые и зеленые	32	$556 \cdot \frac{1}{16} = 35$
Всего	556	556

борке, равно  $\sum_{i=1}^4 \frac{(n_i - np_i)^2}{np_i} = 0,47$ , что меньше 5%-го критического значения  $\chi_{3;0,95}^2 = 7,81$ ,  $p = 0,9$ . Следовательно, теория наследственности Менделя не противоречит полученным экспериментальным данным.

Наряду с количественными статистическими критериями для определения вида распределения по выборочным данным используются *графические методы*.

Простейший способ — построение по имеющейся выборке гистограммы относительных частот и на том же графике и в том же масштабе — кривой плотности нормального распределения с выборочным средним и выборочной дисперсией в качестве параметров. Значительные отклонения от нормальности (сильная асимметрия, бимодальность) легко обнаруживаются на графике.

**Пример 2.10.** Применим графический метод к рассмотренной ранее модельной выборке объема  $n = 50$  (см. пример 2.6), извлеченной из равномерного распределения. На рис. 2.7 приведены гистограмма и кривая нормальной плотности. Визуально заметно отклонение от нормальности в пользу равномерности (хотя, как было показано, статистически значимо при таком числе наблюдений оно не подтверждается).

С точки зрения визуального обнаружения отклонений от нормальности сравнение эмпирической и гипотетической функ-

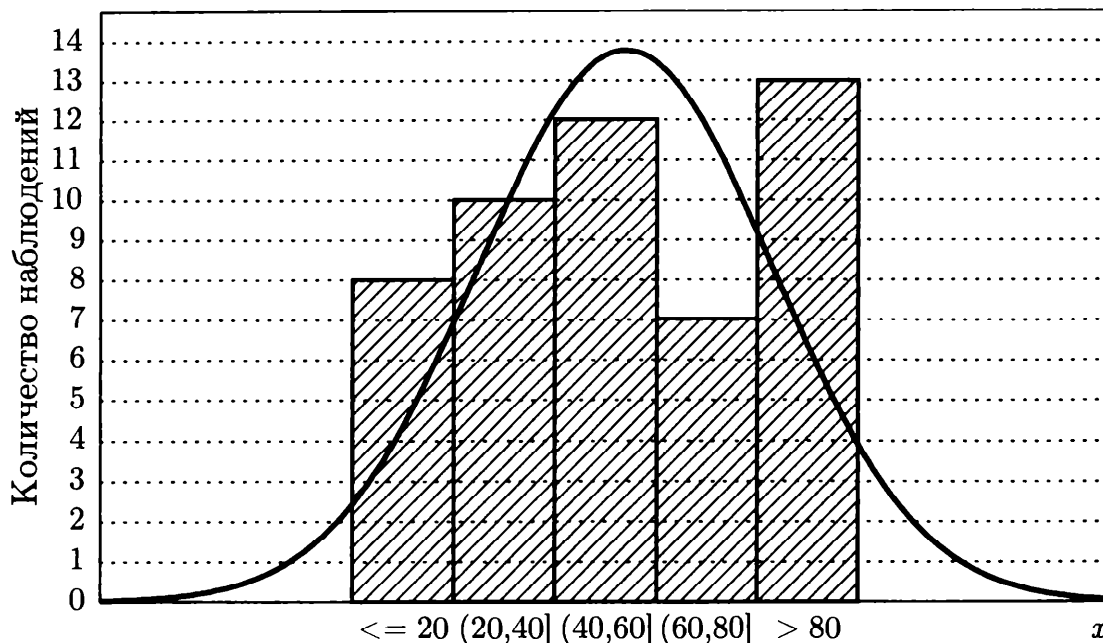


Рис. 2.7. Сравнение гистограммы и кривой нормальной плотности

ций распределения гораздо менее наглядно, чем сравнение гистограммы с графиком плотности. Однако обычно сравнивают не сами функции распределения, а обратные нормальные преобразования от них, так называемые *пробит-графики*. Пробит-график теоретической нормальной функции распределения представляет собой прямую, а пробит-график эмпирической функции распределения тем ближе к прямой, чем ближе она к нормальной. Этот прием позволяет на первом этапе анализа данных выявить их особенности, выдвинуть гипотезы о характере распределения, решить вопрос о целесообразности замены переменной.

### 2.3.8. Непараметрические критерии

Предположим, что для конкретной выборки мы отклоняем гипотезу о нормальном распределении генеральной совокупности. Тогда для получения статистических выводов применяют разные способы. Если объем выборки достаточно велик, то можно использовать параметрические критерии (см. 2.3.2, 2.3.3) как приближенные. Другой способ состоит в подборе замены переменной, приводящей к нормальному распределению. Третий способ — применение *непараметрических критериев* (часто употребляют термин «критерий, свободный от распределения»). Эти критерии не требуют знания вида исходного распределения  $F(x)$ , за исключением предположения о том, что это распределение непрерывно. В качестве статистик этих критериев используются такие функции от наблюдений,

распределение которых, если верна нулевая гипотеза, не зависит от  $F(x)$ . Конечно, нельзя ожидать, что непараметрические критерии будут обладать такой же чувствительностью к различению нулевой и альтернативной гипотез, как и критерии, основанные на знании вида  $F(x)$ . Более подробно с методами непараметрической статистики можно познакомиться в работах [Ю. Н. Благовещенский, В. П. Самсонова, Е. А. Дмитриев, 1989; М. Дж. Кендалл, А. Стьюарт, 1973; Ю. Н. Тюрин, 1978; М. Холлендер, Д. А. Вульф, 1983; J. H. Zar, 1999].

Наибольшее развитие и применение в задачах непараметрической проверки гипотез получили ранговые методы. Статистики таких критериев основаны не непосредственно на численных значениях наблюдений  $x_1, x_2, \dots, x_n$ , а на их **рангах**, т. е. на порядковых номерах наблюдений при их упорядочении по возрастанию (в вариационном ряду),  $R(x_1), R(x_2), \dots, R(x_n)$ . Хотя непараметрические критерии применяются для непрерывных распределений, среди выборочных значений могут быть совпадающие, например из-за неизбежных ошибок округления. В этом случае для определения рангов равных величин используются различные методы. **Метод случайного ранга** состоит в том, что ранги совпадающих наблюдений можно взять в произвольном порядке. Например, для того чтобы решить, какой из двух равных величин  $x_i$  или  $x_k$  приписать больший ранг, можно бросить монету. **Метод среднего ранга** каждому из совпадающих наблюдений приписывает ранг, равный среднему арифметическому их порядковых номеров в вариационном ряду. Например, равным величинам  $x_i = x_k$ , которые в случае их неравенства имели бы ранги  $r$  и  $r + 1$ , приписывается средний ранг  $r + 1/2$ . Третий метод заключается в том, чтобы отбросить равные наблюдения. Ранги наблюдений, будучи функциями выборочных значений, являются случайными величинами с возможными значениями  $1, 2, \dots, n$ . Оказалось, что набор рангов  $R(x_1), R(x_2), \dots, R(x_n)$  содержит значительную долю информации о распределении наблюдаемой случайной величины.

Если статистика рангового критерия  $g(R_1, R_2, \dots, R_n)$  — дискретная случайная величина, то для заданного уровня значимости  $\alpha$  может не существовать значения квантили распределения статистики критерия при справедливости нулевой гипотезы. Поэтому для определения критического множества используются верхнее критическое значение статистики критерия  $g_{1-\alpha}$ , равное наименьшему значению квантили распределения статисти-

ки критерия  $g$ , такому, что  $P(g \geq g_{1-\alpha}) \leq \alpha$ , и нижнее критическое значение  $g_\alpha$ , равное наибольшему значению квантили распределения статистики критерия  $g$ , такому, что  $P(g \leq g_\alpha) \leq \alpha$ . Значения  $g_\alpha$  и  $g_{1-\alpha}$  находят по таблицам. Для всех рассматриваемых далее критериев существуют таблицы критических значений статистики [Л. Н. Большев, Н. В. Смирнов, 1983].

**Замечание.** Ранговые критерии применяются и в тех случаях, когда наблюдения не являются количественными, но допускают упорядочение, что часто имеет место в исследованиях по биологии, медицине, психологии и социологии.

### **Одновыборочные непараметрические критерии**

Одновыборочные критерии предназначены для проверки гипотезы о равенстве медианы заданному значению. Пусть имеется выборка  $x_1, x_2, \dots, x_n$  значений случайной величины  $\xi$  с неизвестной непрерывной функцией распределения  $F(x, Me)$ , где  $Me$  — неизвестная медиана. Требуется проверить гипотезу  $H_0: Me = Me_0$ , о равенстве медианы заданному числу  $Me_0$ . Рассмотрим два наиболее часто используемых критерия для проверки этой гипотезы.

**Критерий знаков.** В качестве статистики критерия  $n^+$  применяют число положительных разностей  $x_i - Me_0$ ,  $i = 1, \dots, n$ . Если верна нулевая гипотеза  $H_0$ , то  $P(x_i > Me_0) = P(x_i < Me_0) = 1/2$  и, следовательно, статистика критерия  $n^+$  — дискретная случайная величина, распределенная по биномиальному закону с параметрами  $n$  и  $p = 1/2$ . Критическая область уровня значимости  $\alpha$  против двусторонней альтернативы  $H_1: Me \neq Me_0$  будет состоять из двух интервалов  $[0, n_{\alpha/2}^+]$  и  $[n_{1-\alpha/2}^+, n]$ , причем  $n_{\alpha/2}^+ = n - n_{1-\alpha/2}^+$ , так как распределение статистики критерия  $n^+$  симметрично относительно своего среднего  $n/2$ , где  $n_{\alpha/2}^+, n_{1-\alpha/2}^+$  — нижнее и верхнее критические значения статистики  $n^+$  порядка  $\alpha/2$  соответственно. Критическая область против односторонней альтернативы  $H_1: Me > Me_0$  имеет вид  $[n_{1-\alpha}^+, n]$  и против односторонней альтернативы  $H_1: Me < Me_0$  —  $[0, n_\alpha^+]$ . При  $n \rightarrow \infty$  случайная величина  $n^+$  распределена асимптотически нормально,  $n^+ \sim N(n/2, n/4)$ , и для нахождения критических значений можно воспользоваться нормальным приближением.

**Замечание.** Критерий знаков обладает недостаточной чувствительностью к различению нулевой и альтернативной гипотез. Но из-за

простоты и наглядности он часто используется для предварительного анализа данных. Более мощным является критерий знаковых рангов.

**Критерий знаковых рангов.** Упорядочим по возрастанию разности  $x_i - Me_0$  и обозначим через  $R(x_i - Me_0)$  — ранг  $i$ -й разности,  $i = 1, \dots, n$ . Статистика критерия знаковых рангов равна сумме рангов положительных разностей  $x_i - Me_0$ :

$$T^+ = \sum_{(x_i - M_0) > 0} R(x_i - Me_0).$$

Если нулевая гипотеза  $H_0$  верна, то вероятность каждого из возможных  $2^n$  исходов для набора рангов положительных разностей равна  $(1/2)^n$ , что и определяет распределение статистики критерия для заданного  $n$ , симметричное относительно среднего  $n(n+1)/4$ , откуда следует, что  $T_{1-\alpha}^+ = n(n+1)/2 - T_\alpha^+$ , где  $T_\alpha^+, T_{1-\alpha}^+$  — нижнее и верхнее критические значения статистики критерия для заданных  $\alpha$  и  $n$  соответственно. Критическая область уровня  $\alpha$  против двусторонней альтернативы будет состоять из двух интервалов  $[0, T_{\alpha/2}^+]$  и  $[T_{1-\alpha/2}^+, n(n+1)/2]$  против правосторонней альтернативы — из одного интервала  $[T_{1-\alpha}^+, n(n+1)/2]$  и против левосторонней альтернативы — также из одного интервала  $[0, T_\alpha^+]$ .

Если верна гипотеза  $H_0$ , то при  $n \rightarrow \infty$  распределение статистики критерия стремится к нормальному распределению,  $T^+ \sim N(n(n+1)/4, n(n+1)(2n+1)/24)$ . При  $n > 25$  этим приближением можно пользоваться для определения критических значений статистики.

### ***Проверка гипотезы об отсутствии сдвига***

Пусть имеются выборки  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_m$  значений случайных величин  $\xi$  и  $\eta$  с неизвестными непрерывными функциями распределения  $F(x)$  и  $G(x)$  одинаковой формы, которые могут различаться лишь параметром сдвига  $\Delta$ , т. е.  $G(x) = F(x - \Delta)$ . Требуется проверить гипотезу  $H_0: \Delta = 0$  об отсутствии сдвига между распределениями случайных величин  $\xi$  и  $\eta$ .

**Случай независимых выборок.** Пусть  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_m$  — независимые выборки.

**Критерий Манна — Уитни.** Пусть  $n \leq m$ , в противном случае выборки поменяем местами. Упорядочим  $n + m$  наблюдений по возрастанию и обозначим через  $R_i$  ранг  $i$ -го наблюдения в объединенном ряду наблюдений,  $i = 1, \dots, n + m$ . Если есть сов-

падающие значения внутри какой-либо из выборок, то их ранги можно взять в произвольном порядке. Если же совпадают значения, принадлежащие разным выборкам, то для определения их рангов применяется метод среднего ранга (см. 2.3.8).

В качестве статистики Манна — Уитни  $U$  используется общее число случаев (инверсий) в упорядоченной по возрастанию последовательности из  $x$  и  $y$ , в которых  $x$  появляется позднее некоторого  $y$ :

$$U = \sum_{i=1}^n \sum_{j=1}^m \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1 & \text{при } x_i > y_j; \\ 0 & \text{при } x_i < y_j. \end{cases}$$

Если  $x_i = y_j$ , то к значению  $U$  прибавляется  $1/2$ . Статистика  $U$  — дискретная случайная величина, принимающая значения от 0 до  $nm$ .

Если нулевая гипотеза верна, то  $C_{n+m}^m$  последовательностей из  $x$  и  $y$  являются равновероятными, что и определяет распределение статистики  $U$ , симметричное относительно своего среднего  $nm/2$ . Критическая область уровня  $\alpha$  против двусторонней альтернативы  $H_1: \Delta \neq 0$  будет состоять из двух интервалов  $[0, U_{\alpha/2}]$  и  $[U_{1-\alpha/2}, nm]$ , где  $U_{\alpha/2}$  и  $U_{1-\alpha/2}$  — нижнее и верхнее критические значения статистики  $U$  порядка  $\alpha/2$  соответственно. Критическая область против односторонней альтернативы  $H_1: \Delta > 0$  — из одного интервала  $[U_{1-\alpha}, nm]$  и против односторонней альтернативы  $H_1: \Delta < 0$  — из одного интервала  $[0, U_{\alpha}]$ . Заметим, что нижнее и верхнее критические значения статистики  $U$  связаны соотношением  $U_{1-\alpha} = mn - U_{\alpha}$ .

При малых  $n$  и  $m$  критическое значение  $U_{\alpha}$  определяется непосредственным подсчетом последовательностей с наименьшим количеством инверсий. Если гипотеза  $H_0$  верна, то при  $n \rightarrow \infty, m \rightarrow \infty$  статистика  $U$  распределена асимптотически нормально:  $U \sim N(nm/2, nm(n+m+1)/12)$ .

**Критерий Вилкоксона.** Статистику критерия  $U$  можно также вычислить по формуле

$$U = W - \frac{n(n+1)}{2},$$

где  $W = \sum_{j=1}^n R_j$ , сумма рангов наблюдений  $x_j, j = 1, \dots, n$ , —

статистика критерия Вилкоксона. Следовательно, критерии, основанные на статистиках  $U$  и  $W$ , эквивалентны.

**Пример 2.11.** Пусть получены выборки значений двух случайных величин  $\xi$  и  $\eta$  объема  $n = 4$  и  $m = 5$ :

$\xi$ .....	174	175	183	174	
$\eta$ .....	187	185	185	179	181

Составим из них общий вариационный ряд (т. е. расположим в порядке возрастания), сохранив информацию о принадлежности к выборке:

Ранг	1	2	3	4	5	6	7	8	9
$\xi$	174	174	175			183			
$\eta$				179	181		185	185	187

Сумма рангов выборки значений случайной величины  $\xi$  равна  $W = 1 + 2 + 3 + 6 = 12$ . Это значение не выходит за двусторонние критические пределы  $W_{0,025} = 11$  и  $W_{0,975} = 29$  уровня значимости 5%. Выборочное значение статистики  $U = 2$  и соответствующее ему  $p = 0,032 < \alpha = 0,05$  и, следовательно, нет оснований отвергнуть нулевую гипотезу о том, что сдвиг между распределениями  $F(x)$  и  $G(x)$  отсутствует.

Заметим, что приведенные данные были получены с помощью датчика нормально распределенных случайных чисел  $\xi \sim N(\mu = 175, \sigma^2 = 25)$  и  $\eta \sim N(\mu = 180, \sigma^2 = 25)$ . Приведенные значения могли бы быть, например, значениями роста четырех случайно выбранных французов и пяти случайно выбранных норвежцев (средний рост взрослых мужчин Франции и Норвегии равен 175 и 180 см соответственно), т. е. в действительности сдвиг  $\Delta$  между распределениями отличен от нуля (он равен  $\Delta = 175 - 180 = -5$ ) и гипотеза  $H_0: \Delta = 0$  неверна, но критерии Вилкоксона и Манна — Уитни не обнаружили различия между распределениями  $F(x)$  и  $G(x)$ . Если применить к приведенным данным  $t$ -критерий Стьюдента для сравнения математических ожиданий двух нормально распределенных случайных величин с неизвестными дисперсиями, то получим выборочное значение  $t = -2,72$  для статистики критерия. Поскольку это значение выходит за 5%-е критические пределы  $t_{7;0,025} = -2,36$  и  $t_{7;0,975} = 2,36$   $t$ -распределения с  $4 + 5 - 2 = 7$  степенями свободы, то гипотеза о равенстве математических ожиданий должна быть отвергнута. Это типичная ситуация — непараметрические критерии обладают меньшей мощностью по сравнению с аналогичными параметрическими критериями, использующими дополнительную информацию о наблюдаемых случайных вели-



чинах. Поэтому, если имеется достоверная дополнительная информация, то предпочтительнее использовать критерий, учитывающий эту информацию.

**Критерий Ван-дер-Вардена.** Обозначим  $N = n + m$ . Статистика критерия имеет вид

$$X = \sum_{i=1}^n \Psi \left( \frac{R(x_i)}{N + 1} \right),$$

где  $R(x_i)$  — ранг наблюдения  $x_i$ ;  $\Psi(p)$  —  $p$ -квантиль стандартного нормального распределения.

Критерий Ван-дер-Вардена является наиболее мощным непараметрическим критерием для проверки гипотезы об отсутствии сдвига. Если обе выборки извлечены из нормальных совокупностей, то при постоянном  $n$  и  $m \rightarrow \infty$  критерий Ван-дер-Вардена имеет такую же мощность, как и двухвыборочный  $t$ -критерий.

**Случай связанных выборок.** Пусть  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_n$  — связанные выборки, например каждая пара наблюдений  $(x_i, y_i), i = 1, \dots, n$ , принадлежит одному объекту, либо  $(x_i, y_i)$  попарно связаны тем, что условия проведения наблюдений менялись от опыта к опыту, но для каждой пары  $(x_i, y_i)$  оставались постоянными, что в практике биологического эксперимента встречается очень часто. Обозначим через  $z_i = x_i - y_i, i = 1, \dots, n$ . Тогда проверка гипотезы об отсутствии сдвига между  $F(x)$  и  $G(x)$  сводится к проверке гипотезы о равенстве нулю медианы для выборки  $z_1, z_2, \dots, z_n$ . Для этого можно применить **критерий знаков** или **критерий знаковых рангов**, рассмотренные ранее.

**Критерии однородности.** Критерии Манна — Уитни (Вилкоксона) и Ван-дер-Вардена позволяют обнаруживать лишь различия в центральных тенденциях непрерывных распределений двух случайных величин. Если важно обнаружить любые расхождения в форме непрерывных распределений, то пользуются **критериями однородности**, например **двухвыборочным критерием Смирнова**. С помощью этого критерия проверяется гипотеза  $H_0: F_\xi(x) = F_\eta(x)$  о том, что функции распределения  $F_\xi(x)$  и  $F_\eta(x)$  случайных величин  $\xi$  и  $\eta$  идентичны против альтернативной гипотезы  $H_1: F_\xi(x) \neq F_\eta(x)$  о том, что они различны.

Статистика критерия Смирнова  $D_{m,n}$  определяется как максимум модуля разности между эмпирической функцией  $\hat{F}_\xi(x)$ , построенной по выборке  $x_1, x_2, \dots, x_n$ , и эмпирической функцией  $\hat{F}_\eta(x)$ , построенной по выборке  $y_1, y_2, \dots, y_m$ ,

$$D_{m,n} = \max_x |\hat{F}_\xi(x) - \hat{F}_\eta(x)|.$$

При справедливости гипотезы  $H_0$  статистика  $\sqrt{\frac{mn}{m+n}} D_{m,n}$  имеет асимптотическое (при  $\min(m, n) \rightarrow \infty$  так, что отношение  $m/n$  остается постоянным) распределение Колмогорова. Критическая область уровня значимости  $\alpha$  против двусторонней альтернативы  $H_1$  будет состоять из одного полуинтервала  $[D_{m,n,1-\alpha}, \infty)$ , где  $D_{m,n,1-\alpha}$  — квантиль распределения статистики  $D_{m,n}$  порядка  $1 - \alpha$ .

*Замечание.* В англоязычной литературе критерий однородности двух выборок Смирнова называют *двухвыборочным критерием Колмогорова — Смирнова*.

**Проверка гипотезы о независимости.** Пусть имеется двумерная выборка  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  из неизвестного двумерного распределения. Причем наблюдаемые признаки могут быть как количественными, так и порядковыми. Найдем ранги  $R(x_i)$  и  $R(y_i)$  в последовательностях  $x_1, x_2, \dots, x_n$  и  $y_1, y_2, \dots, y_n$ , упорядоченных по отдельности. Мерой зависимости двух случайных величин является *коэффициент ранговой корреляции Спирмена*, определяемый формулой

$$r_s = 1 - 6 \sum_{i=1}^n d_i^2 / n(n^2 - 1),$$

где  $d_i = R(x_i) - R(y_i)$ .

Коэффициент  $r_s$  принимает значение от  $-1$  до  $+1$ , причем  $r_s = +1$ , когда  $R(x_i) = R(y_i)$ , и  $r_s = -1$ , когда последовательности рангов полностью противоположны, т. е.  $R(x_i) = (n + 1) - R(y_i)$ ,  $i = 1, \dots, n$ . Для проверки гипотезы независимости признаков,  $H_0$ , против альтернативы, что признаки зависимы, в качестве статистики критерия используется  $r_s$ . Если  $H_0$  верна, то распределение статистики  $r_s$  симметрично относительно  $0$  с  $M(r_s) = 0$  и  $D(r_s) = 1/(n - 1)$ . Критическая область уровня значимости  $\alpha$  имеет вид  $|r_s| \geq r_{s,1-\alpha/2}$ , где  $r_{s,1-\alpha/2}$  — верхнее критическое значение статистики критерия  $r_s$  порядка  $\alpha/2$ .

Если  $H_0$  верна, то при  $n \rightarrow \infty$  статистика  $\sqrt{n-1} r_s$  распределена асимптотически нормально с параметрами  $(0, 1)$ . При  $4 \leq n \leq 10$  критические значения статистики критерия находят по таблицам точного распределения  $r_s$ , а при  $n > 10$  пользуются нормальной аппроксимацией.

## АНАЛИЗ МНОГОМЕРНЫХ ДАННЫХ

3.1. Классификация методов анализа  
многомерных данных

В предыдущей главе рассматривались методы статистического анализа данных, описываемых в основном одной переменной. Здесь приведем методы, позволяющие анализировать данные, описываемые любым числом переменных, т. е. так называемые *многомерные данные*.

Несмотря на большое разнообразие методов анализа многомерных данных, имеется четкая система представления как самих данных, так и результатов их анализа. Данные представляются в виде прямоугольной таблицы (матрицы), строки которой соответствуют различным ситуациям (*наблюдениям*), а столбцы — *переменным*, наблюдаемым в этих ситуациях; результаты — в виде функций, выражающих одни переменные (называемые *зависимыми переменными*, или *откликами*) через другие (называемые *независимыми переменными*, или *факторами*).

Такое представление исходных данных и результатов анализа чрезвычайно удобно, поскольку предлагает ясную схему для сбора данных, их обработки и интерпретации полученных результатов.

Приведенная на рис. 3.1 классификация методов анализа данных проведена по двум основаниям: отсутствию (правая часть схемы) или наличию (левая часть схемы) независимых переменных, а также по типу зависимых и независимых переменных, которые могут быть *качественными* или *количественными*. Деление по первому основанию определяет содержательную постановку задачи, тогда как дальнейшее деление методов по типу зависимых и независимых переменных носит скорее технический характер, детализируя математическую процедуру ее решения.

В случае априорного разделения переменных на зависимые и независимые (левая часть схемы) задача анализа состоит в получении описания зависимости  $Y$  от  $X$ . Выбор метода решения зависит прежде всего от того, являются качественными или ко-



Рис. 3.1. Классификация методов анализа данных

личественными зависимые переменные  $Y$ . Окончательное решение о выборе метода анализа данных принимается в зависимости от типа независимых переменных  $X$ .

Наиболее часто на практике для установления связи между независимыми и зависимыми переменными применяют регрессионный и дисперсионный анализ. В обоих случаях откликами служат количественные переменные, однако факторы в регрессионном анализе количественные, а в дисперсионном — качественные.

В *регрессионном анализе* наиболее явно виден функциональный характер модели анализа данных. Задача регрессионного анализа прямо формулируется как задача поиска функциональной зависимости  $Y$  от  $X$ , причем задача поиска формы связи не менее важна, чем вопросы статистической значимости полученных результатов.

Наиболее широко применяется модель множественного линейного регрессионного анализа, позволяющая получать аналитически все стандартные статистические оценки.

В более сложных ситуациях (например, при использовании пошаговых процедур) для получения статистических выводов приходится применять методы, основанные на стохастическом моделировании, такие как случайная пермутация или бут-стрэп.

Задачей *дисперсионного анализа* является установление связи между независимыми качественными переменными и зависимыми количественными. Но поскольку функциональная структура связи очень проста — отклики представляются как линейные комбинации бинарных переменных — уровней факторов, то основное внимание в дисперсионном анализе уделяется вопросам статистической значимости влияния отдельных факторов.

Если отклики  $Y$  качественные, то для анализа применяют группу методов, известную под общим названием «*распознавание образов*». Наиболее используемым методом распознавания в случае количественных факторов является дискриминантный анализ.

Примерами методов распознавания, ориентированных на случай качественных факторов, могут служить сегментационный анализ и метод обобщенного портрета.

Целью *дискриминантного анализа* является получение правила, позволяющего на основе наблюдаемых значений количественных независимых переменных  $X$  предсказывать значение качественной переменной  $Y$ , указывающей на принадлежность наблюдения к одному из заданных классов.

*Сегментационный анализ* состоит в последовательном разбиении совокупности наблюдений в целях получения в конечном итоге групп, максимально однородных по классовому составу.

Случай отсутствия зависимых переменных (правая часть рис. 3.1) предполагает, что все анализируемые переменные в некотором смысле равноправны, и мы принимаем их за отклики (для простоты будем считать их количественными), значения которых определяются какими-то неизвестными нам факторами.

Примерами могут служить морфологические или генетические характеристики растений, животных или людей, принадлежащих определенному таксону или обитающих на определенной территории. Задача анализа состоит в поиске неизвестных факторов. Выбор метода решения зависит от того, считаем ли мы искомые факторы качественными или количественными.

Для поиска качественных факторов применяется группа методов, известная под названием *кластерный анализ*, среди которых наиболее часто используется так называемый *агломеративно-иерархический метод*, основанный на последовательном объединении многомерных наблюдений сначала в мел-

кие, а затем во все более крупные группы. Результатом кластерного анализа является разбиение всей совокупности наблюдений на классы. Полученной классификации соответствует качественная переменная (или группа переменных, если используются несколько классификаций разной степени дробности или пересекающиеся классификации), категориями которой служат номера классов. Именно эта переменная (или переменные) и будет искомым качественным фактором. Найдя такой фактор (классифицирующую переменную), мы получаем возможность объяснять сходство или различие в значениях откликов для разных наблюдений принадлежностью их к одному или разным классам.

Если же неизвестные факторы ищутся в форме количественных переменных, то применяются методы **факторного анализа**. В этом случае задача состоит в представлении имеющихся откликов  $Y$  в виде линейных комбинаций неизвестных количественных факторов  $X$ . С практической точки зрения применение этого метода оправдано, если удастся с достаточной степенью приближения выразить большое количество откликов через малое число факторов. Одним из наиболее часто используемых методов этого класса является **метод главных компонент**, основанный на ортогональном проектировании исходного многомерного пространства в пространство меньшей размерности, в котором точки-наблюдения имеют наибольший разброс. Метод позволяет записать исходные данные в более компактном виде с сохранением максимума содержащейся в них информации и даже представить их графически на плоскости для случая двух факторов.

Следует еще раз подчеркнуть, что основным является разделение методов анализа на те, в которых переменные делятся на зависимые и независимые (анализ связи), и те, в которых такого деления нет (анализ факторов). Дальнейшее дробление методов по типу откликов и факторов довольно относительно. Дело в том, что уровни качественных факторов можно рассматривать как бинарные переменные, которые, в свою очередь, можно считать количественными переменными со значениями 0 и 1. С другой стороны, непрерывную шкалу значений количественной переменной можно категоризовать и рассматривать эту переменную как качественную. Во всяком случае, такого рода преобразования приходится делать вынужденно, когда по типу различаются не только факторы и отклики, но и разные переменные среди факторов или среди откликов.

## 3.2. Матричная алгебра

Адекватным математическим аппаратом для описания методов статистического анализа многомерных данных является матричная алгебра. Напомним ее основные понятия.

**Матрицей** размера  $n \times m$  (или  $n \times m$ -матрицей) называется любая таблица из  $n$  строк и  $m$  столбцов. Если  $n = m$ , то матрица называется **квадратной**, а число  $n = m$  — ее порядком. Например, матрицей является таблица многомерных статистических данных, содержащая значения  $m$  переменных для  $n$  наблюдений:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}.$$

Другие примеры — **выборочная корреляционная матрица** порядка  $m$ , образованная выборочными коэффициентами корреляции, вычисленными для всех пар столбцов матрицы  $\mathbf{X}$ :

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mm} \end{pmatrix},$$

и **выборочная ковариационная матрица** порядка  $m$ , образованная выборочными ковариациями, вычисленными для всех пар столбцов матрицы  $\mathbf{X}$ :

$$\hat{\mathbf{C}} = \begin{pmatrix} \widehat{\text{COV}}_{11} & \widehat{\text{COV}}_{12} & \dots & \widehat{\text{COV}}_{1m} \\ \widehat{\text{COV}}_{21} & \widehat{\text{COV}}_{22} & \dots & \widehat{\text{COV}}_{2m} \\ \dots & \dots & \dots & \dots \\ \widehat{\text{COV}}_{m1} & \widehat{\text{COV}}_{m2} & \dots & \widehat{\text{COV}}_{mm} \end{pmatrix}.$$

Теоретическая **ковариационная матрица** состоит из ковариаций случайных величин, соответствующих столбцам матрицы данных,

$$\mathbf{C} = \begin{pmatrix} \text{COV}_{11} & \text{COV}_{12} & \text{COV}_{1m} \\ \text{COV}_{21} & \text{COV}_{22} & \text{COV}_{2m} \\ \text{COV}_{m1} & \text{COV}_{m2} & \dots & \text{COV}_{r_{mm}} \end{pmatrix}.$$

Матрица из одного столбца называется также **вектор-столбцом**, или просто **вектором**. Например, совокупность наблюдений зависимой переменной  $Y$  можно представить в виде вектор-столбца

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Для матриц одного размера определена операция **сложения** — элементы матрицы суммы равны сумме соответствующих элементов матриц-слагаемых:

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & & a_{2m} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & & b_{1m} \\ b_{21} & b_{22} & & b_{2m} \\ \dots & \dots & & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{pmatrix} = \\ &= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & & a_{2m} + b_{2m} \\ \dots & \dots & & \dots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & & a_{nm} + b_{nm} \end{pmatrix}. \end{aligned}$$

**Умножение матрицы на число** определяется как умножение каждого элемента матрицы на это число

$$\mathbf{A}\lambda = \begin{pmatrix} a_{11}\lambda & a_{12}\lambda & & a_{1m}\lambda \\ a_{21}\lambda & a_{22}\lambda & & a_{2m}\lambda \\ \dots & \dots & & \dots \\ a_{n1}\lambda & a_{n2}\lambda & \dots & a_{nm}\lambda \end{pmatrix}.$$



Если число столбцов одной матрицы равно числу строк другой (такие матрицы называются *соответственными*), то для них определена операция **умножения матриц** по правилу «строка на столбец»:

$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mk} \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \dots & \sum_{i=1}^m a_{1i}b_{ik} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \dots & \sum_{i=1}^m a_{2i}b_{ik} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \dots & \sum_{i=1}^m a_{ni}b_{ik} \end{pmatrix}.$$

Видим, что умножение матрицы размера  $n \times m$  на матрицу размера  $m \times k$  дает матрицу размера  $n \times k$ .

Операция **транспонирования** матрицы состоит в том, что строки исходной матрицы становятся столбцами транспонированной

$$\mathbf{A}^T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}.$$

Квадратная матрица может иметь **обратную матрицу**. Матрица  $\mathbf{A}^{-1}$  называется обратной к квадратной матрице  $\mathbf{A}$  размера  $n \times n$ , если  $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ , где  $\mathbf{I}$  — *единичная матрица*, т. е. матрица, диагональные элементы которой единицы, а внедиагональные — нули. Единичная матрица обладает тем свойством, что умножение любой квадратной матрицы  $\mathbf{A}$  на единичную матрицу того же порядка не меняет  $\mathbf{A}$ , т. е.  $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ . Обратная матрица для единичной матрицы является также единичной матрицей, т. е.  $\mathbf{I}^{-1} = \mathbf{I}$ .

Обратная матрица может быть вычислена по формуле

$$\mathbf{A}^{-1} = \text{adj}(\mathbf{A}) / \det(\mathbf{A}),$$

где  $\text{adj}(\mathbf{A})$  — так называемая *присоединенная матрица* для  $\mathbf{A}$ ;  $\det(\mathbf{A})$  — ее *определитель*, часто обозначаемый так же, как  $|\mathbf{A}|$ . Определитель матрицы — это связанное с ней числовое значение, определяемое выражением

$$|\mathbf{A}| = \sum (-1)^r a_{1k_1} a_{2k_2} \dots a_{nk_n},$$

в котором суммирование ведется по всем  $n!$  перестановкам индексов  $1, 2, \dots, n$ ;  $r$  — число парных инверсий, необходимых для получения перестановки  $k_1, k_2, \dots, k_n$  из исходного упорядочения  $1, 2, \dots, n$ . Например, для определителя квадратной матрицы  $2 \times 2$

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

получаем следующий результат:

$$|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}.$$

Определитель матрицы порядка 1 равен значению ее единственного элемента.

Квадратная матрица  $\mathbf{A}$  называется *ортогональной*, если  $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}$ , т. е. если обратная матрица может быть получена из исходной просто путем ее транспонирования  $\mathbf{A}^{-1} = \mathbf{A}^T$ .

Очевидно, если определитель матрицы равен нулю (в этом случае она называется *вырожденной*), то для нее не существует обратной матрицы (так как в формуле для вычисления обратной матрицы определитель находится в знаменателе). В частности, если элементы какой-либо строки или столбца матрицы пропорциональны другой строке или столбцу (т. е. получены путем умножения их элементов на одно и то же число), то определитель будет равен нулю, и такая матрица не будет иметь обратной. Например,

$$\det \begin{pmatrix} a_{11} & a_{12} \\ ka_{11} & ka_{12} \end{pmatrix} = a_{11}ka_{12} - a_{12}ka_{11} = 0.$$

Вообще, определитель будет равен нулю в случае, если какая-либо строка или столбец матрицы является линейной комбинацией других ее строк или столбцов.

С любой квадратной матрицей  $\mathbf{A}$  порядка  $n$  связан также набор ее *собственных значений*  $\lambda_1, \lambda_2, \dots, \lambda_n$  — решений уравнения  $n$ -й степени:

$$|\mathbf{A} - \lambda \mathbf{I}| = 0.$$

В свою очередь каждому собственному значению  $\lambda_i$  соответствует *собственный вектор*  $\mathbf{V}_i$ , удовлетворяющий уравнению

$$\mathbf{A}\mathbf{V}_i = \lambda_i \mathbf{V}_i,$$

которое означает, что умножение слева собственного вектора  $\mathbf{V}_i$  на матрицу  $\mathbf{A}$  сводится к умножению его на скаляр — собственное значение  $\lambda_i$ . Отсюда следует, что если из  $n$  собственных вектор-столбцов составить квадратную матрицу  $\mathbf{V}$ , то будет выполняться соотношение

$$\mathbf{A}\mathbf{V} = \mathbf{V}\Lambda,$$

где  $\Lambda$  — матрица, главная диагональ которой образована собственными значениями  $\lambda_1, \lambda_2, \dots, \lambda_n$ , а внедиагональные элементы — нули. Можно показать, что если все собственные значения матрицы  $\mathbf{A}$  различны, то она не вырождена и имеет обратную матрицу  $\mathbf{A}^{-1}$ . В этом случае, умножая полученное соотношение слева на  $\mathbf{V}^{-1}$ , получаем

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{V}^{-1}\mathbf{V}\Lambda = \mathbf{I}\Lambda = \Lambda,$$

т. е. умножая  $\mathbf{A}$  слева на  $\mathbf{V}^{-1}$ , а справа — на  $\mathbf{V}$ , приводим ее к диагональному виду.

### 3.3. Регрессионный анализ

Ранее было отмечено, что одними из самых распространенных методов анализа связи между количественными переменными являются методы регрессионного анализа. Пусть в эксперименте наблюдаются значения  $m + 1$  переменной  $Y, X_1, X_2, \dots, X_m$ . Рассмотрим матрицу экспериментальных данных

$$\begin{pmatrix} y_1 & x_{11} & x_{12} & \dots & x_{1m} \\ y_2 & x_{21} & x_{22} & & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix},$$

где  $y_i$  и  $x_{ij}$  — значения переменных  $Y$  и  $X_j$  соответственно в  $i$ -м эксперименте.

В **регрессионном анализе** (*regression analysis*) рассматривается связь между переменной  $Y$ , называемой **зависимой переменной** (*dependent variable*), и переменными  $X_1, X_2, \dots, X_m$ , называемыми **независимыми переменными** (*independent variables*) (слово «независимые» здесь применяется не в вероятностном смысле). Эта связь описывается некоторой математической моделью, выражаемой функцией

$$Y = f(X_1, X_2, \dots, X_m; \beta_0, \beta_1, \dots, \beta_k) + \varepsilon,$$

где  $\beta_0, \beta_1, \dots, \beta_k$  — неизвестные параметры;  $\varepsilon$  — случайное отклонение  $Y$  от функции регрессии  $f$ .

Если функция регрессии линейна по параметрам (но не обязательно линейна по независимым переменным), то эта модель называется моделью **линейного регрессионного анализа**. Рассмотрим более подробно эту модель. Во многих реальных задачах экспериментальных исследований более подходящей является модель **нелинейного регрессионного анализа** (функция регрессии  $f$  нелинейна по параметрам). Но модель линейного регрессионного анализа часто бывает удовлетворительной в малой области изменения независимых переменных, а также как первое приближение к модели нелинейного регрессионного анализа. В практических исследованиях описание зависимости между переменными с помощью функции регрессии помогает установить наличие возможных причинных связей. Это часто привлекает исследователей к использованию методов регрессионного анализа. Другая причина частого применения методов регрессионного анализа состоит в том, что в ситуации, когда прямые измерения зависимой переменной затруднены, уравнение регрессии позволяет предсказать ее значения по значениям независимых переменных.

Матрицу экспериментальных данных можно получить одним из двух способов. При первом способе значения независимых переменных  $X_1, X_2, \dots, X_m$  надлежащим образом выбираются и устанавливаются без погрешностей экспериментатором в каждом опыте, и при этих значениях измеряется с ошибками значение зависимой переменной  $Y$ . Такой эксперимент называется **активным**. Например, при изучении влияния температуры водной среды  $X$  на скорость дыхания гидробионтов  $Y$  экспериментатор выбирает определенные значения температур: 5 °С, 10 °С и т. д. Затем для этих значений температуры определяется ско-

рость дыхания гидробионтов. При таком подходе только  $Y$  будет случайной величиной. При втором способе одновременно наблюдаются значения всех  $m + 1$  переменных  $Y, X_1, X_2, \dots, X_m$ , причем все эти переменные случайны, т. е. матрица экспериментальных данных в этом случае есть случайная выборка значений многомерной случайной величины  $(Y, X_1, X_2, \dots, X_m)$ . Такой эксперимент называется *пассивным*. Например, в случайно выбранной пробе воды регистрируется число бактерий на 1 мл  $Y$ , температура водной среды ( $X_1$ ), соленость ( $X_2$ ) и изучается влияние независимых переменных  $X_1, X_2$  на численность бактерий в водоеме  $Y$ .

Второй способ позволяет проводить так называемый *корреляционный анализ*, т. е. делать статистические выводы (оценивание, проверка гипотез) о мерах линейной зависимости между переменными (коэффициенте корреляции, множественном коэффициенте корреляции и частном коэффициенте корреляции).

К статистическим проблемам регрессионного анализа относятся: проверка выполнения предположений, лежащих в основе регрессионного анализа, нахождение оценок неизвестных параметров регрессии, построение доверительных интервалов для неизвестных параметров регрессии, проверка гипотез относительно этих параметров, проверка адекватности регрессионной модели.

### 3.3.1. Множественная линейная регрессия

Пусть  $Y$  зависимая, а  $X_1, X_2, \dots, X_m$  независимые переменные. Рассмотрим ситуацию активного эксперимента. Модель *множественной линейной регрессии* записывается в виде

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i, \quad (3.1)$$

где  $\beta_0, \beta_1, \dots, \beta_m$  — неизвестные параметры регрессии;  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  — независимые случайные величины, распределенные по нормальному закону со средним 0 и одинаковой дисперсией  $\sigma^2$ :  $\varepsilon_i \sim N(0, \sigma^2)$ .

#### *Оценивание параметров линейной регрессии*

При построении оценок неизвестных параметров регрессии можно отказаться от предположения о нормальности, а предположение о независимости заменить на предположение о некоррелируемости.

В случае, когда  $m = 1$ , модель множественной линейной регрессии (3.1) записывается в виде

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

и называется **простой линейной регрессией** ( $x_i$  — значение в  $i$ -м эксперименте единственной независимой переменной).

Оценки неизвестных параметров регрессии  $\beta_0, \beta_1, \dots, \beta_m$  получаются с помощью **метода наименьших квадратов** (МНК). Обозначим через  $b_0, b_1, \dots, b_m$  **МНК-оценки** неизвестных параметров ( $b_0$  называется **свободным членом**, или **константой регрессии**, а  $b_1, \dots, b_m$  — **коэффициентами регрессии**), т. е. оценки, обеспечивающие минимум суммы квадратов отклонений значений зависимой переменной от регрессии

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_m x_{im})^2, \quad (3.2)$$

которая является мерой ошибки предсказания зависимой переменной  $Y$  с помощью модели множественной линейной регрессии по независимым переменным  $X_1, X_2, \dots, X_m$ . Оценкой функции регрессии (или прямой наименьших квадратов в случае простой линейной регрессии) будет зависимость

$$\hat{Y} = b_0 + b_1 X_1 + \dots + b_m X_m.$$

На рис. 3.2, иллюстрирующем МНК в случае простой линейной регрессии, изображены прямая наименьших квадратов  $\hat{Y} = b_0 + b_1 X_1$  и модули остатков

$$d_i = |y_i - b_0 - b_1 x_i| = |y_i - \hat{y}_i|,$$

где  $\hat{y}_i$  — предсказанные по регрессионной модели значения зависимой переменной, которые в случае простой регрессии равны

$$\hat{y}_i = b_0 + b_1 x_i,$$

а в случае множественной линейной регрессии —  $\hat{y}_i = b_0 + b_1 x_{i1} + \dots + b_m x_{im}$ .

Можно доказать, что оценки  $b_0, b_1, \dots, b_m$ , получаемые по методу наименьших квадратов, линейно зависят от наблюдений  $y_1, y_2, \dots, y_n$  зависимой переменной  $Y$ . Кроме того, МНК-оценки являются несмещенными оценками неизвестных параметров.

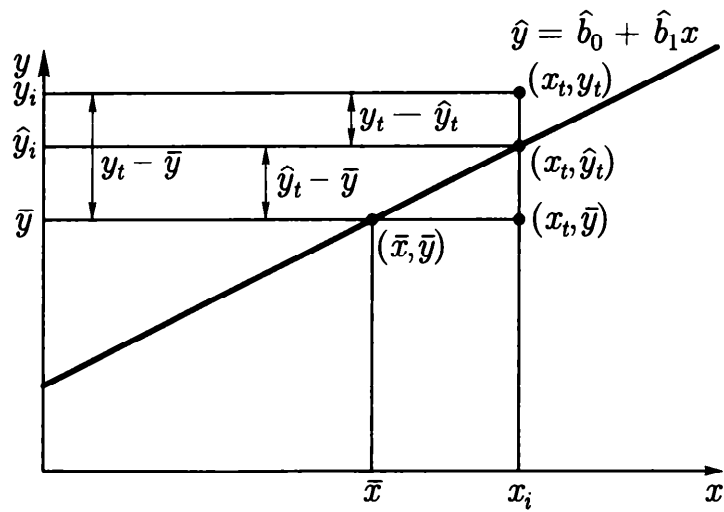


Рис. 3.2. Прямая наименьших квадратов простой линейной регрессии (графическая иллюстрация к определению сумм квадратов таблицы дисперсионного анализа)

Более подробно см. [А. Афифи, С. Эйзен, 1982; М. Дж. Кендалл, А. Стьюарт, 1973].

Приведем формулы для вычисления МНК-оценок для случая простой линейной регрессии ( $m = 1$ ):

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{и} \quad b_0 = \bar{y} - b_1 \bar{x},$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{и} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Программы множественной линейной регрессии всех пакетов статистических программ вычисляют оценки  $b_0, b_1, \dots, b_m$ . Кроме того, составляется таблица дисперсионного анализа для модели множественной линейной регрессии (табл. 3.1), на основании которой можно судить о качестве «подгонки» модели. На рис. 3.2 дана графическая иллюстрация слагаемых сумм квадратов в случае простой линейной регрессии. Если  $SS_R = 0$ , то все экспериментальные точки лежат на линии регрессии — прямой наименьших квадратов  $\hat{Y} = b_0 + b_1 X_1$ , а если  $SS_D = 0$ , то наилучшее предсказание  $Y$  осуществляется моделью  $\hat{Y} = \bar{y}$ .

Можно показать, что  $s_{\text{ост}}^2 = MS_R$  — несмещенная и состоятельная оценка для дисперсии  $\sigma^2$  зависимой переменной. Величину  $s_{\text{ост}} = \sqrt{MS_R}$  называют **стандартной ошибкой оценки зависимой переменной** (*standard error of estimate*).

Таблица дисперсионного анализа для модели множественной линейной регрессии

Источник дисперсии	Сумма квадратов	Степени свободы	Средний квадрат	F-отношение	p-значение
Регрессия	$SS_D = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\gamma_D = m$	$MS_D = \frac{SS_D}{\gamma_D}$	$\frac{MS_D}{MS_R}$	Для F-отношения
Отклонение от регрессии (остатки)	$SS_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\gamma_R = n - m - 1$	$MS_R = s_{\text{ост}}^2 = \frac{SS_R}{\gamma_R}$	—	
Полная	$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$	$\gamma_T = n - 1$	—	—	

*Примечание.*  $SS_D$  — обусловленная регрессией сумма квадратов;  $SS_R$  — сумма квадратов отклонений от линии регрессии, или остаточная сумма квадратов (сумма квадратов остатков);  $SS_T$  — полная сумма квадратов.



Верно замечательное свойство таблицы дисперсионного анализа

$$SS_T = SS_D + SS_R.$$

Отношение  $SS_D/SS_T$  есть доля вариации  $Y$ , объясняемая регрессией  $Y$  по  $X_1, X_2, \dots, X_m$ . Это отношение называется **коэффициентом детерминации**. Коэффициент детерминации является мерой качества предсказаний значений зависимой переменной  $Y$  моделью множественной линейной регрессии. Если он равен 1 (т. е.  $SS_R = 0$ ), то экспериментальные точки в точности лежат на линии регрессии. А если коэффициент детерминации равен 0 (т. е.  $SS_D = 0$ , а  $SS_T = SS_R$ ), то наилучшее предсказание  $Y$  осуществляется моделью  $\hat{Y} = \bar{y}$ , т. е.  $Y$  «не зависит» от  $X_1, X_2, \dots, X_m$ .

Можно показать в случае множественной линейной регрессии, что

$$\frac{SS_D}{SS_T} = R^2,$$

а в случае простой линейной регрессии

$$\frac{SS_D}{SS_T} = r^2,$$

где  $R$  — **выборочный множественный коэффициент корреляции** между  $Y$  и  $X_1, X_2, \dots, X_m$  (определение множественного коэффициента корреляции приведено далее);  $r$  — выборочный коэффициент корреляции между  $Y$  и  $X_1$ .

Модель множественной линейной регрессии удобно записывать в матричной форме. Обозначим через  $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$  вектор-столбец неизвестных параметров регрессии,  $\mathbf{b} = (b_0, b_1, \dots, b_m)^T$  — вектор-столбец МНК-оценок;  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  — вектор-столбец наблюдаемых значений зависимой переменной  $Y$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$  — вектор-столбец ошибок, а также

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}.$$

Тогда **матричная модель множественной линейной регрессии** (3.1) записывается в виде

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon},$$

где  $\varepsilon_i$  — независимые случайные величины, распределенные по нормальному закону  $N(0, \sigma^2)$ .

Можно показать, что вектор МНК-оценок является решением системы так называемых *нормальных уравнений*

$$(\mathbf{X}^T \mathbf{X})\mathbf{b} = \mathbf{X}^T \mathbf{y}$$

и равен

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{y}),$$

а *ковариационная матрица вектора МНК-оценок* равна

$$\text{cov}(\mathbf{b}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}.$$

### Доверительные интервалы

Большинство программ множественной линейной регрессии вычисляют среднеквадратические отклонения МНК-оценок коэффициентов регрессии, обычно называемые их *стандартными ошибками* (*standard error of regression coefficients*), которые будем обозначать через  $se(b_i)$ , где  $i = 0, 1, \dots, m$ . Можно показать, что  $100(1 - \alpha)\%$ -й доверительный интервал для неизвестного коэффициента регрессии  $\beta_i$  равен

$$(b_i - se(b_i)t_{n-m-1, 1-\alpha/2}, b_i + se(b_i)t_{n-m-1, 1-\alpha/2}),$$

где  $t_{n-m-1, 1-\alpha/2}$  — квантиль распределения Стьюдента с  $n-m-1$  степенями свободы порядка  $1 - \alpha/2$ . Более подробно о случае  $m > 1$  см. [М. Дж. Кендалл, А. Стьюарт, 1973].

Для случая простой линейной регрессии ( $m = 1$ ) выражения для стандартной ошибки свободного члена регрессии и углового коэффициента регрессии  $b_1$  имеют вид

$$se(b_1) = \sqrt{\frac{MS_R}{n \sum_{i=1}^n (x_i - \bar{x})^2}}; \quad se(b_0) = \sqrt{\frac{MS_R \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Построим теперь доверительный интервал для неизвестного значения  $Y$ . Предсказанное значение можно интерпретировать двумя способами, поэтому, соответственно способу интерпретации, можно построить два доверительных интервала. При первом способе  $\hat{y}$  — наилучшая оценка  $Y$ , соответствующего значениям  $x_1, x_2, \dots, x_m$  независимых переменных  $X_1, X_2, \dots, X_m$ .

При втором способе  $\hat{y}$  — наилучшая оценка математического ожидания  $Y$ , соответствующего значениям  $x_1, x_2, \dots, x_m$  независимых переменных  $X_1, X_2, \dots, X_m$ .

Обозначим через  $\mathbf{d} = (x_1 - \bar{x}_1, x_2 - \bar{x}_2, \dots, x_m - \bar{x}_m)^T$  вектор-столбец, где  $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ki}$ , а через  $\mathbf{A}$  квадратную матрицу  $m \times m$  с элементами  $a_{ij} = \sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j)$ . Тогда нижняя и верхняя границы  $100(1 - \alpha)\%$ -го доверительного интервала для  $Y$  при заданных  $x_1, x_2, \dots, x_m$  определяются выражением

$$\hat{y} \pm \left[ MS_R \left( 1 + \frac{1}{n} + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{d} \right) \right]^{1/2} t_{n-m-1, 1-\alpha/2}, \quad (3.3)$$

а нижняя и верхняя границы  $100(1 - \alpha)\%$ -го доверительного интервала для математического ожидания  $Y$  при заданных  $x_1, x_2, \dots, x_m$  определяются выражением

$$\hat{y} \pm \left[ MS_R \left( \frac{1}{n} + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{d} \right) \right]^{1/2} t_{n-m-1, 1-\alpha/2}, \quad (3.4)$$

где  $t_{n-m-1, 1-\alpha/2}$  — квантиль распределения Стьюдента с  $n - m - 1$  степенями свободы порядка  $1 - \alpha/2$ .

В случае простой линейной регрессии ( $m = 1$ ) доверительные интервалы (3.3) и (3.4) вычисляются более просто:

$$\hat{y} \pm \left[ MS_R \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right]^{1/2} t_{n-2, 1-\alpha/2}$$

и

$$\hat{y} \pm \left[ MS_R \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right]^{1/2} t_{n-2, 1-\alpha/2}.$$

**Замечание.** Чем дальше  $x$  от среднего значения, тем больше доверительный интервал. Многие программы простой линейной регрессии рисуют эти доверительные интервалы в виде кривых вокруг пря-

мой наименьших квадратов. Выбор типа доверительного интервала зависит от того, как интерпретирует предсказанное значение исследователь.

### ***Проверка гипотез о коэффициентах линейной регрессии***

Рассмотрим проверку гипотезы о том, что вычисленные коэффициенты регрессии значимо не отличаются от нуля. Это равносильно гипотезе о том, что независимые переменные  $X_1, X_2, \dots, X_m$  значимо не улучшают предсказания  $Y$  по сравнению с моделью  $\hat{y}_i = \bar{y}$ . Таким образом, рассмотрим нулевую гипотезу

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$$

при альтернативной гипотезе  $H_1$ : не все  $\beta_i$  равны нулю (альтернативная гипотеза состоит в том, что некоторые из независимых переменных  $X_1, X_2, \dots, X_S$  значимо улучшают предсказание  $Y$  по сравнению с моделью  $\hat{y}_i = \bar{y}$ ).

Зададимся уровнем значимости критерия  $\alpha$ . Статистика критерия

$$F = \frac{MS_D}{MS_R} \quad (3.5)$$

при условии, что верна гипотеза  $H_0$ , имеет  $F$ -распределение с  $m$  и  $n - m - 1$  степенями свободы. Эта статистика вычисляется программами множественной линейной регрессии в таблице дисперсионного анализа для модели множественной линейной регрессии (см. табл. 3.1) вместе с соответствующим  $p$ -значением, по которому, не пользуясь таблицами квантилей (в данном случае  $F$ -распределения), можно судить о том, согласуется ( $p$ -значение  $> \alpha$ ) или не согласуется ( $p$ -значение  $\leq \alpha$ ) гипотеза  $H_0$  с экспериментальными данными.

При отсутствии программы надо построить множество принятия нулевой гипотезы, которое в данном случае имеет вид  $(0, F_{m, n-m-1, 1-\alpha})$ , где  $F_{m, n-m-1, 1-\alpha}$  — квантиль  $F$ -распределения с  $m$  и  $n - m - 1$  степенями свободы порядка  $1 - \alpha$ . И если вычисленная по выборке статистика критерия (3.5) примет значение из множества принятия нулевой гипотезы, то считается, что гипотеза  $H_0$  согласуется с экспериментальными данными. В противном случае гипотеза  $H_0$  не согласуется с экспериментальными данными и отвергается на уровне значимости  $\alpha$ .

Рассмотрим теперь проверку гипотезы о том, что  $k$ -й коэффициент регрессии значимо не отличается от нуля. Это равно-

сильно гипотезе о том, что независимая переменная  $X_k$  значимо не улучшает предсказание  $Y$ . Таким образом, рассмотрим следующие нулевую и альтернативную гипотезы

$$H_0: \beta_k = 0;$$

$$H_1: \beta_k \neq 0.$$

Зададимся уровнем значимости критерия  $\alpha$ . При проверке данной гипотезы обычно рассматривают одну из двух эквивалентных статистик:

$$b_k^2/[se(b_k)]^2 \quad \text{и} \quad b_k/[se(b_k)]. \quad (3.6)$$

При условии, что верна гипотеза  $H_0$ , первая из статистик (3.6) имеет  $F$ -распределение с 1 и  $n - m - 1$  степенями свободы, а вторая — распределение Стьюдента с  $n - m - 1$  степенями свободы. Эти статистики вычисляются программами множественной линейной регрессии вместе с соответствующими  $p$ -значениями.

При отсутствии программы надо построить множество принятия нулевой гипотезы, которое для первой статистики имеет вид  $(0, F_{1, n-m-1, 1-\alpha})$ , а для второй  $(-t_{n-m-1, 1-\alpha/2}, t_{n-m-1, 1-\alpha/2})$ , где  $F_{1, n-m-1, 1-\alpha}$  — квантиль  $F$ -распределения с 1 и  $n - m - 1$  степенями свободы порядка  $1 - \alpha$ , а  $t_{n-m-1, 1-\alpha/2}$  — квантиль распределения Стьюдента с  $n - m - 1$  степенями свободы порядка  $1 - \alpha/2$ .

**Замечание.** Для проверки гипотезы  $H_0: \beta_k = C$ , где  $C$  — некоторое известное число при двусторонней альтернативе  $H_1: \beta_k \neq C$ , используется статистика

$$\frac{b_k - C}{se(b_k)},$$

которая, при условии что верна гипотеза  $H_0$ , распределена по закону Стьюдента с  $n - m - 1$  степенями свободы.

Предположения регрессионного анализа и проверка адекватности регрессионной модели исследуются при рассмотрении остатков  $d_i = y - \hat{y}_i$ . Графики остатков выводятся на печать большинством программ множественной и простой линейной регрессии. В некоторых программах для проверки коррелируемости ошибок  $\varepsilon_i$  вычисляется статистика Дурбина — Ватсона.

### 3.3.2. Множественный и частный коэффициенты корреляции

Пусть  $Y$  — зависимая, а  $X_1, X_2, \dots, X_m$  — независимые переменные. Рассмотрим ситуацию пассивного эксперимента, т. е.

предположим, что матрица экспериментальных данных является случайной выборкой значений многомерной случайной величины  $(Y, X_1, X_2, \dots, X_m)$ .

**Множественным коэффициентом корреляции** (*multiple correlation coefficient*) между  $Y$  и  $X_1, X_2, \dots, X_m$  называется число  $\rho(Y; X_1, X_2, \dots, X_m)$ , равное максимальному значению коэффициента корреляции между  $Y$  и любой невырожденной линейной комбинацией  $X_1, X_2, \dots, X_m$ :

$$\begin{aligned} \rho(Y; X_1, X_2, \dots, X_m) = \\ = \max_{a_1, \dots, a_m} \rho(Y; a_1 X_1 + a_2 X_2 + \dots + a_m X_m). \end{aligned} \quad (3.7)$$

**Частным коэффициентом корреляции** (*partial correlation*) между  $Y$  и  $X_k$  ( $k < m$ ) при фиксированных значениях переменных  $X_1, X_2, \dots, X_p$  (список не содержит переменную  $X_k$ ) называется число

$$\rho(Y; X_k | X_1, X_2, \dots, X_p) = \rho(U_1, U_2),$$

где  $U_1 = Y - m_1$ ;  $U_2 = X_k - m_2$ ;  $m_1$  и  $m_2$  — математические ожидания условных распределений  $Y$  и  $X_k$  соответственно при фиксированных значениях  $X_1, X_2, \dots, X_p$ .

Множественный коэффициент корреляции (3.7) служит мерой линейности зависимости между  $Y$  и набором переменных  $X_1, X_2, \dots, X_m$ , а частный коэффициент корреляции — мерой линейности зависимости между  $Y$  и  $X_k$ , когда  $X_1, X_2, \dots, X_p$  фиксированы.

Предположим теперь, что  $(Y, X_1, X_2, \dots, X_m)$  — многомерная нормально распределенная случайная величина. Обозначим через  $\mu_y, \mu_1, \mu_2, \dots, \mu_m$  и  $\sigma_y^2, \sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$  математические ожидания и дисперсии случайных величин  $Y, X_1, X_2, \dots, X_m$ , а через  $\sigma_{yi}$  и  $\sigma_{ij}$  — коэффициенты ковариации  $Y$  с  $X_i$  и  $X_i$  с  $X_j$  соответственно. Тогда можно доказать, что условное распределение  $Y$  для данных значений  $x_1, x_2, \dots, x_m$  независимых переменных  $X_1, X_2, \dots, X_m$  является нормальным с параметрами

$$\begin{aligned} \mu &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad \text{и} \\ \sigma^2 &= \sigma_y^2 (1 - \rho^2(Y; X_1, X_2, \dots, X_m)), \end{aligned} \quad (3.8)$$

где  $\beta_1, \dots, \beta_m$  — функции дисперсий и ковариаций случайных величин  $Y, X_1, X_2, \dots, X_m$ ;  $\beta_0 = \mu_y - \beta_1 \mu_1 - \beta_2 \mu_2 - \dots - \beta_m \mu_m$ . Случайная величина  $\varepsilon = Y - \mu$  нормально распределена с параметрами 0 и  $\sigma^2$ , поэтому можно записать

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon,$$

где  $\varepsilon \sim N(0, \sigma^2)$ . Таким образом, мы пришли к модели множественной линейной регрессии (3.1), поэтому изложенные ранее результаты, относящиеся к активному эксперименту, применимы и в случае пассивного эксперимента. Заметим также, что из (3.8) следует, что квадрат множественного коэффициента корреляции равен доле дисперсии  $Y$ , «объясненной» линейной регрессионной зависимостью от  $X_1, X_2, \dots, X_m$ :

$$\rho^2(Y; X_1, X_2, \dots, X_m) = \frac{\sigma_y^2 - \sigma^2}{\sigma_y^2}.$$

Аналогично можно показать, что квадрат частного коэффициента корреляции можно понимать как долю остаточной дисперсии  $Y$ , «объясненной» добавлением  $X_k$  к набору  $X_1, X_2, \dots, X_p$ .

### 3.3.3. Пошаговая регрессия

До сих пор рассматривалась множественная линейная регрессия зависимой переменной  $Y$  от всех независимых переменных  $X_1, X_2, \dots, X_m$ . Но часто перед исследователем стоит задача сначала выбрать из этого множества независимых переменных некоторое подмножество (называемое наилучшим подмножеством переменных для предсказания  $Y$ ), такое, что этому подмножеству соответствует наибольший коэффициент множественной корреляции с зависимой переменной  $Y$  среди всех подмножеств той же длины, а затем провести множественный линейный регрессионный анализ (см. 3.3.1), считая независимыми переменные из наилучшего подмножества переменных.

Обычно для выбора подмножества переменных применяется *пошаговая регрессия* (*step-wise regression*), когда независимые переменные включаются в наилучшее подмножество последовательно одна за другой на основании некоторого критерия включения переменной. В некоторый момент на основании правила остановки пошаговая процедура прекращает включение новых переменных, поскольку добавление новых не улучшает значимо предсказание переменной  $Y$ . При этом на каждом шаге процедуры некоторая переменная, уже включенная в наилучшее подмножество переменных, на основании некоторого критерия удаления переменных может быть удалена из подмножества переменных. Таким образом, метод пошаговой регрессии состоит из построения на каждом шаге множественной линей-

ной регрессии (см. 3.3.1) и вычисления критериев включения и исключения переменных и остановки.

Рассмотрим стандартную пошаговую процедуру ( $F$ -метод), которая реализуется в программах пошаговой регрессии.

Критерий включения переменной основывается на статистике  $F$ -включения. На очередном шаге процедуры включается та переменная, для которой значение статистики  $F$ -включения наибольшее, при условии, что оно превосходит некоторый установленный минимум, задаваемый правилом остановки.

Критерий исключения переменной основывается на статистике  $F$ -удаления. На очередном шаге процедуры исключается та переменная, для которой значение статистики  $F$ -исключения наименьшее, при условии, что оно меньше некоторого установленного минимума, задаваемого правилом остановки.

Рассмотрим более подробно статистики  $F$ -включения и  $F$ -исключения. Предположим, что в набор переменных для предсказания  $Y$  уже включено  $k$  переменных  $X_1, X_2, \dots, X_k$  (без потери общности можно считать, что это первые  $k$  переменных исходного набора переменных  $X_1, X_2, \dots, X_m, k < m$ ). Тогда значение статистики  $F$ -включения для переменной  $X_{k+1}$  равно

$$\frac{r^2(Y; X_{k+1} | X_1, X_2, \dots, X_k)}{1 - r^2(Y; X_{k+1} | X_1, X_2, \dots, X_k)} (n - k - 2), \quad (3.9)$$

где  $r^2(Y; X_{k+1} | X_1, X_2, \dots, X_k)$  — квадрат выборочного частного коэффициента корреляции между  $Y$  и  $X_{k+1}$  при фиксированных значениях переменных  $X_1, X_2, \dots, X_k$ . Статистика (3.9) служит для проверки гипотезы  $H_0: \rho(Y; X_{k+1} | X_1, X_2, \dots, X_k) = 0$ , т. е. для проверки гипотезы о том, что включение  $X_{k+1}$  в набор переменных  $X_1, X_2, \dots, X_k$  значительно не улучшает предсказание  $Y$ . При условии, что гипотеза  $H_0$  верна, статистика (3.9) распределена по  $F$ -закону с 1 и  $n - k - 2$  степенями свободы.

Значение статистики  $F$ -исключения для переменной  $X_k$  равно

$$\frac{r^2(Y; X_k | X_1, X_2, \dots, X_{k-1})}{1 - r^2(Y; X_k | X_1, X_2, \dots, X_{k-1})} (n - k - 1), \quad (3.10)$$

где  $r^2(Y; X_k | X_1, X_2, \dots, X_{k-1})$  — квадрат выборочного частного коэффициента корреляции между  $Y$  и  $X_k$  при фиксированных значениях переменных  $X_1, X_2, \dots, X_{k-1}$ . Статистика (3.10) служит для проверки нулевой гипотезы

$$H_0: \rho(Y; X_k | X_1, X_2, \dots, X_{k-1}) = 0,$$



от 0 до 1 (функция  $y(x) = 1/(1 + \exp(-x))$  называется *сигмоидной*, или *логистической*).

Оценки  $b_0, b_1, \dots, b_m$  коэффициентов уравнения определяют методом максимального правдоподобия. Для их получения необходимо максимизировать логарифм функции правдоподобия, который в данном случае имеет вид

$$L(b_0, b_1, \dots, b_m) = \sum_{i=1}^n y_i \ln \left( \frac{1}{1 + \exp(-b_0 - \sum_{j=1}^m b_j x_{ij})} \right) + \\ + \sum_{i=1}^n (1 - y_i) \ln \left( 1 - \frac{1}{1 + \exp(-b_0 - \sum_{j=1}^m b_j x_{ij})} \right).$$

Смысл выражения в правой части уравнения легко понять, если заметить, что слагаемые первой суммы отличны от нуля только для  $y_i = 1$ , а второй — только для  $y_i = 0$  и что выражение под логарифмом в первой сумме содержит оценку вероятности появления 1, а во второй — оценку вероятности появления 0, т. е. это действительно логарифм вероятности появления наблюдаемого набора значений зависимой переменной при заданном наборе значений независимых переменных и заданных значениях оцениваемых параметров уравнения, как это и требуется в методе максимального правдоподобия.

### 3.4. Дисперсионный анализ

Метод *дисперсионного анализа* (*Analysis of Variance, ANOVA*), разработанный Р. Фишером, используется для анализа связи между количественной зависимой переменной  $Y$  и качественными независимыми переменными  $X_1, \dots, X_s$ , или **факторами**, такими, как пол животного, вид растения или животного, тип почвы или удобрений, которые не допускают количественного описания и потому называются качественными переменными. Конкретные значения фактора называют его **уровнями**. В общем случае рассматривается влияние нескольких факторов, каждый из которых имеет не менее двух уровней.

Дисперсионный анализ основан на разложении (анализе) общей дисперсии зависимой переменной  $Y$  на компоненты, каждая из которых соответствует определенному источнику изменчивости измеряемой переменной  $Y$ . Это и объясняет название метода — дисперсионный анализ. Аббревиатура ANOVA, которая используется наряду с полным названием метода, была введена Дж. Тьюки (J. W. Tukey).

В основе каждой задачи дисперсионного анализа лежит *план эксперимента*. Планом эксперимента называют совокупность комбинаций уровней факторов (или ячеек, как их принято называть) и числа наблюдений для каждой из комбинаций. Если в каждой из возможных ячеек проводится хотя бы одно наблюдение, то такой план называют *полным факторным планом*. В противном случае — *неполным факторным планом*. План с равным числом наблюдений в ячейках называют *сбалансированным*, в противном случае — *несбалансированным*.

В зависимости от математической природы факторов различают три типа моделей. Если рассматривается влияние конкретных уровней каждого фактора, то используется *модель с постоянными факторами* (модель I). Если же уровни фактора случайно выбираются из практически бесконечного множества уровней, то применяют *модель со случайными факторами* (модель II). И наконец, если уровни одних факторов являются постоянными, а уровни других факторов — случайными, то такая модель называется *смешанной моделью* (модель III). Задачи дисперсионного анализа различают также по числу анализируемых факторов — дисперсионный анализ может быть однофакторным, двухфакторным и т. д.

Все конкретные модели дисперсионного анализа, которые будут рассмотрены, можно записать в виде линейной модели. Для построения оценок неизвестных параметров, как и в регрессионном анализе, используется метод наименьших квадратов. МНК-оценки неизвестных параметров являются несмещенными, линейно зависящими от наблюдений оценками с наименьшей дисперсией среди всех несмещенных линейных оценок.

Дисперсионный анализ опирается на предположение нормальности распределения ошибок исходных данных. Отметим, что нормальность нужна только для проверки гипотез и построения интервальных оценок. Само построение оценок и их оптимальные свойства не опираются на предположения нормальности ошибок наблюдений. Критерии проверки гипотез в модели I будут относительно устойчивы при нарушении нормальности, но

это не так для модели II. Заметим также, что при планировании эксперимента нужно, где это возможно, использовать равное число наблюдений в ячейках для обеспечения свойств устойчивости к нарушению предпосылок.

### 3.4.1. Однофакторный дисперсионный анализ

**Однофакторный дисперсионный анализ** используется для исследования влияния одного фактора на количественную переменную. Такие задачи возникают, например, при сравнении урожайности нескольких сортов пшеницы, эффективности действия различных лекарств и т. п. Если исследователя интересует урожайность определенных сортов пшеницы, то сорт является постоянным фактором, и для представления результатов эксперимента используется модель I. И наоборот, сорт будет случайным фактором, если рассматриваемые сорта были выбраны случайно из большого (практически бесконечного) числа сортов. В этом случае применяют модель II.

#### *Модель с постоянными уровнями фактора*

Пусть фактор  $A$  варьируется в эксперименте на заранее выбранных  $I$  уровнях, а  $y_{i1}, \dots, y_{in_i}$  — результаты  $n_i$  наблюдений, соответствующих  $i$ -му уровню фактора, которые являются одной из  $I$  независимых выборок из генеральной совокупности с математическими ожиданиями (генеральными средними)  $\mu_i$ ,  $i = 1, \dots, I$ . Фактор  $A$  является основанием для классификации всей совокупности наблюдений, поэтому однофакторный дисперсионный анализ называют еще **однофакторной классификацией**.

Математическая модель однофакторного дисперсионного анализа имеет вид

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i, \quad (3.11)$$

где  $\varepsilon_{ij}$  — ошибки наблюдений, независимые нормально распределенные случайные величины с нулевым математическим ожиданием и одинаковой дисперсией  $\sigma^2$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ .

Введем принятые в дисперсионном анализе обозначения:  $\mu = \frac{1}{n} \sum_{i=1}^I n_i \mu_i$  — общее среднее;  $n = \sum_{i=1}^I n_i$  — общее число наблюдений;  $\alpha_i = \mu_i - \mu$  — эффект, обусловленный влиянием  $i$ -го уров-

ня фактора. Заметим, что для моделей с постоянными уровнями факторов эффекты уровней обозначаются греческими буквами.

Используя введенные обозначения, модель (3.11) можно переписать в виде

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i. \quad (3.12)$$

Неизвестными параметрами в (3.11) и (3.12) являются  $\mu_i$ ,  $\alpha_i$  и  $\mu$ . МНК-оценки этих параметров при дополнительных ограни-

чениях  $\sum_{i=1}^I n_i \alpha_i = 0$ , которые обеспечивают их единственность, находят по формулам:

$$\hat{\mu} = y_{..}, \quad \hat{\mu}_i = y_{i.}, \quad \hat{\alpha}_i = \hat{\mu}_i - \hat{\mu}, \quad i = 1, \dots, I, \quad (3.13)$$

где точкой обозначены результаты усреднений  $y_{ij}$  по соответствующим индексам

$$y_{i.} = \sum_{j=1}^{n_i} y_{ij} / n_i; \quad y_{..} = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}.$$

На следующем этапе анализа проверяется гипотеза  $H_0: \mu_1 = \mu_2 = \dots = \mu_I = \mu$  (или  $\alpha_1 = \alpha_2 = \dots = \alpha_I = 0$ ) о том, что фактор  $A$  не влияет на зависимую переменную  $Y$  при альтернативе, что, по крайней мере, одно из средних отличается от общего среднего:  $\mu_k \neq \mu$  (или  $\alpha_k \neq 0$ ). Критерий для проверки гипотезы  $H_0$  строится (как и в регрессионном анализе) на основе следующего тождества — разложения суммы квадратов  $SS_T$  отклонений результатов наблюдений от общего среднего на две компоненты (два источника дисперсии) —  $SS_A$  и  $SS_R$ :

$$\begin{aligned} SS_T &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{..})^2 = \\ &= \sum_{i=1}^I n_i (y_{i.} - y_{..})^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2 = SS_A + SS_R. \end{aligned}$$

Слагаемое  $SS_A$  характеризует степень разброса наблюдений между уровнями фактора и обусловлено их влиянием. Заметим, что обозначение  $SS_A$  является традиционным для дисперсионного анализа, в отличие от обозначения  $SS_D$  в регрессионном анализе. Слагаемое  $SS_R$ , называемое остаточной суммой

квадратов, характеризует степень разброса наблюдений внутри уровней фактора и является следствием случайных причин и неучтенных факторов. Результаты проверки гипотезы  $H_0$  принято заносить в *таблицу однофакторного дисперсионного анализа* (табл. 3.2).

Можно показать, что средние квадраты  $MS_A$  и  $MS_R$  (другое, часто используемое обозначение этих величин, —  $s_A^2$  и  $s_R^2$ ), определение которых дано в табл. 3.2, при выполнении  $H_0$  являются двумя различными и независимыми оценками дисперсии единичного наблюдения  $\sigma^2$  с математическими ожиданиями, равными

$$M(MS_R) = \sigma^2; \quad M(MS_A) = \sigma^2 + \frac{1}{I} \sum_{i=1}^I n_i (\mu_i - \mu)^2 = \sigma^2.$$

Заметим, что  $MS_R$  является *несмещенной оценкой*  $\sigma^2$  для всех моделей дисперсионного анализа.

Если гипотеза  $H_0$  верна, то можно показать, что справедливы следующие утверждения:

- $SS_A$  и  $SS_R$  независимы и распределены как:  $SS_A \sim \sigma^2 \chi_{I-1}^2$ ,  $SS_R \sim \sigma^2 \chi_{n-I}^2$ ;

- $M(MS_R) = M(MS_A) = \sigma^2$ ;

- статистика  $F = MS_A/MS_R$  имеет  $F$ -распределение с  $I - 1$  и  $n - I$  степенями свободы. Критическая область уровня значимости  $\alpha$  состоит из одного полуинтервала  $[F_{I-1, n-I, 1-\alpha}, \infty)$ , где  $F_{I-1, n-I, 1-\alpha}$  — квантиль  $F$ -распределения с  $I - 1$  и  $n - I$  степенями свободы порядка  $1 - \alpha$ .

Если вычисленное по наблюдениям значение статистики критерия ( $F_{\text{набл}}$ ) меньше значения  $F_{I-1, n-I, 1-\alpha}$ , то нет оснований для отклонения гипотезы  $H_0$ . Обычно на этом анализ может быть завершён. Но если такой вывод противоречит, например, результатам предыдущих исследований, тогда следует выяснить:

- не нарушаются ли предпосылки дисперсионного анализа;
- нет ли других факторов, влияющих на зависимую переменную, и либо перейти к двухфакторному анализу, либо спланировать эксперимент так, чтобы по возможности исключить влияние неоднородностей условий проведения наблюдений, т. е. использовать *план со случайными блоками*.

Если же  $F_{\text{набл}} \geq F_{I-1, n-I, 1-\alpha}$ , то нулевая гипотеза отвергается на уровне значимости  $\alpha$ , т. е. существует, по крайней мере, одна пара средних  $y_k$  и  $y_l$ , которые различаются статистически значимо. Часто исследователю желательно также знать, какое из выборочных средних можно считать максимальным. Ес-

Таблица 3.2

## Однофакторный дисперсионный анализ

Источник дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат	F-отношение
Фактор A	$SS_A = \sum_{i=1}^I n_i (y_{i.} - y_{..})^2$	$\nu_A = I - 1$	$MS_A = \frac{SS_A}{\nu_A}$	$F_A = \frac{MS_A}{MS_R}$
Остаток (ошибка)	$SS_R = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2$	$\nu_R = n - I$	$MS_R = \frac{SS_R}{\nu_R} = \hat{\sigma}^2$	—
Полная	$SS_T = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_{..})^2$	$\nu_T = n - 1$	—	—

ли нельзя выделить пары средних, представляющие интерес для сравнения, то приходится проверять гипотезы о равенстве всевозможных пар средних. Чтобы при этом сохранить  $\alpha$  в качестве общего уровня значимости, следует использовать **методы множественного сравнения**.

Кроме сравнения двух средних для исследователя могут представлять интерес сравнения вида  $\mu_1 - \mu_4$ ,  $\mu_1 - (\mu_2 + \mu_3 + \mu_4)/3$  и т. п. Линейные комбинации  $\sum_{i=1}^I c_i \mu_i$ , где  $\sum_{i=1}^I c_i = 0$ , называются **контрастами**. Выдвижение гипотез о контрастах

( $H_0: \sum_{i=1}^I c_i \mu_i = 0$ ) и их проверка позволяют получить допол-

нительную информацию. Если линейная комбинация задана до получения экспериментальных данных, то такие сравнения называют **планируемыми** (*planned comparisons*). Например, если врач хочет выявить эффективность нескольких лечебных препаратов по сравнению с контролем, то такое сравнение является планируемым.

На практике сравнения, представляющие интерес, определяются часто после получения результатов эксперимента, когда уже известны оценки средних  $\mu_i$ , и вид наиболее важных сравнений подсказывают сами результаты эксперимента. Такие сравнения называют **непланируемыми** (*unplanned comparisons*). Сравнение всевозможных пар средних является примером непланируемых сравнений. Число таких сравнений для модели (3.11) равно  $k = I(I - 1)/2$ . С рекомендациями по использованию методов множественного сравнения можно познакомиться в работах [Г. Шеффе, 1980; R. R. Sokal, F. J. Rohlf, 1995; А. Т. Teriokhin, Т. de Meeus, J.-F. Guegan, 2007].

**Замечание.** Аналогично проводятся исследования сравнений любой модели дисперсионного анализа, а не только однофакторной.

**Пример 3.1.** [R. R. Sokal, F. J. Rohlf, 1995]. В восьми различных местах обитания (в США) были собраны 130 жуков-скакунов (*Cicindela circumpecta*). У каждого жука измерялась ширина головки (мм). Требовалось выяснить, влияет ли место обитания на этот признак. Для ответа на данный вопрос используем однофакторный дисперсионный анализ, предполагая, что все его предпосылки выполнены. Обозначим через  $y_{ij}$  результат измерения ширины головки  $j$ -го жука из  $i$ -й местности,  $j = 1, \dots, n_i$ ;  $i = 1, \dots, 8$ . В табл. 3.3 приведены средние значения  $y_i$  резуль-

**Средние значения ширины головки жука  $y_i$  в  $i$ -м месте обитания**

$I$	1	2	3	4	5	6	7	8
$y_i$	3,512	3,694	3,722	3,723	3,738	3,748	3,784	3,845
$n_i$	20	20	20	14	10	15	11	20

Таблица 3.4

**Результаты однофакторного дисперсионного анализа для примера 3.1**

Источник разброса	Сумма квадратов	Степени свободы	Средний квадрат	$F$ -отношение $MS_A/MS_R$
Фактор $A$	1,2451	7	0,1779	$F = 6,724$ $p = 0,000$
Остаток (ошибка)	3,2270	122	0,0264	—
Полная	4,4721	129	—	—

Таблица 3.5

**Нижний и верхний концы сравнительных интервалов**

$I$	1	2	3	4	5	6	7	8
$s_i$	0,0363	0,0132	0,0132	0,0434	0,0514	0,0420	0,0490	0,0363
$l_i$	3,430	3,612	3,640	3,655	3,621	3,653	3,673	3,763
$u_i$	3,594	3,775	3,804	3,820	3,853	3,842	3,894	3,927

татов наблюдений, упорядоченные по возрастанию, а также значения  $n_i$ .

Результаты проверки гипотезы  $H_0$  о том, что место обитания не влияет на размеры головки жука, при альтернативе  $H_1$ , что хотя бы в двух районах жуки отличаются по этому признаку, занесем в таблицу дисперсионного анализа (табл. 3.4).

Так как  $F_{\text{набл}} = 6,724 > F_{7,122;0,95} = 2,085$  и  $p \ll \alpha = 0,05$ , гипотезу  $H_0$  на уровне значимости  $\alpha = 0,05$  отвергаем.

Чтобы выяснить, какие из средних различаются статистически значимо, используем метод множественного сравнения — GT2-метод. Для наглядного представления результатов сравнения построим 95%-е сравнительные интервалы для каждого вы-



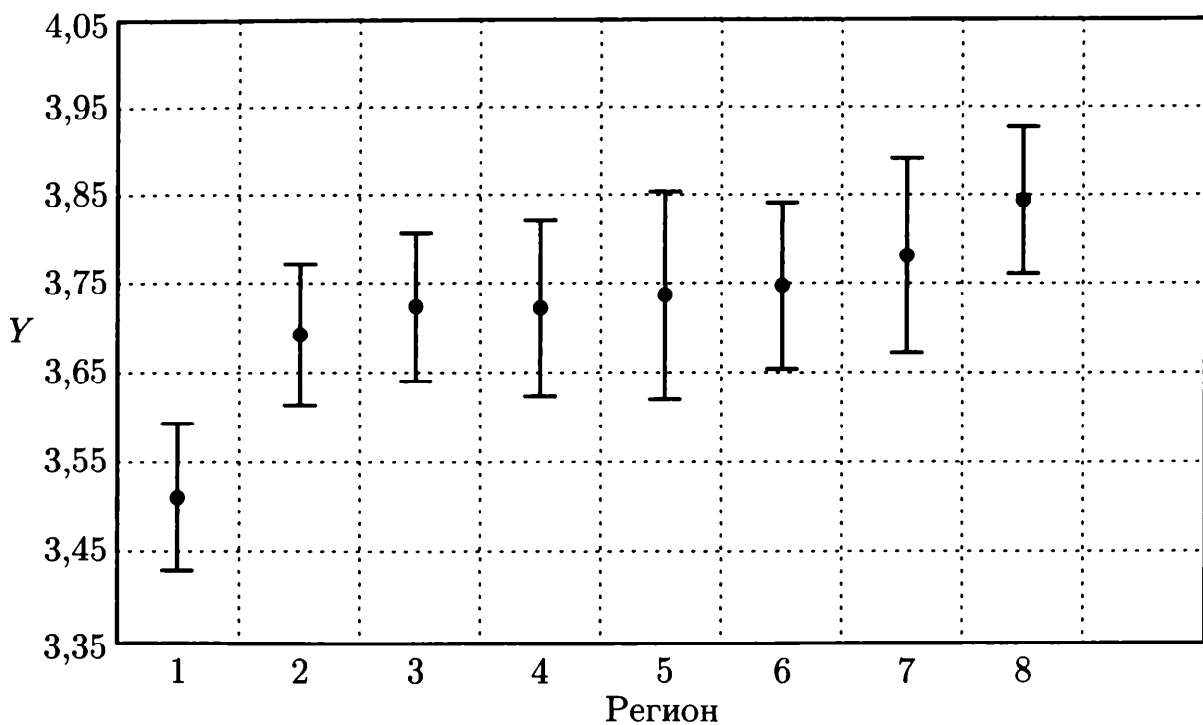


Рис. 3.3. 95%-е сравнительные интервалы для средних значений ширины головки жука-скакуна

борочного среднего  $y_i$  по методу Габриэла (рис. 3.3). Выборочные средние, для которых сравнительные интервалы не пересекаются, различаются статистически значимо. Значения нижних и верхних концов  $(l_i, u_i)$  этих интервалов для GT2-метода вычисляются по формулам

$$l_i = y_i - \frac{1}{\sqrt{2}} m_{k, \nu_R, 1-\alpha/2} s_i; u_i = y_i + \frac{1}{\sqrt{2}} m_{k, \nu_R, 1-\alpha/2} s_i,$$

где  $s_i = \sqrt{MS_R/n_i}$ ;  $m_{8, 122; 0,975} = 3,183$  — квантиль порядка 0,975 студентизированного максимального абсолютного отклонения [М. Дж. Кендалл, А. Стьюарт, 1973; R. R. Sokal, F. J. Rohlf, 1995] с числом степеней свободы  $k = 8$  и  $n - I = 122$ . Значения  $s_i$ ,  $l_i$ ,  $u_i$  приведены в табл. 3.5. На рис. 3.3 видно, что выборочное среднее  $y_1$  значительно отличается от всех остальных средних, среди которых нет значимого различия.

### *Модель со случайными уровнями фактора*

Рассмотрим ситуацию, когда исследователя интересует, какой вклад в общую изменчивость зависимой переменной  $Y$  вносит фактор  $A$ . Такие задачи возникают, например, в популяционной биологии, когда изменчивость зависимой переменной связана с генетическими различиями. В этом случае предполагаем, что  $I$  уровней фактора  $A$  были выбраны случайно из гене-

ральной совокупности (например, с помощью таблицы случайных чисел).

Пусть на  $i$ -м уровне фактора  $A$  производится  $n_i$  наблюдений  $y_{i1}, \dots, y_{in_i}$  и  $y_{ij} \sim N(\mu_i, \sigma_y^2)$ , а  $\{\mu_i\}$ ,  $i = 1, \dots, I$ , независимы и одинаково распределены со средним  $\mu$  и дисперсией  $\sigma_A^2$ . Тогда уравнение модели  $\Pi$  (**модели дисперсионного анализа со случайным фактором**) имеет вид

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i, \quad (3.14)$$

где  $\varepsilon_{ij} \sim N(0, \sigma^2)$  — ошибки наблюдений, независимые случайные величины.

Обозначим через  $a_i$  эффект  $i$ -го уровня фактора  $A$ ,  $a_i = \mu_i - \mu$ . Заметим, что для моделей со случайными уровнями факторов эффекты уровней принято обозначать латинскими буквами. Тогда уравнение модели (3.14) и предположения можно переписать в виде

$$y_{ij} = \mu + a_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i, \quad (3.15)$$

где  $\{a_i\}$  и  $\{\varepsilon_{ij}\}$  независимы в совокупности;  $a_i \sim N(\mu, \sigma_A^2)$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ .

Дисперсия наблюдения  $y_{ij}$  будет равна  $\sigma_y^2 = \sigma_A^2 + \sigma^2$ , поэтому  $\sigma_A^2$  и  $\sigma^2$  называют **компонентами дисперсии** единичного наблюдения. Компонента  $\sigma_A^2$  обусловлена вкладом фактора в общую изменчивость  $Y$ . Долю этого вклада, отношение  $\rho = \sigma_A^2 / (\sigma_A^2 + \sigma^2)$ , называют **коэффициентом внутриклассовой корреляции**.

Неизвестными параметрами в (3.15) являются  $\mu$  и две компоненты дисперсии  $\sigma^2$  и  $\sigma_A^2$ . Оценка для  $\mu$  находится, как и для модели I, по формуле (3.13). Несмещенные оценки для компонент дисперсии имеют вид

$$\hat{\sigma}^2 = MS_R; \quad \hat{\sigma}_A^2 = (MS_A - MS_R)/k,$$

где  $k = \frac{1}{I-1} \left( n - \frac{1}{n} \sum_{i=1}^I n_i^2 \right)$  (если  $n_1 = \dots = n_I = K$ , то  $k = K$ ).

На следующем этапе анализа проверяется гипотеза  $H_0: \sigma_A^2 = 0$  о том, что фактор не вносит никакого вклада в дисперсию наблюдений, при альтернативе  $H_1: \sigma_A^2 \neq 0$ . Критерий для проверки этой гипотезы строится, как и для модели I. Если верна гипотеза  $H_0$ , то статистика  $F = MS_A / MS_R$  имеет  $F$ -распределение с  $\nu_A = I - 1$  и  $\nu_R = I(k - 1)$  степенями свободы.

Заметим, что  $F$ -критерий для проверки гипотезы  $H_0$  для однофакторной модели II является только приближенным для неравных  $n_i$  и точным для случая равных  $n_i$ . Результаты анализа заносятся в таблицу дисперсионного анализа, аналогичную табл. 3.2.

Если гипотеза  $H_0$  отвергается, то оценкой доли общей дисперсии, обусловленной фактором  $A$ , является значение выборочного коэффициента внутриклассовой корреляции  $\hat{\rho} = \hat{\sigma}_A^2 / (\hat{\sigma}_A^2 + \hat{\sigma}^2)$ .

### 3.4.2. Двухфакторный дисперсионный анализ

Для исследования влияния двух одновременно действующих факторов применяется двухфакторный дисперсионный анализ. Пусть факторы  $A$  и  $B$  имеют  $I$  и  $J$  уровней соответственно, а  $y_{ij1}, \dots, y_{ijK_{ij}}, K_{ij}$  наблюдений зависимой переменной  $Y$  на  $i$ -м уровне фактора  $A$  и  $j$ -м уровне фактора  $B$  — это  $IJ$  независимых выборок с математическим ожиданием  $\mu_{ij}, i = 1, \dots, I, j = 1, \dots, J$ . Если в каждой ячейке имеется, по крайней мере, одно наблюдение, то план эксперимента называется **полным двухфакторным планом**.

Для двухфакторной модели рассмотрим только случай постоянных уровней факторов.

1.  $K_{ij} = K > 1$  — равные наблюдения в ячейках. Математическая модель двухфакторного дисперсионного анализа, называемая моделью **эффектов уровней**, имеет вид:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad (3.16)$$

где  $\varepsilon_{ijk} \sim N(0, \sigma^2)$  и независимы,  $i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K; \mu$  — общее среднее;  $\alpha_i$  — эффект  $i$ -го уровня фактора  $A$ ;  $\beta_j$  — эффект  $j$ -го уровня фактора  $B$ ;  $\gamma_{ij} = \mu_{ij} - \alpha_i - \beta_j + \mu$  — **эффект взаимодействия**  $i$ -го уровня фактора  $A$  с  $j$ -м уровнем фактора  $B$ , т. е. эффект совместного неаддитивного вклада факторов  $A$  и  $B$  в изменчивость  $Y$ , когда влияние фактора  $A$  на переменную  $Y$  зависит от того, на каком уровне находится фактор  $B$ .

МНК-оценки неизвестных параметров  $\mu, \alpha_i, \beta_j, \gamma_{ij}$  при дополнительных ограничениях

$$\sum_{i=1}^I \alpha_i = 0; \quad \sum_{j=1}^J \beta_j = 0; \quad \sum_{i=1}^I \gamma_{ij} = 0, \quad j = 1, \dots, J;$$

$$\sum_{j=1}^J \gamma_{ij} = 0, \quad i = 1, \dots, I,$$

которые обеспечивают единственность оценок, находятся по формулам

$$\begin{aligned} \hat{\mu} &= y_{...}; & \hat{\mu}_{ij} &= y_{ij.}; & \hat{\alpha}_i &= y_{i..} - y_{...}; & \hat{\beta}_j &= y_{.j.} - y_{...}; \\ \hat{\gamma}_{ij} &= y_{ij.} - y_{i..} - y_{.j.} + y_{...}, & i &= 1, \dots, I; & j &= 1, \dots, J. \end{aligned}$$

В этих формулах, как и ранее, точкой обозначено усреднение по соответствующему индексу.

Для определения влияния факторов и их взаимодействия на результаты эксперимента проверяются следующие гипотезы:

- $H_0: \gamma_{11} = \gamma_{12} = \dots = \gamma_{IJ} = 0$  о том, что влияние взаимодействия уровней факторов  $A$  и  $B$  отсутствует при альтернативе, что не все  $\gamma_{ij} = 0$ ;
- $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$  о том, что фактор  $A$  не влияет при альтернативе, что не все  $\alpha_i = 0$ ;
- $H_0: \beta_1 = \beta_2 = \dots = \beta_J = 0$  о том, что фактор  $B$  не влияет при альтернативе, что не все  $\beta_j = 0$ .

Критерии для проверки этих гипотез строятся на основе следующего тождества — разложения полной суммы квадратов на четыре компоненты (четыре источника дисперсии):

$$\begin{aligned} SS_T &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{...})^2 = \\ &= SS_A + SS_B + SS_{AB} + SS_R. \end{aligned} \quad (3.17)$$

Слагаемое  $SS_A$  характеризует степень разброса наблюдений между уровнями фактора  $A$ ,  $SS_B$  — между уровнями фактора  $B$ . Компонента  $SS_{AB}$  обусловлена влиянием взаимодействия уровней факторов  $A$  и  $B$ . Слагаемое  $SS_R$  характеризует степень разброса наблюдений внутри ячеек и является следствием влияния случайных причин и неучтенных факторов. Результаты анализа заносятся в таблицу двухфакторного дисперсионного анализа (табл. 3.6).

Можно показать, что если верны нулевые гипотезы, то справедливы следующие утверждения:

- суммы квадратов  $SS_A, SS_B, SS_{AB}, SS_R$  независимы и распределены как  $\sigma^2 \chi_{I-1}^2, \sigma^2 \chi_{J-1}^2, \sigma^2 \chi_{(I-1)(J-1)}^2, \sigma^2 \chi_{IJ(K-1)}^2$  соответственно;
- каждый из средних квадратов  $MS_A, MS_B, MS_{AB}, MS_R$  (определены в табл. 3.6) является несмещенной оценкой  $\sigma^2$ ;
- статистика  $F_{AB} = MS_{AB}/MS_R$  имеет  $F$ -распределение с  $\nu_{AB} = (I-1)(J-1)$  и  $\nu_R = IJ(K-1)$  степенями свободы;

Двухфакторный дисперсионный анализ

Источник разброса	Сумма квадратов	Число степеней свободы	Средний квадрат	F-отношение
Фактор A	$SS_A = JK \sum_{i=1}^I (y_{i..} - y_{...})^2$	$\nu_A = I - 1$	$MS_A = \frac{SS_A}{\nu_A}$	$F_A = \frac{MS_A}{MS_R}$
Фактор B	$SS_B = IK \sum_{j=1}^J (y_{.j.} - y_{...})^2$	$\nu_B = J - 1$	$MS_B = \frac{SS_B}{\nu_B}$	$F_B = \frac{MS_B}{MS_R}$
Взаимодействие AB	$SS_{AB} = K \sum_{i=1}^I \sum_{j=1}^J (y_{ij.} - y_{i..} - y_{.j.} - y_{...})^2$	$\nu_{AB} = (I - 1)(J - 1)$	$MS_{AB} = \frac{SS_{AB}}{\nu_{AB}}$	$F_{AB} = \frac{MS_{AB}}{MS_R}$
Остаток (ошибка)	$SS_R = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{ij.})^2$	$\nu_R = IJ(K - 1)$	$MS_R = \frac{SS_R}{\nu_R}$	—
Полная	$SS_T = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{...})^2$	$\nu_T = IJK - 1$	—	—

- статистика  $F_A = MS_A/MS_R$  имеет  $F$ -распределение с  $\nu_A = I - 1$  и  $\nu_R$  степенями свободы;
- статистика  $F_B = MS_B/MS_R$  имеет  $F$ -распределение с  $\nu_B = J - 1$  и  $\nu_R$  степенями свободы.

Критическая область для проверки гипотез  $H_0$  уровня значимости  $\alpha$  состоит из одного полуинтервала  $[F_{\nu_i, \nu_R, 1-\alpha}, \infty)$ , где  $F_{\nu_i, \nu_R, 1-\alpha}$  — квантиль  $F$ -распределения с  $\nu_i$  и  $\nu_R$  степенями свободы порядка  $1 - \alpha$ , где  $i = A, B, AB$ , соответственно.

**Замечание.** Если гипотеза об отсутствии взаимодействий отвергается, а гипотезы об отсутствии влияния фактора  $A$  и фактора  $B$  не отвергаются, то правильный вывод состоит в том, что нельзя сделать заключение об отсутствии влияния обоих факторов.

2.  $K_{ij} = K = 1$  (в каждой ячейке имеется одно наблюдение). Такие планы эксперимента используются, если априори известно, что взаимодействие факторов мало. Тогда модель двухфакторного дисперсионного анализа является *аддитивной* и имеет вид

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad (3.18)$$

где  $\varepsilon_{ij} \sim N(0, \sigma^2)$  и независимы,  $i = 1, \dots, I; j = 1, \dots, J$ .

Проверка гипотез об отсутствии влияния факторов  $A$  и  $B$  проводится так же, как и для модели (3.16), только роль  $SS_R$  играет  $SS_{AB}$ . Вычеркнув в табл. 3.6 строку «остаток» и заменив обозначения:  $SS_{AB}$  — на  $SS_R$ ,  $\nu_{AB}$  — на  $\nu_R$ , получим таблицу дисперсионного анализа для аддитивной модели (3.18).

3.  $K_{ij} > 1$  (неравные наблюдения в ячейках). В этом случае не существует разложения, подобного (3.17). Для проверки гипотез (см. случай 1) здесь также используется  $F$ -критерий, но построение его становится сложным и громоздким [Г. Шеффе, 1980].

Таблица 3.7

### Результаты наблюдений для примера 3.2

Год	Юноши		Девушки	
	$N$	$Y$	$N$	$Y$
2003	34	180,6	79	168,1
2004	53	171,1	53	169,9

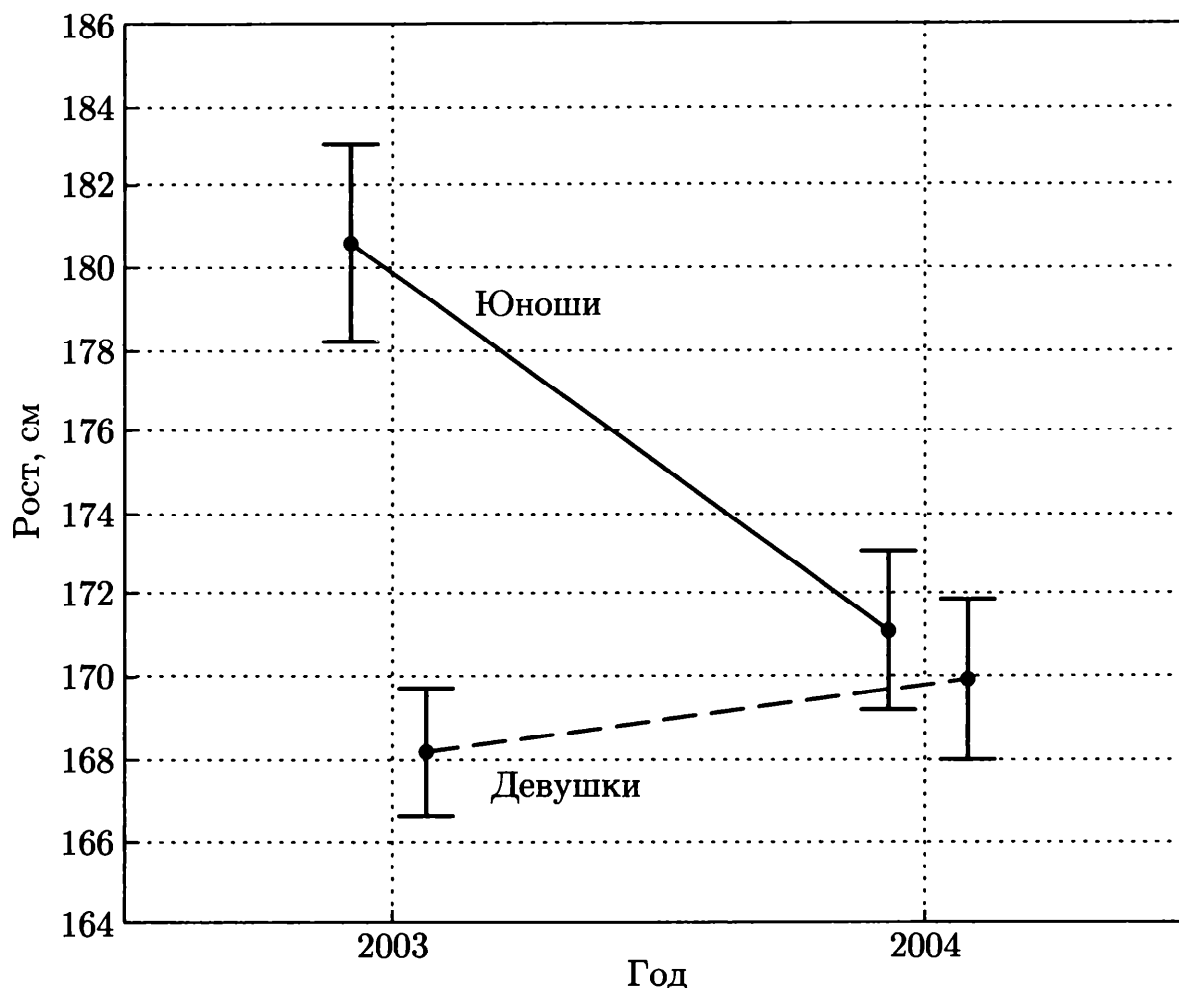


Рис. 3.4. Изменение средних значений переменной «рост» от факторов «пол» и «год»

**Пример 3.2.** Проводился анализ влияния на рост студентов второго курса биологического факультета МГУ им. М. В. Ломоносова 2003 и 2004 гг. факторов «пол» и «год». В табл. 3.7 для всех сочетаний уровней факторов приведены выборочные значения среднего роста юношей и девушек (см),  $\bar{Y}$ , и число наблюдений

Таблица 3.8

**Результаты двухфакторного дисперсионного анализа**

Источник дисперсии	Суммы квадратов $SS$	Число степеней свободы $df$	Средние квадраты $MS$	$F$ -отношение	$p$ -значение
Фактор 1 «пол»	2 342	1	2 342	45,3	0,000
Фактор 2 «год»	738	1	738	14,3	0,000
Взаимодействие 1 * 2	1 598	1	1 598	30,9	0,000
Ошибка	11 109	215	52	—	—

**Результаты множественных сравнений**

Пол	Год	Рост, см	Группа
Юноши	2003	180,5	1
Юноши	2004	171,1	2
Девушки	2004	169,9	2
Девушки	2003	168,1	2

ний,  $N$ . В табл. 3.8 приведены результаты двухфакторного дисперсионного анализа.

Анализ табл. 3.8 и рис. 3.4 показывает, что взаимодействие факторов «пол» и «год», а также влияние каждого из факторов статистически значимы на уровне значимости 0,05. Чтобы выяснить, какие из средних значений различаются значимо, используем метод множественного сравнения — метод Тьюки (HSD) для неравных  $n_i$ . Результаты таких сравнений, приведенные в табл. 3.9, показали, что в отдельную группу (1 группа) выделяются юноши-второкурсники 2003 г., средний рост которых значимо отличается от среднего роста юношей-второкурсников 2004 г., и девушек 2004 г. и 2003 г., среди которых нет значимого различия (2 группа).

**3.4.3. Смешанные модели дисперсионного анализа**

Рассмотрим свойства планов эксперимента для смешанных моделей, которые наиболее часто используются при проведении биологических исследований.

***План со случайными блоками***

Пусть, например, исследователь хочет сравнить урожайности  $I$  сортов некоторой культуры. Опыты проводятся на поле, плодородие которого может меняться от одного участка к другому. Если сорта по участкам распределять случайно, то один сорт может «получить», например, два «хороших» участка и ни одного «плохого», а другой сорт — наоборот. В этом случае ошибками наблюдений будут в основном различия между участками, а не ошибки наблюдения. Это может привести к тому, что не будет выявлено различия между сортами, даже если оно существует. Поэтому рассмотренный ранее однофакторный план может ока-



заться неэффективным. *План со случайными блоками* позволяет уменьшить влияние на результат эксперимента тех или иных неоднородностей условий, при которых проводятся отдельные наблюдения.

Покажем на этом примере построение плана со случайными блоками. Однотипные по плодородию участки группируются в  $J$  блоков (случайный фактор  $B$ ) по  $I$  участков, каждый из которых засевают одним из  $I$  сортов (постоянный фактор  $A$ ), причем внутри каждого блока сорта распределяются по участкам случайно. Таким образом, получаем *план со случайными блоками без повторений*, т. е.  $K = 1$ .

Обозначим через  $y_{ij}$  урожайность участка  $(i, j)$ , засеянного  $i$ -м сортом и принадлежащего  $j$ -му блоку, и положим

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, J, \quad (3.19)$$

где  $\mu$  — общее среднее;  $\alpha_i$  — эффект  $i$ -го уровня фактора  $A$ ;  $b_j$  — эффект  $j$ -го уровня фактора  $B$ ;  $b_j$  и  $\varepsilon_{ij}$  — независимые в совокупности нормально распределенные случайные величины с нулевым средним и дисперсиями  $\sigma_B^2$  и  $\sigma^2$  соответственно. Модель (3.19) называется *смешанной двухфакторной моделью*. Использование аддитивной модели объясняется тем, что оценить взаимодействие «сорт-блок» невозможно, ведь каждым сортом засеян только один участок внутри каждого блока.

Неизвестными параметрами в (3.19) являются  $\mu$ ,  $\alpha_i$  и две компоненты дисперсии  $\sigma^2$  и  $\sigma_B^2$ . МНК-оценки  $\mu$  и  $\alpha_i$ , как и для одно-

факторной модели I, при дополнительном ограничении  $\sum_{i=1}^I \alpha_i = 0$ ,

равны:  $\hat{\mu} = y_{..}$ ,  $\hat{\alpha}_i = y_{i.} - y_{..}$ ,  $i = 1, \dots, I$ . Несмещенные оценки для компонент дисперсии имеют вид:  $\hat{\sigma}^2 = MS_R$ ,  $\hat{\sigma}_B^2 = (MS_B - MS_R)/I$ . Вычеркнув в табл. 3.6 строку «остаток» и заменив обозначения:  $SS_{AB}$  — на  $SS_R$ ,  $\nu_{AB}$  — на  $\nu_R$  и  $MS_{AB}$  — на  $MS_R$ , получим таблицу дисперсионного анализа для модели (3.19).

Для этой модели проверяются две гипотезы:

- $H_0$ : все  $\alpha_i = 0$ , при альтернативе  $H_1$ : не все  $\alpha_i = 0$ . Критерий для проверки этой гипотезы строится, как и для модели I однофакторного дисперсионного анализа. Если верна гипотеза  $H_0$ , то статистика  $F_A = MS_A/MS_R$  имеет  $F$ -распределение с  $\nu_A = I - 1$  и  $\nu_R = (I - 1)(J - 1)$  степенями свободы;

- $H_0$ :  $\sigma_B^2 = 0$ , о том, что фактор блока  $B$  не вносит никакого вклада в дисперсию наблюдений, при альтернативе  $H_1$ :  $\sigma_B^2 \neq 0$ .

## План со случайными блоками с повторениями

Сорт	1-й блок из $I$ однородных участков	...	$J$ -й блок из $I$ однородных участков
1-й	1-я делянка	...	1-я делянка
	...		...
	$K$ -я делянка		$K$ -я делянка
...	...	...	...
$I$ -й	1-я делянка	...	1-я делянка
	...		...
	$K$ -я делянка		$K$ -я делянка

Критерий для проверки этой гипотезы строится, как и для модели II однофакторного дисперсионного анализа. Если верна гипотеза  $H_0$ , то статистика  $F_B = MS_B/MS_R$  имеет  $F$ -распределение с  $\nu_B = J - 1$  и  $\nu_R = (I - 1)(J - 1)$  степенями свободы.

Если каждый блок делится на  $KI$  делянок и каждый сорт случайно засеивается на  $K$  делянках внутри блока (табл. 3.10), то такой план называется **планом со случайными блоками с повторениями**. Использование плана с повторениями позволяет оценить взаимодействие «сорт-блок» [Г. Шеффе, 1980].

Необходимо отметить, что уменьшить влияние неоднородностей условий проведения эксперимента может только разбиение на однородные блоки. Любое свойство экспериментальных объектов, которое может быть определено до начала эксперимента, должно быть использовано для определения блока. Примерами однородных блоков являются следующие: в агробиологических исследованиях соседние площадки часто можно считать близкими по плодородию, а следовательно, группы смежных участков можно принимать за блоки; листья на одном растении при анти-вирусной прививке, животные одного помета при опытах с питанием, пробы крови одного человека при сравнении нескольких методов подсчета числа лейкоцитов. Отметим, что более полное исключение влияния неоднородностей условий достигается с помощью планов латинских квадратов [Г. Шеффе, 1980].

**Планы с группировкой**

Пусть необходимо выбрать план эксперимента для сравнения  $I$  методов (постоянный фактор  $A$ ) определения некоторой харак-

## План с группировкой

Метод	1-е животное	...	J-е животное	...	1-е животное	...	J-е животное
1-й	1-е измерение	...	1-е измерение	...	Нет	...	Нет
	... K-е измерение	...	... K-е измерение	...	Нет	...	Нет
...	...	...	...	...	...	...	...
I-й	Нет	...	Нет	...	1-е измерение	...	1-е измерение
	Нет	...	Нет	...	... K-е измерение	...	... K-е измерение

теристики крови (зависимая переменная  $Y$ ) у животного. Рассмотрим план, для проведения которого потребуется  $IJ$  животных. Для каждого метода случайно выбираются «свои»  $J$  животных (случайный фактор  $B$  с  $J$  уровнями), т. е. у каждого из  $J$  животных одной выборки проводится  $K$  независимых измерений характеристики крови первым методом, для другой выборки — вторым методом и т. д. (табл. 3.11). В этом случае каждый уровень фактора  $B$  сочетается не более чем с одним уровнем фактора  $A$ . Такой план эксперимента называется **двухфакторным планом с группировкой**, или **иерархическим**, или **гнездовым** (*nested*). Фактор  $A$  называют **группирующим фактором**, а фактор  $B$  — **сгруппированным фактором**  $A$ , что обозначается  $B(A)$ . Фактор  $A$  может быть как постоянным, так и случайным, фактор  $B$  — случайным. Одна из основных областей применения иерархических планов — генетика. Эти планы используются, например, при изучении наследственного влияния родительских поколений на продуктивность или другие признаки потомства.

В качестве примера рассмотрим **смешанную модель для двухфакторного плана с группировкой**, когда фактор  $A$  является постоянным, а фактор  $B(A)$  — случайным:

$$y_{ijk} = \mu + \alpha_i + b_{j(i)} + \varepsilon_{ijk}, \quad (3.20)$$

$$i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K,$$

где  $\mu$  — общее среднее;  $\alpha_i$  — эффект  $i$ -го уровня фактора  $A$ ;  $b_{j(i)}$  — эффект  $j$ -го уровня фактора  $B$ , сгруппированного  $i$ -м уровнем фактора  $A$ ;  $b_{j(i)}$  и  $\varepsilon_{ijk}$  — независимые в совокупности, нормально распределенные случайные величины с нулевым средним и дисперсиями  $\sigma_{B(A)}^2$  и  $\sigma^2$  соответственно. Компонента  $\sigma_{B(A)}^2$  обусловлена вкладом случайного фактора  $B(A)$ , сгруппированного фактором  $A$ , в общую изменчивость  $Y$ .

Неизвестными параметрами в (3.20) являются  $\mu$ ,  $\alpha_i$  и две компоненты дисперсии  $\sigma^2$  и  $\sigma_{B(A)}^2$ . МНК-оценки  $\mu$  и  $\alpha_i$  при допол-

нительном ограничении  $\sum_{i=1}^I \alpha_i = 0$  равны:  $\hat{\mu} = y_{...}$ ;  $\hat{\alpha}_i = y_{i..} - y_{...}$ ,

$i = 1, \dots, I$ . Несмещенные оценки для компонент дисперсии имеют вид:  $\hat{\sigma}^2 = MS_R$ ,  $\hat{\sigma}_{B(A)}^2 = (MS_{B(A)} - MS_R)/K$ . Результаты анализа для модели (3.20) заносятся в таблицу дисперсионного анализа для плана с группировкой (табл. 3.12).

Для этой модели проверяются две гипотезы:

- $H_0$ : все  $\alpha_i = 0$ , при альтернативе  $H_1$ : не все  $\alpha_i = 0$ . Если верна гипотеза  $H_0$ , то статистика  $F_A = MS_A/MS_{B(A)}$  (см. табл. 3.12) имеет  $F$ -распределение с  $\nu_A = I - 1$  и  $\nu_{B(A)} = I(J - 1)$  степенями свободы;

- $H_0$ :  $\sigma_{B(A)}^2 = 0$ , при альтернативе  $H_1$ :  $\sigma_{B(A)}^2 \neq 0$ . Если верна гипотеза  $H_0$ , то статистика  $F_{B(A)} = MS_{B(A)}/MS_R$  (см. табл. 3.12) имеет  $F$ -распределение с  $\nu_{B(A)} = I(J - 1)$  и  $\nu_R = IJ(K - 1)$  степенями свободы.

Рассмотрим пример смешанной модели для трехфакторного плана с группировкой, когда постоянный фактор  $A$  является группирующим, фактор  $B(A)$  — сгруппированным фактором  $A$ , а фактор  $C(B)$  — сгруппированным фактором  $B$ .

**Пример 3.3.** Для сравнения трех методов (фактор  $A$ ) определения содержания гликогена в печени крысы проводились два повторных измерения для каждой из трех случайных проб печени (фактор  $C$ ), взятых у каждой из двух крыс (фактор  $B$ ) [R. R. Sokal, F. J. Rohlf, 1995]. В этом случае фактор  $A$  — постоянный группирующий фактор с тремя уровнями; случайный фактор  $B$  с двумя уровнями сгруппирован фактором  $A$ ; случайный трехуровневый фактор  $C$  сгруппирован фактором  $B$  (рис. 3.5).

Дисперсионный анализ двухфакторного плана с группировкой

Источник разброса	Сумма квадратов	Число степеней свободы	Средний квадрат	F-отношение
Фактор A	$SS_A = JK \sum_{i=1}^I (y_{i..} - y_{...})^2$	$\nu_A = I - 1$	$MS_A = \frac{SS_A}{\nu_A}$	$F_A = \frac{MS_A}{MS_{B(A)}}$
Фактор B (внутри A)	$SS_{B(A)} = K \sum_{i=1}^I \sum_{j=1}^J (y_{ij.} - y_{i..})^2$	$\nu_{B(A)} = I(J - 1)$	$MS_{B(A)} = \frac{SS_{B(A)}}{\nu_{B(A)}}$	$F_{B(A)} = \frac{MS_{B(A)}}{MS_R}$
Остаток (ошибка)	$SS_R = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{ij.})^2$	$\nu_R = IJ(K - 1)$	$MS_R = \frac{SS_R}{\nu_R}$	—
Полная	$SS_T = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - y_{...})^2$	$\nu_T = IJK - 1$	—	—

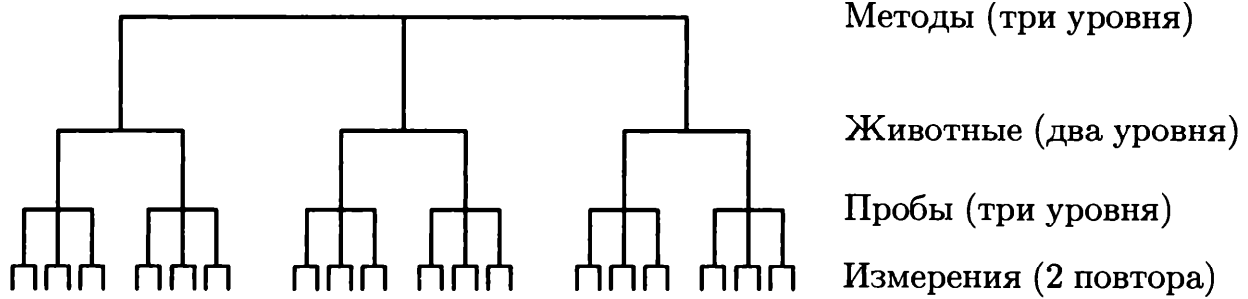


Рис. 3.5. Пример трехфакторного плана с группировкой

Для анализа результатов эксперимента использовалась *смешанная модель для трехфакторного плана с группировкой*

$$y_{ijk} = \mu + \alpha_i + b_{j(i)} + c_{l(j)} + \varepsilon_{ijkl}, \quad I = 3, J = 2, L = 3, K = 2.$$

Значения оценок общего среднего и среднего содержания гликогена в печени крысы, определенного каждым из трех методов, равны:

$$\hat{\mu} = y_{\dots} = 142,2; \quad \hat{\mu}_1 = 132,5; \quad \hat{\mu}_2 = 151,0; \quad \hat{\mu}_3 = 135,2.$$

Анализ табл. 3.13 показывает, что гипотеза  $H_0$ : все  $\alpha_i = 0$  не отвергается на уровне  $\alpha = 0,05$ , т. е. различия между методами определения содержания гликогена в печени крысы не обнаружено; гипотеза  $H_0$ :  $\sigma_{B(A)}^2 = 0$  отвергается на уровне  $\alpha = 0,05$ , как и гипотеза  $H_0$ :  $\sigma_{C(B)}^2 = 0$  — на уровне  $\alpha = 0,05$ .

Приведем значения оценок компонент дисперсии:

$$\hat{\sigma}_{B(A)}^2 = 36,1; \quad \hat{\sigma}_{C(B)}^2 = 14,2; \quad \hat{\sigma}^2 = MS_R = 21,2.$$

Так как гипотеза  $H_0$ : все  $\alpha_i = 0$  не отвергается, а гипотезы  $H_0$ :  $\sigma_{B(A)}^2 = 0$  и  $H_0$ :  $\sigma_{C(B)}^2 = 0$  отвергаются, оценка дисперсии

Таблица 3.13

**Результаты дисперсионного анализа для примера 3.3 (план с группировкой)**

Источник разброса	Сумма квадратов $SS$	Число степеней свободы $\nu$	Средний квадрат $MS = SS/\nu$	$F$ -отношение	$p$ -значение
A	1557,5	2	778,8	2,93	0,197
B(A)	797,7	3	265,9	5,37	0,014
C(B)	594,0	12	49,5	2,34	0,050
Остаток	381,0	18	21,2	—	—
Полная	3330,2	35	—	—	—

единичного наблюдения и доля каждой компоненты дисперсии определяются так:

$$\hat{\sigma}_y^2 = \hat{\sigma}_{B(A)}^2 + \hat{\sigma}_{C(B)}^2 + \hat{\sigma}^2 = 71,4.$$

Вклад фактора  $B(A)$  в общую дисперсию равен  $(36,1/71,4) \times 100 \% = 50,5 \%$ ; вклад фактора  $C(B)$  равен  $(14,2/71,4)100 \% = 19,8 \%$ ; доля дисперсии ошибки наблюдения равна  $(21,2/71,4) \times 100 \% = 29,6 \%$ .

На основании проведенного анализа можно сделать следующие выводы: так как наибольший вклад в общую дисперсию вносит фактор  $B$ , то при планировании подобных экспериментов желательно увеличить число уровней фактора  $B$ , т. е. число крыс, а также число повторных измерений.

### **3.4.4. Непараметрические методы анализа связи между количественной и качественными переменными**

Дисперсионный анализ опирается на предположение нормальности распределения исходных данных. Возможны два подхода к ситуациям, когда это предположение не выполняется. Первый подход заключается в том, чтобы подобрать преобразование, которое сделало бы распределение зависимой переменной  $Y$  более близким к нормальному. Второй подход состоит в использовании методов анализа, не зависящих от вида распределения наблюдаемых случайных величин, из которых наиболее разработаны ранговые методы. Рассмотрим некоторые из них. Более подробно с этими методами можно познакомиться в [М. Холлендер, Д. А. Вульф, 1983; J. H. Zar, 1999].

#### ***Однофакторный анализ***

Пусть фактор  $A$  варьируется в эксперименте на  $I$  уровнях и на  $i$ -м уровне производится  $n_i$  наблюдений  $y_{i1}, \dots, y_{in_i}$ ,  $i = 1, \dots, I$ ,  $N = n_1 + \dots + n_I$ . Предполагается, что  $I$  выборки являются независимыми и каждая из них извлечена из непрерывной совокупности с неизвестным распределением  $F_i(x)$ , причем распределения  $F_i(x)$  имеют одинаковую форму и могут различаться лишь параметром сдвига.

Для проверки гипотезы  $H_0$  о том, что  $I$  выборки извлечены из одной и той же совокупности, можно воспользоваться свободным от распределения **критерием Крускала — Уоллиса** (*Kruskal — Wallis test*). Этот критерий является обобщением

двухвыборочного критерия Вилкоксона на случай  $I$  выборок,  $I \geq 2$ . Упорядочим все  $N$  наблюдений по возрастанию, заменим величины  $y_{ij}$  их рангами и найдем для каждой выборки  $i$  сумму рангов  $R_i, i = 1, \dots, I$ . Статистика Крускала — Уоллиса имеет вид:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^I \frac{R_i^2}{n_i} - 3(N+1).$$

Заметим, что среди  $y_{ij}$  могут быть равные по величине наблюдения. Если совпадения встречаются внутри одной выборки, то при ранжировании применяется метод случайного ранга. Такого типа совпадения не влияют на значение статистики критерия. Когда же совпадения принадлежат разным выборкам, то используется метод среднего ранга (см. 2.3.8). В этом случае значение статистики критерия  $H$  следует разделить на величину

$$1 - \frac{\sum_{k=1}^K (t_k - 1)t_k(t_k + 1)}{N^3 - N}, \quad (3.21)$$

где  $K$  — общее число выборок с равными наблюдениями;  $t_k$  — число совпадений в  $k$ -й выборке,  $k = 1, \dots, K$ .

Если гипотеза  $H_0$  верна, то статистика  $H$  распределена асимптотически при  $n \rightarrow \infty$  как  $\chi^2$  с  $I - 1$  степенями свободы. Этим приближением можно пользоваться при  $n_i > 5$  или  $I > 3$ . Критическое множество уровня значимости  $\alpha$  при альтернативе  $H_1$ , что не все выборки извлечены из одной совокупности, состоит из одного полуинтервала  $[\chi_{I-1, 1-\alpha}^2, +\infty)$ , где  $\chi_{I-1, 1-\alpha}^2$  — квантиль  $\chi^2$ -распределения с числом степеней свободы  $I - 1$  порядка  $1 - \alpha$ .

Для проверки гипотезы о том, что медианы  $I$  выборок совпадают, можно воспользоваться **медианным критерием Брауна — Муда** (*Brown — Mood test*). Статистика этого критерия имеет вид

$$X^2 = \sum_{i=1}^I \frac{(m_i - n_i/2)^2}{n_i/4},$$

где  $m_i$  — число наблюдений  $i$ -й выборки, значения которой не меньше значения выборочной медианы для объединенной выборки из  $N$  наблюдений.

Если верна нулевая гипотеза, то статистика  $X^2$  распределена асимптотически как  $\chi^2$  с  $I - 1$  степенями свободы. Критическое



множество уровня значимости  $\alpha$  при альтернативе  $H_1$ , что хотя бы две выборки имеют различные медианы, состоит из одного полуинтервала  $[\chi_{I-1,1-\alpha}^2, +\infty)$ , где  $\chi_{I-1,1-\alpha}^2$  — квантиль  $\chi^2$ -распределения с  $I - 1$  степенями свободы порядка  $1 - \alpha$ .

Если гипотеза  $H_0$  отвергается, то чтобы выделить заведомо различные выборки, применяются **свободные от распределения методы множественного сравнения**. Для сравнения всевозможных пар выборок (непараметрический аналог непланируемых сравнений, которые рассматривались в 3.4.1) можно использовать **приближенный метод Данна (Dunn test)**. Две выборки различаются параметром сдвига значимо на уровне значимости  $\alpha$ , если

$$|R_{i.} - R_{j.}| \geq u_{1-\alpha/(I(I-1))} \sqrt{\frac{N(N+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)},$$

где  $R_{i.} = R_i/n_i$ ;  $R_i$  — сумма рангов  $i$ -й выборки;  $R_{j.} = R_j/n_j$ ;  $R_j$  — сумма рангов  $j$ -й выборки,  $i < j = 1, \dots, I$ ;  $u_{1-\alpha/(I(I-1))}$  — квантиль стандартного нормального распределения порядка  $1 - \alpha/(I(I - 1))$ .

В случае равных объемов выборок,  $n_1 = \dots = n_I = n \geq 8$ , можно использовать **приближенный метод Дуосса (Dwass method)**. Для каждой пары выборок вычисляется статистика Манна — Уитни,  $U_{ij}$ ,  $i < j$ ,  $i, j = 1, \dots, I$ , критическое значение которой равно

$$U_{I,n;1-\alpha} = \frac{n^2}{2} + Q_{I,\infty;1-\alpha} n \sqrt{\frac{2n+1}{24}},$$

где  $Q_{I,\infty;1-\alpha}$  — квантиль порядка  $1 - \alpha$  студентизированного размаха с числом степеней свободы  $I, \infty$  [Л. Н. Большев, Н. В. Смирнов, 1983; R. R. Sokal, F. J. Rohlf, 1995]. Выборки, для которых  $U_{ij} \geq U_{I,n;1-\alpha}$ , статистически значимо различаются параметром сдвига.

**Пример 3.4.** В работе Хачатурьян и Панченко<sup>1</sup> проводилось исследование влияния факторов внешней среды, одним из которых является время суток, на устойчивость организма к гипоксии. Характеристикой устойчивости к острой гипоксии у крыс-самцов Вистар являлось время жизни животного, т. е. время пребывания в барокамере после «подъема» на высоту 11,5 км до обратимой остановки дыхания, после которой осуществляли

<sup>1</sup>Хачатурьян М. Л., Панченко Л. А. Влияние сезона года на устойчивость крыс к гипоксии // Бюл. экспер. биол. — 2002. — Т. 133. — № 3. — С. 348 — 352.

## Результаты однофакторного анализа для примера 3.4

$I$	1	2	3
$n_i$	29	43	54
$R_i$	2313,5	2054,5	3633,0
$R_{i.}$	79,77	47,78	62,28
$H(2, N = 126) = 14,52; p = 0,0007$			

Таблица 3.15

## Результаты сравнения пар выборок

	$ R_{i.} - R_{j.} $	$SE$	$ R_{i.} - R_{j.} /SE$	$u_{крит}$
$ R_{1.} - R_{2.} $	31,99	8,78	3,65	2,39
$ R_{1.} - R_{3.} $	17,49	8,40	2,08	2,39
$ R_{2.} - R_{3.} $	14,50	7,30	1,99	2,39

«спуск». Приведем пример анализа данных для низкоустойчивых к острой гипоксии крыс. Измерения проводились в 13, 17 и 21 ч (табл. 3.14). Проверка гипотезы о нормальности распределения каждой из трех выборок показала, что распределение исходных данных отлично от нормального. Поэтому для проверки гипотезы  $H_0$  о том, что время суток не влияет на устойчивость к гипоксии, при альтернативе  $H_1$ , что фактор влияет, использовался критерий Крускала — Уоллиса. Анализ табл. 3.14 показывает, что гипотеза  $H_0$  отвергается на уровне значимости  $\alpha = 0,05$ . Для сравнения всех пар выборок использовался метод Данна для  $1 - \alpha = 0,95$ ,  $I = 3$ ,  $u_{крит} = u_{0,9917} = 2,39$  (табл. 3.15).

Результаты анализа, приведенные в табл. 3.15, позволяют сделать вывод о том, что только две выборки (для 13 и 17 ч) отличаются значимо параметром сдвига, так как  $|R_{1.} - R_{2.}|/SE = 3,65$ , что больше  $u_{крит} = 2,39$ .

**Двухфакторный анализ. Критерий Фридмана**

Метод Фридмана предназначен для анализа результатов наблюдений по схеме плана случайных блоков без повторений. Пусть в каждом из  $J$  блоков имеется по одному наблюдению на каждом из  $I$  уровней постоянного фактора  $A$ . Предполагается, что каждая из  $I$  выборок извлечена из непрерывной совокупности с неизвестным распределением  $F_i(x)$ . Распределения

$F_i(x)$ ,  $i = 1, \dots, I$ , имеют одинаковую форму и могут различаться лишь параметром сдвига.

Результаты наблюдений можно представить в виде таблицы из  $I$  строк и  $J$  столбцов. Наблюдениям каждого столбца сопоставим соответствующие им ранги. Обозначим через  $R_i$  сумму рангов  $i$ -й строки. Для проверки гипотезы  $H_0$  о том, что фактор  $A$  не влияет, т. е. все  $I$  выборок извлечены из одной и той же совокупности, против альтернативы  $H_1$ , что хотя бы  $F_i(x)$  и  $F_j(x)$ ,  $i \neq j$ , различаются параметром сдвига, можно воспользоваться свободным от распределения **критерием Фридмана** (*Friedman test*). Статистика критерия Фридмана имеет вид:

$$\chi_r^2 = \frac{12}{IJ(I+1)} \sum_{i=1}^I R_i^2 - 3J(I+1).$$

Если верна гипотеза  $H_0$ , то статистика  $\chi_r^2$  распределена асимптотически при  $n \rightarrow \infty$  как  $\chi^2$  с  $I - 1$  степенями свободы. Критическое множество уровня значимости  $\alpha$  при альтернативе  $H_1$  состоит из одного полуинтервала  $[\chi_{I-1, 1-\alpha}^2, +\infty)$ , где  $\chi_{I-1, 1-\alpha}^2$  — квантиль  $\chi^2$ -распределения с числом степеней свободы  $I - 1$  порядка  $1 - \alpha$ .

Если в столбцах таблицы встречаются совпадающие значения, то при переходе к таблице рангов применяют метод средних рангов (см. 2.3.8), а значение статистики критерия  $\chi_r^2$  следует разделить на величину (3.21), заменив  $N$  на  $IJ$ .

Для сравнения всевозможных пар выборок можно использовать **свободный от распределения метод множественного сравнения Дюосса**, заменив в (3.21)  $N$  на  $J$ .

### **Двухфакторный анализ. Критерий Пейджа**

Часто исследователь может предположить монотонное влияние фактора на зависимую переменную. В этом случае для проверки гипотезы  $H_0$  об отсутствии влияния фактора  $A$  против альтернативы  $H_1$  об упорядоченном влиянии фактора  $A$  можно воспользоваться **критерием Пейджа** (*Page test*). Этот критерий является более мощным, чем критерий Фридмана, для данной альтернативы  $H_1$ . Статистика Пейджа имеет вид

$$L = \sum_{i=1}^I IR_i = R_1 + 2R_2 + \dots + IR_I,$$

где  $R_i$  — сумма рангов  $i$ -го столбца, как и ранее. Если верна гипотеза  $H_0$ , то при  $n \rightarrow \infty$  статистика

$$L^* = \frac{L - IJ(I + 1)/4}{(J(I^3 - J^3)/144(I - 1))^{1/2}}$$

распределена асимптотически нормально  $N(0, 1)$ . Критическое множество уровня значимости  $\alpha$  при альтернативе  $H_1$  состоит из одного полуинтервала  $[u_{1-\alpha}, +\infty)$ , где  $u_{1-\alpha}$  — квантиль порядка  $1 - \alpha$  стандартного нормального распределения. Если в строках таблицы встречаются совпадающие значения, то при переходе к таблице рангов используются средние ранги. Число совпадений влияет на точность полученных выводов, с увеличением числа совпадений выводы становятся менее точными.

### 3.5. Кластерный анализ

Кластерный анализ в некотором смысле похож на многомерный дисперсионный анализ — количественные отклики и качественные факторы. Существенное различие состоит в том, что факторы неизвестны и надо каким-то образом их сконструировать. Идея этого конструирования проста — наблюдения разбиваются на однородные группы, которые объявляются разными классами, а переменная, указывающая на принадлежность каждого наблюдения одному из этих классов, объявляется искомым качественным фактором. Проблема, следовательно, состоит только в том, как разбить совокупность наблюдений на однородные группы.

Имеется много методов, решающих эту задачу, наиболее распространен среди них *агломеративно-иерархический кластерный анализ*, который последовательно объединяет наблюдения, начиная с самых близких, во все более крупные группы. Результаты этого последовательного объединения представляются в виде *дендрограммы*. Можно по-разному определять как близость отдельных наблюдений, так и сформированных из них групп — *кластеров* (от англ. *cluster* — гроздь, скопление).

Исходные данные в задачах кластерного анализа задаются в виде таблицы (матрицы), строки которой соответствуют различным наблюдениям, а столбцы — различным переменным

$$\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{pmatrix}.$$

Если рассматривать строки-наблюдения, как точки в пространстве переменных, то естественно использовать в качестве **меры различия** между наблюдениями  $i$  и  $k$  **евклидово расстояние** между соответствующими им точками в  $m$ -мерном евклидовом пространстве

$$d_{ik} = \sqrt{\sum_{j=1}^m (y_{ij} - y_{kj})^2}.$$

В случае если переменные измерены в разных единицах, то суммирование квадратов их разностей будет неправомерным. Сначала необходимо сделать переменные соизмеримыми, например путем их **стандартизации**, т. е. вычитания среднего и деления на среднеквадратическое отклонение:

$$\tilde{y}_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j},$$

где

$$\bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij} \quad \text{и} \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}.$$

Можно использовать и другие меры различия между наблюдениями, например **манхеттенское**, или **сити-блок, расстояние**, равное сумме абсолютных значений разностей между значениями переменных:

$$\Delta_{ik} = \sum_{j=1}^n |y_{ij} - y_{kj}|.$$

Для бинарных переменных манхеттенское расстояние между двумя наблюдениями равно просто сумме несовпадающих значений в них и обычно называется **хэмминговым**.

Вычислив каким-либо способом расстояния между всеми парами наблюдений, получают **матрицу расстояний**

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix},$$

которая и служит основой для дальнейших расчетов. Главная диагональ матрицы состоит из нулей, и матрица симметрична относительно этой диагонали.

Имея матрицу расстояний, можно начать процесс последовательного объединения наблюдений в кластеры. Алгоритм этого объединения работает следующим образом.

Сначала все наблюдения считаются отдельными кластерами. На первом шаге в матрице расстояний ищется минимальный внедиагональный элемент (поскольку матрица симметрична, достаточно просмотреть ее поддиагональную часть). Пусть этот элемент находится на пересечении  $i$ -й строки и  $j$ -го столбца, что означает, что среди всех пар кластеров, каждый из которых состоит из одного наблюдения, наиболее близки между собой  $i$ -й и  $j$ -й кластеры. Эти два кластера объединяются в один, и общее число кластеров уменьшается на единицу — их становится  $n - 1$  вместо исходных  $n$ . При этом возникает проблема определения расстояния от вновь образованного кластера  $i + j$  до остальных кластеров.

Рассмотрим некоторые возможные варианты.

1. Можно определить расстояние от  $i + j$  до любого другого кластера  $k$  как минимум из расстояний от  $i$  до  $k$  и от  $j$  до  $k$ :

$$d_{i+j,k} = \min\{d_{ik}, d_{jk}\}.$$

Этот метод называется методом *минимальной связи*, или *методом ближайшего соседа*.

2. Определить расстояние от  $i + j$  до любого другого кластера  $k$  как максимум из расстояний от  $i$  до  $k$  и от  $j$  до  $k$ :

$$d_{i+j,k} = \max\{d_{ik}, d_{jk}\}$$

— это *метод максимальной связи*, или *метод дальнего соседа*.

3. Определить расстояние от  $i + j$  до любого другого кластера  $k$  как среднее арифметическое из расстояний от  $i$  до  $k$  и от  $j$  до  $k$ :

$$d_{i+j,k} = \frac{d_{ik} + d_{jk}}{2}$$

— *метод средней связи*.

Второй и последующий шаги аналогичны первому — каждый раз объединяются два самых близких кластера и вычисляются расстояния от нового объединенного кластера до остальных. На

каждом шаге общее число кластеров сокращается на единицу, и процесс заканчивается, когда все кластеры объединятся в один.

Метод средней связи имеет модификацию, называемую *методом взвешенной средней связи*. Она отличается тем, что если на некотором шаге объединяются кластеры от  $i$  и  $j$ , состоящие соответственно из  $N_i$  и  $N_j$  наблюдений, то расстояние от их объединения до любого другого кластера  $k$  вычисляется не как простое среднее, а как среднее взвешенное из расстояний от  $i$  до  $k$  и от  $j$  до  $k$ :

$$d_{i+j,k} = \frac{N_i d_{ik} + N_j d_{jk}}{N_i + N_j}.$$

Метод дальнего соседа имеет тенденцию формировать кластеры одинакового размера и может не учитывать реальной неоднородности расположения наблюдений в пространстве признаков. Метод ближайшего соседа, наоборот, хорошо отслеживает локальные неоднородности и может выявлять кластеры довольно сложной формы и разного размера. Однако возможно объединение очень разных кластеров, если между ними случайно имеется цепочка близких наблюдений. Компромиссом между этими двумя методами является метод взвешенной средней связи.

Результаты кластерного анализа удобно представлять в виде дендрограммы, графически показывающей последовательность объединений с учетом расстояний между объединяющимися кластерами. Визуальный анализ дендрограммы позволяет определить число кластеров, на которые естественным образом разбивается совокупность наблюдений — объединение сильно различающихся кластеров отражается на дендрограмме скачкообразным увеличением межкластерного расстояния.

На рис. 3.6 приведен пример работы агломеративно-иерархического кластерного анализа. Данные состоят из 12 наблюдений, характеризующихся двумя переменными  $Y_1$  и  $Y_2$  (они представлены на рис. 3.6, а в виде точек на плоскости). Здесь наглядно иллюстрируется процесс объединения наблюдений в кластеры: сначала объединяются наиболее близкие друг к другу наблюдения 6 и 4, затем — 3 и 1, потом к  $3 + 1$  присоединяется 2 и т. д. (рис. 3.6, а). На рис. 3.6, б представлена дендрограмма — результат работы алгоритма, на которой ясно видно, что совокупность наблюдений четко разделена на два кластера. Конечно, это еще лучше видно из представления точек на плоскости (см. рис. 3.6, а), но это возможно только потому, что рассматривается случай двух переменных. Дендрограмму же можно получить

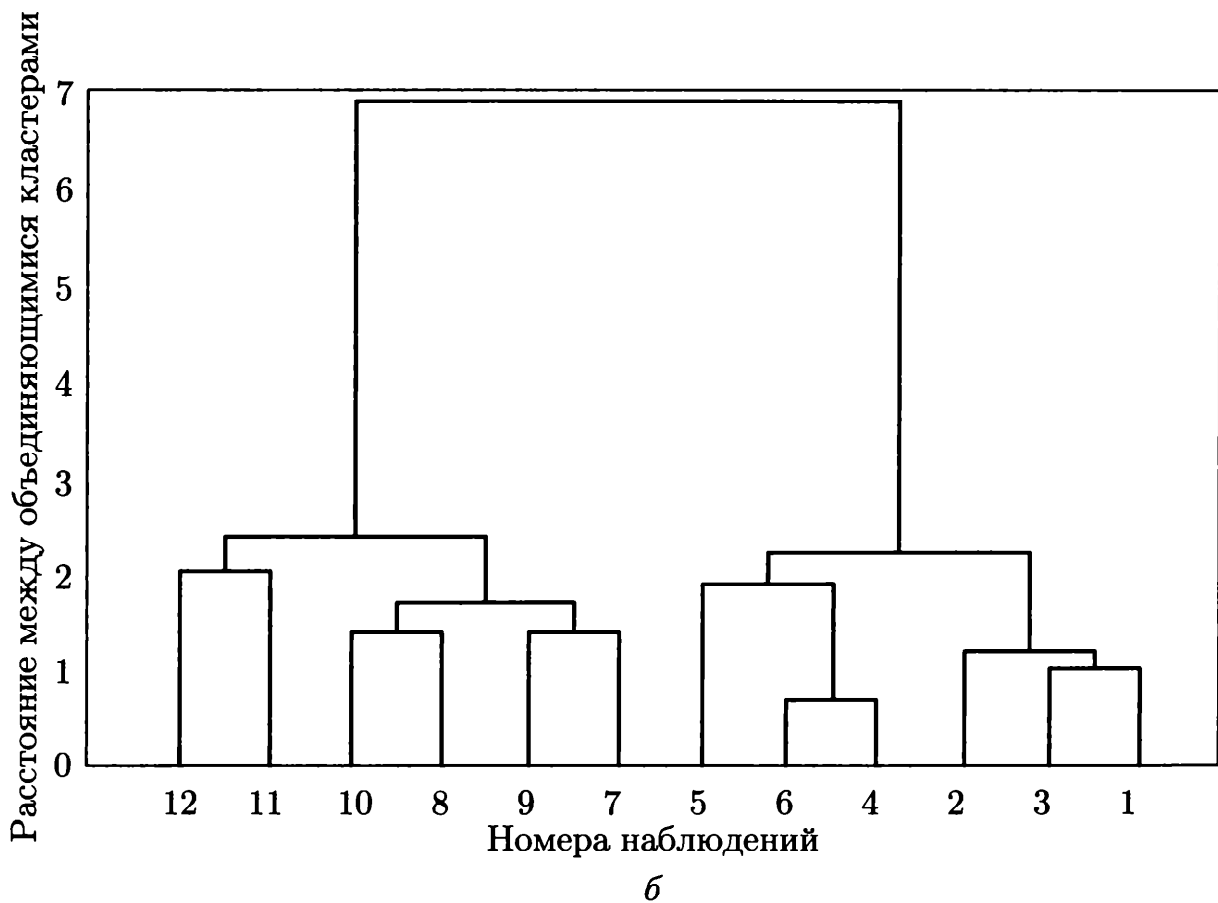
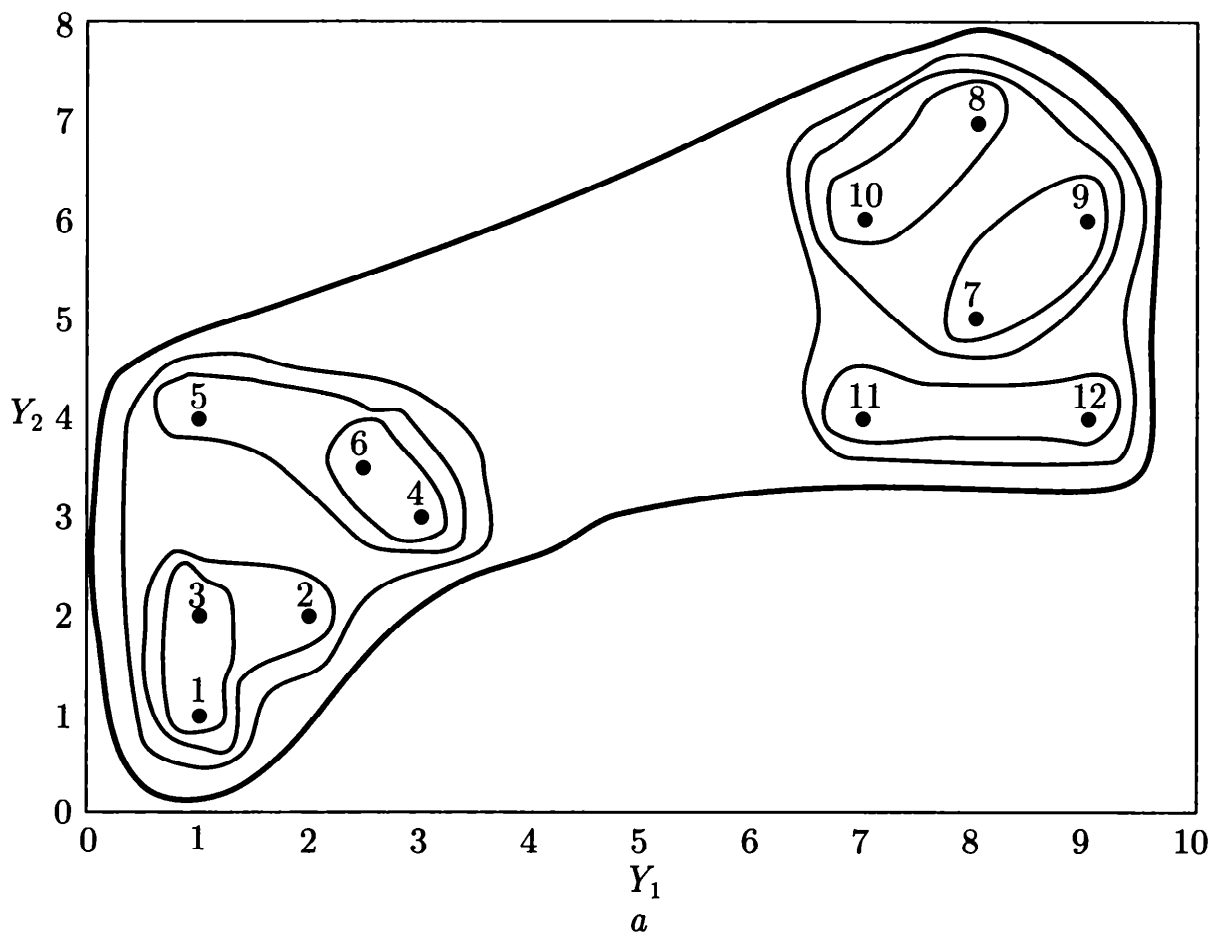
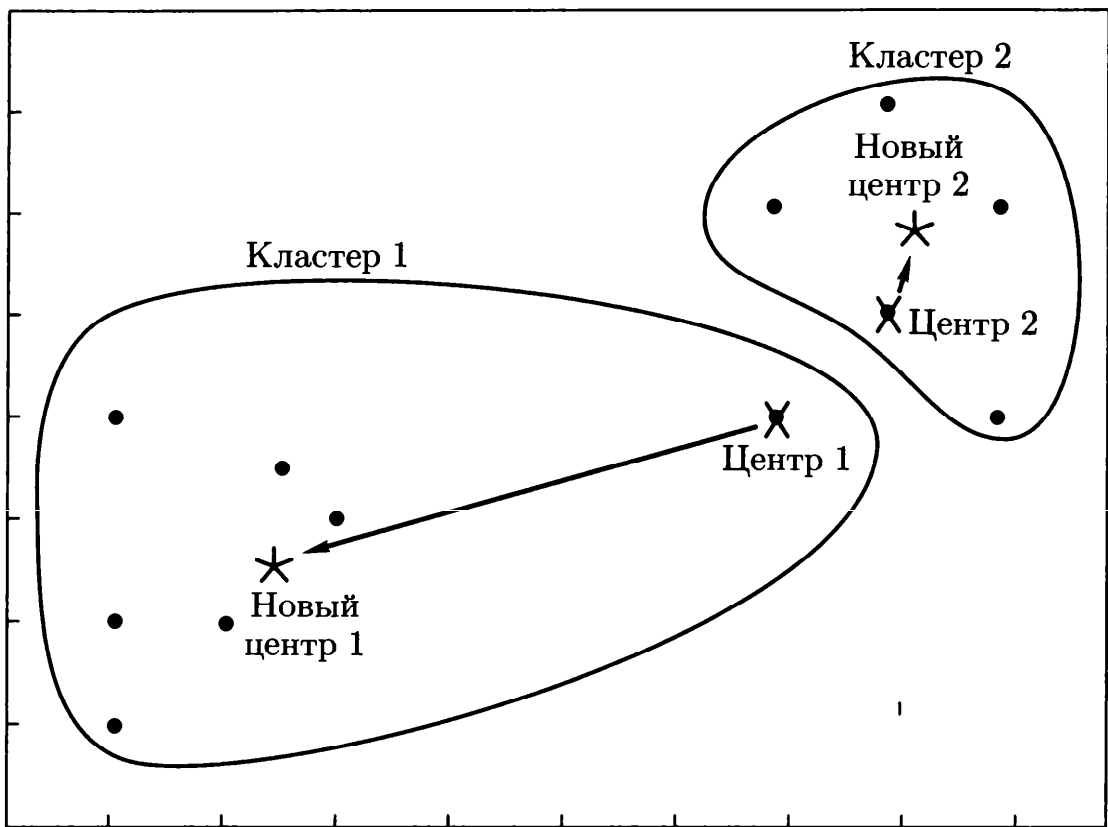


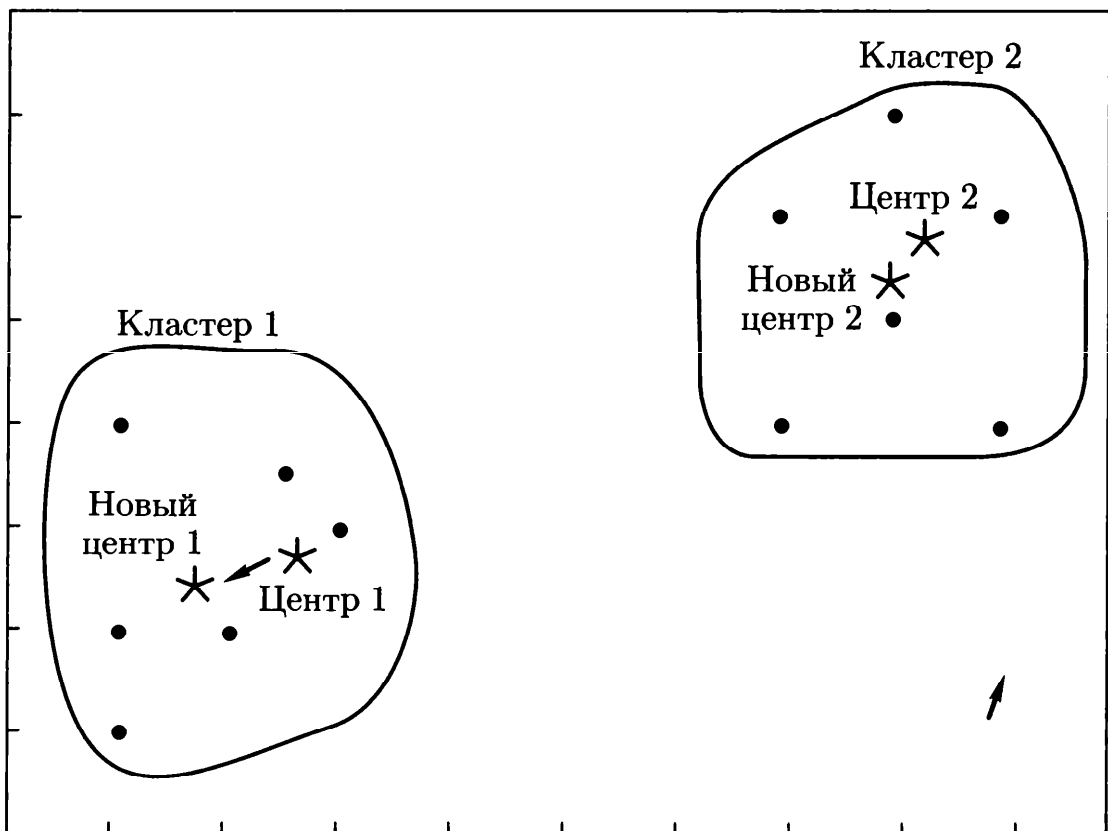
Рис. 3.6. Пример работы агломеративно-иерархического кластерного анализа:

*a* — процесс объединения наблюдений в кластеры; *b* — дендрограмма





*a*



*b*

Рис. 3.7. Пример работы метода *K*-средних:

*a* — 1-й шаг алгоритма; *b* — 2-й шаг алгоритма

при любом числе переменных, поскольку она строится на основе анализа расстояний между наблюдениями, а их можно вычислять для любого числа переменных (см. формулы для евклидова или манхеттеновского расстояний).

Среди других широко известных алгоритмов кластерного анализа следует отметить *метод  $K$ -средних*, при реализации которого среди исходных наблюдений выбирают наудачу  $K$  наблюдений — начальных центров будущих кластеров. Первый шаг алгоритма начинается с распределения всех наблюдений между этими центрами — каждое наблюдение относится к ближайшему центру. После этого центры кластеров пересчитываются заново — каждая координата центра кластера вычисляется как среднее из координат, входящих в этот кластер наблюдений. Второй и последующие шаги аналогичны первому. Процедура заканчивается, когда центры кластеров перестанут смещаться при новом пересчете. На рис. 3.7 показано два шага работы алгоритма. Очевидно, что на третьем шаге не будет изменения положения центров кластеров, так как их состав не изменится. Таким образом, в данном случае для разбиения совокупности наблюдений на кластеры оказалось достаточно двух шагов.

Несомненным преимуществом метода  $K$ -средних является возможность классифицировать наборы из многих сотен и даже тысяч наблюдений, тогда как с помощью агломеративно-иерархического метода, хотя теоретически и возможно, но слишком громоздко обрабатывать более 100 — 200 наблюдений. Недостатки метода состоят в необходимости априорного задания числа кластеров, отсутствии наглядного представления результатов и зависимости результатов от начального выбора центров — при неудачном их выборе алгоритм может дать классификацию, не соответствующую реально имеющейся.

### 3.6. Факторный анализ

Целью факторного анализа (*factor analysis*) является поиск представления, возможно, приближенного, но с наименьшей потерей информации, исходного набора  $m$  переменных  $Y_1, Y_2, \dots, Y_m$  в виде линейных комбинаций меньшего числа  $r$  переменных  $V_1, V_2, \dots, V_r$ , называемых *факторами*. Для упрощения изложения будем считать  $Y_1, Y_2, \dots, Y_m$  центрированными, т. е. полученными из непосредственно измеренных переменных вычитанием из них их средних. Тогда переменные можно выра-

зить в виде линейных комбинаций (без свободных членов) центрированных факторов  $V_1, V_2, \dots, V_r$ :

$$Y_i = \sum_{j=1}^r a_{ij} V_j, \quad i = 1, 2, \dots, m.$$

В матричных обозначениях эти соотношения можно записать следующим образом:

$$\mathbf{Y} = \mathbf{AV} \quad (3.22)$$

или в развернутом матричном виде

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{1r} \\ a_{21} & a_{22} & a_{2r} \\ a_{31} & a_{32} & a_{3r} \\ \dots & & \dots \\ a_{m1} & a_{m2} & a_{mr} \end{pmatrix} \times \begin{pmatrix} V_1 \\ V_2 \\ \dots \\ V_r \end{pmatrix}. \quad (3.23)$$

Элементы  $a_{ij}$  матрицы  $\mathbf{A}$  размера  $m \times r$  называют **факторными нагрузками** (*factor loadings*).

Поскольку каждой переменной  $Y_i$  исходных данных соответствует вектор-строка ее наблюдаемых значений  $y_{ik}, k = 1, 2, \dots, n$ , то (3.23) можно записать в еще более развернутом виде

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{41} & y_{1n} \\ y_{21} & y_{22} & y_{23} & y_{42} & y_{2n} \\ y_{31} & y_{32} & y_{33} & y_{43} & y_{3n} \\ \dots & & & & \dots \\ y_{m1} & y_{m2} & y_{m3} & y_{m4} & \dots & y_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{1r} \\ a_{21} & a_{22} & a_{2r} \\ a_{31} & a_{32} & a_{3r} \\ \dots & & \dots \\ a_{m1} & a_{m2} & \dots & a_{mr} \end{pmatrix} \times \begin{pmatrix} v_{11} & v_{12} & v_{13} & v_{14} & \dots & v_{1n} \\ v_{21} & v_{22} & v_{23} & v_{24} & \dots & v_{2n} \\ \dots & & & & & \dots \\ v_{r1} & v_{r2} & v_{r3} & v_{r4} & \dots & v_{rn} \end{pmatrix}.$$

Это соотношение в матричной форме выражает значения наблюдений  $y_{ik}$ ,  $i = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, n$ , через **факторные значения** (*factor scores*)  $v_{jk}$ ,  $j = 1, 2, \dots, r$ ;  $k = 1, 2, \dots, n$ ,

$$y_{ik} = \sum_{j=1}^r a_{ij}v_{jk}, \quad i = 1, 2, \dots, m; \quad k = 1, 2, \dots, n.$$

Факторный анализ используется для выявления скрытой структуры данных. Например, оценки, полученные группой испытуемых по большому числу тестовых задач, можно приближенно попытаться представить как линейные комбинации нескольких факторов, которые можно интерпретировать как различные способности испытуемых — общие, математические, лингвистические и др. Нагрузки  $a_{ij}$  в представлении задачи  $Y_i$  будут выражать степень важности способности  $V_j$  для решения этой задачи.

Из соотношения (3.22) можно получить обратное для выражения факторов  $\mathbf{V}$  через исходные переменные  $\mathbf{Y}$ . Умножив слева обе части (3.22) на  $\mathbf{A}^T$ , а затем обе части полученного равенства слева на  $(\mathbf{A}^T \mathbf{A})^{-1}$  (заметим, что матрица  $\mathbf{A}^T \mathbf{A}$  порядка  $r$  является квадратной и поэтому может иметь обратную матрицу), получаем

$$(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} \mathbf{V},$$

откуда

$$\mathbf{V} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y},$$

или

$$\mathbf{V} = \mathbf{B} \mathbf{Y}, \quad (3.24)$$

где  $\mathbf{B} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  — матрица размера  $r \times m$ .

Соотношение (3.24) можно также записать в нематричной форме

$$V_j = \sum_{i=1}^m b_{ji} Y_i, \quad j = 1, 2, \dots, r,$$

где  $b_{ji}$  — элементы матрицы  $\mathbf{B}$ ,  $j = 1, 2, \dots, r$ ;  $i = 1, 2, \dots, m$ .

Непосредственно факторные значения выражаются через значения наблюдений следующим образом:

$$\begin{pmatrix} v_{11} & v_{12} & v_{13} & v_{14} & \dots & v_{1n} \\ v_{21} & v_{22} & v_{23} & v_{24} & \dots & v_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ v_{r1} & v_{r2} & v_{r3} & v_{r4} & \dots & v_{rn} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1m} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ b_{r1} & b_{r2} & b_{r3} & \dots & b_{rm} \end{pmatrix} \times \begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{14} & \dots & y_{1n} \\ y_{21} & y_{22} & y_{23} & y_{24} & \dots & y_{2n} \\ y_{31} & y_{32} & y_{33} & y_{34} & \dots & y_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ y_{m1} & y_{m2} & y_{m3} & y_{m4} & \dots & y_{mn} \end{pmatrix},$$

или

$$v_{jk} = \sum_{i=1}^m b_{ji} y_{ik}, \quad k = 1, 2, \dots, n; \quad j = 1, 2, \dots, r.$$

Из соотношения (3.24) можно (при  $r = m$ ) снова получить соотношение (3.22) для выражения исходных переменных  $\mathbf{Y}$  через факторы  $\mathbf{V}$ . Умножив слева обе части (3.24) на  $\mathbf{B}^T$ , а затем обе части полученного равенства слева на  $(\mathbf{B}^T \mathbf{B})^{-1}$  (матрица  $\mathbf{B}^T \mathbf{B}$  порядка  $m$  является квадратной и поэтому может иметь обратную матрицу), получаем

$$(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{B} \mathbf{V},$$

откуда

$$\mathbf{Y} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{V}, \quad (3.25)$$

или

$$\mathbf{Y} = \mathbf{A} \mathbf{V},$$

где  $\mathbf{A} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T$  — матрица размера  $m \times r$ . Таким образом, для решения задачи факторного анализа достаточно найти любую из матриц  $\mathbf{A}$  или  $\mathbf{B}$ .

Однако очевидно, что представить, тем более приближенно, матрицу наблюдаемых значений в виде (3.22), т. е. в виде произведения матрицы нагрузок на матрицу факторных значений, можно бесчисленным множеством способов, так как заранее неизвестны обе матрицы-сомножители в правой части. Поэтому,

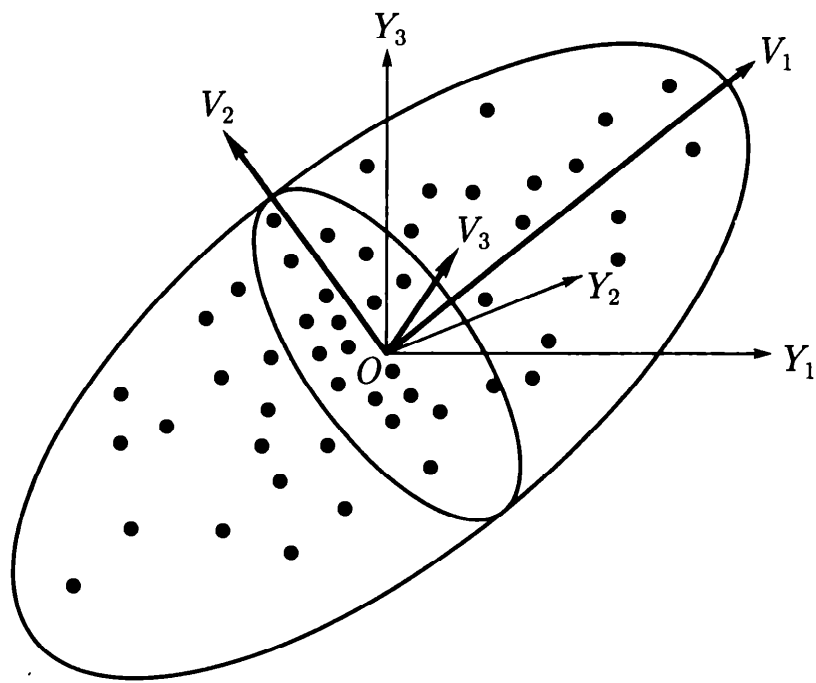


Рис. 3.8. Геометрическая иллюстрация метода главных компонент

чтобы сделать задачу факторного анализа определенной, необходимо наложить какие-то ограничения на факторные нагрузки и факторные значения. В зависимости от вида этих ограничений получаются различные методы факторного анализа. Базовым можно считать *метод главных компонент* (*principal component analysis*), в котором факторы, называемые *главными компонентами* (*principal components*), должны удовлетворять двум требованиям — быть ортогональными и содержать максимум информации об исходных данных.

Метод главных компонент имеет простую геометрическую интерпретацию. На рис. 3.8 в трехмерном пространстве ( $Y_1, Y_2, Y_3$ ) представлена совокупность точек-наблюдений, образующая конфигурацию, напоминающую удлинённый сплюснутый эллипсоид, который для наглядности также изображен на рисунке. Пусть, для простоты, начало координат  $O$  находится в центре этой совокупности.

На рис. 3.8 изображены также оси главных компонент —  $V_1, V_2$  и  $V_3$ . Первая главная компонента  $V_1$  имеет направление, соответствующее наибольшему разбросу проекций наблюдений среди всех направлений, проходящих через  $O$ . Вторая главная компонента  $V_2$  находится в плоскости, ортогональной к  $V_1$ , и имеет направление наибольшего разброса среди всех направлений этой плоскости. Третья главная компонента  $V_3$  в случае трех измерений однозначно определяется условием ортогональности к первым двум. Она должна находиться на той же плоскости, что и  $V_2$ , и соответственно разброс проекций наблюдений на  $V_3$

меньше, чем на  $V_2$  (поскольку, по построению, направление наибольшего разброса уже поставлено в соответствие второй главной компоненте). Как видим, число главных компонент в данном примере равно числу исходных переменных. Однако целью факторного анализа является сокращение числа переменных, описывающих данные. В методе главных компонент этого можно добиться простым отбрасыванием последних компонент и заменой исходных наблюдений их проекциями на оставшиеся компоненты. Например, в ситуации, представленной на рис. 3.8, можно спроектировать наблюдения на плоскость первой и второй главных компонент и отбросить третью, что не приведет к большой потере информации о взаимном расположении наблюдений.

Опишем метод главных компонент в более точных терминах. **Общий разброс** совокупности  $n$  наблюдений в  $m$ -мерном пространстве ортогональных наблюдаемых переменных  $Y_1, Y_2, \dots, Y_m$  определяется как сумма квадратов евклидовых расстояний от центра совокупности до точек-наблюдений

$$S_{\text{общ}}^2 = \sum_{k=1}^n d_{0k}^2.$$

В свою очередь квадрат евклидова расстояния между двумя точками равен сумме квадратов разностей (расстояний) между проекциями этих точек на каждую координатную ось пространства, т. е., в частности, расстояние от начала координат до  $k$ -го наблюдения равно

$$d_{0k}^2 = \sum_{i=1}^m (0 - y_{ki})^2.$$

Отсюда следует, что общий разброс можно представить как сумму разбросов проекций наблюдений на координатные оси

$$S_{\text{общ}}^2 = \sum_{i=1}^m S_{Y_i}^2, \quad (3.26)$$

где

$$S_{Y_i}^2 = \sum_{k=1}^n (0 - y_{ki})^2.$$

Из равенства (3.26) путем деления обеих его частей на  $n$  (воспользуемся, для простоты, смещенной оценкой) получаем также соотношение

$$s_{\text{общ}}^2 = \sum_{i=1}^m s_{Y_i}^2,$$

показывающее, что **общая дисперсия наблюдений**  $s_{\text{общ}}^2 = S_{\text{общ}}^2/n$  равна сумме дисперсий  $s_{Y_i}^2 = S_{Y_i}^2/n$  проекций наблюдений на координатные оси.

При переходе к любой новой системе координат, в частности к системе координат, определяемой главными компонентами  $V_1, V_2, \dots, V_m$ , общая дисперсия не изменится и по-прежнему будет равна сумме дисперсий проекций наблюдений на координатные оси (уже новые)

$$s_{\text{общ}}^2 = \sum_{j=1}^m s_{V_j}^2. \quad (3.27)$$

Теперь можно точно определить условия, налагаемые на главные компоненты и позволяющие их вычислить: первая главная компонента — это направление  $V_1$ , которому соответствует максимальное значение дисперсии  $s_{V_1}^2$  проекций наблюдений на это направление; вторая главная компонента — направление  $V_2$ , ортогональное направлению первой главной компоненты и максимизирующее дисперсию  $s_{V_2}^2$  проекций наблюдений на это направление; третья главная компонента — направление  $V_3$ , ортогональное направлению первой и второй главных компонент и максимизирующее дисперсию  $s_{V_3}^2$  проекций наблюдений на это направление, и т. д. Поскольку в соответствии с (3.27) общая дисперсия равна сумме дисперсий проекций наблюдений на главные компоненты, можно принять  $s_{\text{общ}}^2$  за 100 %, и тогда каждой главной компоненте будет соответствовать своя **доля (процент) объясненной дисперсии**.

Можно доказать, что условие максимизации дисперсий проекций наблюдений на подпространство нескольких первых компонент вызывает одновременную минимизацию суммы квадратов расстояний от наблюдений до этого подпространства. Например, если исходные данные являются наблюдениями двух переменных  $Y_1$  и  $Y_2$ , то первая главная компонента  $V_1$  будет для них не только направлением наибольшего разброса точек, но и направлением, сумма квадратов расстояний до которого от наблюдений минимальна. В этом смысле первая главная компонента похожа на линию регрессии  $Y_2$  по  $Y_1$ , которая также находится минимизацией суммы квадратов расстояний от наблюдений до линии регрессии. Разница состоит в том, что в методе глав-



ных компонент минимизируется сумма квадратов длин перпендикуляров, опущенных из точек-наблюдений на главную компоненту, а в регрессионном анализе расстояние измеряется вдоль оси  $Y_2$  (по этой причине регрессия  $Y_2$  по  $Y_1$  может не совпадать с регрессией  $Y_1$  по  $Y_2$ ).

Можно также показать, что условие максимизации дисперсий проекций наблюдений на подпространство нескольких первых компонент вызывает одновременную минимизацию суммы разностей между квадратами расстояний между всеми парами наблюдений в исходном пространстве и квадратами расстояний между проекциями этих наблюдений в подпространстве нескольких первых главных компонент. Например, если спроектировать все наблюдения на плоскость первых двух главных компонент, то можно быть уверенным, что это именно та плоскость, на которой взаимные расстояния между наблюдениями в среднем наименее искажены. Таким образом, первые две главные компоненты дают возможность наглядного графического представления данных на плоскости, сохраняя максимум информации о них.

Отметим, что если направление какой-либо оси поменять на противоположное, т. е. поменять на противоположные знаки всех проекций на эту ось, то значение дисперсии проекций наблюдений на это направление не изменится. Это означает, что любое из двух противоположных направлений главных компонент можно выбирать произвольно. Иногда это бывает полезно для более удобной их интерпретации.

Рассмотрим кратко некоторые вычислительные аспекты нахождения главных компонент.

Ранее отмечалось, что для решения задачи факторного анализа достаточно найти любую из матриц  $\mathbf{A}$  или  $\mathbf{B}$ , выражающих соответственно либо наблюдения через факторы, либо, наоборот, факторы — через наблюдения. Поскольку координатные оси главных компонент получаются вращением исходной системы координат с сохранением их ортогональности, матрица  $\mathbf{B}$  в преобразовании  $\mathbf{V} = \mathbf{B}\mathbf{Y}$  исходных переменных в главные компоненты должна быть ортогональной, т. е.  $\mathbf{B}\mathbf{B}^T = \mathbf{B}^T\mathbf{B} = \mathbf{I}$ . Из этого, в частности, следует, что в методе главных компонент упрощается переход от соотношения (3.24), выражающего факторы через исходные переменные, к соотношению (3.22), выражающему исходные переменные через факторы. Действительно, преобразуя (3.25), получаем

$$\mathbf{Y} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{V} = (\mathbf{I})^{-1}\mathbf{B}^T\mathbf{V} = \mathbf{I}\mathbf{B}^T\mathbf{V} = \mathbf{B}^T\mathbf{V},$$

т. е. матрица факторных нагрузок  $\mathbf{A}$ , фигурирующая в (3.22), совпадает с транспонированной матрицей  $\mathbf{B}$ . Элементы матриц  $\mathbf{B}$  и  $\mathbf{B}^T$  — это косинусы углов между старыми и новыми осями, и суммы их квадратов по строкам и столбцам равны единице.

Можно показать, что условие максимизации проекций наблюдений приводит к уравнению

$$|\mathbf{C} - \lambda \mathbf{I}| = 0, \quad (3.28)$$

где  $\mathbf{C}$  — выборочная матрица ковариаций переменных  $Y_1, Y_2, \dots, Y_m$  порядка  $m$ , равная деленному на  $n$  произведению матрицы наблюдений на ее транспонированную матрицу (напомним, что переменные  $Y_1, Y_2, \dots, Y_m$  центрированы):

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & & c_{2m} \\ \dots & & & \dots \\ c_{r1} & c_{r2} & \dots & c_{rm} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{14} & \dots & y_{1n} \\ y_{21} & y_{22} & y_{23} & y_{24} & & y_{2n} \\ y_{31} & y_{32} & y_{33} & y_{34} & & y_{3n} \\ \dots & & & & & \dots \\ y_{m1} & y_{m2} & y_{m3} & y_{m4} & \dots & y_{mn} \end{pmatrix} \times$$

$$\times \begin{pmatrix} y_{11} & y_{21} & y_{31} & \dots & y_{m1} \\ y_{12} & y_{22} & y_{32} & \dots & y_{m2} \\ y_{13} & y_{23} & y_{33} & & y_{m3} \\ y_{14} & y_{24} & y_{34} & & y_{m4} \\ \dots & & & & \dots \\ y_{1n} & y_{2n} & y_{3n} & \dots & y_{mn} \end{pmatrix}.$$

Уравнение (3.28), левая часть которого является полиномом степени  $m$ , в данном случае имеет  $m$  действительных корней, собственных чисел матрицы  $\mathbf{C}$ . Эти собственные значения, расположенные в порядке убывания, равны дисперсиям соответствующих главных компонент, а совокупность собственных векторов матрицы  $\mathbf{C}$  образует искомую матрицу  $\mathbf{B}$ , преобразующую исходные переменные в главные компоненты.

Часто наблюдаемые переменные  $Y_1, Y_2, \dots, Y_m$  измерены в разных единицах, и тогда их непосредственный совместный анализ является бессмысленным. В этом случае переменные нормируются путем деления на свои выборочные среднеквадратические значения (их центрирование путем вычитания выборочных

средних мы предположили с самого начала), что делает их соизмеримыми по величине. Матрица ковариаций  $S$  при этом превращается в корреляционную матрицу.

Если факторный анализ используется не просто для редукции громоздких данных или их наглядного представления — эта задача достаточно успешно решается методом главных компонент, а нацелен, действительно, на поиск неких скрытых факторов, лежащих в основе наблюдаемых данных, то используются более сложные методы. Для многих из них метод главных компонент служит начальным шагом анализа. Наиболее часто применяемым является *метод варимаксного вращения* (*varimax rotation method*), при котором в качестве отправных берутся факторы, совпадающие с первыми главными компонентами. Предположим, что на первую главную компоненту приходится 60 % дисперсии, а на вторую — 30 %. В сумме они объясняют 90 % дисперсии — это достаточно много и можно предположить, что за наблюдаемыми значениями большого числа переменных, возможно нескольких десятков, стоят всего два базовых фактора. Желательно понять их смысл, что можно сделать только анализируя содержательно представление этих факторов-компонент через исходные переменные. Это легче сделать, если нагрузки «контрастны», т. е. близки либо к нулю, либо к единице. Однако главные компоненты в этом отношении могут оказаться не лучшим выбором. Можно попробовать улучшить ситуацию, вращая первые две главные компоненты в их собственной плоскости. При этом суммарная дисперсия, приходящаяся на эти главные компоненты, не уменьшится, так что потери информации не произойдет. Может уменьшиться дисперсия первой главной компоненты, например соотношение между дисперсиями, приходящимися на новые оси (уже не главные компоненты), будет 50 и 40 %. Однако требование максимальной информативности первого фактора может быть совсем не важным или, по крайней мере, менее важным, чем желательность хорошей интерпретируемости факторов. Контрастность нагрузок можно определить разными способами. В методе варимаксного вращения она определяется как усредненная по всем факторам дисперсия квадратов нагрузок в выражениях факторов через исходные переменные. Чаще используется *нормализованный метод варимаксного вращения* (*normalized varimax rotation method*), в котором факторные нагрузки предварительно нормируются путем деления каждой из них на соответствующую ей сумму квадратов нагрузок по факторам.

### 3.7. Дискриминантный анализ

Дискриминантный анализ (*discriminant analysis*) — один из методов *распознавания образов* (*pattern recognition*). Задачей распознавания образов является построение правила, позволяющего с наименьшей ошибкой предсказывать принадлежность объекта к одному из заданных классов по его описанию. Для построения этого правила используется предварительная информация о наборе объектов, для которых известны как описание, так и принадлежность к классу. Например, имеется группа пациентов, для которых известны набор симптомов и точный диагноз. Требуется, используя эти данные, построить правило, дающее возможность в дальнейшем ставить диагноз по симптомам новым пациентам.

Рассмотрим задачу распознавания образов, пользуясь вероятностными терминами. Пусть рассматриваемые потенциальные объекты (наблюдения) характеризуются значениями  $m$  переменных  $X_1, X_2, \dots, X_m$ , являющихся непрерывными случайными величинами, распределенными по одному из законов  $F_1(x_1, x_2, \dots, x_m), F_2(x_1, x_2, \dots, x_m), \dots$  или  $F_k(x_1, x_2, \dots, x_m)$  в зависимости от того, к какому из классов  $k = 1, 2, \dots, K$  относится объект.

Пусть, например, имеется два класса, заданных двумя одномерными нормальными распределениями  $N(\mu = 3, \sigma^2 = 1)$  и  $N(\mu = 6, \sigma^2 = 1)$ , причем объекты из этих классов встречаются с равными *априорными вероятностями*  $p = q = 0,5$ . На рис. 3.9 изображены плотности этих двух распределений с рав-

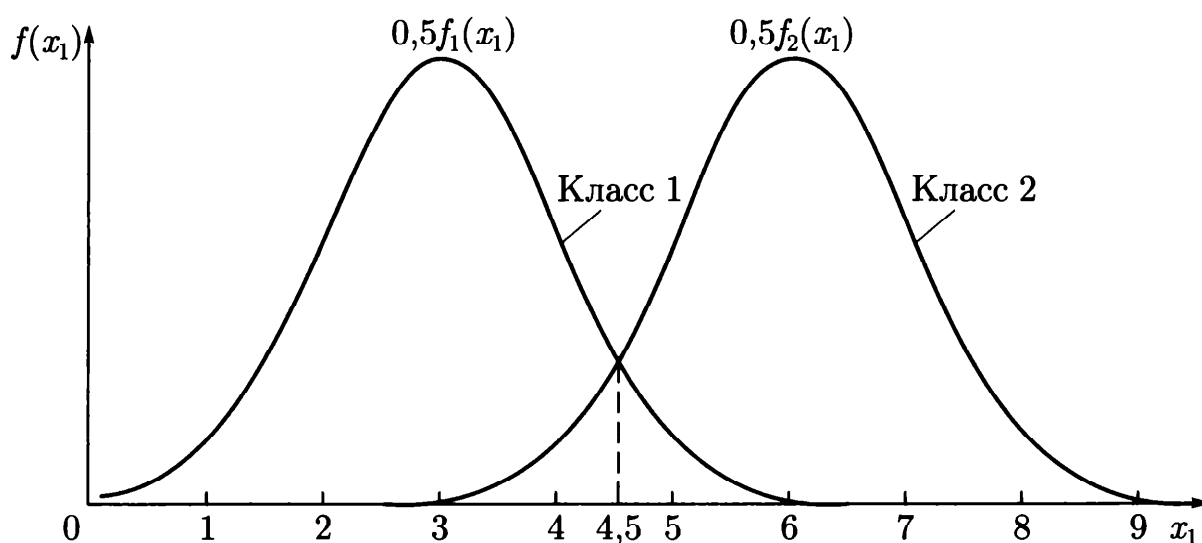


Рис. 3.9. Распознавание между двумя классами, заданными двумя одномерными нормальными распределениями с равными априорными вероятностями

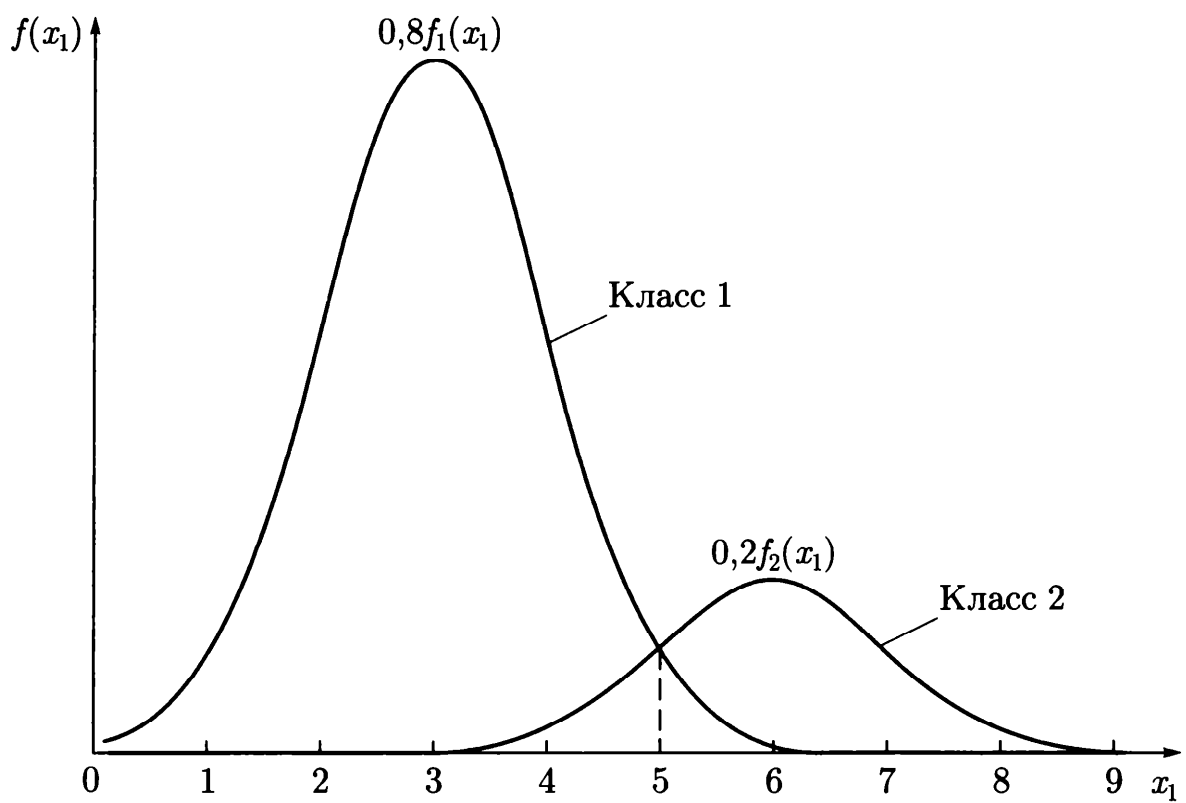


Рис. 3.10. Распознавание между двумя классами, заданными двумя одномерными нормальными распределениями с неравными априорными вероятностями

ными весами. Видим, что левее точки  $x_1 = 4,5$  плотность  $f_1(x_1)$  выше плотности  $f_2(x_1)$ , а правее точки — наоборот. Правило, состоящее в том, чтобы любое новое наблюдение, оказавшееся левее точки  $x_1 = 4,5$ , относить к первому классу, а оказавшееся правее точки — ко второму, обеспечивает минимальную ошибку классификации.

На рис. 3.10 рассмотренная ситуация несколько модифицирована — объекты класса 1 встречаются при случайном выборе в четыре раза чаще, чем объекты класса 2. Это приводит к тому, что граница разделения между классами сдвигается правее — примерно в точку  $x_1 = 5$ .

Иначе, мы пользуемся *байесовским правилом* — относим наблюдение к классу, имеющему наибольшую *апостериорную вероятность*.

В двух рассмотренных примерах предполагалось, что известны теоретические распределения, определяющие классы. Если имеются только выборки наблюдений из них, то по ним можно построить оценки распределений. В частности, если есть основания считать, что неизвестные распределения нормальны, то достаточно вычислить средние и дисперсии двух выборок. Априорные вероятности можно взять равными долям встречаемости наблюдений двух классов в выборке.

На рис. 3.11 показана ситуация, когда классы заданы двумя двумерными нормальными распределениями с равными дисперсиями и корреляциями и одинаковыми априорными вероятностями, различаясь математическими ожиданиями. Каждая плотность схематично представлена одной из своих линий уровня (взято одно и то же значение уровня для обеих плотностей). Легко показать, что в этом случае линия разделения между двумя классами, т. е. линия, вдоль которой значения обеих плотностей равны, является прямой (заметим, что в данном примере на одномерных проекциях на оси  $x_1$  и  $x_2$  распределения разделялись бы гораздо хуже, чем в двумерном пространстве). Действительно, условие равенства плотностей имеет вид

$$pk_1 \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1^{(1)})^2}{\sigma_1^2} + \frac{(x_2 - \mu_2^{(1)})^2}{\sigma_2^2} - 2\rho \frac{(x_1 - \mu_1^{(1)})(x_2 - \mu_2^{(1)})}{\sigma_1\sigma_2} \right] \right\} / qk_1 \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1^{(2)})^2}{\sigma_1^2} + \frac{(x_2 - \mu_2^{(2)})^2}{\sigma_2^2} - 2\rho \frac{(x_1 - \mu_1^{(2)})(x_2 - \mu_2^{(2)})}{\sigma_1\sigma_2} \right] \right\} = 1,$$

где  $(\mu_1^{(1)}, \mu_2^{(1)})$  и  $(\mu_1^{(2)}, \mu_2^{(2)})$  — координаты центров первого и второго распределений;  $\sigma_1$  и  $\sigma_2$  — их дисперсии по первой и второй оси;  $\rho$  — коэффициент корреляции, а коэффициент  $k_1$  определяется выражением

$$k_1 = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

Сокращая на  $k_1$  и логарифмируя, получаем соотношение

$$\left\{ \log p - \frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1^{(1)})^2}{\sigma_1^2} + \frac{(x_2 - \mu_2^{(1)})^2}{\sigma_2^2} - 2\rho \frac{(x_1 - \mu_1^{(1)})(x_2 - \mu_2^{(1)})}{\sigma_1\sigma_2} \right] \right\} - \left\{ \log q - \frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1^{(2)})^2}{\sigma_1^2} + \frac{(x_2 - \mu_2^{(2)})^2}{\sigma_2^2} - 2\rho \frac{(x_1 - \mu_1^{(2)})(x_2 - \mu_2^{(2)})}{\sigma_1\sigma_2} \right] \right\} = 0, \quad (3.29)$$

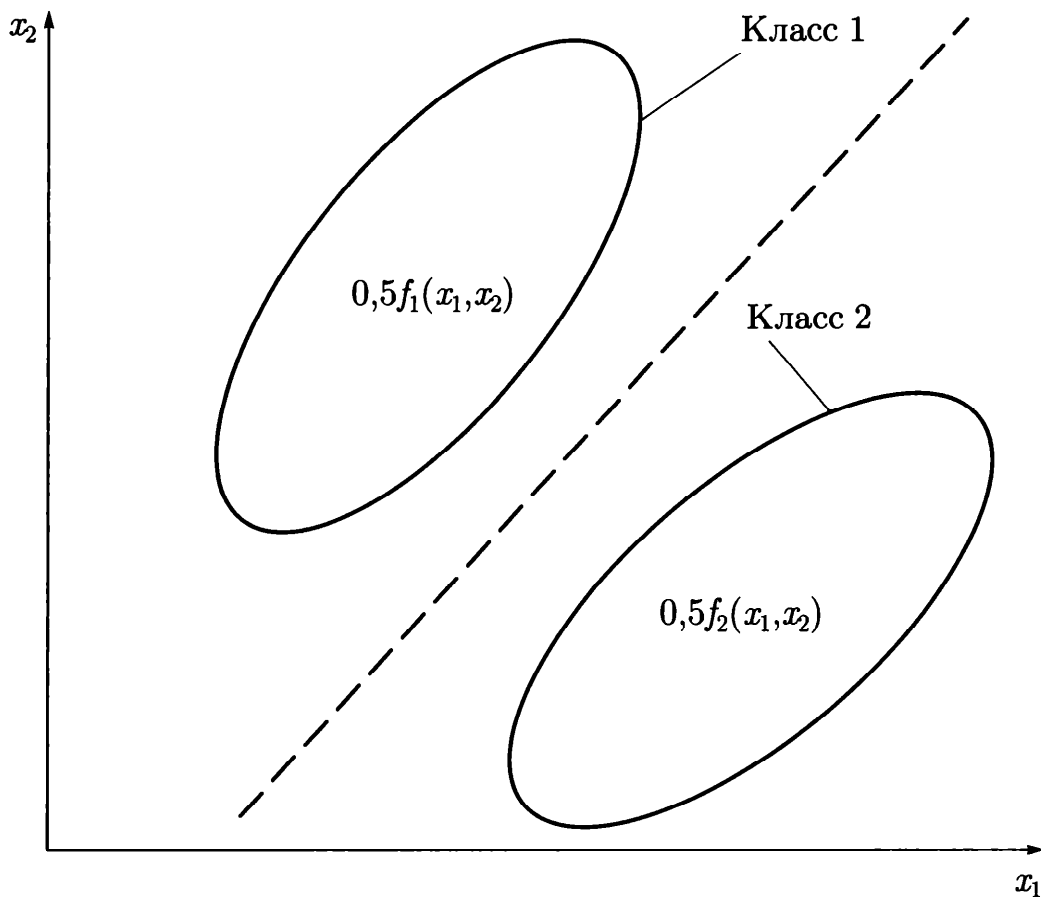


Рис. 3.11. Распознавание между двумя классами, заданными двумя двумерными нормальными распределениями с равными априорными вероятностями

из которого видно, что после возведений в квадрат и перемножений выражений в скобках члены второго порядка  $x_1^2$ ,  $x_2^2$  и  $x_1x_2$  сократятся и в левой части уравнения останется выражение первого порядка относительно  $x_1$  и  $x_2$ , т. е. это уравнение действительно определяет прямую на плоскости  $(x_1, x_2)$ . Заметим, что этот вывод существенно использует предположение о равенстве соответствующих дисперсий и коэффициентов корреляций двух распределений, что эквивалентно равенству их ковариационных матриц  $C_1$  и  $C_2$ :

$$C_1 = C_2 = C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

На рис. 3.12 показан случай, когда надо разделить три класса. Здесь также предполагается, что соответствующие дисперсии и корреляции трех распределений равны, поэтому каждая пара разделяется прямой, равной плотности. Совокупность этих прямых формирует, как видно из рисунка, результирующую кусочно-линейную границу разделения классов на плоскости.

В предположении равенства ковариационных матриц и многомерной нормальности всех распределений рассмотренная ситуация обобщается и на большее число измерений. В трехмерном пространстве классы будут разделяться плоскостями, в четырехмерном — трехмерными гиперплоскостями и т. д. Именно предположение о нормальности и равенстве ковариационных матриц распределений, соответствующих разным классам, определяет условия применимости дискриминантного анализа, точнее *линейного дискриминантного анализа*. Эти условия довольно жестки, однако, благодаря своей статистической прозрачности и легкости численной реализации, метод линейного дискриминантного анализа реализован во всех пакетах программ статистической обработки данных и широко применяется при анализе данных даже в ситуациях, когда его предпосылки выполняются лишь приближенно.

Рассмотрим еще некоторые понятия, связанные с линейным дискриминантным анализом.

Левую часть равенства (3.29), которую после преобразований можно записать в виде

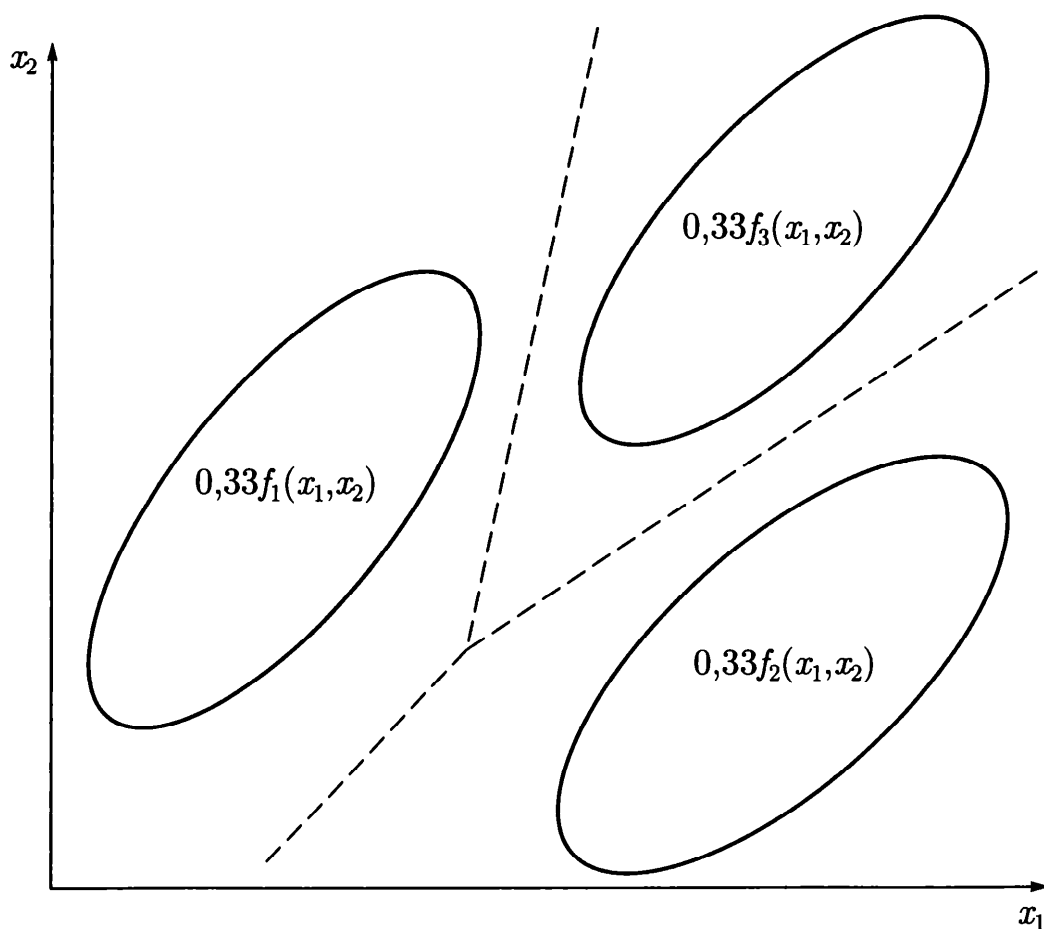


Рис. 3.12. Распознавание между тремя классами, заданными тремя двумерными нормальными распределениями с равными априорными вероятностями



$$L_{12}(x_1, x_2) = a_{12,0} + a_{12,1}x_1 + a_{12,2}x_2,$$

называют *линейной дискриминантной функцией* для классов 1 и 2. Если для некоторого нового наблюдения эта функция больше нуля, то наблюдение следует отнести к классу 1, а если меньше — к классу 2.

Очевидно, для случая  $K$  классов имеется  $K(K - 1)/2$  дискриминантных функций и все их надо сравнить с нулем. В случае большого числа классов процедура становится слишком громоздкой, поэтому поступают иначе. С каждым классом  $k$  связывают функцию  $L_k(x_1, x_2)$ , называемую *линейным дискриминантом* этого класса, и для нового наблюдения вычисляют значения дискриминантов для всех классов. Полученные  $K$  значения сравниваются, и наблюдение относится к тому классу, которому соответствует максимальное значение дискриминанта. Если вернуться к уравнению (3.29), то легко видеть, что его левую часть можно представить как разность двух линейных дискриминантов, соответствующих первому и второму членам в левой части (после раскрытия скобок и взаимного уничтожения квадратичных членов).

В пространстве наблюдаемых переменных для любого наблюдения можно вычислить его евклидовы расстояния от центра каждого класса. Однако эти расстояния могут не отражать истинные близости наблюдения к разным классам. Например, на рис. 3.11 точка на пересечении трех разделяющих прямых находится на разном расстоянии от центров трех классов, однако с точки зрения ее отнесения к одному из классов она равноудалена от них.

Адекватной мерой расстояния в ситуации дискриминантного анализа (при равных априорных расстояниях) является *расстояние Махаланобиса*. Расстояние Махаланобиса между двумя точками в пространстве наблюдаемых переменных определяется как евклидово расстояние между этими точками в новом пространстве, получаемом из исходного путем преобразования, при котором первоначальные произвольные нормальные распределения становятся нормальными распределениями с единичными дисперсиями и нулевыми коэффициентами корреляции.

Например, для случая, приведенного на рис. 3.10, этого можно достичь путем поворота координатных осей в направлении осей эллипсов, представляющих линии уровня, и последующего соответствующего изменения масштаба новых осей.

Если предпосылки линейного дискриминантного анализа не выполняются даже приближенно, то применяют нелинейные методы распознавания образов. Одним из таких методов является *метод  $k$  ближайших соседей*. Он состоит в том, что находятся  $k$  ближайших соседей нового наблюдения, и оно относится к тому классу, представителей которого больше всего среди этих ближайших соседей. Выбор числа  $k$  не формализован — алгоритм работает даже при  $k = 1$ , но очевидно, что при большом числе наблюдений его можно брать большим, поскольку это должно повысить надежность правильного распознавания. Может оказаться полезным перед вычислением расстояний от нового наблюдения до всех остальных (для выявления ближайших) произвести преобразование пространства с тем, чтобы новые расстояния лучше соответствовали реальной близости наблюдений к своим классам. Во всяком случае обязательно стоит пронормировать наблюдаемые переменные путем деления их на свои выборочные среднеквадратические отклонения, если эти переменные измерены в разных единицах.

# ЧАСТЬ II

## МОДЕЛИРОВАНИЕ ДИНАМИКИ БИОЛОГИЧЕСКИХ ПРОЦЕССОВ

---

### ГЛАВА 4

#### МОДЕЛИ, ОПИСЫВАЕМЫЕ ОДНИМ УРАВНЕНИЕМ

В этой главе будут рассмотрены дифференциальные математические модели, описываемые только одним *дифференциальным уравнением*.

#### 4.1. Основные понятия

Ограничимся изучением свойств дифференциального уравнения только *первого порядка*

$$\frac{dx}{dt} = f(x, t),$$

поскольку дифференциальные уравнения более высокого порядка могут быть путем замены переменных сведены к системам двух или более дифференциальных уравнений первого порядка, которые будут рассматриваться в гл. 5. Например, уравнение второго порядка

$$\frac{d^2x}{dt^2} = f(x, t)$$

путем замены

$$y = \frac{dx}{dt}$$

сводится к следующей системе двух уравнений первого порядка:

$$\begin{cases} \frac{dx}{dt} = y; \\ \frac{dy}{dt} = f(x, y). \end{cases}$$

Дифференциальное уравнение называется *автономным*, если его правая часть не зависит явно от времени, т. е. уравнение имеет вид

Решения уравнения (4.1) можно представить графически на плоскости  $(t, x)$  в виде *интегральных кривых*  $x = x(t)$ . Если на плоскости  $(t, x)$  задана точка с координатами  $(t_0, x_0)$  и для уравнения (4.1) выполнены условия *теоремы Коши*, то имеется единственная интегральная кривая  $x(t)$  уравнения (4.1), проходящая через точку  $(t_0, x_0)$ , т. е. интегральные кривые уравнения (4.1) не могут пересекаться. Они монотонны и либо уходят в бесконечность, либо асимптотически приближаются к *стационарному состоянию* (называемому также *состоянием равновесия, точкой покоя, особой точкой*), т. е. состоянию, в котором скорость изменения фазовой переменной равна нулю:

$$\frac{dx}{dt} = 0.$$

Стационарные состояния  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$  являются корнями алгебраического уравнения

$$f(x) = 0.$$

В дальнейшем мы будем пользоваться понятиями *фазового пространства* и *фазовых переменных*. Фазовым называется пространство, на координатных осях которого отложены значения переменных системы, которые в данном случае называются фазовыми переменными. В соответствии с дифференциальными уравнениями, описывающими систему, задание всех координат системы для одного момента времени (точки в фазовом пространстве) определяет состояние системы для всех других моментов времени.

Таким образом, изменение состояния системы во времени можно представить как движение системы вдоль некоторой линии в фазовом пространстве. Эта линия называется *фазовой траекторией*. Метод фазового пространства широко используется в статистической физике, теории колебаний, качественной теории дифференциальных уравнений. Нам он будет полезен при качественном рассмотрении базовых моделей математической биологии.

Для одного уравнения фазовое пространство представляет собой прямую.

На фазовой прямой стационарное состояние  $\bar{x}_i$  — это точка, к которой стремится величина  $x$ , а на плоскости  $(t, x)$  горизонтальная прямая  $x = \bar{x}_i$  — это асимптота, к которой приближаются интегральные кривые (рис. 4.1).

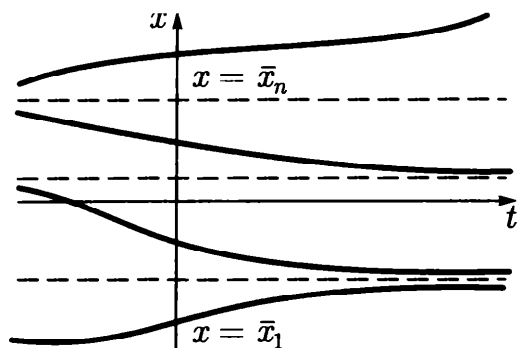


Рис. 4.1. Примеры интегральных кривых автономного дифференциального уравнения

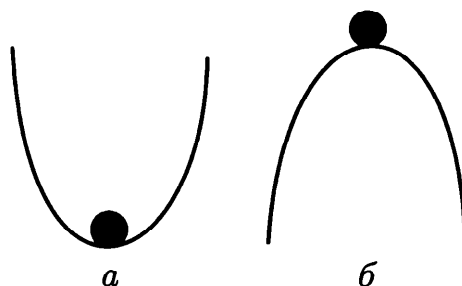


Рис. 4.2. Иллюстрация устойчивого и неустойчивого равновесия:

*a* — стационарное состояние устойчиво; *b* — стационарное состояние неустойчиво

Реальные биологические системы испытывают многочисленные флуктуации и неизбежно отклоняются от стационарного состояния. Если после небольшого отклонения состояние системы не удаляется от стационарного, то оно называется **устойчивым**, в противном случае — **неустойчивым**. Поэтому важно знать, устойчивы ли стационарные состояния модели.

Строгое математическое определение устойчивости стационарного состояния дано А. М. Ляпуновым: состояние равновесия **устойчиво по Ляпунову**, если для любого сколь угодно малого положительного числа  $\epsilon$  всегда можно найти такое  $\delta$ , что если в начальный момент времени  $t_0$  отклонение от состояния равновесия меньше  $\delta$ , т. е.  $|x(t_0) - \bar{x}| < \delta$ , то и в любой последующий момент времени  $t > t_0$  отклонение решения системы от состояния равновесия будет также мало, т. е. будет выполняться неравенство  $|x(t) - \bar{x}| < \epsilon$ . Иначе, малое начальное отклонение от равновесия не может вызвать больших последующих отклонений.

Устойчивое равновесие называется **асимптотически устойчивым**, если можно найти достаточно малое  $\delta$ , такое, что если в начальный момент времени  $t_0$  отклонение от состояния равновесия меньше  $\delta$ , т. е.  $|x(t_0) - \bar{x}| < \delta$ , то  $\lim_{t \rightarrow \infty} |x(t) - \bar{x}| = 0$ , т. е. при не слишком большом начальном отклонении значение фазовой переменной возвращается к стационарному.

Ситуации устойчивого и неустойчивого равновесия на примере шарика, находящегося внизу вогнутой и сверху выпуклой поверхности, проиллюстрированы на рис. 4.2. Легко проверить, что в первом случае условие устойчивости по Ляпунову и условие асимптотической устойчивости выполняются, а во втором — нет.

Асимптотически устойчивое стационарное состояние представляет собой простейший тип *аттрактора*, т. е. множества в фазовом пространстве, к которому с течением времени стремится состояние динамической системы.

## 4.2. Исследование уравнения в окрестности стационарного состояния

### 4.2.1. Исследование устойчивости стационарного состояния методом линеаризации

Для исследования поведения решений автономного дифференциального уравнения первого порядка в окрестности стационарного состояния применим метод линеаризации Ляпунова. Пусть  $\bar{x}$  — стационарное решение уравнения (4.1). Пусть система, первоначально находившаяся в стационарном состоянии, отклонилась от него и перешла в близкую точку  $x = \bar{x} + \xi$ . Перейдем в уравнении (4.1) от переменной  $x$  к переменной  $\xi$ , являющейся отклонением системы от стационарного состояния:

$$\frac{d(\bar{x} + \xi)}{dt} = \frac{dx}{dt} = f(\bar{x} + \xi).$$

Учитывая, что  $f(\bar{x}) = 0$  (по определению стационарного состояния), и разложив  $f(\bar{x} + \xi)$  в ряд Тейлора в точке  $\bar{x}$ , получаем

$$\frac{d\xi}{dt} = f(\bar{x}) + f'(\bar{x})\xi + \frac{1}{2}f''(\bar{x})\xi^2 + \dots$$

или

$$\frac{d\xi}{dt} = a_1\xi + a_2\xi^2 + \dots$$

где  $a_1 = f'(\bar{x})$ ,  $a_2 = f''(\bar{x})$ , ... Отбросив члены порядка выше первого, получим линейное уравнение

$$\frac{d\xi}{dt} = a_1\xi,$$

называемое *линеаризованным уравнением*, или уравнением первого приближения. Его решение имеет вид

$$\xi(t) = c \exp(\lambda t),$$

где  $\lambda = a_1 = f'(\bar{x})$ ;  $c$  — произвольная постоянная.

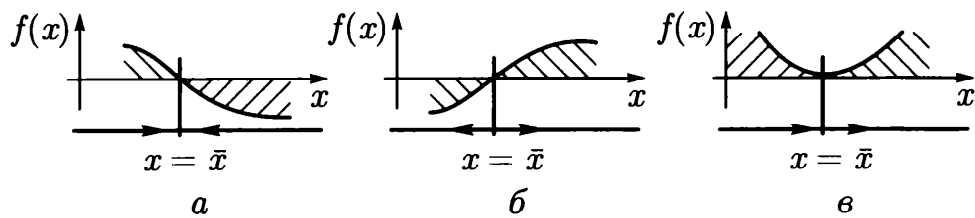


Рис. 4.3. Определение устойчивости стационарного состояния по графику функции  $f(x)$ :

$a$  — стационарное состояние  $\bar{x}$  устойчиво;  $b, v$  — стационарное состояние  $\bar{x}$  неустойчиво

Если  $\lambda < 0$ , то  $\xi \rightarrow 0$  при  $t \rightarrow \infty$  и, следовательно, первоначальное отклонение от состояния равновесия со временем затухает. Это означает, по определению, что состояние равновесия устойчиво. Если  $\lambda > 0$ , то  $\xi \rightarrow \infty$  при  $t \rightarrow \infty$ , т. е. исходное состояние равновесия неустойчиво. Если  $\lambda = 0$ , то уравнение первого приближения не может дать ответа на вопрос об устойчивости состояния равновесия системы. Необходимо рассматривать члены более высокого порядка в разложении в ряд Тейлора.

Таким образом, устойчивость стационарного состояния автономного дифференциального уравнения (4.1) определяется знаком производной правой части в стационарной точке.

Вопрос об устойчивости состояния равновесия одного уравнения нетрудно решить, рассматривая график функции  $f(x)$ . По определению, в стационарной точке правая часть уравнения (4.1) — функция  $f(x)$  — обращается в нуль. При этом возможны три случая (рис. 4.3,  $a, b, v$ ).

1. Вблизи состояния равновесия функция  $f(x)$  меняет знак с плюса на минус при возрастании  $x$  (см. рис. 4.3,  $a$ ).

Отклоним изображающую точку системы в сторону  $x < \bar{x}$ . В этой области скорость изменения  $x$ :  $dx/dt = f(x)$  положительна. Следовательно,  $x$  увеличивается, т. е. возвращается к  $\bar{x}$ . При  $x > \bar{x}$  скорость изменения величины  $x$  отрицательна, так как функция  $f(x) < 0$ . Следовательно, здесь  $x$  уменьшается и опять стремится к  $\bar{x}$ . Таким образом, отклонения от стационарного состояния в обе стороны затухают. Стационарное состояние устойчиво.

2. Вблизи состояния равновесия функция  $f(x)$  меняет знак с минуса на плюс при возрастании  $x$  (см. рис. 4.3,  $b$ ).

Проведем рассуждения, аналогичные случаю 1, помещая изображающую точку сначала в область  $x < \bar{x}$ , затем — в область  $x > \bar{x}$ . В обоих случаях изображающая точка удаляется от состояния равновесия. Стационарное состояние неустойчиво.

3. Вблизи состояния равновесия функция  $f(x)$  не меняет знак (см. рис. 4.3, в)

Поскольку  $f(\bar{x}) = 0$ , это означает, что изображающая точка, помещенная достаточно близко к состоянию равновесия с одной стороны, будет приближаться к нему, помещенная с другой стороны — удаляться. В соответствии с определением устойчивости по Ляпунову состояние равновесия является неустойчивым.

#### 4.2.2. Модель роста колонии микроорганизмов

За время  $\Delta t$  прирост численности  $\Delta x$  равен

$$\Delta x = R - S,$$

где  $R$  и  $S$  — численности родившихся и умерших за время  $\Delta t$  особей, которые пропорциональны этому промежутку времени:

$$R(\Delta t, x) = R(x)\Delta t;$$

$$S(\Delta t, x) = S(x)\Delta t,$$

т. е.

$$\Delta x = [R(x) - S(x)]\Delta t.$$

Разделив на  $\Delta t$  и переходя к пределу при  $t \rightarrow 0$ , получим дифференциальное уравнение

$$\frac{dx}{dt} = R(x) - S(x).$$

В простейшем случае, когда рождаемость и смертность пропорциональны численности, уравнение принимает вид:

$$\frac{dx}{dt} = \alpha x - \beta x, \quad \alpha - \beta = r,$$

или

$$\frac{dx}{dt} = rx.$$

Проинтегрируем полученное уравнение методом разделения переменных:

$$\frac{dx}{rx} = dt,$$

$$\ln x = rt + C.$$



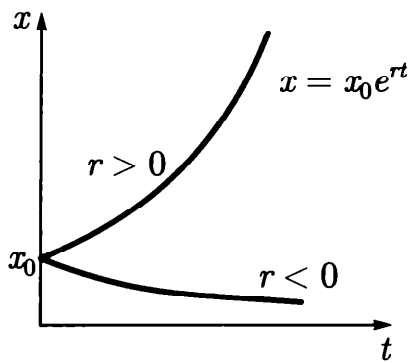


Рис. 4.4. Экспоненциальная форма динамики роста численности колонии микроорганизмов

Переходя от логарифмов к значениям переменной  $x$  и определяя произвольную постоянную  $C$  из начальных условий, получим экспоненциальную форму динамики роста:

$$x = x_0 e^{rt},$$

$$x_0 = x(0).$$

Здесь  $x(0)$  соответствует численности популяции в начальный момент времени.

График полученной экспоненциальной функции при положительных (размножение) и отрицательных (вымирание) значениях константы скорости роста  $r$  представлен на рис. 4.4. Роль этой модели в развитии математической биологии и экологии будет рассмотрена далее.

### 4.2.3. Модель перехода вещества в раствор

Пусть количество вещества, переходящего в раствор, пропорционально интервалу времени и разности между максимально возможной концентрацией  $P$  и концентрацией  $x$  в данный момент времени:  $\Delta x = (P - x)\Delta t$ .

В форме дифференциального уравнения этот закон выглядит следующим образом:

$$\frac{dx}{dt} = k(P - x).$$

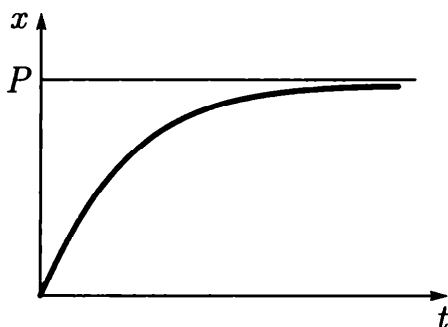


Рис. 4.5. Концентрация переходящего в раствор вещества в зависимости от времени

Разделяя переменные и интегрируя, получаем

$$\begin{aligned}\frac{dx}{k(P-x)} &= dt, \\ -\ln(P-x) &= kt + C, \\ x &= P - C_1 e^{-kt},\end{aligned}$$

где  $C_1$  — произвольная постоянная. Если  $x(0) = 0$ , т. е. в начальный момент времени концентрация вещества в растворе равна нулю, то  $C_1 = P$ ,  $x = P(1 - e^{-kt})$ .

График этой функции приведен на рис. 4.5, он представляет собой кривую с насыщением.

В дальнейшем мы, как правило, не будем искать аналитическое решение для наших моделей. Для более сложных нелинейных уравнений это и невозможно. Важные заключения относительно свойств моделей можно сделать и на основании качественного их исследования, в первую очередь путем исследования устойчивости стационарных состояний и типов поведения системы вблизи этих состояний. Следует иметь в виду, что с помощью одного автономного дифференциального уравнения могут быть описаны только монотонные изменения переменной и, следовательно, ни периодические, ни хаотические процессы не могут быть описаны. Однако, вводя в правую часть дифференциального уравнения первого порядка запаздывание, можно описать колебания, а делая его дискретным, получить динамический хаос.

### 4.3. Непрерывные модели популяций

Численность популяции может меняться во времени различным образом: расти, совершать колебания, падать, и причины этого могут быть различны. Здесь рассмотрим модели роста популяций и математический аппарат, позволяющий описывать динамику численности разных популяций.

#### 4.3.1. Уравнение экспоненциального роста

Широко известной математической моделью, в основу которой положена задача о динамике численности популяции, является классическая модель неограниченного роста. Это — геометрическая прогрессия в дискретном представлении

$$A_{n+1} = qA_n$$

или экспонента — в непрерывном представлении:

$$\frac{dx}{dt} = rx. \quad (4.2)$$

Модель была предложена известным английским демографом и экономистом Т. Р. Мальтусом (1798) в его классическом труде «О законе роста народонаселения», где он обратил внимание на тот факт, что численность популяции растет по экспоненте (в геометрической прогрессии), в то время как производство питания увеличивается со временем линейно (в арифметической прогрессии), из чего сделал справедливый вывод, что рано или поздно экспонента обязательно «обгонит» линейную функцию, и наступит голод. Можно говорить о том, что Мальтус был первым ученым-«алармистом», который на основании результатов моделирования «бил тревогу» и предупреждал человечество об опасности следования развитию по используемым ранее сценариям прогресса.

Обсуждению важности вывода Мальтуса для популяционной динамики великий Ч. Дарвин посвятил несколько страниц своего дневника, указывая, что поскольку ни одна популяция не размножается до бесконечности, должны существовать факторы, препятствующие такому неограниченному размножению. Среди этих факторов может быть нехватка ресурса (продовольствия), вызывающая конкуренцию внутри популяции за ресурс, хищничество, конкуренция с другими видами. Результатом является замедление скорости роста популяции и выход ее численности на стационарный уровень. Модели ограниченного роста будут рассмотрены далее (см. 4.3.2). Что касается отбора, то в нем большое значение наряду с конкуренцией близких по своим потреб-

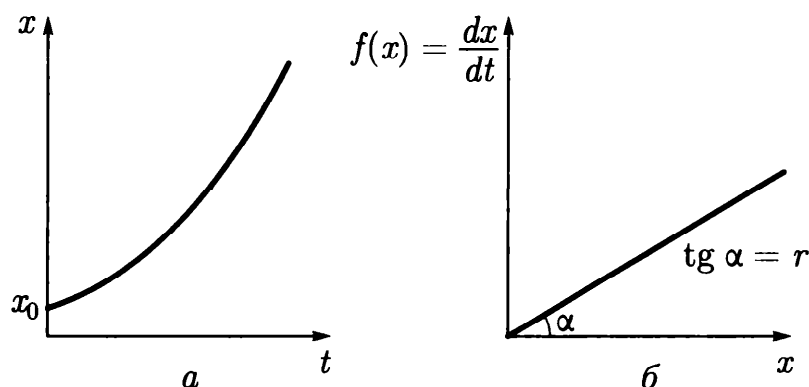


Рис. 4.6. Экспоненциальный рост:

*a* — зависимость численности от времени; *б* — зависимость скорости роста от численности

ностям особей играет территориальная изоляция, которая ведет к вымиранию близких форм и обеспечивает процесс дивергенции. Модели отбора также будут рассмотрены далее (см. 5.5).

График зависимости численности от времени в соответствии с законом экспоненциального роста изображен на рис. 4.6, а, а на рис. 4.6, б представлена зависимость скорости роста популяции (правая часть уравнения (4.2)) от ее численности.

В соответствии с экспоненциальным законом изолированная популяция развивалась бы в условиях неограниченных ресурсов. В природе такие условия встречаются крайне редко. Примером может служить размножение видов, завезенных в места, где имеется много пищи и отсутствуют конкурирующие виды и хищники (кролики в Австралии).

### 4.3.2. Ограниченный рост

Впервые системный фактор, ограничивающий рост популяции, описал Ферхюльст в уравнении логистического роста

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right), \quad (4.3)$$

которое можно также переписать в виде

$$\frac{dx}{dt} = rx - \delta x^2. \quad (4.4)$$

где  $\delta$  — коэффициент внутривидовой конкуренции (за пищевой ресурс, убежища и др.);  $K$  — величина, называемая в биологии емкостью экологической ниши популяции.

*Логистическое уравнение* обладает двумя важными свойствами. При малых значениях  $x$  численность возрастает экспоненциально (как в уравнении (4.2)), а при больших — приближается к некоторому пределу  $K$ . Эта величина определяется ограниченностью пищевых ресурсов, мест для гнездования и многими другими факторами, которые могут быть различными для разных видов. Таким образом, емкость экологической ниши представляет собой системный фактор, который определяет ограниченность роста популяции в данном ареале обитания.

Уравнение (4.3) можно решить аналитически. Ход решения следующий. Произведем деление переменных:

$$\frac{K dt}{x(K - x)} = r dt.$$

Представим левую часть в виде суммы и проинтегрируем:

$$\left(\frac{1}{x} + \frac{1}{K-x}\right) dx = r dt,$$

$$\ln x - \ln(K-x) = rt + \ln C.$$

Переходя от логарифмов к исходным переменным, получим

$$\frac{x}{K-x} = Ce^{rt}, \quad (4.5)$$

где  $C$  — произвольная постоянная, которая определяется начальным значением численности  $x_0$ :

$$x(0) = x_0,$$

$$C = \frac{x_0}{K-x_0}.$$

Подставляя полученное значение  $C$  в формулу (4.5):

$$\frac{x}{K-x} = \frac{x_0}{K-x_0} e^{rt},$$

получаем решение — зависимость численности от времени:

$$x(t) = \frac{x_0 K e^{rt}}{K - x_0 + x_0 e^{rt}}.$$

График этой функции при разных начальных значениях численности популяции представлен на рис. 4.7, а.

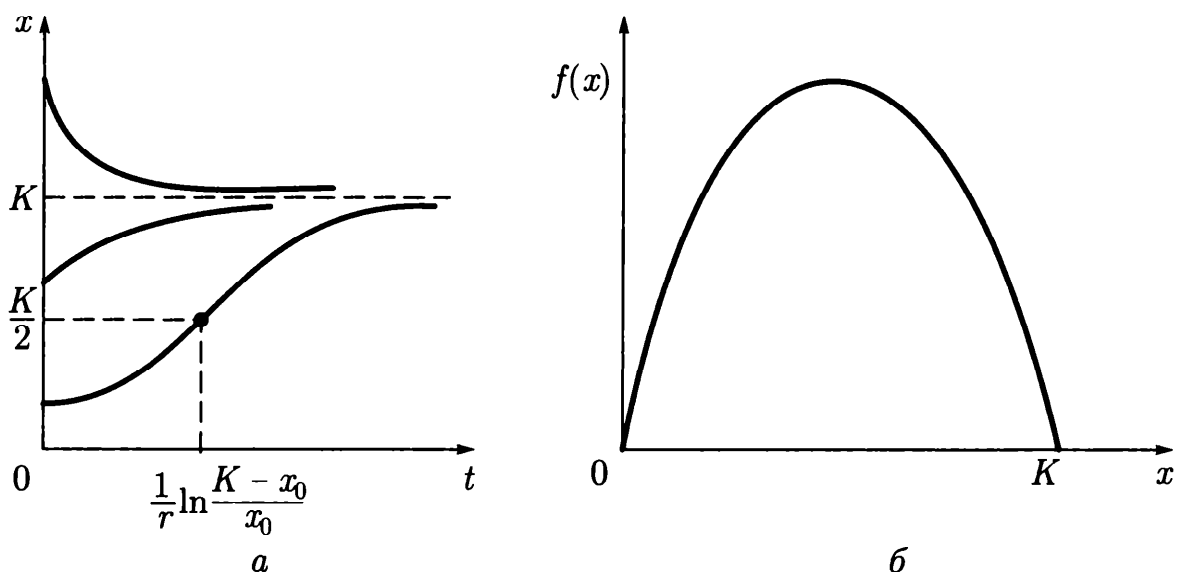


Рис. 4.7. Логистическая модель (4.4):

$a$  — зависимость численности от времени при разных начальных значениях численности;  $b$  — зависимость скорости роста от численности

В случае если начальная численность меньше половины максимальной, кривая  $x(t)$  имеет точку перегиба с координатами

$$\left( \frac{1}{r} \ln \frac{K - x_0}{x_0}; \frac{K}{2} \right).$$

Ордината точки перегиба представляет собой половину максимальной численности, а абсцисса зависит как от емкости популяции  $K$ , так и от константы собственной скорости роста  $r$  — чем выше генетические возможности популяции, тем скорее наступает перегиб на кривой численности (рис. 4.7, б).

При изучении более сложных систем мы не будем искать решение для  $x(t)$  в явном виде, а ограничимся исследованием устойчивости их стационарных состояний. Проведем такое исследование для логистического уравнения. Уравнение стационарных состояний  $f(\bar{x}) = 0$  в данном случае имеет два корня

$$\bar{x}_1 = 0, \quad \bar{x}_2 = K.$$

Проверим, будут ли эти корни устойчивыми. Для этого воспользуемся методом линеаризации Ляпунова. Введем новую переменную  $\xi$ , обозначающую отклонение переменной  $x$  от ее стационарного значения:

$$\xi = x - \bar{x}$$

и запишем линеаризованное уравнение для (4.4.)

$$d\xi/dt = a\xi, \quad \text{где } a = f'(x) \Big|_{x=\bar{x}_i}, \quad i = 1, 2.$$

Напомним, что знак величины  $a(\bar{x}_1)$  определяет устойчивость соответствующей особой точки  $\bar{x}_i$

$$a(\bar{x}_i) = f'(x) \Big|_{x=\bar{x}_i} = r - \frac{2r\bar{x}_i}{K}, \quad i = 1, 2.$$

Подставив в это выражение значение первого корня  $\bar{x}_i = 0$ , получим  $a(\bar{x}_1) = r$ . Если считать, что  $r$  (коэффициент естественной скорости роста популяции) — величина положительная, то  $\bar{x}_1 = 0$  — неустойчивая особая точка. Если подставить в выражение  $a(\bar{x}_i)$  значение  $\bar{x}_2 = K$ , то получим  $a(\bar{x}_2) = -r$ , т. е. отрицательную величину. Это дает возможность утверждать, что стационарное решение уравнения  $\bar{x}_2 = K$  соответствует устойчивому стационарному режиму существования популяции в ограниченной среде.

Проведем теперь исследование устойчивости стационарных решений этого уравнения, исходя из графика функции правой части. На рис. 4.7, б видно, что при переходе от отрицательных к положительным значениям  $x$  в точке  $\bar{x}_1 = 0$  функция  $f(x)$  меняет знак с минуса на плюс, т. е. особая точка неустойчива. Наоборот, в точке  $\bar{x}_2 = K$  имеет место изменение знака  $f(x)$  с роста  $x$  с плюса на минус, следовательно, эта особая точка устойчива.

Несмотря на схематичность положенных в ее основу представлений, логистическая кривая оказалась очень хорошим приближением для описания кривых роста численности многих популяций. В природе внутривидовая конкуренция не удерживает численность естественных популяций на строго неизменном уровне, но, действуя в широком диапазоне начальных значений плотности, приводит к гораздо более узкому диапазону конечных (стационарных) значений численности. Таким образом, про-

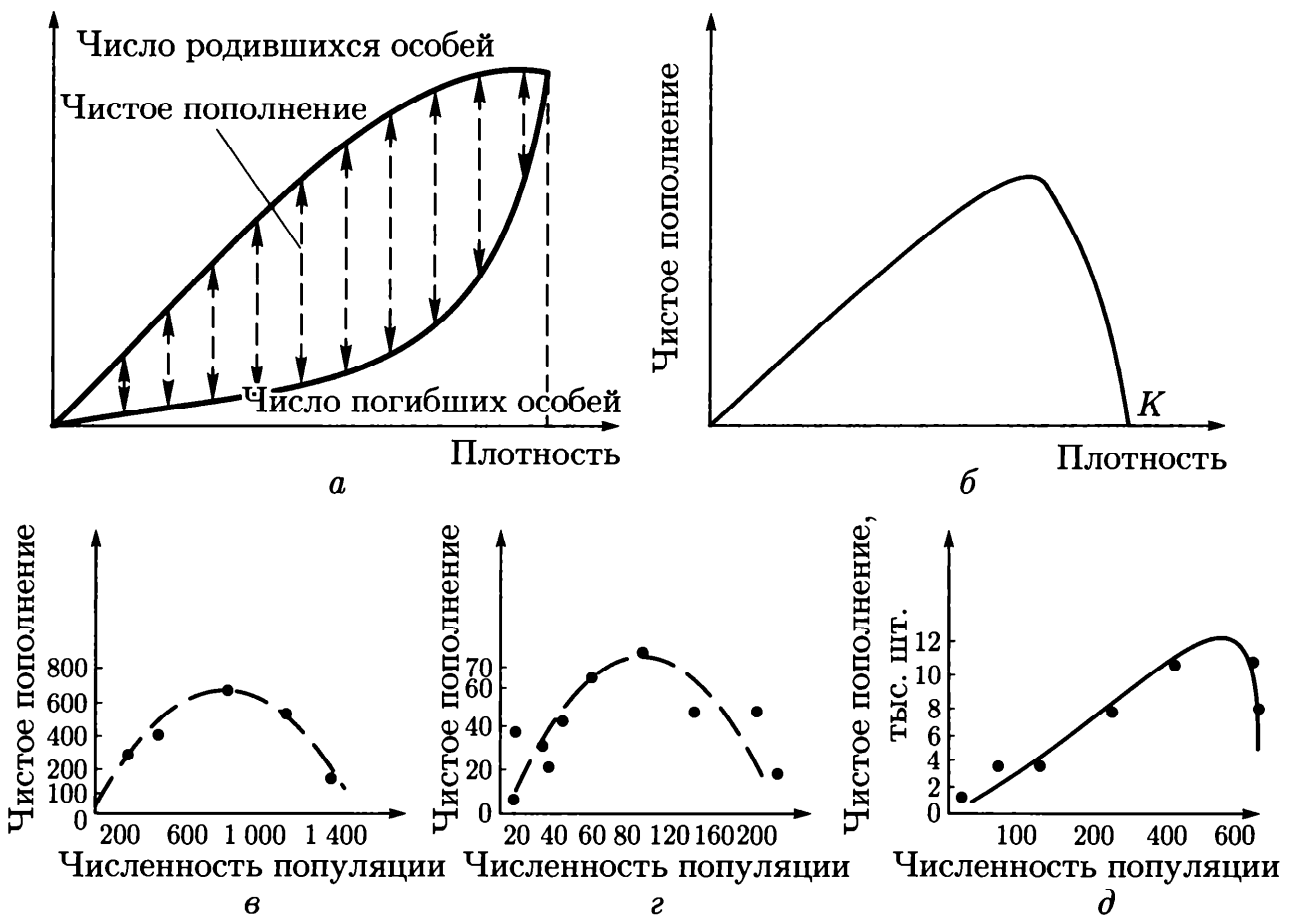


Рис. 4.8. Кривые пополнения, соответствующие в модели кривым зависимости скорости роста от численности популяции:

*а, б* — схема расчета кривых пополнения; *в, г, д* — кривые пополнения для разных видов: *в* — численность обыкновенного фазана на о. Протекши-Айленд после его интродукции в 1937 г. [A. S. Einarsen, 1945], *г* — экспериментальная популяция плодовой мушки *Drosophyla melanogaster* [R. Pearl, 1927], *д* — оценка численности арктического финвала [J. R. Allen, 1972]

является тенденция к поддержанию плотности каждой популяции в определенных пределах.

На рис. 4.8, *а*, *б* показана схема расчета экспериментальных *кривых пополнения*. Разность между двумя кривыми (число рожденных минус число погибших) представляет собой число особей, на которое изменяется численность популяции в течение какой-либо стадии развития или за какой-нибудь промежуток времени. Пополнение популяции невелико при низких значениях плотности, возрастает по мере ее увеличения, снова снижается при достижении предельной плотности насыщения и становится отрицательным (смертность превышает рождаемость), когда начальная плотность превышает  $K$ .

Конкретный вид связи между скоростью пополнения популяции и ее плотностью меняется в зависимости от биологии вида (см., например, графики для фазанов, мух и китов на рис. 4.8, *в*, *г*, *д*). Поскольку скорость пополнения популяции зависит от множества факторов, эмпирические точки никогда не ложатся точно на определенную кривую. Однако во всех случаях наблюдается колоколообразная кривая, форма которой отражает общую природу зависимых от плотности изменений рождаемости и смертности всякий раз, когда возникает внутривидовая конкуренция.

### 4.3.3. Модель популяции с наименьшей критической численностью

В рассмотренных моделях прирост численности (биомассы) популяции представлен линейным членом  $rx$ , пропорциональным численности. Строго говоря, это соответствует лишь тем популяциям, размножение которых происходит неполовым путем (микроорганизмы). Если в основе размножения лежит скрещивание, предполагающее встречи между особями разных полов одного и того же вида, то прирост будет тем выше, чем больше количество встреч между особями, а последнее пропорционально второй степени  $x$ . Таким образом, для разнополой популяции в условиях неограниченных ресурсов можно записать:

$$\frac{dx}{dt} = rx^2.$$

Это уравнение хорошо описывает тот факт, что при низких плотностях популяций скорость размножения резко падает, так как вероятность встречи двух особей разных полов уменьшается



при понижении плотности популяции пропорционально квадрату плотности. Однако при больших плотностях популяций скорость размножения лимитирует уже не число встреч особей противоположного пола, а число самок в популяции. Формула, учитывающая оба эти эффекта, имеет вид:

$$\frac{dx}{dt} = \alpha \frac{\beta x^2}{\beta + \tau x}.$$

В действительности плотность популяции не должна опускаться ниже некоторого критического значения. При падении плотности популяции ниже критической среднее время, в течение которого может состояться оплодотворение, становится больше времени жизни отдельной особи, точнее времени, в течение которого особь способна к размножению. В этом случае популяция вымирает.

Этот эффект может быть учтен, если в последнее уравнение ввести член, пропорциональный численности и описывающий смертность. Зависимость скорости роста популяции от ее численности при этом примет вид

$$\frac{dx}{dt} = \alpha \frac{\beta x^2}{\beta + \tau x} - \gamma x. \quad (4.6)$$

Уравнение (4.6) имеет два стационарных решения:  $\bar{x} = 0$  и  $\bar{x} = d\beta(\alpha\beta - \gamma\tau) = L$ . (Последнее решение может быть реализовано при соотношении параметров:  $\alpha\beta - \gamma\tau > 0$ , поскольку численность — положительная величина.) Соответствующие графики  $x(t)$  и  $f(x)$  приведены на рис. 4.9, а, б. Из рис. 4.9, б видно, что решение  $\bar{x} = 0$  — устойчивое, а  $\bar{x} = L$  — неустойчивое. При начальных численностях  $x_{\text{нач}} < L$  популяция вырождается,  $x \rightarrow 0$ , причем тем быстрее, чем меньше  $x_{\text{нач}}$ . Кривые  $x(t)$  при разных  $x_{\text{нач}}$  даны на рис. 4.9, а. При  $x_{\text{нач}} > L$ , в соответствии с уравнением (4.6), популяция неограниченно размножается.

Величина нижней критической плотности  $L$  различна для разных видов. Наблюдения биологов показали, что это всего лишь одна пара особей на тысячу квадратных километров в случае ондатры и сотни тысяч особей для американского странствующего голубя. Для голубых китов критическая граница общей численности оказалась равной десяткам — сотням. Хищническое истребление этих гигантских животных привело к тому, что их осталось слишком мало в Мировом океане. И хотя охота на них запрещена, надежды на восстановление популяции голубых китов практически нет.

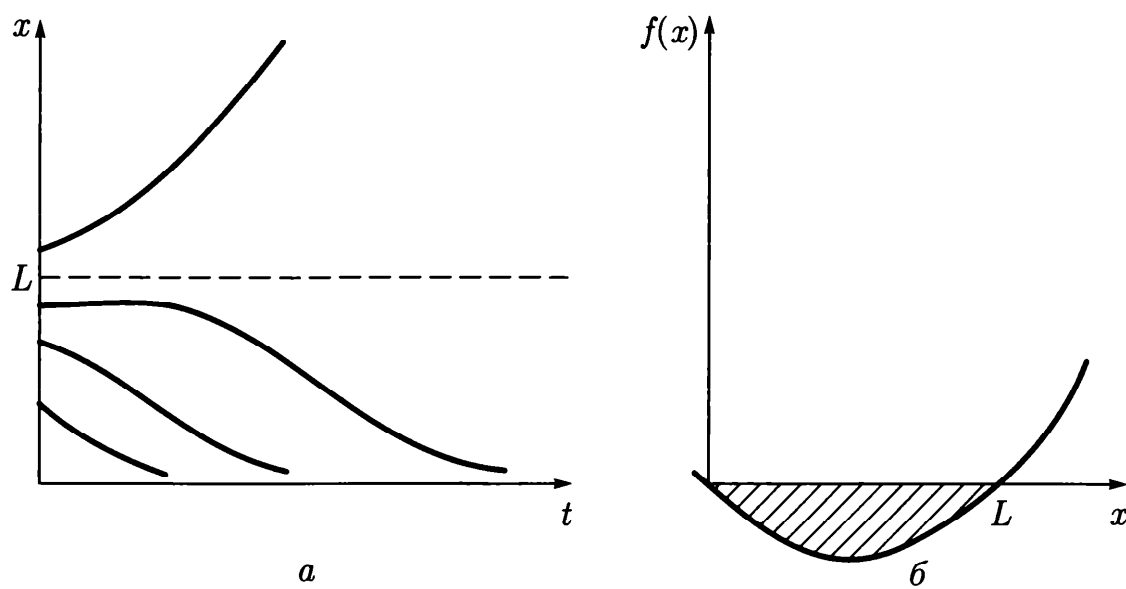


Рис. 4.9. Зависимость численности популяции от времени (а) и скорости роста от численности (б) для уравнения (4.6). Штриховкой обозначена область вырождения популяции

Наиболее общая формула, учитывающая как нижнюю границу численности, так и внутривидовую конкуренцию, имеет вид

$$\frac{dx}{dt} = \alpha \frac{\beta x^2}{\beta + \tau x} - \gamma x - \delta x^2. \quad (4.7)$$

Зависимости численности от времени и скорости прироста от численности представлены на рис. 4.10, а, б. Корни  $\bar{x} = 0$  и  $\bar{x} = K$  — устойчивые стационарные состояния;  $\bar{x} = L$  — неустойчивое, разделяющее области притяжения устойчивых состояний равновесия. Величины  $L$  и  $K$  различны для разных популяций и могут быть определены только из наблюдений и экспериментов. Ясно, что их определение представляет значительные трудности. Кривые 1—4 на рис. 4.10, а соответствуют различным начальным значениям численности популяции.

При любых промыслах особый интерес представляет величина нижней критической границы, при переходе через которую популяция уже не сможет восстановиться. Модель позволяет дать некий методический рецепт определения если не самой критической границы, то степени близости к ней численности вида.

Обратимся к рис. 4.10, а. Пусть численность вида в начальный момент времени была близка к максимально возможной. При  $t = 0$  происходит одноразовое выбивание популяции. Если численность осталась значительно больше нижней критической  $L$ , то восстановление происходит сначала быстро, а затем с монотонным замедлением (кривая 1). Если же оставшаяся по-

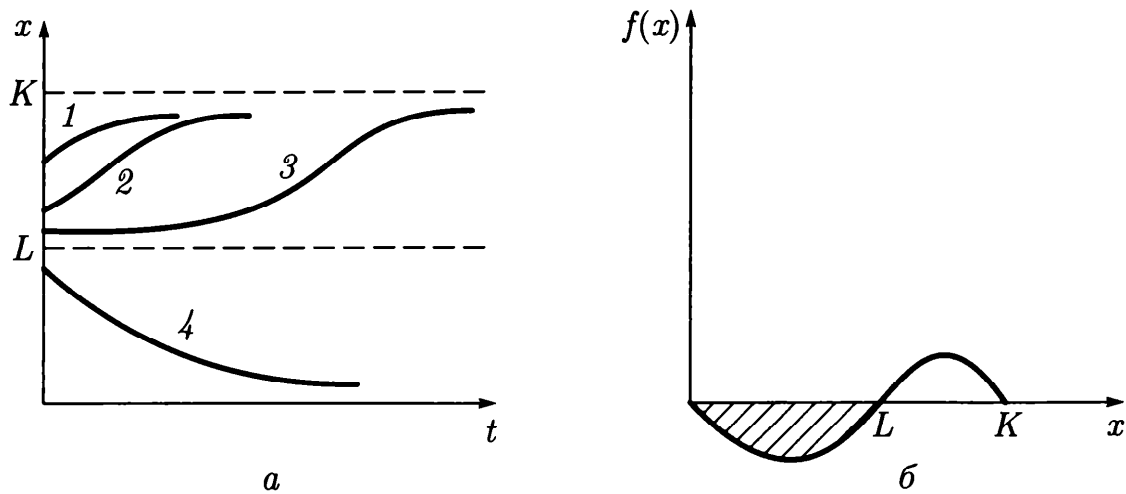


Рис. 4.10. Зависимость численности популяции от времени (а) и скорости роста от численности (б) для уравнения (4.7). Штриховкой обозначена область вырождения популяции

пуляция близка к критической точке, то восстановление происходит сначала очень медленно, численность популяции надолго «застревает» вблизи критической точки, а затем уже, «набрал силы», более быстро приближается к стационарному уровню (кривая 3). Кривая 2 представляет промежуточный случай. Таким образом, наблюдая реакцию системы на возмущение, можно предсказать приближение ее к опасным границам. Кривая 4 соответствует вырождению популяции.

#### 4.4. Дискретные модели популяций

Численность реальной популяции принимает только целые значения, она не изменяется непрерывно, а является дискретной величиной. Данные переписи реальных популяций представляют собой наборы целых чисел. Если предположить, что численность  $N$  зависит от численностей в некоторые предшествующие моменты времени, то для описания динамики численности популяций можно применить аппарат **разностных уравнений (отображений)**.

Если при этом внешние и внутренние факторы, определяющие развитие популяции, остаются во времени неизменными, то численность популяции в момент времени  $t + 1$  может быть описана с помощью разностного уравнения в виде:

$$N_{t+1} = F(N_t, N_{t-1}, N_{t-2}, \dots, N_{t-k}).$$

Здесь функция  $F$  зависит от численности популяции в момент времени  $t$  и  $k$  предшествующих моментов времени.

#### 4.4.1. Модели популяций с неперекрывающимися поколениями

Особенно просто выглядит разностное уравнение в случае, когда численность каждого следующего поколения  $N_{t+1}$  в популяции зависит лишь от численности предыдущего поколения  $N_t$ . Это справедливо для многих видов насекомых. Их взрослые особи живут непродолжительное время, достаточное для откладывания яиц, и к моменту появления на свет нового поколения (на стадии взрослой особи) предшествующее поколение прекращает свое существование. То же имеет место для некоторых видов зоопланктона, рыб, птиц. Про эти виды можно сказать, что поколения в них не перекрываются во времени, и уравнение для численности популяции в момент времени  $t + 1$  может быть записано в виде

$$N_{t+1} = F(N_t). \quad (4.8)$$

Опыт показывает, что в таких системах при малых численностях  $N$  растет от одной генерации к другой, а при высоких — падает. В таких случаях функция  $N$  имеет один максимум (одноэкстремальная функция) рис. 4.12, *a*. В зависимости от крутизны графика  $F(N_t)$  в системе могут возникать различные типы поведения (*динамические режимы*): монотонное и колебательное приближение к равновесию, колебательные изменения — циклы разной длины и квазистохастическое поведение — *хаос*.

Если поколения перекрываются — появляются дискретные возрастные группы (стадии развития). Для их описания обычно используются тоже дискретные, но более сложные матричные модели.

Модели вида (4.8) являются простейшими детерминированными объектами, демонстрирующими хаотическое поведение. В частности, такое поведение может возникнуть в дискретном логистическом уравнении.

Вспомним логистическое уравнение, которое описывает развитие популяции в непрерывном времени:

$$\frac{dN}{dt} = rN \left( 1 - \frac{N}{K} \right).$$

Заменив  $dN/dt$  на  $\Delta N/\Delta t$ , где  $\Delta N = N_{t+1} - N_t$ ,  $\Delta t = 1$ , получим

$$N_{t+1} = N_t \left[ 1 + r \left( 1 - \frac{N_t}{K} \right) \right] \quad (4.9)$$

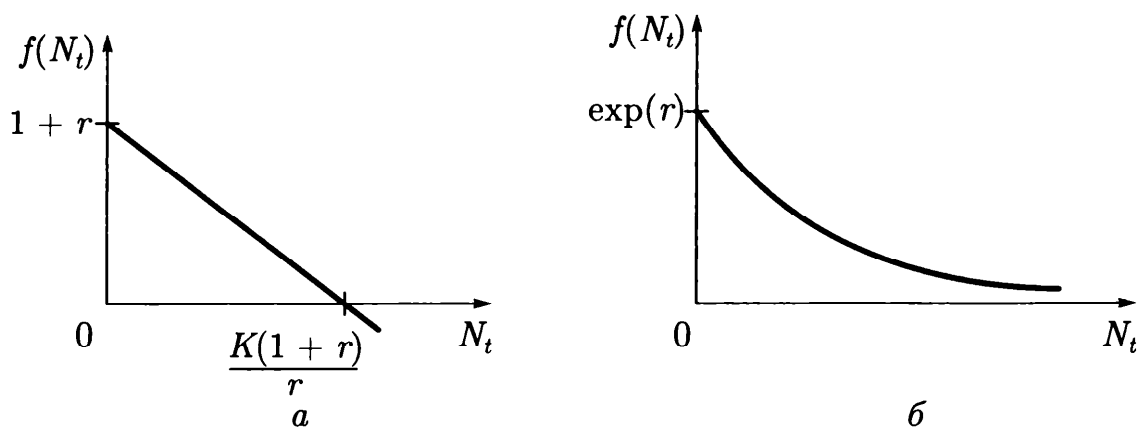


Рис. 4.11. Вид функции  $f(N_t)$  для дискретного аналога логистического уравнения:

$$a - f(N_t) = r(1 - N_t/K); \quad б - f(N_t) = \exp\{r(1 - N_t/K)\}$$

или

$$N_{t+1} = N_t f(N_t). \quad (4.10)$$

Однако это уравнение биологически некорректно. Если в какой-то момент времени значение численности  $N_t$  становится больше  $K(1+r)/r$ , то уравнение (4.9) дает отрицательное значение  $N_{t+1}$  (этот недостаток отсутствует у непрерывного логистического уравнения (4.3)). Это связано с видом функции  $f(N_t)$  (рис. 4.11, а).

Чтобы исправить положение в дискретном уравнении, в качестве  $f(N_t)$  следует взять функцию, асимптотически стремящуюся к нулю при  $N_t \rightarrow \infty$ . Вид такой функции изображен на рис. 4.11, б.

Экспоненциальная форма такой зависимости была предложена Г. Мораном (1950) для численности насекомых и В. Рикером (1952) для рыбных популяций:

$$N_{t+1} = N_t \exp \left\{ r \left( 1 - \frac{N_t}{K} \right) \right\}. \quad (4.11)$$

Точка пересечения биссектрисы первого координатного угла  $N_{t+1} = N_t$  и функции  $F(N_t)$  определяет равновесное состояние системы, аналогичное стационарному состоянию дифференциального уравнения (рис. 4.12, а).

#### 4.4.2. Диаграмма Ламерея

Ход решения можно наглядно продемонстрировать графически с помощью диаграммы, называемой *лестницей Ламерея*. На рис. 4.12, б показан способ нахождения значений  $N_t$  в последовательные моменты времени. Пусть в начальный момент

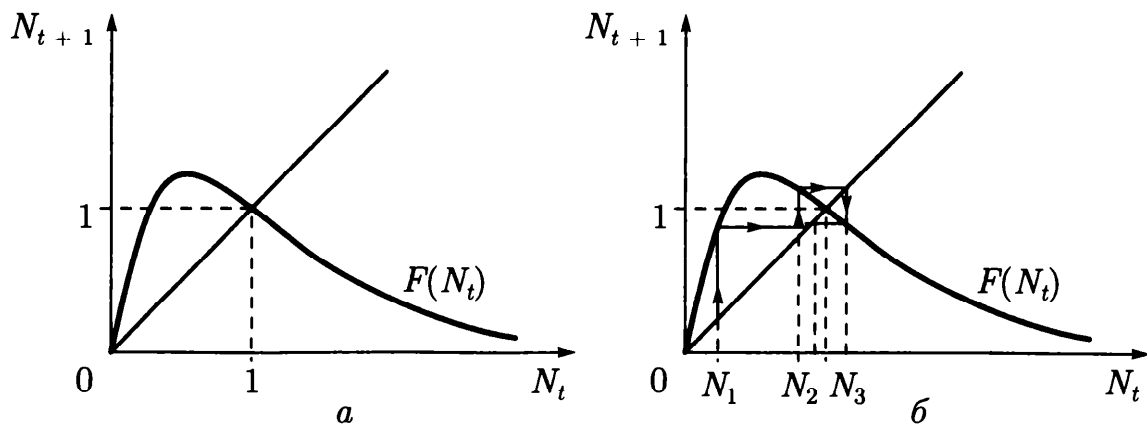


Рис. 4.12. Дискретная логистическая модель:

*a* — определение равновесного состояния в дискретной модели популяции с неперекрывающимися поколениями; *б* — лестница Ламерея

времени  $N = N_0$ ,  $F(N_0) = N_1$  задает значение численности в последующий момент времени  $t = 1$ . Величина  $N_1$  в свою очередь определяет значение  $F(N_1) = N_2$  и т. д. На рис. 4.12, б изображен случай, когда траектория сходится к равновесному состоянию, совершая затухающие колебания.

### 4.4.3. Понятия равновесия и его устойчивости для разностных уравнений

Определим для разностных уравнений (отображений) некоторые понятия, аналогичные основным понятиям теории дифференциальных уравнений.

**Решение (траектория)** — любая последовательность значений  $\{N_t\}$ , ( $t = 0, 1, \dots$ ), удовлетворяющая данному разностному уравнению при любом  $t$ . Разным начальным значениям соответствуют разные решения.

**Равновесием** называется решение вида:  $N_t = \text{const} = N^*$ , удовлетворяющее соотношению  $N^* = F(N^*)$ . Возвращаясь к формуле (4.10), видно, что для равновесного состояния  $f(N^*) = 1$ .

Как и в случае дифференциальных уравнений, для исследования устойчивости применим линейный анализ. Положим

$$N_t = N^* + x_t.$$

Линеаризуя уравнение (4.8), т. е. разлагая  $F$  в ряд по степеням  $x_t$  и отбрасывая члены порядка  $x_t^2$  и выше, получаем

$$x_{t+1} = \left( \frac{dF}{dN} \right)_{N^*} x_t + O(x_t^2).$$

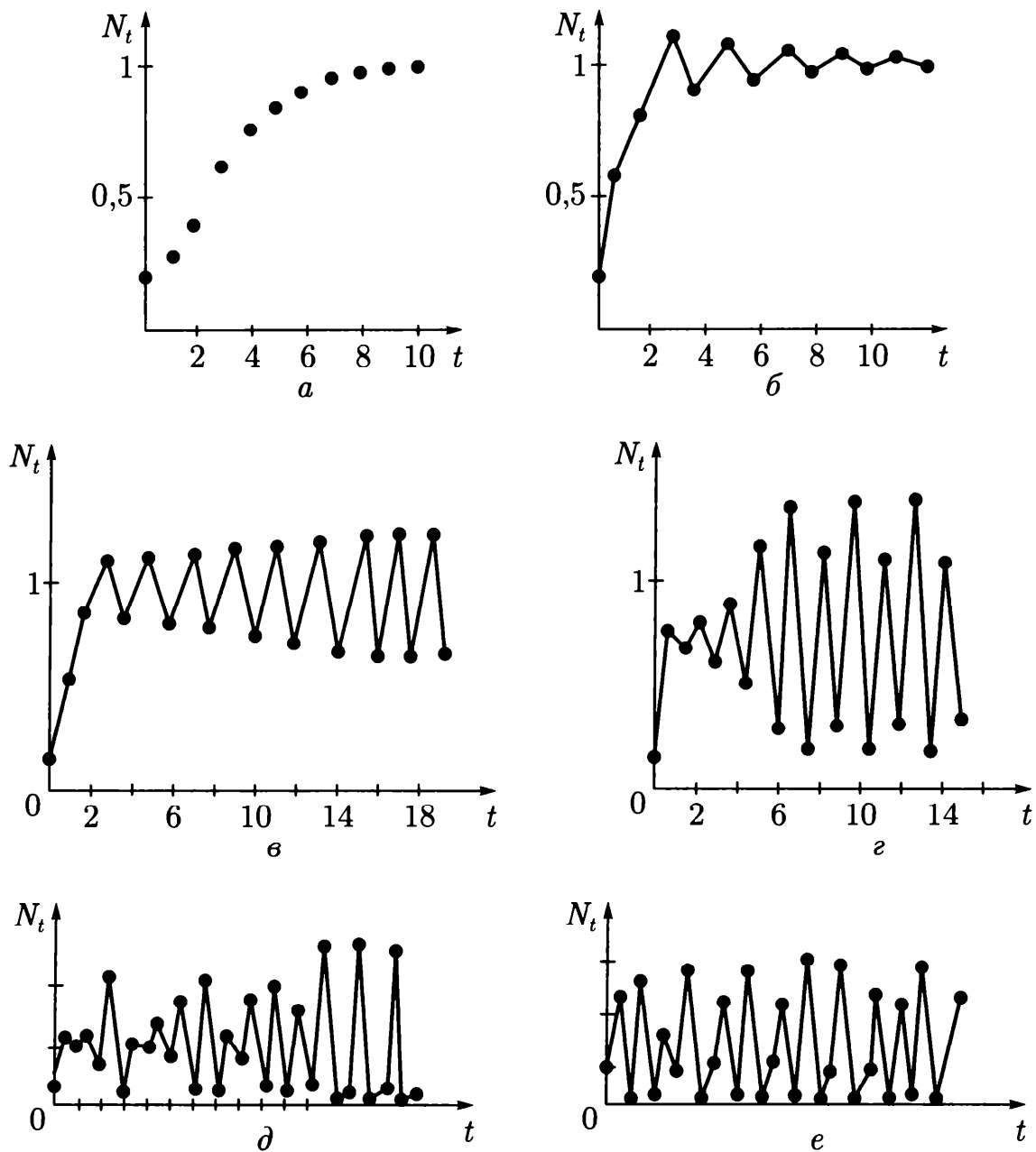


Рис. 4.13. Графики численности популяции, составленные по формуле (4.11):

*a* — монотонный рост; *б* — затухающие колебания; *в* — двухточечный цикл; *г* — четырехточечный цикл; *д*, *е* — примеры хаотических решений

Из условий сходимости геометрической прогрессии следует, что для линеаризованного уравнения

$$x_t \rightarrow 0 \text{ при } t \rightarrow \infty, \text{ если } \left| \frac{dF}{dN} \right|_{N^*} \leq 1 \text{ (равновесие устойчиво),}$$

и

$$x_t \rightarrow \infty \text{ при } t \rightarrow \infty, \text{ если } \left| \frac{dF}{dN} \right|_{N^*} > 1 \text{ (равновесие неустойчиво).}$$

Обозначим  $\left| \frac{dF}{dN} \right|_{N^*} = A.$

### ***В случае устойчивого равновесия:***

- $0 < A < 1$  — отклонения от равновесия исчезают монотонно;
- $-1 < A < 0$  — затухающие колебания вокруг  $N^*$ .

Графики монотонного и немонотонного стремления численности к равновесному состоянию в модели (4.11) представлены на рис. 4.13.

### ***В случае неустойчивого равновесия:***

- $A > 1$  — отклонение от равновесия монотонно растет;
- $A < -1$  — отклонение от равновесия в виде нарастающих колебаний.

При  $|dF/dN| = 1$  геометрическая прогрессия сходится, однако для выяснения характера сходимости (наличия монотонности) требуется дополнительное исследование членов более высокого порядка.

Для уравнения (4.11) равновесие находится из выражения

$$\exp \left\{ r \left( 1 - \frac{N^*}{K} \right) \right\} = 1.$$

Единственное равновесное значение  $N^* = K > 0$  существует при любом  $r$ . Равновесие устойчиво, если  $0 < r < 2$ , решение монотонно при  $0 < r < 1$  и представляет собой затухающие колебания при  $1 < r < 2$ .

#### **4.4.4. Циклы и динамический хаос**

В уравнении (4.11) возможны и более сложные решения — циклы, являющиеся аналогом предельных циклов для систем дифференциальных уравнений.

Решение называется циклом длины  $T$ , если

$$\begin{aligned} N_t^* &= N_{t+T}^*, t = 0, 1, 2, \dots; \\ N_{t+j}^* &\neq N_t^*, j = 1, 2, \dots, T - 1. \end{aligned}$$

В уравнении (4.11) циклы (колебательное поведение переменных) наблюдаются при следующих значениях параметра:

- при  $2 < r = r_2 < 2,526$  — двухточечные циклы;
- при  $r_2 < r < r_c$  появляются циклы длины  $4, 8, 16, \dots, 2^k$ ;
- при  $r > r_c = 3,102$  решение зависит от начальных условий.

Существуют трехточечные циклы и ***квазистохастические решения***.

Иногда в качестве дискретного аналога используется более простая формула, в которой численность выражена в безразмерном виде (при  $N_t > 1$  численность принимается равной нулю):



$$N_{t+1} = aN_t(1 - N_t).$$

Аналогичными свойствами обладает уравнение

$$N_{t+1} = \lambda_t N_t (1 + aN_t)^{-\beta},$$

которое хорошо описывает динамику численности сезонно размножающихся насекомых с неперекрывающимися поколениями.

Наблюдения показывают, что большинство популяций демонстрирует монотонный переход к состоянию равновесия. Эти результаты подтверждают правомерность использования непрерывного логистического уравнения в большом числе случаев для описания динамики численности.

#### 4.5. Вероятностные модели популяций

Рассмотренные модели популяций являлись детерминистическими. Однако существуют два аспекта, по которым детерминистическая модель не может служить точным отражением реальных экологических систем, поскольку она не учитывает вероятностный характер процессов размножения и гибели, а также случайных колебаний, происходящих в среде во времени и приводящих к случайным флуктуациям параметров моделей. Учет этих факторов приводит к существенному усложнению математического аппарата, поэтому обычно исследователи стараются строить детерминистические модели, ограничиваясь упоминанием о возможных последствиях учета стохастики. Если детерминистическая модель свидетельствует об устойчивом равновесии, то стохастическая модель предскажет длительное выживание. Если детерминистическая модель предсказывает периодические снижения численности одного или нескольких видов, то стохастическая модель даст некоторую положительную вероятность вымирания этих видов. Наконец, если детерминистическая модель не выявляет равновесия или равновесие неустойчивое, то стохастическая модель предскажет высокую вероятность вымирания.

В качестве простейшего примера рассмотрим вероятностное описание процесса роста популяции с учетом только размножения. При детерминистическом подходе считалось, что существует определенная скорость размножения  $\epsilon$ , такая, что численность популяции  $n$  за время  $dt$  увеличивается на  $dn = \epsilon n dt$ . Это приводит к экспоненциальному закону

$$n = ae^{\epsilon t},$$

где  $a$  — численность популяции в начальный момент времени.

Подойдем к процессу размножения с вероятностной точки зрения. Пусть вероятность появления одного потомка у данной особи в интервале времени  $dt$  равна  $\epsilon dt$ . Тогда вероятность появления одной новой особи в целой популяции за время  $dt$  равна  $\epsilon ndt$ . Обозначим через  $p_n(t)$  вероятность того, что в момент  $t$  в популяции имеется ровно  $n$  особей. Предположим, что в каждый момент времени может произойти только одно событие, а именно за время  $dt$  численность популяции может либо увеличиться на 1, либо остаться неизменной. Размер популяции в момент  $t$  можно связать с размером популяции в момент  $t + dt$  с помощью следующих рассуждений. Если число особей в момент  $t + dt$  равно  $n$ , это означает, что либо в момент  $t$  их было  $n - 1$  и за время  $dt$  появилась еще одна, вероятность этого события равна  $\epsilon$ , либо в момент  $t$  было  $n$  особей и за время  $dt$  это число не изменилось, вероятность этого события равна  $(1 - \epsilon)$ . Складывая вероятности, получим соотношение

$$p_n(t + dt) = p_{n-1}(t)\epsilon(n - 1)dt + p_n(t)(1 - \epsilon)ndt,$$

откуда путем перестановки членов и деления на  $dt$  получим

$$\frac{p_n(t + dt) - p_n(t)}{dt} = p_{n-1}(t)\epsilon(n - 1) - p_n(t)\epsilon n$$

или

$$\frac{dp_n}{dt} = \epsilon(n - 1)p_{n-1} - \epsilon np_n. \quad (4.12)$$

Уравнение (4.12) справедливо при  $n > a$ , где  $a$  — начальная численность популяции. Соответствующее уравнение для  $n = a$  имеет вид

$$\frac{dp_a}{dt} = -\epsilon a p_a, \quad (4.13)$$

так как в случае, когда процесс начинается при значении  $n = a$ , отсутствует член, содержащий  $p_{n-1}$ .

Системы дифференциально-разностных уравнений, аналогичные полученным, которые можно рассматривать как динамические уравнения для случайного процесса, обычно бывает трудно разрешить в общем виде. Однако в нашем примере это возможно. Проинтегрируем уравнение (4.13) с учетом того обстоятельства, что  $p_a(0) = 1$ :

$$p_a(t) = e^{-a\epsilon t}.$$

Затем подставляем это решение в уравнение для  $n = a + 1$ , интегрируем, используя начальное условие  $p_{a+1}(0) = 0$ , и находим

$$p_{a+1}(t) = ae^{-(a+1)\epsilon t}(e^{\epsilon t} - 1).$$

В свою очередь этот результат подставляем в последующее уравнение, и весь процесс повторяется. После вычисления нескольких последовательных членов можно записать результат в общем виде:

$$p_n(t) = C_{n-1}^{a-1} e^{-n\epsilon t} (e^{\epsilon t} - 1)^{n-a}, n \geq a. \quad (4.14)$$

Выражение (4.14) определяет распределение вероятностей для любого момента времени, заменяющее то единственное значение, которое рассматривалось в детерминистической модели. Оно является частным случаем биномиального распределения с математическим ожиданием

$$m(t) = ae^{\epsilon t}$$

и дисперсией

$$\sigma^2(t) = ae^{2\epsilon t}(1 - e^{-\epsilon t}).$$

Легко заметить, что математическое ожидание совпадает с детерминистическим средним. Таким образом, при большом числе особей детерминистическое описание будет удовлетворительно заменять любую стохастическую модель, в которой основное внимание уделяется нахождению средних значений. Когда же число особей мало, например, когда начальный размер популяции составляет всего лишь несколько единиц, дисперсия, т. е. среднее квадратичное отклонение численности отдельно взятой популяции от математического ожидания, может быть довольно значительной. При этом при  $t \rightarrow \infty$  коэффициент вариации величины  $n$ , равный  $\sigma/m$ , стремится к  $1/\sqrt{a}$ .

При изучении какой-либо определенной популяции мы наблюдаем только одно численное значение. График роста обнаруживает значительные колебания. Смысл выражения (4.14) состоит в том, что если имеется некоторое большое число популяций и в начальный момент времени  $t = 0$  численность каждой из них равна  $a$ , то доля этих популяций, имеющих в момент  $t$  численность  $n$ , теоретически равна  $p_n(t)$ . При этом математическое ожидание численности популяции (соответствует средней численности, определяемой в детерминистической модели) составляет  $m(t)$ , а дисперсия  $\sigma^2(t)$ .

МОДЕЛИ, ОПИСЫВАЕМЫЕ СИСТЕМОЙ  
ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

5.1. Основные понятия

5.1.1. Фазовая плоскость и фазовый портрет

Наиболее интересные результаты по качественному моделированию свойств биологических систем получены на моделях из двух дифференциальных уравнений, которые допускают качественное исследование с помощью метода *фазовой плоскости*. Напомним некоторые понятия качественной теории дифференциальных уравнений.

Рассмотрим систему двух автономных обыкновенных дифференциальных уравнений общего вида

$$\begin{cases} \frac{dx}{dt} = P(x, y); \\ \frac{dy}{dt} = Q(x, y), \end{cases} \quad (5.1)$$

где  $P(x, y)$ ,  $Q(x, y)$  — непрерывные функции, определенные в некоторой области  $G$  евклидовой плоскости ( $x, y$  — декартовы координаты) и имеющие в этой области непрерывные производные порядка не ниже первого.

Область  $G$  может быть как неограниченной, так и ограниченной. Если переменные  $x, y$  имеют конкретный биологический смысл (концентрации веществ, численности видов), то чаще всего область  $G$  представляет собой положительный квадрант правой полуплоскости:

$$0 \leq x < \infty, \quad 0 \leq y < \infty.$$

Концентрации веществ или численности видов также могут быть ограничены сверху объемом сосуда или площадью ареала обитания. Тогда область значений переменных имеет вид:

$$0 \leq x < x_0, \quad 0 \leq y < y_0.$$

Переменные  $x, y$  во времени изменяются в соответствии с системой уравнений (5.1) так, что каждому состоянию системы соответствует пара значений переменных  $(x, y)$ . Обратно, каждой

паре переменных  $(x, y)$  соответствует определенное состояние системы.

Рассмотрим плоскость с осями координат, на которых отложены значения переменных  $x, y$ . Каждая точка  $M$  этой плоскости соответствует определенному состоянию системы. Такая плоскость называется фазовой плоскостью и изображает совокупность всех состояний системы. Точка  $M(x, y)$  называется *изображающей*, или *представляющей*, *точкой*.

Пусть в начальный момент времени  $t = t_0$  координаты изображающей точки  $M_0(x(t_0), y(t_0))$ . В каждый следующий момент времени  $t$  изображающая точка будет смещаться в соответствии с изменениями значений переменных  $x(t), y(t)$ . Совокупность точек  $M(x(t), y(t))$  на фазовой плоскости, положение которых соответствует состояниям системы в процессе изменения во времени переменных  $x(t), y(t)$  согласно уравнениям (5.1), называется *фазовой траекторией*.

Совокупность фазовых траекторий при различных начальных значениях переменных дает легко обозримый «портрет» системы. Построение *фазового портрета* позволяет сделать выводы о характере изменений переменных  $x, y$  без знания аналитических решений исходной системы уравнений (5.1).

Для изображения фазового портрета необходимо построить векторное поле направлений траекторий системы в каждой точке фазовой плоскости. Задавая приращение  $\Delta t > 0$ , получим соответствующие приращения  $\Delta x$  и  $\Delta y$  из выражений:

$$\Delta x = P(x, y)\Delta t; \Delta y = Q(x, y)\Delta t.$$

Задача построения векторного поля упрощается, если получить выражение для фазовых траекторий в аналитическом виде. Для этого разделим второе уравнение системы (5.1) на первое:

$$\frac{dy}{dx} = \frac{Q(x, y)}{P(x, y)}. \quad (5.2)$$

Решение этого уравнения  $y = y(x, c)$ , или в неявном виде  $F(x, y) = c$ , где  $c$  — постоянная интегрирования, дает семейство интегральных кривых уравнения (5.2) — фазовых траекторий системы (5.1) на плоскости  $x, y$ .

### 5.1.2. Метод изоклин

Для построения фазового портрета применяют *метод изоклин* — на фазовую плоскость наносят линии, которые пере-

секают интегральные кривые под одним определенным углом. Уравнение изоклин легко получить из (5.2). Положим

$$\frac{dy}{dx} = A,$$

где  $A$  — определенная постоянная величина. Значение  $A$  представляет собой тангенс угла наклона касательной к фазовой траектории и может принимать значения от  $-\infty$  до  $+\infty$ . Подставляя вместо  $dy/dx$  в (5.2) величину  $A$ , получим уравнение изоклин:

$$A = \frac{Q(x, y)}{P(x, y)}. \quad (5.3)$$

Уравнение (5.3) определяет в каждой точке плоскости единственную касательную к соответствующей интегральной кривой, за исключением точки, где  $P(x, y) = 0$ ,  $Q(x, y) = 0$ , в которой направление касательной становится неопределенным, так как при этом становится неопределенным значение производной:

$$\left. \frac{dy}{dx} \right|_{x=\bar{x}, y=\bar{y}} = \frac{Q(\bar{x}, \bar{y})}{P(\bar{x}, \bar{y})} = \frac{0}{0}.$$

Эта точка является точкой пересечения всех изоклин — **особой точкой** (*точкой покоя*). В ней одновременно обращаются в нуль производные по времени переменных  $x$  и  $y$ :

$$\left. \frac{dx}{dt} \right|_{\bar{x}, \bar{y}} = P(\bar{x}, \bar{y}) = 0; \quad \left. \frac{dy}{dt} \right|_{\bar{x}, \bar{y}} = Q(\bar{x}, \bar{y}) = 0.$$

Следовательно, особая точка дифференциальных уравнений фазовых траекторий соответствует *стационарному состоянию системы*, а ее координаты — суть стационарные значения переменных  $x$ ,  $y$ . (В исключительных случаях особая точка может не быть точкой пересечения изоклин, например начало координат при  $P(x, y) = Q(x, y) = x^2 + y^2$ .)

Особый интерес представляют *главные изоклины*:

- $dy/dx = \infty$ ,  $P(x, y) = 0$  — изоклина *вертикальных касательных*;
- $dy/dx = 0$ ,  $Q(x, y) = 0$  — изоклина *горизонтальных касательных*.

Построив на фазовой плоскости главные изоклины и найдя точку их пересечения  $(x, y)$ , координаты которой удовлетворяют условиям:

$$P(\bar{x}, \bar{y}) = 0; \quad Q(\bar{x}, \bar{y}) = 0,$$

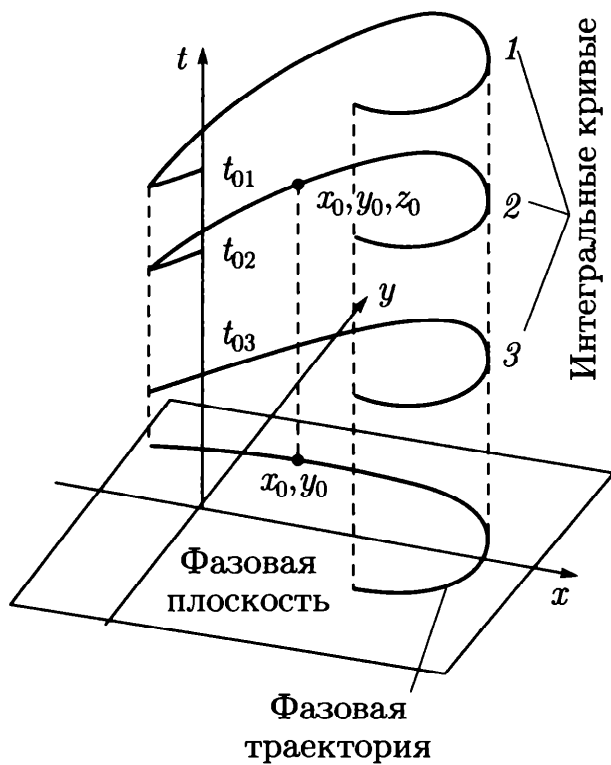


Рис. 5.1. Траектория системы в пространстве  $x, y, t$

находим стационарное состояние системы. Система (5.1) обладает столькими стационарными состояниями, сколько точек пересечения главных изоклин имеется на фазовой плоскости.

Каждая фазовая траектория соответствует совокупности движений динамической системы, проходящих через одни и те же состояния и отличающихся друг от друга только началом отсчета времени.

Таким образом, фазовые траектории системы — это проекции интегральных кривых в пространстве всех трех измерений  $x, y, t$  на плоскость  $x, y$  (рис. 5.1).

Если условия теоремы Коши выполнены, то через каждую точку пространства  $x, y, t$  проходит единственная интегральная кривая. То же справедливо, благодаря автономности, для фазовых траекторий: через каждую точку фазовой плоскости проходит единственная фазовая траектория.

### 5.1.3. Устойчивость стационарного состояния

Пусть система находится в состоянии равновесия. Тогда изображающая точка находится в одной из особых точек системы, в которых по определению:

$$\frac{dx}{dt} = 0; \quad \frac{dy}{dt} = 0.$$

Устойчива или нет особая точка, определяется тем, уйдет или нет изображающая точка при малом отклонении от стационар-

ного состояния. Применительно к системе двух уравнений определение устойчивости на языке  $\epsilon$ ,  $\delta$  выглядит следующим образом:

*состояние равновесия устойчиво, если для любой заданной области отклонений от состояния равновесия ( $\epsilon$ ) можно указать область  $\delta(\epsilon)$ , окружающую состояние равновесия и обладающую тем свойством, что ни одна траектория, которая начинается внутри области  $\delta$ , никогда не достигнет границы  $\epsilon$  (рис. 5.2).*

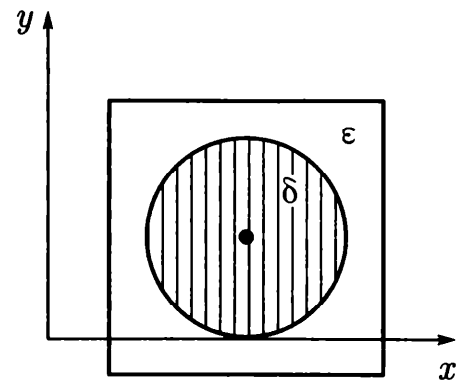


Рис. 5.2. Иллюстрация к определению устойчивости области  $\epsilon$  и  $\delta$  на плоскости  $x, y$

Для большого класса систем — *грубых систем*, поведения которых не меняется при малом изменении вида уравнений, информацию о типе поведения в окрестности стационарного состояния можно получить, исследуя не исходную, а упрощенную *линеаризованную* систему.

## 5.2. Исследование систем двух линейных уравнений

### 5.2.1. Характеристическое уравнение

Рассмотрим систему двух линейных уравнений:

$$\begin{cases} \frac{dx}{dt} = ax + by; \\ \frac{dy}{dt} = cx + dy, \end{cases} \quad (5.4)$$

где  $a, b, c, d$  — константы;  $x, y$  — декартовы координаты на фазовой плоскости.

Общее решение будем искать в виде:

$$x = Ae^{\lambda t}; \quad y = Be^{\lambda t}.$$

Подставим эти выражения в (5.4) и сократим на  $e^{\lambda t}$ :

$$\begin{cases} \lambda A = aA + bB; \\ \lambda B = cA + dB. \end{cases} \quad (5.5)$$

Такая алгебраическая система уравнений с неизвестными  $A, B$  имеет ненулевое решение лишь в том случае, если ее определи-



тель, составленный из коэффициентов при неизвестных, равен нулю:

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0.$$

Раскрывая этот определитель, получим *характеристическое уравнение* системы:

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0. \quad (5.6)$$

Решение уравнения (5.6) дает значения показателя  $\lambda_{1,2}$ , при которых возможны ненулевые для  $A$  и  $B$  решения уравнения (5.5). Эти значения суть

$$\lambda_{1,2} = \frac{a + d}{2} \pm \sqrt{\frac{(a + d)^2 - 4(ad - bc)}{4}}.$$

Если подкоренное выражение отрицательно, то  $\lambda_{1,2}$  комплексно сопряженные числа. Предположим, что уравнение (5.6) не имеет кратных корней. Тогда общее решение системы (5.4) можно представить в виде линейной комбинации экспонент с показателями  $\lambda_1, \lambda_2$ :

$$\begin{cases} x = c_{11}e^{\lambda_1 t} + c_{12}e^{\lambda_2 t}; \\ y = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t}. \end{cases}$$

Для анализа характера возможных траекторий системы на фазовой плоскости используем *линейное однородное преобразование координат*, которое позволит привести систему к *каноническому виду*:

$$\begin{cases} \frac{d\xi}{dt} = \lambda_1 \xi; \\ \frac{d\eta}{dt} = \lambda_2 \eta, \end{cases} \quad (5.7)$$

допускающее более удобное представление на фазовой плоскости по сравнению с исходной системой (5.4). Введем новые координаты  $\xi, \eta$  по формулам:

$$\xi = \alpha x + \beta y; \quad \eta = \gamma x + \delta y. \quad (5.8)$$

Коэффициенты преобразований (5.7), (5.8), вообще говоря, комплексные.

Из курса линейной алгебры известно, что в случае  $\lambda_1 \neq \lambda_2$  исходную систему (5.4) с помощью формул (5.8) можно преобразовать к каноническому виду (5.7) и изучать ее поведение на фазовой плоскости  $\xi, \eta$ . Рассмотрим различные случаи, которые могут здесь представиться.

### 5.2.2. Корни $\lambda_1, \lambda_2$ действительны и одного знака

Если корни  $\lambda_1, \lambda_2$  действительны и одного знака, то коэффициенты преобразования действительны, и мы переходим от действительной плоскости  $x, y$  к действительной плоскости  $\xi, \eta$ . Разделив второе уравнение системы (5.7) на первое, получим

$$\frac{d\eta}{d\xi} = \frac{\lambda_2 \eta}{\lambda_1 \xi}. \quad (5.9)$$

Интегрируя это уравнение, находим

$$\eta = C' |\xi|^r, \quad \text{где } r = \frac{\lambda_2}{\lambda_1}.$$

Условимся понимать под  $\lambda_2$  корень характеристического уравнения с бóльшим модулем, что не нарушает общности рассуждений. Тогда, поскольку в рассматриваемом случае корни  $\lambda_1, \lambda_2$  действительны и одного знака,  $r > 1$ , и мы имеем дело с интегральными кривыми параболического типа.

Все интегральные кривые (кроме оси  $\eta$ , которой соответствует  $C = \infty$ ) касаются в начале координат оси  $\xi$ , которая также является интегральной кривой уравнения (5.9). Начало координат является особой точкой.

Выясним теперь направление движений изображающей точки вдоль фазовых траекторий. Если  $\lambda_1, \lambda_2$  отрицательны, то  $|\xi|, |\eta|$  убывают с течением времени. Изображающая точка приближается к началу координат, никогда, однако, не достигая его. В противном случае это противоречило бы теореме Коши, которая утверждает, что через каждую точку фазовой плоскости проходит лишь одна фазовая траектория.

Такую особую точку, через которую проходят интегральные кривые, подобно тому, как семейство парабол  $y = cx^r$  ( $r > 0$ ) проходит через начало координат, называют узлом (рис. 5.3).

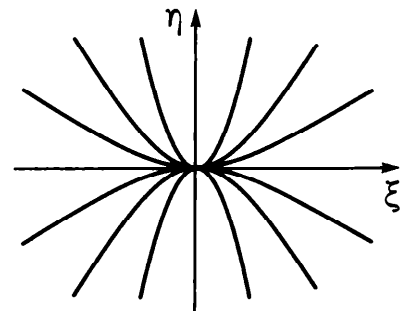


Рис. 5.3. Особая точка типа «узел» на плоскости канонических координат  $\xi, \eta$

Состояние равновесия типа «узел» при  $\lambda_1, \lambda_2 < 0$  устойчиво по Ляпунову, так как изображающая точка по всем интегральным кривым движется по направлению к началу координат. Это **устойчивый узел**. Если же  $\lambda_1, \lambda_2 > 0$ , то  $|\xi|, |\eta|$  возрастают с течением времени и изображающая точка удаляется от начала координат. В этом случае особая точка — **неустойчивый узел**.

На фазовой плоскости  $x, y$  общий качественный характер поведения интегральных кривых сохранится, но касательные к интегральным кривым не будут совпадать с осями координат. Угол наклона этих касательных будет определяться соотношением коэффициентов  $\alpha, \beta, \gamma, \delta$  в преобразовании (5.8).

### 5.2.3. Корни $\lambda_1, \lambda_2$ действительны и разных знаков

Если корни  $\lambda_1$  и  $\lambda_2$  действительны и разных знаков, то преобразование от координат  $x, y$  к координатам  $\xi, \eta$  опять действительное. Уравнения для канонических переменных снова имеют вид (5.7), а уравнение фазовых траекторий имеет вид:

$$\frac{d\eta}{d\xi} = -\frac{\eta}{\xi}, \text{ где } r = \left| \frac{\lambda_1}{\lambda_2} \right|.$$

Интегрируя, находим

$$\eta = C|\xi|^{-r}.$$

Это уравнение определяет семейство кривых гиперболического типа, где обе оси координат — асимптоты (при  $r = 1$  получается семейство равнобочных гипербол). Оси координат и в этом случае являются интегральными кривыми — это будут единственные интегральные кривые, проходящие через начало координат. Каждая из них состоит из трех фазовых траекторий: из двух движений к состоянию равновесия (или от состояния равновесия) и из состояния равновесия. Все остальные интегральные кривые — суть гиперболы, не проходящие через начало координат (рис. 5.4). Такая особая точка носит название «седло». Линии уровня вблизи горной седловины ведут себя подобно фазовым траекториям в окрестности седла.

Рассмотрим характер движения изображающей точки по фазовым траекториям вблизи состояния равновесия. Пусть, например,  $\lambda_1 > 0, \lambda_2 < 0$ . Тогда изображающая точка, помещенная на ось  $\xi$ , будет удаляться от начала координат, а помещенная на ось  $\eta$  — неограниченно приближаться к началу координат, не достигая его за конечное время. Где бы ни находилась изображающая

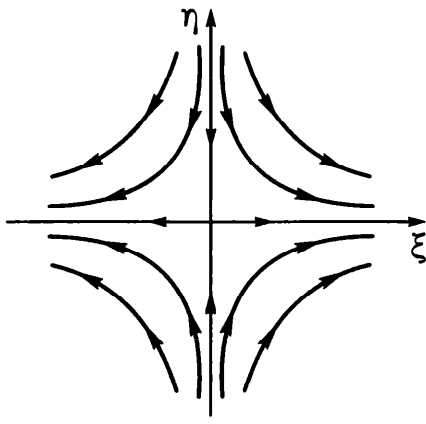


Рис. 5.4. Особая точка типа «седло» на плоскости канонических координат  $\xi, \eta$

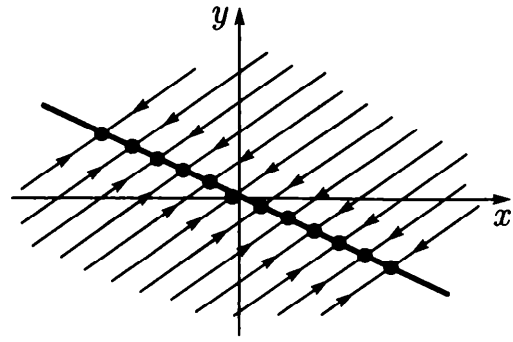


Рис. 5.5. Фазовый портрет системы, один из характеристических корней которой равен нулю, а второй отрицателен

точка в начальный момент (за исключением особой точки и точек на асимптоте  $\eta = 0$ ), она в конечном счете будет удаляться от состояния равновесия, даже если вначале она движется по одной из интегральных кривых по направлению к особой точке.

Очевидно, что особая точка типа «седла» всегда неустойчива. Только при специально выбранных начальных условиях на асимптоте  $\eta = 0$  система будет приближаться к состоянию равновесия. Однако это не противоречит утверждению о неустойчивости системы. Любая малая флуктуация приведет изображающую точку к уходу с асимптоты, поэтому всякое реальное движение будет удалять систему от состояния равновесия. Переходя обратно к координатам  $x, y$ , получим ту же качественную картину характера движения траекторий вокруг начала координат.

Пограничным между рассмотренными случаями узла и седла является случай, когда один из характеристических показателей, например  $\lambda_1$ , обращается в нуль, что имеет место, когда определитель системы  $ad - bc = 0$ . В этом случае коэффициенты правых частей уравнений (5.4) пропорциональны друг другу:

$$\frac{a}{c} = \frac{b}{d},$$

и система имеет своими состояниями равновесия все точки прямой:

$$ax + by = 0.$$

Остальные интегральные кривые представляют собой семейство параллельных прямых с угловым коэффициентом  $\chi = c/d$ , по которым изображающие точки либо приближаются к состоянию

равновесия, либо удаляются от него в зависимости от знака второго корня характеристического уравнения  $\lambda_2 = a+d$  (рис. 5.5). В этом случае координаты состояния равновесия зависят от начального значения переменных.

#### 5.2.4. Корни $\lambda_1, \lambda_2$ комплексные сопряженные

Если корни  $\lambda_1$  и  $\lambda_2$  комплексные сопряженные, то при действительных  $x$  и  $y$  имеем комплексные сопряженные  $\xi, \eta$ . Однако, вводя еще одно промежуточное преобразование, можно и в этом случае свести рассмотрение к действительному линейному однородному преобразованию. Положим

$$\begin{cases} \lambda_1 = a_1 + ib_1; & \lambda_2 = a_1 - ib_1; \\ \xi = u + iv; & \eta = u - iv, \end{cases} \quad (5.10)$$

где  $a_1, b_1$  и  $u, v$  — действительные величины. Можно показать, что преобразование от  $x, y$  к  $u, v$  является при допущенных предположениях действительным, линейным, однородным, с детерминантом, отличным от нуля. В силу уравнений (5.7), (5.10) имеем

$$\begin{aligned} \frac{du}{dt} + i \frac{dv}{dt} &= (a_1 + ib_1)(u + iv); \\ \frac{dv}{dt} - i \frac{du}{dt} &= (a_1 - ib_1)(u - iv), \end{aligned}$$

откуда

$$\begin{cases} \frac{du}{dt} = a_1 u - b_1 v; \\ \frac{dv}{dt} = a_1 v + b_1 u. \end{cases} \quad (5.11)$$

Разделив второе уравнение на первое, получим

$$\frac{dv}{du} = \frac{a_1 v + b_1 u}{a_1 u - b_1 v},$$

которое легче интегрируется, если перейти к полярной системе координат  $(r, \varphi)$ . После подстановки  $u = r \cos \varphi$ ,  $v = r \sin \varphi$  получим  $\frac{dr}{d\varphi} = \frac{a_1}{b_1 r}$ , откуда

$$r = C e^{\frac{a_1}{b_1} \varphi}.$$

Таким образом, на фазовой плоскости  $u, v$  мы имеем дело с семейством логарифмических спиралей, каждая из которых имеет асимптотическую точку в начале координат. Особая точка, которая является асимптотической точкой всех интегральных кривых, имеющих вид спиралей, вложенных друг в друга, называется **фокусом** (рис. 5.6).

Рассмотрим характер движения изображающей точки по фазовым траекториям. Умножив первое уравнение системы (5.11) на  $u$ , а второе на  $v$  и сложив полученные уравнения, имеем

$$\frac{1}{2} \frac{d\rho}{dt} = a_1 \rho, \text{ где } \rho = u^2 + v^2.$$

Пусть  $a_1 < 0$  ( $a_1 = \operatorname{Re} \lambda_{1,2}$ ). Изображающая точка тогда непрерывно приближается к началу координат, не достигая его в конечное время. Это означает, что фазовые траектории представляют собой скручивающиеся спирали и соответствуют затухающим колебаниям переменных. Это — **устойчивый фокус**.

В случае устойчивого фокуса, как и в случае устойчивого узла, выполнено не только условие Ляпунова, но и более жесткое требование асимптотической устойчивости (см. 4.1). Именно при любых начальных отклонениях система по прошествии времени вернется как угодно близко к положению равновесия.

Если  $a_1 > 0$ , то изображающая точка удаляется от начала координат, и мы имеем дело с **неустойчивым фокусом**. При переходе от плоскости  $u, v$  к фазовой плоскости  $x, y$  спирали также останутся спиралями, однако будут деформированы.

При  $a_1 = 0$  фазовыми траекториями на плоскости  $u, v$  будут окружности  $u^2 + v^2 = \text{const}$ , которым на плоскости  $x, y$  соответствуют эллипсы:

$$by^2 + (a - d)xy - cx^2 = \text{const},$$

т. е. при  $a_1 = 0$  через особую точку  $x = 0, y = 0$  не проходит ни одна интегральная кривая. Такая изолированная особая точка, вблизи которой интегральные кривые представляют собой замкнутые кривые, в частности эллипсы, вложенные друг в друга и охватывающие особую точку, называется **центром**.

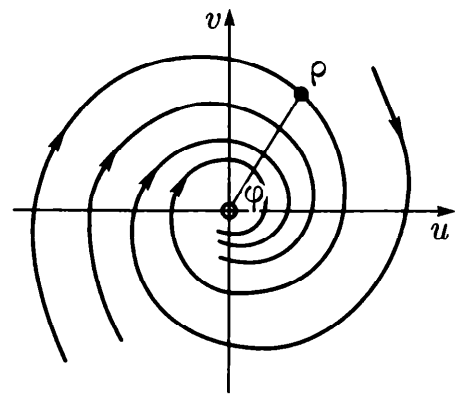
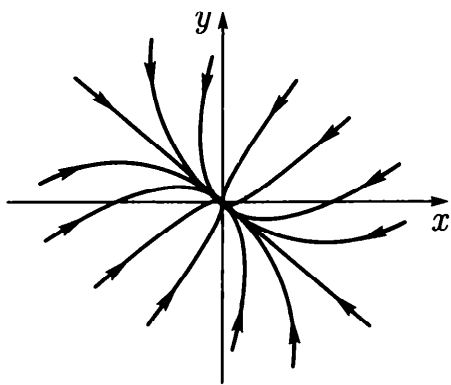
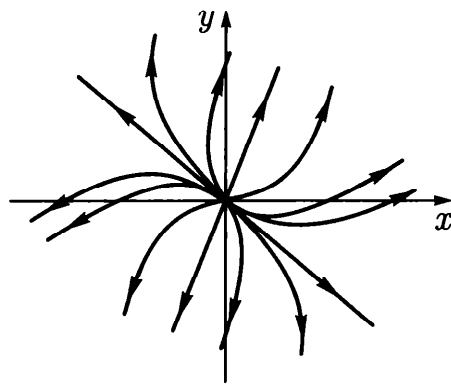


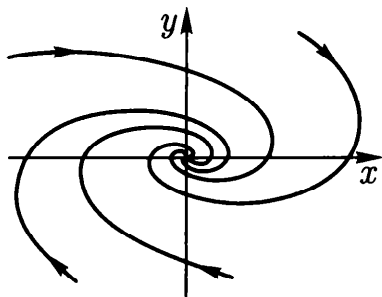
Рис. 5.6. Фазовый портрет системы в окрестности особой точки типа «фокус» на плоскости координат  $u, v$



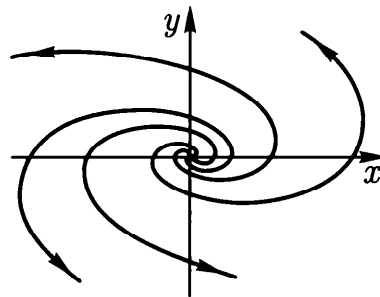
Устойчивый узел  
( $\lambda_1, \lambda_2$  действительны  
и отрицательны)



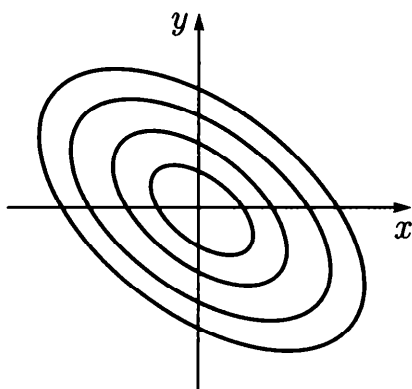
Неустойчивый узел  
( $\lambda_1, \lambda_2$  действительны  
и положительны)



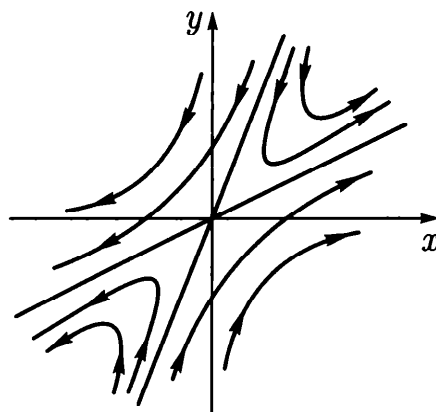
Устойчивый фокус  
( $\lambda_1, \lambda_2$  комплексны,  
 $\text{Re } \lambda_{1,2} < 0$ )



Неустойчивый фокус  
( $\lambda_1, \lambda_2$  комплексны,  
 $\text{Re } \lambda_{1,2} > 0$ )



Центр  
( $\lambda_1, \lambda_2$  чисто мнимое)



Седло  
( $\lambda_1, \lambda_2$  действительны  
и разных знаков)

Рис. 5.7. Типы фазовых портретов в окрестности стационарного состояния для системы линейных уравнений (5.4)

Таким образом, мы рассмотрели шесть типов состояния равновесия в зависимости от характера корней характеристического уравнения (5.6). Вид фазовых траекторий на плоскости  $x, y$  для этих шести случаев изображен на рис. 5.7. Эти типы состояния равновесия относятся к случаям, когда корни характеристического уравнения (5.6) не обращаются в нуль и не совпадают между собой.

Пять типов состояния равновесия грубые, их характер не изменяется при достаточно малых изменениях правых частей уравнений (5.4). При этом малыми должны быть изменения не только правых частей, но и их производных первого порядка. Шестое состояние равновесия — центр — негрубое. При малых изменениях параметров правой части уравнений он переходит в устойчивый или неустойчивый фокус.

### 5.2.5. Бифуркационная диаграмма

Введем обозначения:

$$\sigma = -(a + d); \Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Тогда характеристическое уравнение (5.6) запишется в виде:

$$\lambda^2 + \sigma\lambda + \Delta = 0.$$

Рассмотрим плоскость с прямоугольными декартовыми координатами  $\sigma$ ,  $\Delta$  и отметим на ней области, соответствующие тому или иному типу состояния равновесия, который определяется характером корней характеристического уравнения

$$\lambda_{1,2} = \frac{-\sigma \pm \sqrt{\sigma^2 - 4\Delta}}{2}.$$

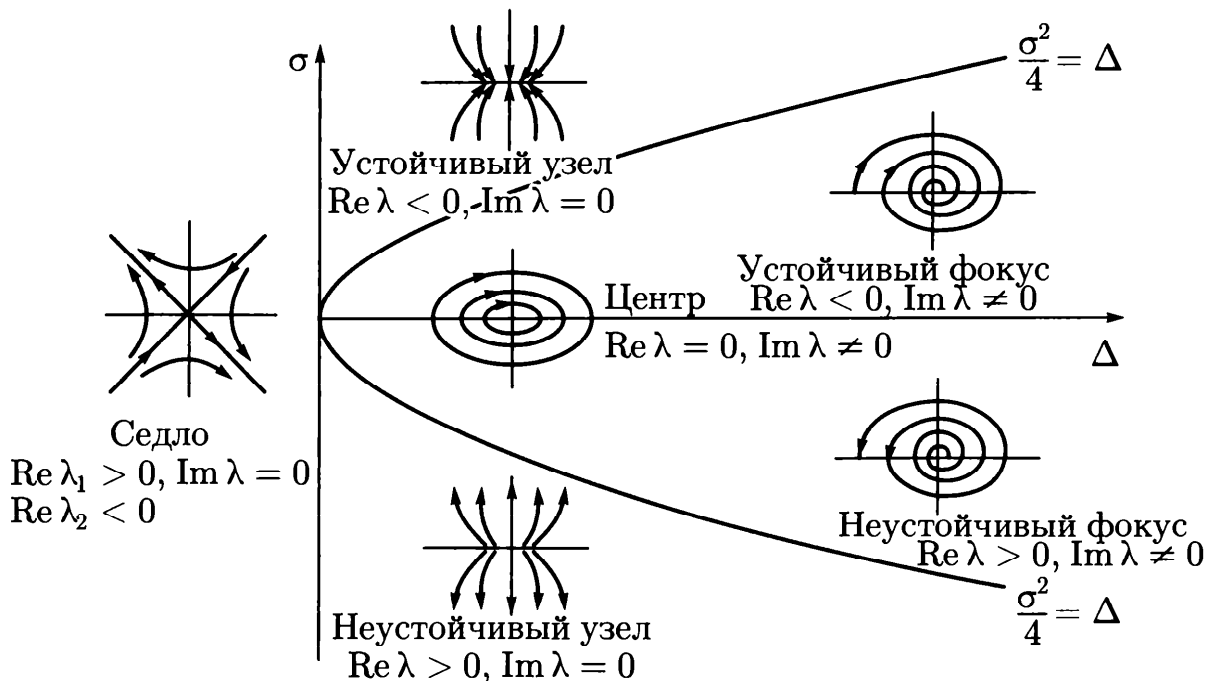


Рис. 5.8. Бифуркационная диаграмма для системы линейных уравнений (5.4)



Условием устойчивости состояния равновесия будет наличие отрицательной действительной части у  $\lambda_1$  и  $\lambda_2$ . Необходимое и достаточное условие этого — выполнение неравенств  $\sigma > 0$ ,  $\Delta > 0$ . На диаграмме (рис. 5.8) этому условию соответствуют точки, расположенные в первой четверти плоскости параметров. Особая точка будет фокусом, если  $\lambda_1$  и  $\lambda_2$  комплексны. Этому условию соответствуют те точки плоскости, для которых  $\sigma^2 - 4\Delta < 0$ , т. е. точки между двумя ветвями параболы  $\sigma^2 = 4\Delta$ . Точки полуоси  $\sigma = 0$ ,  $\Delta > 0$  соответствуют состояниям равновесия типа «центр». Если  $\lambda_1$  и  $\lambda_2$  действительны, но разных знаков, то  $\Delta < 0$ , т. е. особая точка будет седлом. В итоге получаем диаграмму разбиения плоскости параметров  $\sigma$ ,  $\Delta$  на области, соответствующие различным типам состояния равновесия.

Если коэффициенты линейной системы  $a$ ,  $b$ ,  $c$ ,  $d$  зависят от некоторого параметра, то при изменении этого параметра будут меняться и величины  $\sigma$ ,  $\Delta$ . При переходе через границы характер фазового портрета качественно меняется. В случае когда меняется характер устойчивости особой точки, такие границы называются *бифуркационными* — по разные стороны от границы система имеет два различных фазовых портрета и соответственно два разных типа поведения.

На диаграмме видно, как могут проходить такие изменения. Если исключить особые случаи — начало координат, то легко видеть, что седло может переходить в узел, устойчивый или неустойчивый при пересечении оси ординат. Устойчивый узел может перейти либо в седло, либо в устойчивый фокус и т. д.

Отметим, что переходы устойчивый узел — устойчивый фокус и неустойчивый узел — неустойчивый фокус не являются бифуркационными, так как характер устойчивости при этом не меняется. При этом переходе выполняется условие  $\sigma^2 = 4\Delta$ , а корни характеристического уравнения равны между собой:  $\lambda_{1,2} = -(a + d)$ . В случае отрицательных корней особая точка устойчива, в случае положительных — неустойчива.

При бифуркационных переходах меняется характер устойчивости особой точки. Например, устойчивый фокус через центр может переходить в неустойчивый фокус. Эта бифуркация называется *бифуркацией Андронова — Хопфа* по именам исследовавших ее ученых. При этой бифуркации в нелинейных системах происходит рождение предельного цикла и система становится автоколебательной (см. 5.6). Пограничная бифуркационная ситуация при переходе седло — узел рассматривались в 5.2.3.

## 5.3. Системы двух нелинейных дифференциальных уравнений

### 5.3.1. Метод Ляпунова линеаризации системы в окрестности стационарного состояния

Пусть биологическая система описывается системой двух автономных дифференциальных уравнений второго порядка общего вида:

$$\begin{cases} \frac{dx}{dt} = P(x, y); \\ \frac{dy}{dt} = Q(x, y). \end{cases} \quad (5.12)$$

Стационарные значения переменных системы определяются из алгебраических уравнений:

$$\begin{cases} P(\bar{x}, \bar{y}) = 0; \\ Q(\bar{x}, \bar{y}) = 0. \end{cases} \quad (5.13)$$

*Стационарные состояния* соответствуют особым точкам дифференциального уравнения первого порядка, определяющие интегральные кривые:

$$\frac{dy}{dx} = \frac{Q(x, y)}{P(x, y)}. \quad (5.14)$$

Математический анализ поведения траекторий этой системы на фазовой плоскости связан с именами французского математика А. Пуанкаре (1854—1912) и русского математика и механика А. М. Ляпунова (1857—1918).

Ляпунов показал, что в большом числе случаев анализ устойчивости стационарного состояния нелинейной системы можно заменить анализом устойчивости системы, линеаризованной в окрестности стационарного состояния.

Рассмотрим поведение переменных при некотором небольшом отклонении системы от состояния равновесия. Введем вместо переменных  $x$ ,  $y$  новые независимые переменные  $\xi$ ,  $\eta$ , определив их как смещения относительно равновесных значений переменных

$$\begin{cases} x = \bar{x} + \xi; \\ y = \bar{y} + \eta. \end{cases} \quad (5.15)$$

Подставив эти выражения в (5.12), получим

$$\begin{cases} \frac{d\bar{x}}{dt} + \frac{d\xi}{dt} = P(\bar{x} + \xi, \bar{y} + \eta); \\ \frac{d\bar{y}}{dt} + \frac{d\eta}{dt} = Q(\bar{x} + \xi, \bar{y} + \eta) \end{cases} \quad (5.16)$$

$\left(\frac{d\bar{x}}{dt} = \frac{d\bar{y}}{dt} = 0\right)$ , так как  $x, y$  — координаты особой точки).

Предположим, что функции  $P$  и  $Q$  непрерывны и имеют непрерывные производные не ниже первого порядка. Тогда можно разложить правые части уравнений (5.16) в ряд Тейлора по переменным  $\xi, \eta$ :

$$\begin{cases} \frac{d\xi}{dt} = P(\bar{x}, \bar{y}) + a\xi + b\eta + (p_{11}\xi^2 + 2p_{12}\xi\eta + p_{22}\eta^2 + \dots) + \dots; \\ \frac{d\eta}{dt} = Q(\bar{x}, \bar{y}) + c\xi + d\eta + (q_{11}\xi^2 + 2q_{12}\xi\eta + q_{22}\eta^2 + \dots) + \dots, \end{cases} \quad (5.17)$$

где

$$\begin{aligned} a &= P'_x(\bar{x}, \bar{y}), b = P'_y(\bar{x}, \bar{y}); \\ c &= Q'_x(\bar{x}, \bar{y}), d = Q'_y(\bar{x}, \bar{y}). \end{aligned} \quad (5.18)$$

Учтем, что по определению особой точки

$$\begin{cases} P(\bar{x}, \bar{y}) = 0; \\ Q(\bar{x}, \bar{y}) = 0, \end{cases}$$

и отбросим в уравнениях (5.17) нелинейные члены. Получим систему линейных уравнений с постоянными коэффициентами — *систему первого приближения*:

$$\begin{cases} \frac{d\xi}{dt} = a\xi + b\eta; \\ \frac{d\eta}{dt} = c\xi + d\eta. \end{cases} \quad (5.19)$$

Решение этой системы определяется корнями характеристического уравнения системы (см. 5.2):

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0. \quad (5.20)$$

Ляпунов показал, что если оба корня уравнения (5.20)

$$\lambda_{1,2} = \frac{(a + d) \pm \sqrt{(a + d)^2 - 4(ad - bc)}}{2} \quad (5.21)$$

имеют отличные от нуля действительные части, то исследование уравнений первого приближения (5.19) всегда дает правильный ответ на вопрос о типе устойчивости состояния равновесия в системе (5.12), а именно:

- если оба корня имеют отрицательную действительную часть и, следовательно, все решения уравнений первого приближения (5.19) затухают, то состояние равновесия устойчиво;
- если хотя бы один корень имеет положительную действительную часть, т. е. система (5.19) имеет нарастающие решения, то состояние равновесия неустойчиво.

Когда действительные части обоих корней характеристического уравнения равны нулю и корни чисто мнимые, линеаризованное уравнение имеет особую точку типа «центр». Если один корень равен нулю, а другой отрицателен, то уравнения (5.18), (5.19) не дают ответа на вопрос об устойчивости состояния равновесия, и необходимо рассматривать члены более высокого порядка малости в разложении в ряд Тейлора правых частей уравнений (5.12).

В случае когда оба корня характеристического уравнения имеют отличные от нуля действительные части (грубые системы), уравнение первого приближения определяет не только устойчивость стационарного состояния, но и характер фазовых траекторий в достаточно малой его окрестности.

Как и в случае линейных уравнений, здесь возможны пять типов грубых состояний равновесия: устойчивый узел, неустойчивый узел, устойчивый фокус, неустойчивый фокус и седло. Для исследования типов состояний равновесия удобно пользоваться диаграммой, изображенной на рис. 5.8. Для системы (5.12):

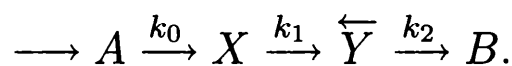
$$\begin{aligned} \sigma &= [P'_x(\bar{x}, \bar{y}) + Q'_y(\bar{x}, \bar{y})]; \\ \Delta &= \begin{vmatrix} P'_x(\bar{x}, \bar{y}) & Q'_x(\bar{x}, \bar{y}) \\ P'_y(\bar{x}, \bar{y}) & Q'_y(\bar{x}, \bar{y}) \end{vmatrix}. \end{aligned} \quad (5.22)$$

Грубым состояниям равновесия соответствуют все точки плоскости параметров  $\sigma$ ,  $\Delta$ , лежащие вне оси  $\Delta = 0$  и полуоси  $\sigma = 0$ ,  $\Delta > 0$ .

Точкам оси  $\Delta = 0$  и полуоси  $\sigma = 0$ ,  $\Delta > 0$  соответствуют негрубые состояния равновесия (негрубые особые точки). Их свойства могут быть изменены сколь угодно малыми изменениями правых частей уравнений (5.12) за счет сколь угодно малых изменений функций  $P(x, y)$ ,  $Q(x, y)$  и их производных. Поэтому характер негрубых состояний равновесия (в частности, устойчивость) уже не определяется значениями коэффициентов в правых частях уравнений первого приближения (5.19). В отличие от линейных систем уже при небольших изменениях в правых частях содержащихся там нелинейных членов может произойти качественное изменение фазового портрета — **бифуркация**.

### 5.3.2. Кинетические уравнения Лотки

А. Лоткой [А. J. Lotka. Elements of Physical Biology, 1925] была исследована гипотетическая химическая реакция:



Применим для исследования стационарного состояния этой модели метод линеаризации.

Пусть в некотором объеме находится в избытке вещество  $A$ . Молекулы вещества  $A$  с некоторой постоянной скоростью  $k_0$  превращаются в молекулы вещества  $X$  (реакция нулевого порядка). Вещество  $X$  может превращаться в вещество  $Y$ , причем скорость этой реакции тем больше, чем больше концентрации как вещества  $X$ , так и вещества  $Y$  (реакция второго порядка). В схеме это отражено обратной стрелкой над символом  $Y$ . Молекулы  $Y$  в свою очередь необратимо распадаются со скоростью, пропорциональной концентрации вещества  $Y$ . В результате образуется вещество  $B$  (реакция первого порядка).

Запишем систему уравнений, описывающих реакцию:

$$\begin{cases} \frac{dx}{dt} = k_0 - k_1xy; \\ \frac{dy}{dt} = k_1xy - k_2y; \\ \frac{dB}{dt} = k_2y, \end{cases} \quad (5.23)$$

где  $x, y, B$  — концентрации химических компонентов. Первые два уравнения этой системы не зависят от  $B$ , поэтому их мож-

но рассматривать отдельно. Рассмотрим стационарное решение системы:

$$\frac{dx}{dt} = 0; \quad \frac{dy}{dt} = 0.$$

Из этих условий получим систему алгебраических уравнений, связывающих стационарные концентрации  $\bar{x}, \bar{y}$ :

$$\begin{cases} k_0 - k_1 \bar{x} \bar{y} = 0; \\ k_1 \bar{x} \bar{y} - k_2 \bar{y} = 0. \end{cases}$$

Координаты особой точки:

$$\bar{x} = \frac{k_2}{k_1}; \quad \bar{y} = \frac{k_0}{k_2}.$$

Исследуем устойчивость этого стационарного состояния методом Ляпунова. Введем новые переменные  $\xi, \eta$ , характеризующие отклонения переменных от равновесных концентраций  $\bar{x}, \bar{y}$ :

$$\begin{aligned} x(t) &= \bar{x} + \xi(t); \\ y(t) &= \bar{y} + \eta(t). \end{aligned}$$

Линеаризованная система в новых переменных имеет вид:

$$\begin{cases} \frac{d\xi}{dt} = -k_2 \eta - \frac{k_1 k_0}{k_2} \xi; \\ \frac{d\eta}{dt} = \frac{k_1 k_0}{k_2} \xi. \end{cases}$$

Отметим, что величины отклонений от стационарных значений переменных  $\xi, \eta$  могут менять знак, в то время как исходные переменные  $x, y$ , являющиеся концентрациями, могут быть только положительными.

Запишем характеристическое уравнение системы:

$$\begin{vmatrix} -\frac{k_1 k_0}{k_2} - \lambda & -k_2 \\ \frac{k_1 k_0}{k_2} & -\lambda \end{vmatrix} = 0, \quad \text{или} \quad \lambda^2 + \lambda \frac{k_1 k_0}{k_2} + k_0 k_1 = 0.$$

Корни характеристического уравнения:

$$\lambda_{1,2} = \frac{1}{2} \left[ \frac{k_1 k_0}{k_2} \pm \sqrt{\left( \frac{k_1 k_0}{k_2} \right)^2 - 4k_0 k_1} \right].$$

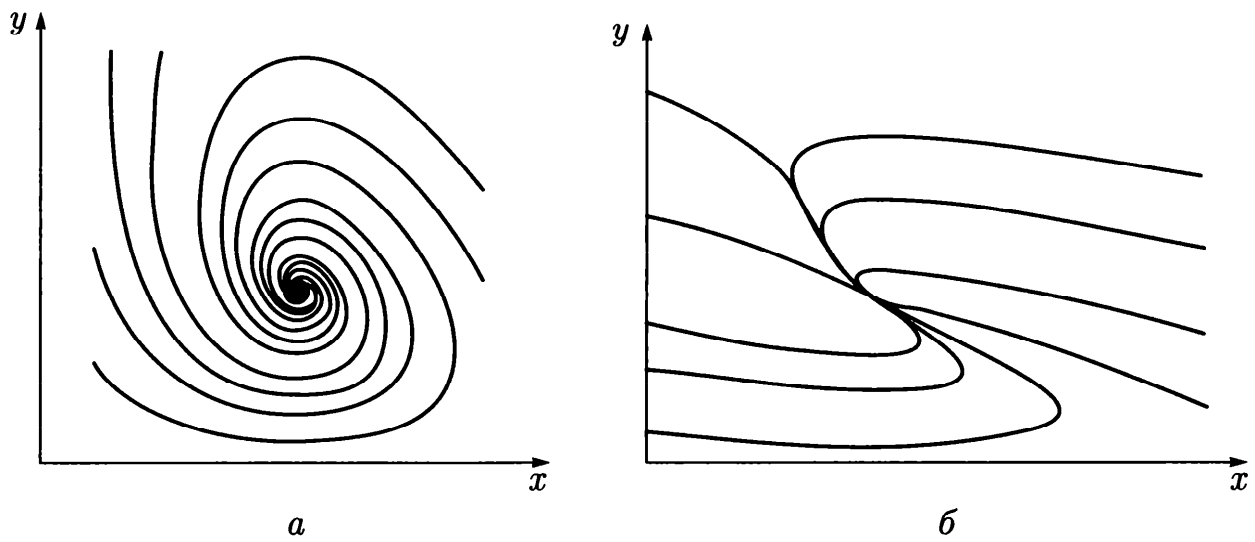


Рис. 5.9. Фазовый портрет системы первых двух уравнений системы (5.23):

*a* — устойчивый фокус,  $k_0 = 2$ ,  $k_1 = 10$ ,  $k_2 = 2$ ; *б* — устойчивый узел,  $k_0 = 2$ ,  $k_1 = 2$ ,  $k_2 = 4$

Фазовый портрет системы (5.23) изображен на рис. 5.9.

При  $4k_2^2 > k_0k_1$  подкоренное выражение отрицательно и особая точка — фокус, при обратном соотношении — узел. И в том и в другом случае особая точка устойчива, так как действительная часть обоих корней характеристического уравнения отрицательна.

Таким образом, в описанной химической реакции (5.23) возможны разные режимы изменения переменных в зависимости от соотношения величин констант скоростей. Если  $4k_2^2 > k_0k_1$ , имеют место затухающие колебания концентраций компонентов, при  $4k_2^2 < k_0k_1$  — бесколебательное приближение концентраций к стационарным. Соотношение параметров  $4k_2^2 = k_0k_1$  соответствует изменению типа особой точки системы уравнений (5.23).

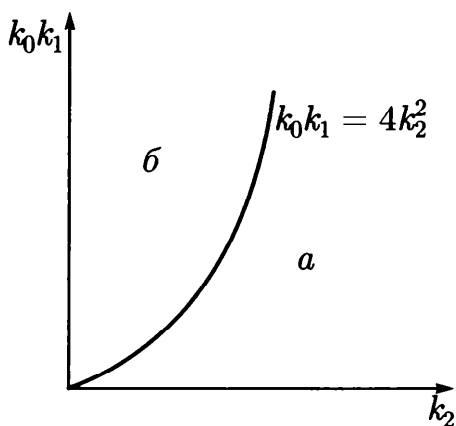


Рис. 5.10. Плоскость параметров для системы (5.23):

*a* — область устойчивого фокуса; *б* — область устойчивого узла

Рассмотрим плоскость параметров (рис. 5.10), где по оси абсцисс отложены значения константы  $k_2$ , а по оси ординат — произведение  $k_0k_1$ . Парабола  $k_0k_1 = 4k_2^2$  делит плоскость параметров на две области — устойчивых узлов и устойчивых фокусов. Задавая те или иные значения параметров, можно получить колебательный и бесколебательный режимы изменения концентраций веществ  $x$  и  $y$ . Фазовый портрет системы при этом бу-

дет представлять собой фокус или узел, изображенные соответственно на рис. 5.10, *а* и *б*.

Если в системе установятся стационарные концентрации веществ  $x$  и  $y$ , то это приведет к установлению постоянной скорости прироста концентрации вещества  $B$  в третьем уравнении системы (5.23):

$$\frac{dB}{dt} = k_2 y.$$

Ясно, что в действительности такая система реализоваться не может, так как в ней при  $t \rightarrow \infty$  концентрация вещества  $B$  стремится к бесконечности. Однако система, подобная системе реакций Лотки, может представлять собой фрагмент более сложной химической системы. Исследованные нами уравнения правильно описывают поведение компонентов  $x$  и  $y$ , если приток вещества  $x$  (скорость его постоянна и равна  $k_0$ ) осуществляется из большого «резервуара», а отток вещества  $y$  — в большой «резервуар» (значение  $B$  очень велико). При этих предположениях на малых промежутках времени (по сравнению с временем существенного изменения заполненности емкости веществом  $B$ ) наше рассмотрение вполне правомерно.

### 5.3.3. Модель Вольтерра

В качестве второго примера рассмотрим классическую *вольтерровскую модель «хищник — жертва»*, которая была предложена В. Вольтерра для объяснения периодических изменений числа особей. Модель впервые опубликована в работе V. Volterra «Variazioni e fluttuazioni del numero d'individui in specie animali conviventi». Mem. Acad. Lincei. Vol. 2. — P. 31 — 113, 1924<sup>1</sup>.

Пусть в некотором замкнутом районе живут хищники и жертвы, например волки и зайцы. Зайцы питаются растительной пищей, имеющейся всегда в достаточном количестве. Волки могут питаться лишь зайцами. Обозначим число зайцев (жертв) —  $x$ , а число волков (хищников) —  $y$ . Так как количество пищи у зайцев неограниченно, можно предположить, что они размножаются со скоростью, пропорциональной их числу:

$$\epsilon_x x. \tag{5.24}$$

---

<sup>1</sup>Вольтерра В. Математическая теория борьбы за существование. — М.: Наука, 1976.



Если рождаемость зайцев превышает их смертность, то  $\epsilon > 0$ . Выражение (5.24) соответствует автокаталитической реакции первого порядка.

Пусть убыль зайцев пропорциональна вероятности встречи зайца с волком, т. е. пропорциональна произведению  $xу$ . Можно предположить по аналогии с бимолекулярными реакциями, где вероятность появления новой молекулы пропорциональна вероятности встречи двух молекул, что и количество волков нарастает тем быстрее, чем чаще происходят их встречи с зайцами, а именно пропорционально  $xу$ .

Кроме того, имеет место процесс естественной смертности волков, причем скорость смертности пропорциональна их количеству.

Эти рассуждения приводят к системе уравнений для изменений численности зайцев-жертв  $x$  и волков-хищников  $y$ :

$$\begin{cases} \frac{dx}{dt} = x(\epsilon_x - \gamma y); \\ \frac{dy}{dt} = -y(\epsilon_y - \delta x). \end{cases} \quad (5.25)$$

Система имеет два стационарных решения. Одно из них — нулевое:  $\bar{x}_0 = 0, \bar{y}_0 = 0$ . Линейный анализ устойчивости показывает, что эта точка всегда представляет собой седло. Покажем, что система уравнений (5.25) также имеет на фазовой плоскости переменных  $xу$  ненулевую особую точку типа «центр». Координаты этой точки:

$$\bar{x} = \frac{\epsilon_y}{\delta}; \quad \bar{y} = \frac{\epsilon_x}{\gamma}.$$

Так как все параметры  $\epsilon_x, \epsilon_y, \gamma, \delta$  положительны, точка  $(\bar{x}, \bar{y})$  расположена в положительном квадранте фазовой плоскости. Линеаризация системы вблизи этой точки дает:

$$\begin{cases} \frac{d\xi}{dt} = a\xi + b\eta; \\ \frac{d\eta}{dt} = c\xi + d\eta. \end{cases}$$

Коэффициенты в этих уравнениях выражаются в соответствии с (5.18) через частные производные правых частей уравнений системы (5.25) формулами:

$$a = \epsilon_x - \gamma\bar{y}; \quad b = -\gamma\bar{x}; \quad c = \delta\bar{x}; \quad d = -\epsilon_y + \delta\bar{x}.$$

Здесь  $\xi, \eta$  — отклонения численностей от их стационарных значений  $\bar{x}, \bar{y}$ :

$$\xi(t) = x(t) - \bar{x};$$

$$\eta(t) = y(t) - \bar{y}.$$

Для нулевого стационарного состояния характеристический определитель имеет вид:

$$\begin{vmatrix} \varepsilon_x - \lambda & 0 \\ 0 & -\varepsilon_y - \lambda \end{vmatrix}.$$

Корни соответствующего характеристического уравнения:  $\lambda_1 = \varepsilon_x$ ,  $\lambda_2 = -\varepsilon_y$ . Тип нулевой особой точки — седло.

Характеристический определитель матрицы коэффициентов системы для ненулевого стационарного состояния имеет вид:

$$\begin{vmatrix} -\lambda & -\frac{\varepsilon_y \gamma_y}{\gamma_x} \\ \frac{\varepsilon_x \gamma_x}{\gamma_y} & -\lambda \end{vmatrix}.$$

Характеристическое уравнение:

$$\lambda^2 + \varepsilon_x \varepsilon_y = 0.$$

Корни этого уравнения чисто мнимые:

$$\lambda_{1,2} = \pm i \sqrt{\varepsilon_x \varepsilon_y}.$$

Таким образом, исследование системы показывает, что траектории вблизи ненулевой особой точки являются концентрическими эллипсами, а сама особая точка — центром. Можно показать, что изученная модель Вольтерра и вдали от особой точки имеет замкнутые траектории, хотя форма этих траекторий уже отличается от эллипсоидальной, и определяется параметрами системы (рис. 5.11).

Изменения численности жертв и хищников во времени представляют собой колебания, причем колебания численности хищников отстают по фазе от колебаний жертв.

Особая точка типа «центр» устойчива по Ляпунову, но не асимптотически. Покажем на данном примере, в чем это проявляется. Пусть колебания  $x(t)$  и  $y(t)$  происходят таким образом, что изображающая точка движется по фазовой траектории 1 (см. рис. 5.11).

В момент, когда точка находится в положении  $M_1$ , в систему добавляется извне некоторое число особей  $y$ , такое, что изображающая точка переходит скачком из точки  $M_1$  в точку  $M_2$ .

Если после этого систему предоставить самой себе, то колебания  $x(t)$ ,  $y(t)$  уже будут происходить с большими амплитудами, чем прежде, и изображающая точка будет двигаться по траектории 2. Это и означает, что колебания в системе неустойчивы: они навсегда изменяют свои характеристики при внешнем воздействии.

В дальнейшем будут рассмотрены модели, описывающие устойчивые колебательные режимы, и показано, что на фазовой плоскости такие асимптотически устойчивые периодические движения описываются предельными циклами.

На рис. 5.12 приведены кривые колебаний численности пушных зверей по данным компании Гудзонова залива о числе заготовленных шкурок. Во всех классических учебниках в течение многих лет колебательный характер этих изменений приводили как подтверждение гипотез, положенных в основу модели Вольтерра, которая только что была рассмотрена. Действительно, периоды колебаний численности зайцев (жертв) и рысей (хищников) примерно одинаковы и составляют порядка 9—10 лет. При этом максимум численности зайцев опережает, как правило, максимум численности рысей на один год. Можно полагать, что мы видим регулярные колебания, осложненные случайными факторами, связанными с погодой и пр.

Однако возможна и другая интерпретация этих данных наблюдений на основе моделей детерминированного хаоса. Дискретные модели такого типа рассматривались в гл. 1 (см. 1.4).

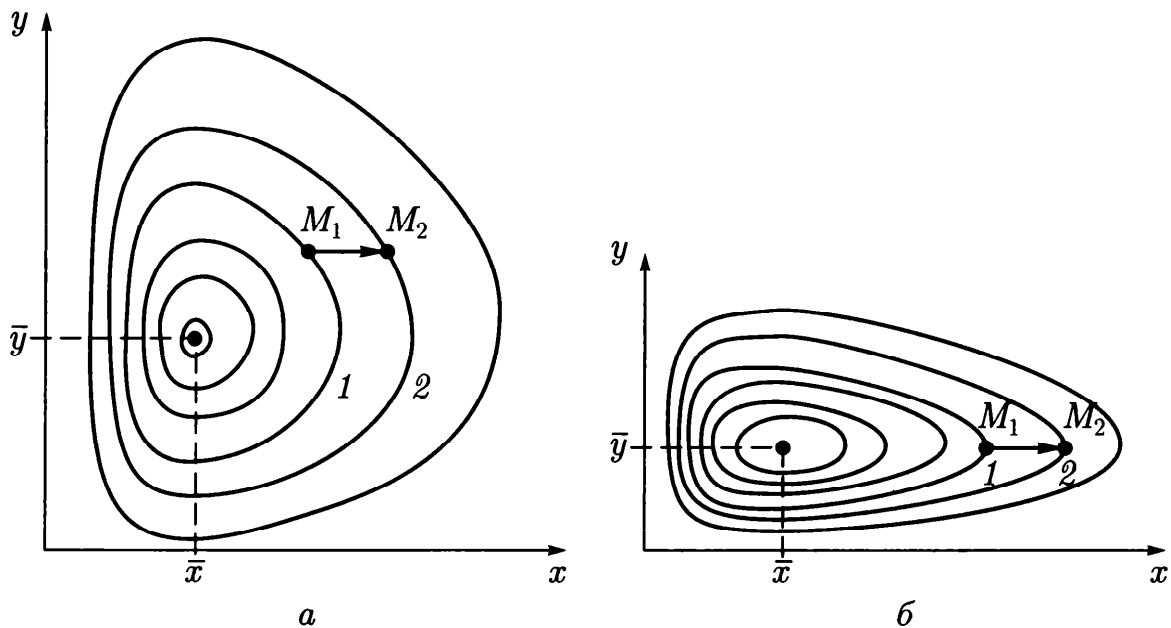


Рис. 5.11. Фазовый портрет системы (5.25). Особая точка типа «центр»:  
 $a$  — параметры системы:  $\epsilon_x = 4$ ,  $\gamma_{xy} = 0,3$ ,  $\epsilon_y = \gamma_{yx} = 0,4$ ;  $b$  — параметры системы:  $\epsilon_x = 2$ ,  $\gamma_{xy} = 0,3$ ,  $\epsilon_y = \gamma_{yx} = 0,4$

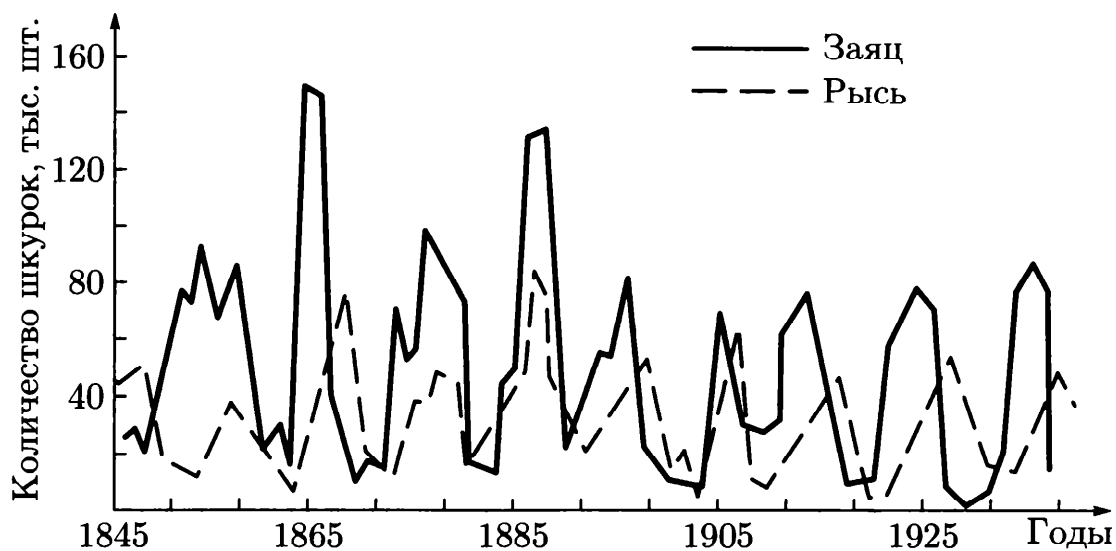


Рис. 5.12. Кривые численности зайца и рыси в Канаде [К. Вилли, В. Детье, 1974]

Непрерывные модели популяционной динамики, приводящие к детерминированному хаосу, будут изучены далее.

Серьезным недостатком модели Вольтерра является неустойчивость решений по отношению к малым случайным воздействиям, приводящим к изменению переменных. Кроме того, в силу «негрубости» этой системы произвольно малое изменение вида правых частей уравнений (величин параметров системы) приведет к изменению типа особой точки и, следовательно, к изменению характера фазовых траекторий.

Поскольку природные системы подвергаются огромному количеству случайных воздействий, реалистическая модель должна быть по отношению к ним устойчивой. Поэтому негрубые системы не могут давать адекватное описание природных явлений.

Различные модификации рассмотренной системы, изученные самим Вольтерра и другими авторами, лишены этих недостатков. Наиболее широко известные из них будут изучены в 5.6. Здесь остановимся на модели, которая учитывает самоограничение в росте обеих популяций. На ее примере видно, как может меняться характер решений при изменении параметров системы.

Итак, рассмотрим систему:

$$\begin{cases} \frac{dx}{dt} = x(\varepsilon_x - \gamma_{xy}y - \delta_x x^2); \\ \frac{dy}{dt} = y(-\varepsilon_y + \gamma_{yx}x - \delta_y y^2), \end{cases} \quad (5.26)$$

которая отличается от системы (5.25) наличием в правых частях членов:  $-\delta_x x^2$  и  $-\delta_y y^2$ .

Эти члены отражают тот факт, что численность популяции жертв не может увеличиваться до бесконечности даже в отсутствие хищников в силу ограниченности пищевых ресурсов, ареала существования и пр. Такие же «самоограничения» накладываются на популяцию хищников.

Система (5.26) имеет четыре стационарных решения, для трех из них либо численность обоих видов, либо численность одного из видов равна нулю. Анализ показывает, что нулевое решение представляет собой неустойчивый узел. Рассмотрим систему алгебраических уравнений, решение которых дает координаты стационарного состояния, ненулевого для численностей обоих видов. Эта система получается приравниванием нулю выражений в скобках в правых частях системы (5.26):

$$\begin{aligned}\delta_x \bar{x} + \gamma_{xy} \bar{y} &= \epsilon_x; \\ \gamma_{yx} \bar{x} - \delta_y \bar{y} &= \epsilon_y.\end{aligned}$$

Стационарное решение:

$$\bar{x} = \frac{\epsilon_x \delta_y + \epsilon_y \gamma_{xy}}{\delta_x \delta_y + \gamma_{xy} \gamma_{yx}}; \quad \bar{y} = \frac{\delta_x \epsilon_y - \gamma_{yx} \epsilon_x}{\delta_x \delta_y + \gamma_{xy} \gamma_{yx}}.$$

Анализ корней характеристического уравнения показывает, что стационарное решение в зависимости от соотношения коэффициентов может быть либо устойчивым фокусом, либо устойчивым узлом (рис. 5.13).

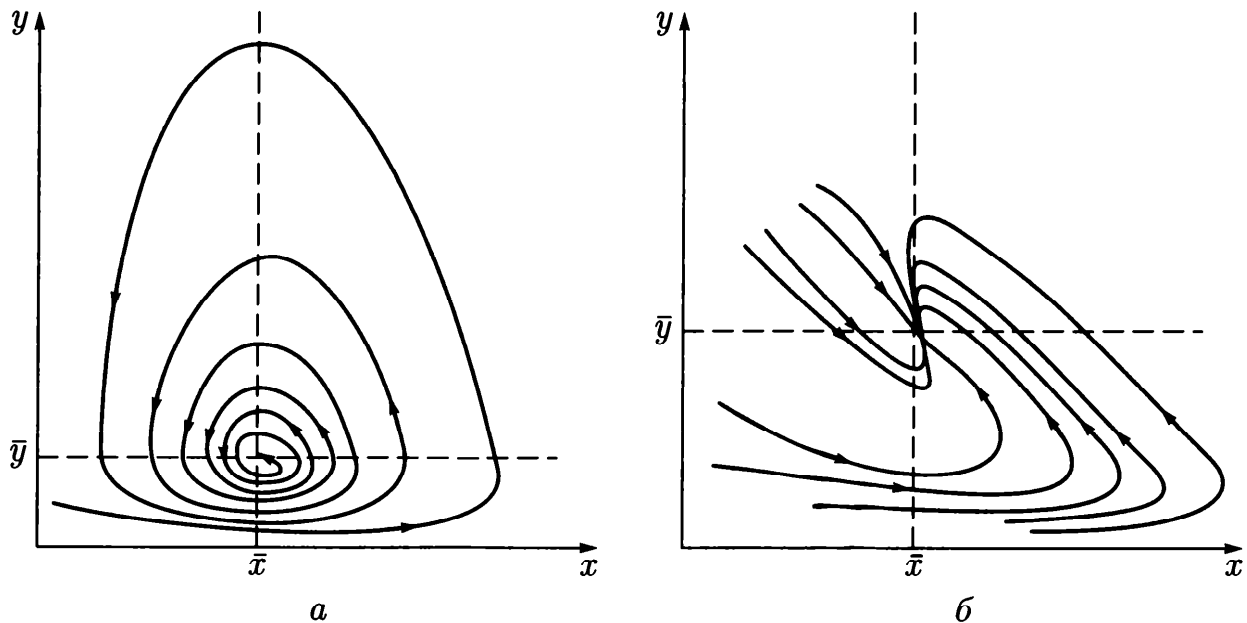


Рис. 5.13. Фазовый портрет системы (5.26):

*a* — устойчивый фокус, параметры системы:  $\epsilon_x = 2$ ,  $\gamma_{xy} = 18$ ,  $\delta_x = 1$ ,  $\epsilon_y = 3$ ,  $\gamma_{yx} = 5$ ,  $\delta_y = 1$ ; *б* — устойчивый узел, параметры системы:  $\epsilon_x = 2$ ,  $\gamma_{xy} = 1$ ,  $\delta_x = 1$ ,  $\epsilon_y = 3$ ,  $\gamma_{yx} = 1$ ,  $\delta_y = 1$

И в том и в другом случае решение устойчиво к малым изменениям правых частей уравнений. Таким образом, самоограничение популяции приводит к устойчивости ее численности.

Важно отметить, что рассмотренные простейшие вольтерровские модели не могут описывать устойчивые колебания с постоянным периодом и амплитудой. Для описания таких колебаний необходимы нелинейные модели, имеющие на фазовой плоскости предельный цикл (см. 5.6).

## **5.4. Проблема быстрых и медленных переменных**

Степень подробности моделирования изучаемых явлений зависит от цели моделирования. Однако в любом случае задача моделирования заключается в том, чтобы построить модель явления, содержащую, возможно, меньшее число переменных и произвольных параметров и в то же время правильно отражающую свойства явления.

Биологические системы включают большое число процессов с разными характерными временами, причем иерархия этих времен такова, что они различаются на много порядков.

### **5.4.1. Иерархия времен. Метод квазистационарных концентраций**

Учет временной иерархии процессов позволяет сократить число дифференциальных уравнений. «Совсем медленные» переменные не меняются на времени рассматриваемых процессов, и их можно считать постоянными параметрами. Для «быстрых» переменных можно вместо дифференциальных уравнений записать алгебраические уравнения для их стационарных значений, поскольку «быстрые» переменные достигают своих стационарных значений практически мгновенно по сравнению с «медленными».

Пусть имеется три группы переменных с различными характерными временами:

$$\frac{dx}{dt} = P(x, y, z);$$

$$\frac{dy}{dt} = Q(x, y, z);$$

$$\frac{dz}{dt} = F(x, y, z).$$

Переменные изменяются с разными характерными временами, причем

$$T_x \ll T_y \ll T_z.$$

Пусть мы наблюдаем за переменной  $y$ , характерное время изменения которой  $T_y$  («среднее» характерное время). Тогда за время  $T_y$  «совсем медленная» переменная  $z$  практически не будет изменяться и ее можно считать постоянным параметром, обозначим его  $z^*$ .

Система дифференциальных уравнений с учетом этого обстоятельства будет содержать два уравнения и может быть записана в виде

$$\begin{cases} \frac{dx}{dt} = P(x, y, z^*); \\ \frac{dy}{dt} = Q(x, y, z^*). \end{cases}$$

Отметим, что  $z^*$  не является истинно стационарным значением, «медленная» переменная  $z$  будет продолжать меняться и «вести» за собой более быстрые переменные  $x$  и  $y$ . В этом смысле медленная переменная является ведущей, или «параметром порядка».

Рассмотрим теперь уравнение для  $x$ . Эта «быстрая» переменная изменяется значительно быстрее, чем  $y$ , и за время  $T_y$  успеет достичь своего стационарного значения. Значит, для переменной  $x$  дифференциальное уравнение можно заменить алгебраическим:

$$P(x, y, z^*) = 0, \quad \text{или} \quad x = \bar{x}(y, z^*).$$

Таким образом, благодаря учету иерархии времен, исходную систему из трех дифференциальных уравнений удастся свести к одному дифференциальному уравнению для переменной  $y$ :

$$\frac{dy}{dt} = Q(\bar{x}(y, z^*), y, z^*),$$

В химической кинетике метод такой редукции системы был впервые предложен Боденштейном и носит название *метода квазистационарных концентраций* (КСК).

Обычно он применяется для систем химических реакций, промежуточные продукты которых являются частицами с высокой реакционной способностью. К ним относятся каталитические процессы, свободно радикальные и цепные реакции.

В процессах с участием активных промежуточных частиц разность скоростей образования  $v_o$  и расхода  $v_p$  частиц мала по сравнению с этими скоростями. Режим называется **квазистационарным режимом**, а отвечающие ему концентрации активных промежуточных веществ — **квазистационарными концентрациями**.

Дифференциальные уравнения для промежуточных соединений

$$\frac{dR_i}{dt} = v_o^i - v_p^i, i = 1, 2, \dots, l,$$

можно заменить алгебраическими:

$$v_o^i = v_p^i, i = 1, 2, \dots, l.$$

Из  $l$  алгебраических уравнений можно выразить  $l$  квазистационарных концентраций промежуточных химических соединений. По мере расходования исходных веществ квазистационарные концентрации промежуточных соединений будут меняться, но если время установления квазистационарного режима мало, он не будет нарушаться в течение всего процесса.

Конечно, такое рассмотрение не правомерно для начальных стадий процесса, когда  $R_i$  меняются от нуля до своих квазистационарных значений. Этот период носит название **периода индукции**.

Аналогичная ситуация имеет место в биохимических ферментативных процессах, где процессы образования и распада фермент-субстратного комплекса происходят значительно быстрее, чем процессы расходования субстрата и образования продукта.

#### 5.4.2. Теорема Тихонова

Математически строгое обоснование применения метода КСК (редукции системы в соответствии с иерархией времен) и формулировка условий его применимости дана в работе крупнейшего российского математика А. Н. Тихонова [А. Н. Тихонов, 1952].

Рассмотрим простейший случай двух дифференциальных уравнений

$$\begin{cases} \frac{dx}{dt} = \varphi(x, y); \\ \frac{dy}{dt} = G(x, y). \end{cases} \quad (5.27)$$



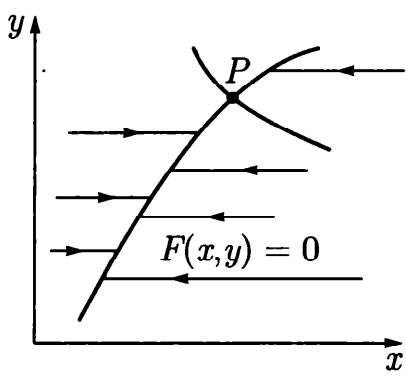


Рис. 5.14. Фазовый портрет вырожденной системы (5.28) ( $P$  — точка пересечения главных изоклин)

Пусть  $y$  — медленная, а  $x$  — быстрая переменная. Это означает, что отношение приращений  $\Delta y$  и  $\Delta x$  за короткий промежуток времени  $\Delta t$  много меньше единицы:

$$\Delta y / \Delta x \ll 1.$$

Скорость изменения  $x$  значительно превосходит скорость изменения  $y$ , поэтому правую часть первого уравнения можно записать в виде

$$\varphi(x, y) = AF(x, y), \text{ где } A \gg 1.$$

Первое уравнение системы представим в виде

$$\frac{dx}{dt} = AF(x, y).$$

Разделив левую и правую части уравнения на  $A$  и обозначив  $\varepsilon = 1/A$ , получим *полную* систему уравнений, тождественную исходной:

$$\begin{cases} \varepsilon \frac{dx}{dt} = F(x, y); \\ \frac{dy}{dt} = G(x, y), \end{cases}$$

где  $\varepsilon \ll 1$  — *малый параметр*.

Если характер решения не изменится при устремлении малого параметра  $\varepsilon$  к нулю (условия этого обстоятельства и составляют содержание теоремы Тихонова), можно устремить  $\varepsilon$  к нулю и получить для быстрой переменной  $x$  вместо дифференциального уравнения — алгебраическое:

$$\begin{cases} F(x, y) = 0; \\ \frac{dy}{dt} = G(x, y). \end{cases} \quad (5.28)$$

В отличие от полной системы (5.27) такая система называется *вырожденной* (ее фазовый портрет представлен на рис. 5.14).

Фазовые траектории в любой точке фазовой плоскости, за исключением  $\varepsilon$ -окрестности кривой  $F(x, y) = 0$ , имеют наклон, определяемый уравнением:  $\frac{dy}{dx} = \varepsilon \frac{G(x, y)}{F(x, y)}$  порядка величины  $\varepsilon \ll 1$ , т. е. расположены почти горизонтально. Это области быстрых движений, при которых вдоль фазовой траектории  $y = \text{const}$ , а  $x$  быстро меняется. Достигнув по одной из таких горизонталей  $\varepsilon$ -окрестности кривой  $F(x, y) = 0$ , изображающая точка потом будет двигаться по этой кривой.

Скорость движения по горизонтальным участкам траектории  $dx/dt \approx 1/\varepsilon = A$ , т. е. очень велика по сравнению со скоростью движения в окрестности кривой  $F(x, y) = 0$ . Поэтому общее время достижения некоего состояния на кривой  $F(x, y)$  определяется лишь характером движения вдоль этой кривой, т. е. зависит лишь от начальных значений медленной переменной  $y$  и не зависит от начальных значений быстрой переменной  $x$ .

Отметим, что квазистационарные значения быстрых переменных являются функциями не окончательных стационарных значений медленных переменных, а лишь их мгновенных значений. В этом смысле говорят о том, что быстрая переменная «подчинена» медленной.

Теорема Тихонова устанавливает условия возможности редукции системы дифференциальных уравнений с малым параметром (условия замены для быстрых переменных дифференциальных уравнений алгебраическими).

Запишем систему  $N$  уравнений, часть из которых содержит малый параметр  $\varepsilon$  перед производной:

$$\varepsilon \frac{dx_p}{dt} = F_p(x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_N), \quad p = 1, \dots, r; \quad (5.29)$$

$$\frac{dx_q}{dt} = F_q(x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_N) \quad q = r + 1, \dots, N. \quad (5.30)$$

Назовем систему (5.29) *присоединенной*, а систему (5.30) — *вырожденной*.

Решение *полной* системы (5.29), (5.30) стремится к решению вырожденной системы (5.30) при  $\varepsilon \rightarrow 0$ , если выполняются следующие условия:

а) решение полной и присоединенной системы единственно, а правые части непрерывны;

б) решение  $x_1 = \varphi_1(x_1, x_2, \dots, x_N), \dots, x_r = \varphi_r(x_1, x_2, \dots, x_N)$  представляет собой изолированный корень алгебраической системы

(в окрестности этого корня нет других корней);

в) решение  $x_1, x_2, \dots, x_r$  — устойчивая изолированная особая точка присоединенной системы (5.29) при всех значениях  $x_{r+1}, x_{r+2}, \dots, x_N$ ;

г) начальные условия  $x_1^0, x_2^0, \dots, x_r^0$  попадают в область влияния устойчивой особой точки присоединенной системы. Это означает, что траектории, начинающиеся в этой области, со временем приходят в особую точку присоединенной системы, которая для них является притягивающей точкой — аттрактором.

Число начальных условий вырожденной системы меньше, чем полной: начальные значения быстрых переменных не используются в вырожденной системе. Согласно теореме Тихонова, если выполняется условие «в», то результат не зависит от начальных условий для переменных присоединенной системы.

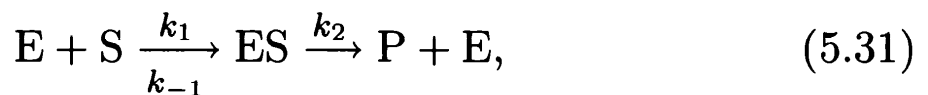
Таким образом, необходимым условием редукции является наличие малого параметра в уравнениях (5.29).

Теорема Тихонова явно или неявно применяется при исследовании практически любых моделей биологических систем, в этом мы убедимся в дальнейшем.

### 5.4.3. Фермент-субстратная реакция Михаэлиса — Ментен

Классическим примером базовой ферментативной реакции является модель, предложенная М. Михаэлисом и М. Ментен (1913)

Схема реакции может быть представлена в виде:



что означает следующее: субстрат  $S$  соединяется с ферментом  $E$  в комплекс  $ES$ , в котором происходит химическое превращение и который затем распадается на фермент  $E$  и продукт  $P$ . По *закону действующих масс* скорость реакции пропорциональна произведению концентраций. Обозначив концентрации реагентов малыми буквами:

$$s = [S], e = [E], c = [ES], p = [P],$$

получим систему уравнений:

$$\left\{ \begin{array}{l} \frac{ds}{dt} = -k_1 es + k_{-1}c; \\ \frac{de}{dt} = -k_1 es + (k_{-1} + k_2)c; \\ \frac{dc}{dt} = k_1 es - (k_{-1} + k_2)c; \\ \frac{dp}{dt} = k_2c. \end{array} \right. \quad (5.32)$$

В системе (5.32) учтены следующие процессы:

- субстрат S расходуется, образуя фермент-субстратный комплекс ES (бимолекулярная реакция с константой скорости  $k_1$ ), и его концентрация увеличивается при распаде комплекса в обратной реакции (константа  $k_{-1}$ );
- фермент E расходуется на образование комплекса ES ( $k_1$ ), его концентрация увеличивается при распаде комплекса в обратной реакции (константа  $k_{-1}$ ) и при образовании продукта ( $k_2$ );
- комплекс ES образуется из фермента E и субстрата S (бимолекулярная реакция с константой скорости  $k_1$ ), а также распадается на субстрат S и фермент E ( $k_{-1}$ ) или продукт P и фермент E ( $k_2$ );
- продукт P образуется при распаде комплекса (константа  $k_2$ ).

Для полной математической формулировки задачи Коши необходимо задать начальные условия:

$$s_0(0) = s_0, \quad e(0) = e_0, \quad c(0) = 0, \quad p(0) = 0.$$

Уравнения (5.32) не являются независимыми. Кроме того, последнее уравнение отделяется от первых трех. Если система первых трех уравнений решена, то концентрация продукта может быть рассчитана по формуле

$$p(t) = k_2 \int_0^t c(t') dt'.$$

В соответствии со схемой реакций (5.31) общее количество фермента, свободного и связанного в комплекс, сохраняется:

$$e(t) + c(t) = e_0.$$

Это условие позволяет одно из дифференциальных уравнений системы (5.32) заменить алгебраическим, и модель сводится к двум дифференциальным уравнениям:

$$\begin{cases} \frac{ds}{dt} = -k_1 e_0 s + (k_1 s + k_{-1}) c; \\ \frac{dc}{dt} = k_1 e_0 s - (k_1 s + k_{-1} + k_2) c \end{cases} \quad (5.33)$$

с начальными условиями:  $s_0(0) = s_0$ ,  $c(0) = 0$ .

Введем безразмерные переменные и параметры:

$$\begin{aligned} \tau = k_1 e_0 t, \quad x(\tau) = \frac{s(t)}{s_0}, \quad y(\tau) = \frac{c(t)}{e_0}, \\ \lambda = \frac{k_2}{k_1 s_0}, \quad K = \frac{k_{-1} + k_2}{k_1 s_0}, \quad \varepsilon = \frac{e_0}{s_0}. \end{aligned} \quad (5.34)$$

Запишем уравнения (5.33) в безразмерном виде:

$$\begin{aligned} \frac{dx}{d\tau} = -x + (x + K - \lambda)y, \quad \varepsilon \frac{dy}{d\tau} = x - (x + K)y, \\ x(0) = 1, \quad y(0) = 0. \end{aligned} \quad (5.35)$$

Из (5.34) следует, что  $(K - \lambda) > 0$ .

Поскольку реакция превращения фермент-субстратного комплекса необратима, уже из схемы реакций (5.31) ясно, что с течением времени весь субстрат будет превращен в продукт и в стационарном состоянии концентрации субстрата и комплекса станут равны нулю:  $x = 0$ ,  $y = 0$ .

Систему (5.35) нельзя решить аналитически. Проанализируем качественно, как ведут себя  $x(t)$  и  $y(t)$ .

Вблизи  $\tau = 0$   $dx/d\tau < 0$ . Это означает, что  $x$  уменьшается от  $x = 1$ , расчеты показывают, что величина  $x$  (безразмерная концентрация субстрата) уменьшается монотонно, при  $t \rightarrow \infty$  весь субстрат будет исчерпан. В то же время  $dy/d\tau > 0$ ,  $y$  растет от  $y = 0$  до величины  $y = x/(x + K)$ , при которой правая часть уравнения для  $dy/d\tau$  обращается в нуль. После этого  $y$  будет уменьшаться до нуля. Таким образом, концентрация фермент-субстратного комплекса  $y$  проходит через максимум. Относительная концентрация свободного фермента  $e/e_0$  сначала убывает, а затем снова возрастает до величины  $e/e_0 = 1$ , поскольку с течением времени субстрат исчерпывается и все меньшая доля фермента оказывается связанной. Кинетические кривые изображены на рис. 5.15.

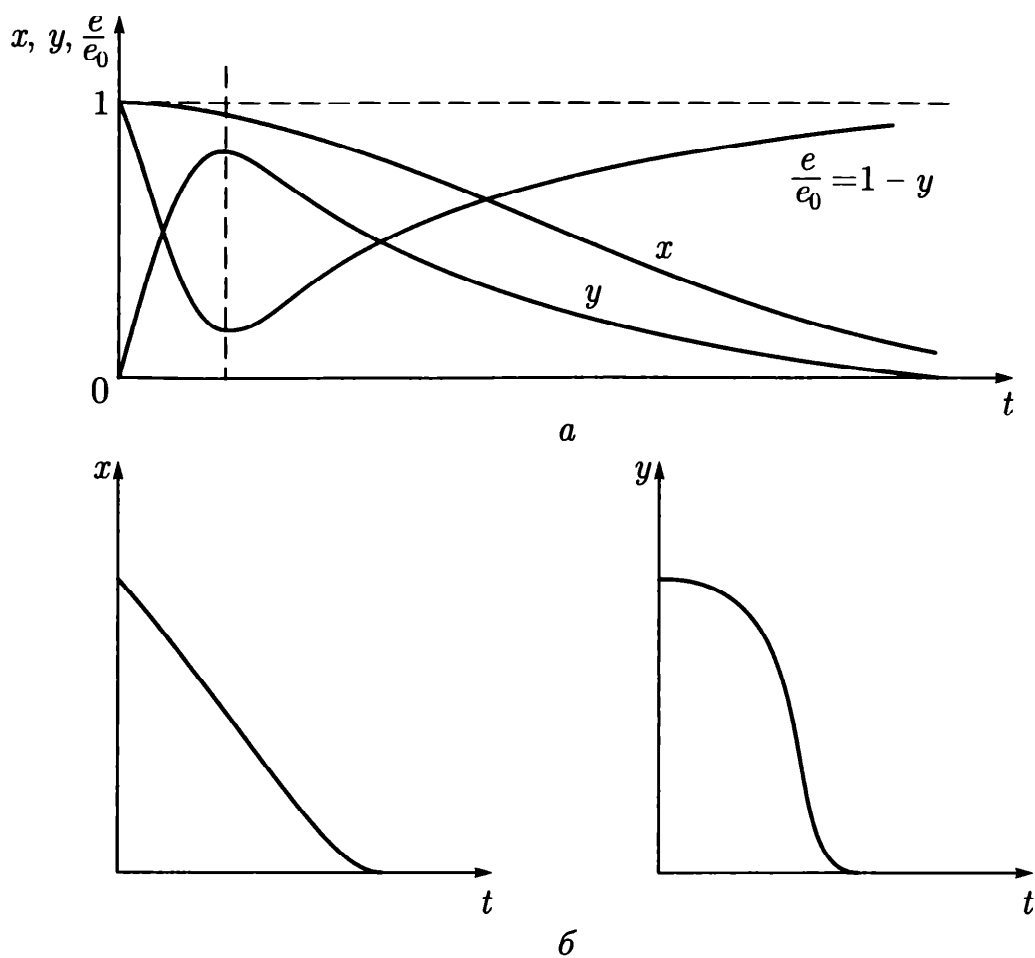


Рис. 5.15. Кинетика изменения безразмерных переменных в уравнении Михаэлиса — Ментен:

*a* — с учетом области переходных процессов на малых временах (полная система (5.35)). Вертикальной штриховой линией отделен «начальный период индукции»; *б* — без учета области переходных процессов (редуцированная система (5.36))

Предположим, что концентрация субстрата значительно превышает концентрацию фермента:  $s_0 \gg e_0$ . Тогда из соотношений (5.34) следует, что  $\epsilon \ll 1$ . Если условия теоремы Тихонова выполняются (для уравнений Михаэлиса — Ментен это можно показать), мы имеем право заменить второе из уравнений (5.35) алгебраическим и найти «квазистационарную концентрацию» фермент-субстратного комплекса:

$$\bar{y}^* = \frac{x^*}{x^* + K}.$$

По терминологии Тихонова мы получим вырожденную систему:

$$\begin{cases} \frac{dx}{d\tau} = -x + (x + K - \lambda)y; \\ y = \frac{x}{x + K}, x(0) = 1. \end{cases} \quad (5.36)$$

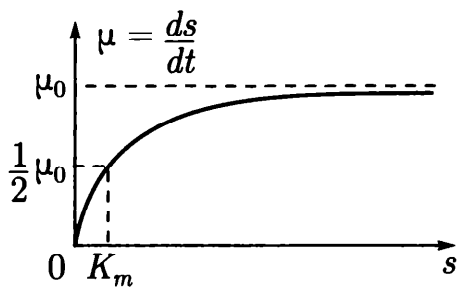


Рис. 5.16. Иллюстрация закона Михаэлиса — Ментен. Зависимость скорости реакции  $\mu$  от начальной концентрации субстрата  $s$ :  $\mu_0$  — максимальная скорость;  $K_m$  — константа Михаэлиса

Подставив выражение для  $y$  в дифференциальное уравнение для  $x$ , получим

$$\frac{dx}{d\tau} = -x + (x + K - \lambda) \frac{x}{x + K}, \quad \text{или}$$

$$\frac{dx}{d\tau} = -\frac{\lambda x}{x + K},$$

$$x(0) = 1.$$

В размерном виде — это классическая формула Михаэлиса — Ментен для кинетики изменения субстрата в ферментативной реакции:

$$\frac{ds}{dt} = -\frac{\mu_0 s}{K_m + s}. \quad (5.37)$$

Таким образом, уравнение (5.37) верно описывает изменение концентрации субстрата, но не характеризует изменение концентраций свободного фермента и фермент-субстратного комплекса, которые на малых временах ведут себя немонотонно (см. рис. 5.15, а).

Величина  $K_m$  называется константой Михаэлиса и имеет размерность концентрации. Она соответствует концентрации субстрата, при которой скорость  $\mu(S)$  равна половине максимальной. При  $s \ll K_m$  скорость пропорциональна концентрации:  $-\mu_0 s / K_m$ . Максимальная скорость ферментативной реакции  $\mu_0$  равна  $k_2 e_0$ , т. е. линейно зависит от константы скорости стадии распада ферментативного комплекса  $k_2$ , которую называют *лимитирующей стадией*.

В эксперименте для оценки параметров ферментативной реакции используют кривую зависимости скорости реакции от концентрации субстрата (рис. 5.16). В ферментативных реакциях возможны гораздо более сложные типы динамического поведения: два или несколько устойчивых стационарных состояния, автоколебания, квазистохастические режимы. Эти типы поведения будут рассмотрены далее.

## 5.5. Мультистационарность

Важная особенность биологических систем — переключение из одного режима функционирования в другой. Системы, обладающие двумя или несколькими устойчивыми стационарными состояниями, называются *мультистационарными*, или *триггерными*, системами.

### 5.5.1. Фазовый портрет мультистационарной системы

На фазовой плоскости триггерной системе в простейшем случае соответствуют два устойчивых стационарных решения, разделенных сепаратрисой. Напомним, что особые точки (устойчивые и седло) лежат на пересечении главных изоклин — изоклин вертикальных и горизонтальных касательных.

На рис. 5.17 представлен фазовый портрет относительно простой триггерной системы, описывающей явление конкуренции двух равноправных видов.

Соответствующая система уравнений имеет вид:

$$\begin{cases} \frac{dx_1}{dt} = x_1 - x_1x_2 - ax_1^2; \\ \frac{dx_2}{dt} = x_2 - x_1x_2 - ax_2^2. \end{cases} \quad (5.38)$$

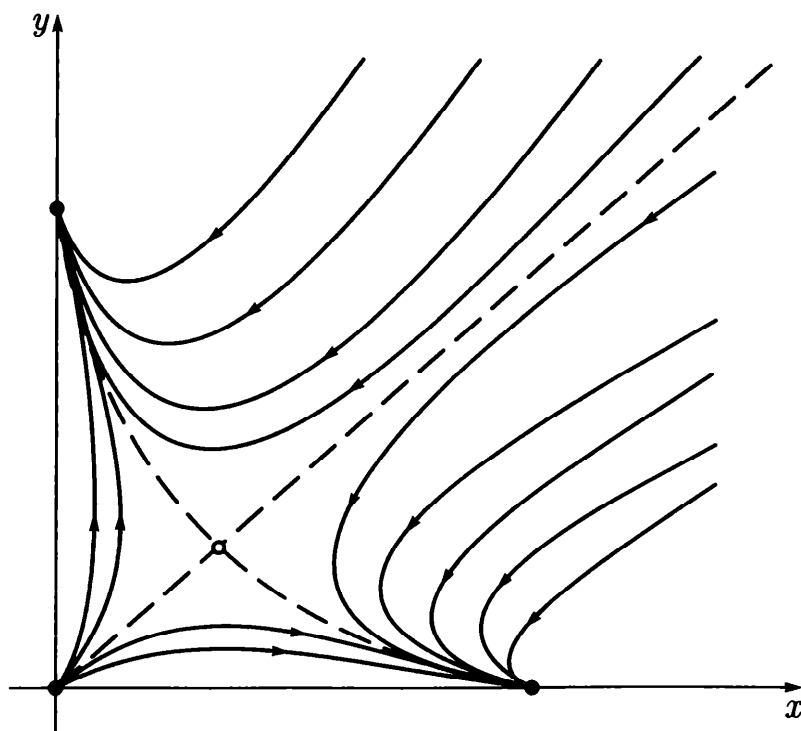


Рис. 5.17. Фазовый портрет триггерной системы (5.38), описывающей явление конкуренции между двумя равноправными видами;  $x, y$  — численности конкурирующих видов



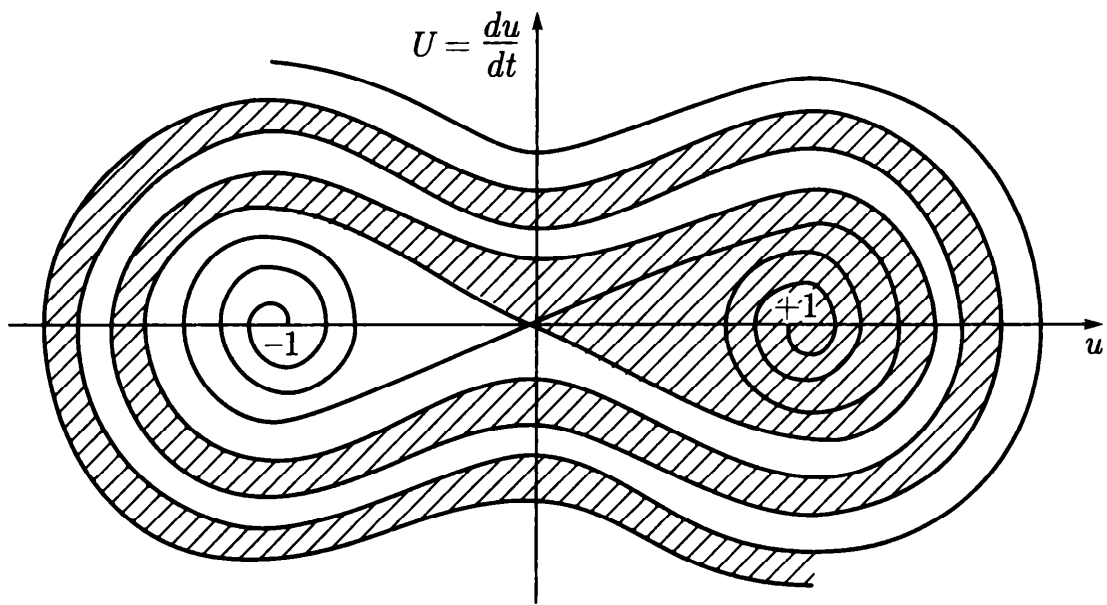


Рис. 5.18. Фазовый портрет системы (5.39) (шарик в ложбине с двумя лунками). Штриховкой обозначена область притяжения стационарного состояния (+1)

Такая система имеет четыре стационарных решения:

- 1)  $x_1 = x_2 = 0$  — неустойчивый узел;
- 2)  $x_1 = x_2 = \frac{1}{1+a}$  — седло;
- 3)  $x_1 = \frac{1}{a}, x_2 = 0$  — устойчивый узел;
- 4)  $x_1 = 0, x_2 = \frac{1}{a}$  — устойчивый узел.

Бистабильная система может иметь более сложную структуру фазового портрета. Пример такой системы — движение шарика в ложбине с двумя лунками в присутствии трения [Д. С. Чернавский, 2004].

На рис. 5.18 дана система, описываемая уравнениями:

$$\begin{cases} \frac{dx}{dt} = y; \\ \frac{dy}{dt} = -ay + b(x - x^3). \end{cases} \quad (5.39)$$

В такой системе три стационарных состояния:

- 1)  $x = y = 0$  — седло; 2, 3) — устойчивые фокусы.

Вблизи этих стационарных состояний траектории представляют собой закручивающиеся спирали. Вдали от стационарных состояний области притяжения имеют слоистую структуру. Толщина слоев уменьшается при уменьшении параметра  $a$ .

Как видно из приведенных примеров, в триггерных системах и поведение во времени, и стационарное решение зависят не только от параметров, но и от начальных условий.

### 5.5.2. Типы переключения триггера

Слово «*триггер*» означает переключатель. Возникает вопрос: «Как можно переключить триггер из одного стационарного состояния в другое?»

Рассмотрим фазовый портрет системы, обладающей двумя устойчивыми стационарными состояниями (рис. 5.19). Здесь  $a$ ,  $c$  — устойчивые стационарные состояния;  $b$  — седло.

Если начальное положение изображающей точки размещено левее сепаратрисы седла (штриховая линия), то система находится в области притяжения особой точки  $a$  и со временем стремится к этому устойчивому стационарному состоянию. Из точек, лежащих правее сепаратрисы, система будет двигаться к особой точке  $c$ . Рассмотрим возможные способы переключения системы из режима  $a$  в режим  $c$ .

**Силовое переключение.** Можно изменить значения концентраций, например добавить определенное количество вещества  $x$ , так что система «перепрыгнет» через сепаратрису в некоторую точку  $c_1$ , которая находится по правую сторону сепаратрисы в области влияния устойчивого стационарного состояния  $c$ , к которому система перейдет сама с течением времени. На фазовом портрете (см. рис. 5.19) силовое (специфическое) переключение показано стрелкой. Кинетика переменных при таком переключении представлена на рис. 5.20.

**Параметрическое переключение.** Другой (неспецифический) способ переключения показан на рис. 5.21. При таком способе непосредственному воздействию подвергаются не переменные, а параметры системы. Это может быть достигнуто разными способами, например изменением скорости поступления субстрата, температуры, pH.

Сущность такого способа переключения состоит в использовании зависимости фазового портрета системы от некоторого управляющего параметра. Пусть с изменением этого параметра

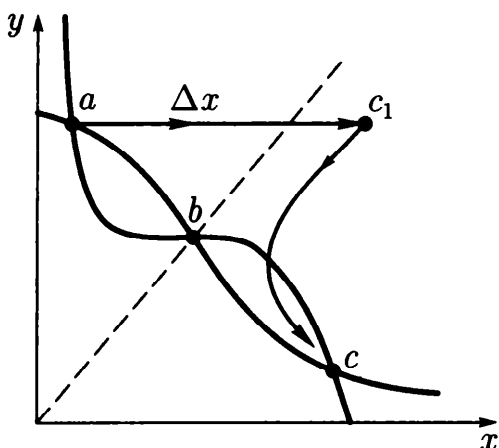
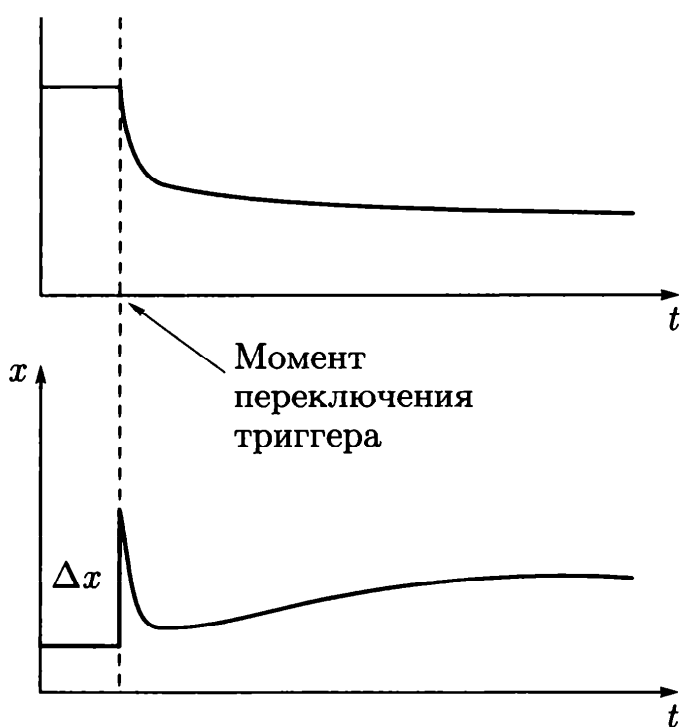


Рис. 5.19. Триггерная система (сплошными линиями показаны главные изоклины; штриховой линией обозначена сепаратриса, отделяющая области влияния двух устойчивых стационарных состояний  $a$  и  $c$ ; линией со стрелкой указан процесс силового переключения триггера)



ных во времени при силовом переключении триггера после добавления в систему вещества  $x$  в количестве, достаточном для переключения системы из режима  $a$  в режим  $c$  (см. рис. 5.19)

фазовый портрет претерпевает последовательность превращений, показанных на рис. 5.21 ( $a - z$ ). На стадии (см. рис. 5.21,  $в$ ) устойчивый узел (см. рис. 5.21,  $a$ ) и седло (см. рис. 5.21,  $б$ ) сливаются в одну полуустойчивую точку седло — узел. На стадии (см. рис. 5.21,  $г$ ) в системе остается лишь одно устойчивое стационарное состояние, к которому и сходятся все фазовые траектории.

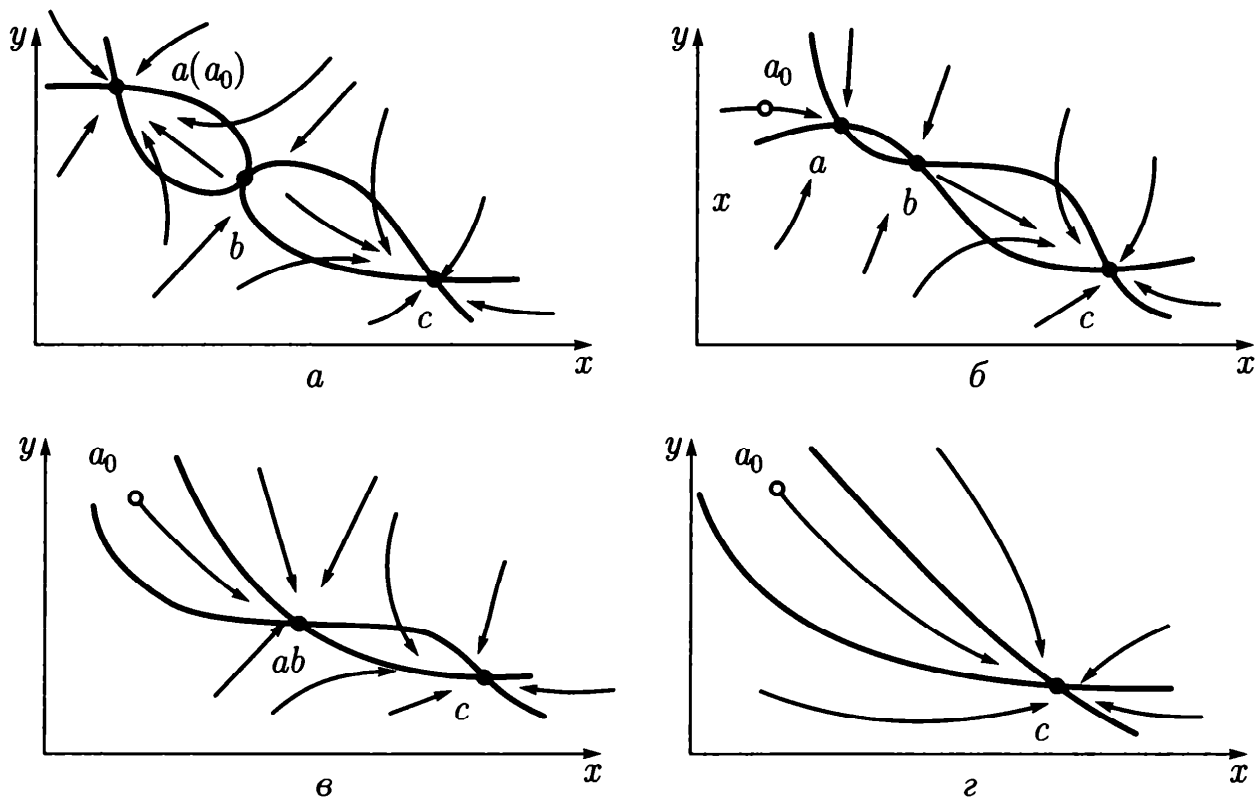


Рис. 5.21. Параметрическое переключение триггера. Последовательные стадии ( $a - z$ ) трансформации фазового портрета (стрелками обозначено направление фазовых траекторий)

Тогда система, находившаяся в начале процесса переключения в стационарном режиме (см. рис. 5.21, *a*), в результате параметрического переключения окажется в области притяжения единственного устойчивого стационарного режима *c*, куда с течением времени и перейдет.

Параметрический способ переключения реализуется при изменении любой генетической программы, он может также иметь место при изменении внешних условий, приводящих к изменению управляющего параметра системы.

Как было показано, в триггерной системе изображающая точка «выбирает» (в зависимости от параметров и начальных условий) стационарный режим функционирования. Триггерные модели могут быть использованы при описании процесса отбора и потому применимы к процессам эволюции.

### 5.5.3. Отбор одного из равноправных видов

Общая модель отбора одного из нескольких равноправных видов можно записать в следующей форме:

$$\frac{dx_i}{dt} = aX_i - \gamma \sum_{j=1, j \neq i}^N X_i X_j, \quad i = 1, 2, \dots, N,$$

где *a* — эффективный коэффициент репродукции;  $\gamma$  — вероятность гибели в результате встречи.

Пусть  $N = 2$ ,  $X_1 = x$ ,  $X_2 = y$ . Система уравнений имеет вид:

$$\begin{cases} \frac{dx}{dt} = ax - \gamma xy; \\ \frac{dy}{dt} = ay - \gamma xy. \end{cases} \quad (5.40)$$

Стационарные решения находят из алгебраических уравнений, полученных приравниванием правых частей нулю:

$$ax - \gamma xy = 0; \quad ay - \gamma xy = 0.$$

Система имеет два стационарных решения:

- 1)  $x = 0, y = 0$ ;
- 2)  $x = \frac{a}{\gamma}, y = \frac{a}{\gamma}$ .

Характеристическое уравнение линеаризованной системы имеет вид

$$\begin{vmatrix} a - \gamma y - \lambda & -\gamma x \\ -\gamma y & a - \gamma x - \lambda \end{vmatrix} = 0.$$

Выражения для характеристических чисел определяют из уравнения

$$\lambda^2 - \lambda(2a - \gamma x - \gamma y) + (a - \gamma y)(a - \gamma x) - \gamma^2 xy = 0. \quad (5.41)$$

Для исследования устойчивости симметричного стационарного состояния 2 подставим в это уравнение значения  $x = a/\gamma$ ,  $y = a/\gamma$  и получим

$$\lambda^2 - a^2 = 0;$$

$$\lambda_{1,2} = \pm a.$$

Корни действительны и разных знаков. Это означает, что симметричное стационарное состояние представляет собой седло.

При подстановке в (5.41) нулевых значений для  $x$ ,  $y$  (стационарное состояние 1) получим

$$\lambda^2 - 2a\lambda + a^2 = 0;$$

$$\lambda_{1,2} = a.$$

Это означает, что нулевая особая точка представляет собой неустойчивый узел (два положительных действительных корня).

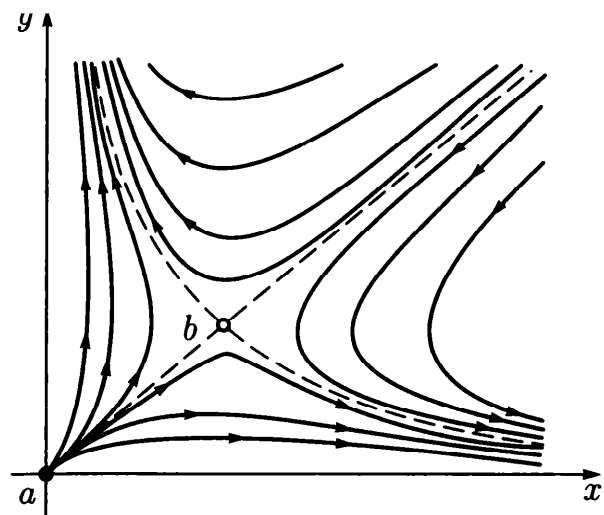
Фазовый портрет системы изображен на рис. 5.22. Изоклины горизонтальных касательных:  $y = 0$  — ось абсцисс и вертикальная прямая  $x = a/\gamma$ ; изоклины вертикальных касательных:  $x = 0$  — ось ординат и горизонтальная прямая  $y = a/\gamma$ .

Все траектории уходят на бесконечность, так как самоограничение роста популяции в данной модели не учитывается.

Модель (5.40) демонстрирует принципиальную возможность отбора в системе равноправных видов, где симметричное состояние сосуществования является неустойчивым. Вот один из примеров такой системы.

Известно, что сахара и аминокислоты являются оптически активными соединениями, причем сахара — левовращающие плоскость поляризации света, аминокислоты — правовращающие. Противоположные изомеры не только не встречаются в

Рис. 5.22. Фазовый портрет системы (5.40), описывающий отбор одного из двух равноправных видов в отсутствие ограничений роста:  $a$  (начало координат) — неустойчивый узел;  $b$  — седло



живых организмах и не усваиваются ими, но являются ядами. В этом заключается одна из сложностей искусственного синтеза.

Ясно, что «зеркальные» организмы не лучше и не хуже. В неживой природе распространены рацемические смеси, содержащие равное количество зеркальных изомеров. Такая же рацемическая смесь получается при небиологическом синтезе. По-видимому, и первичный бульон был рацемической смесью.

Рассмотренная модель описывает выживание одних видов и уничтожение других. Условие, которое обеспечивает при этом отбор одного вида, заключается в том, что при встрече они взаимно отравляются и гибнут. Причина отбора здесь — не преимущество одного из видов, а их взаимный антагонизм.

Однако модель (5.40) не может характеризовать реальную систему, так как описывает неограниченный рост биомассы с течением времени. Этот недостаток может быть исправлен несколькими способами. Один из них — введение самоограничения численности вида в форме ферхюльстовских членов. Тогда приходим к модели (5.38). Другой способ — ввести в модель переменную, описывающую поступающий в систему с определенной скоростью питательный ресурс, общий для обоих видов.

Учтем ограниченность питательных ресурсов. Пусть  $s$  — лимитирующий субстрат (световая энергия, минеральное питание и т. п.). Сам субстрат не является оптически активным, но преобразуется в оптически активные продукты.

Выразим скорость роста каждой популяции  $a$  через  $s$  в соответствии с формулой Моно:

$$a = \frac{a_0 s}{K_s + s}.$$

График этой функции приведен на рис. 5.16.

Пусть  $v$  — интенсивность притока субстрата. Расход субстрата пропорционален поглощению его организмами, т. е. сумме их концентраций. Уравнение для скорости изменения концентрации субстрата во времени имеет вид:

$$\frac{ds}{dt} = -\alpha a_0 \frac{s}{K_s + s} (X + Y) + v.$$

Здесь  $\alpha$  — экономический коэффициент, указывающий, сколько субстрата идет на образование единицы биомассы. Обычно  $\alpha > 10$ , и поэтому скорость поглощения субстрата можно считать значительно большей, чем скорость прироста биомассы.

Уравнения для концентраций объектов типа  $x$  и  $y$  принимают вид:

$$\frac{dX}{dt} = a_0 \frac{s}{K_s + s} X - \beta X - \gamma XY;$$

$$\frac{dY}{dt} = a_0 \frac{s}{K_s + s} Y - \beta Y - \gamma XY.$$

Введем безразмерные переменные:

$$t' = \beta t; \quad x = \frac{\gamma X}{\beta}; \quad y = \frac{\gamma Y}{\beta};$$

$$z = \frac{\gamma s}{\beta}; \quad v' = \frac{\gamma v}{\beta^2};$$

$$f(z) = \frac{a_0 z}{K_z + z}; \quad K_z = \frac{\gamma K_s}{\beta}.$$

Опустив штрихи у  $t'$ ,  $v'$ , запишем систему в безразмерном виде:

$$\begin{cases} \frac{dx}{dt} = f(z)x - x - xy; \\ \frac{dy}{dt} = f(z)y - y - xy; \\ \frac{dz}{dt} = -\frac{\alpha}{\beta} f(z)(x + y) + v'. \end{cases}$$

Допустим, что процессы поглощения субстрата (третье уравнение) существенно более быстрые, чем процессы репродукции (первое и второе уравнения). В этом случае может быть использован метод квазистационарных концентраций (см. 5.4.1) и дифференциальное уравнение для быстрой переменной  $z$  — концентрации субстрата — заменено алгебраическим. Тогда субстрат

на интересующих нас временах достигнет квазистационарной концентрации:  $dz/dt = 0$ .

Отсюда

$$f(z) = \frac{\nu}{\alpha(x+y)} = \frac{\nu_0}{x+y}.$$

Здесь  $\nu_0 = \nu/\alpha$ .

В итоге получается система двух безразмерных уравнений:

$$\begin{cases} \frac{dx}{dt} = x \left[ \frac{\nu_0}{x+y} - (1+y) \right]; \\ \frac{dy}{dt} = y \left[ \frac{\nu_0}{x+y} - (1+x) \right]. \end{cases} \quad (5.42)$$

Построим фазовый портрет системы (рис. 5.23). Изоклины вертикальных касательных:  $x = 0$  (ось ординат) и гипербола  $x = \frac{\nu_0}{1+y} - y$ ; изоклины горизонтальных касательных:  $y = 0$  (ось абсцисс) и гипербола  $y = \frac{\nu_0}{1+x} - x$ .

Переменные  $x$  и  $y$  симметричны, поэтому изоклина вертикальных касательных симметрична изоклине горизонтальных касательных относительно биссектрисы первого координатного угла.

Система имеет четыре особые точки:

- 1)  $x = 0, y = 0$  — неустойчивый узел;
- 2)  $x = 0, y = \nu_0$  — устойчивый узел;
- 3)  $x = \nu_0, y = 0$  — устойчивый узел;
- 4)  $x = y = \frac{\sqrt{1+2\nu_0}-1}{2}$  — симметричную точку — седло.

В такой системе выживет один из видов:  $x$  или  $y$ . Его стационарная концентрация определяется скоростью притока субстрата-

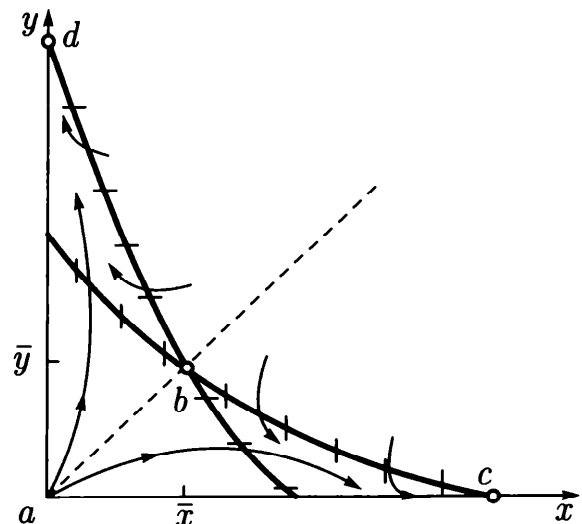


Рис. 5.23. Фазовый портрет системы (5.42), описывающий отбор одного из двух равноправных видов, когда субстрат поступает в систему с постоянной скоростью:  $a$  (начало координат) — неустойчивый узел;  $b$  — седло;  $c, d$  — устойчивые узлы (штриховой линией обозначена сепаратриса седла)



та и экономическим коэффициентом  $\alpha$ . Как и в системе (5.40), здесь причина отбора — неустойчивость симметричного состояния.

## 5.6. Колебания в биологических системах

Биологические системы отличаются периодическим изменением различных характеристик. Период этих колебаний может быть связан с изменениями условий жизни на Земле (смена времен года, дня и ночи), геофизических ритмов (солнечные, лунные, связанные с периодами атмосферных явлений). Но многие периодические процессы имеют частоту изменения, не связанную очевидным образом с внешними геокосмическими циклами. Это так называемые «биологические часы» различной природы, начиная от колебаний биомакромолекул, биохимических колебаний вплоть до популяционных волн.

Внутриклеточные колебания задают эндогенные биологические ритмы, которые свойственны всем живым системам. Именно они определяют периодичность деления клеток, отмеряют время рождения и смерти живых организмов. Модели колебательных систем используются в ферментативном катализе, теории иммунитета, в теории трансмембранного ионного переноса, микробиологии и биотехнологии.

### 5.6.1. Автоколебания и предельные циклы

Некоторые типы периодических движений рассматривались при изучении особых точек типа «центр» и затухающих или нарастающих колебаний в случае устойчивого и неустойчивого фокуса. Однако «биологические часы» имеют свойство, отличающее их от рассмотренных типов колебаний — неизменность во времени периода и амплитуды таких колебаний, означающую стационарность и устойчивость колебательного режима.

В данном случае периодическое изменение величин представляет собой один из типов стационарного поведения системы. Если колебания в системе имеют постоянные период и амплитуду, устанавливаются независимо от начальных условий и поддерживаются благодаря свойствам самой системы, а не вследствие воздействия периодической силы, то система называется *автоколебательной*.

К классу автоколебательных систем относятся колебания в гликолизе и других метаболических системах, периодические

процессы фотосинтеза, колебания концентрации кальция в клетке, колебания численности животных в популяциях.

В фазовом пространстве такому типу поведения соответствует притягивающее множество (аттрактор), называемое *предельным циклом*.

Предельный цикл есть изолированная замкнутая кривая на фазовой плоскости. Если предельный цикл устойчив, то при  $t \rightarrow \infty$  к нему стремятся все фазовые траектории, начинающиеся в некоторой  $\varepsilon$ -окрестности этого предельного цикла. Предельный цикл представляет стационарный режим с определенной амплитудой, не зависящий от начальных условий, а определяющийся только организацией системы. Существование предельного цикла на фазовой плоскости — основной признак автоколебательной системы. Очевидно, что при автоколебательном процессе фаза колебаний может быть любой.

Остановимся на общих характеристиках автоколебательных систем. Рассмотрим систему уравнений общего вида:

$$\begin{cases} \frac{dx}{dt} = P(x, y); \\ \frac{dy}{dt} = Q(x, y). \end{cases}$$

Если  $T (T > 0)$  — наименьшее число, для которого при всяком  $t$

$$x(t + T) = x(t); \quad y(t + T) = y(t),$$

то изменение переменных  $x = x(t)$ ,  $y = y(t)$  называется периодическим изменением с периодом  $T$ .

Периодическому изменению соответствует замкнутая траектория на фазовой плоскости, и обратно: всякой замкнутой траектории соответствует бесконечное множество периодических изменений, отличающихся друг от друга выбором начала отсчета времени.

Если периодическому изменению на фазовой плоскости соответствует изолированная замкнутая кривая, к которой с внешней и внутренней стороны приближаются (при возрастании  $t$ ) соседние траектории по спиральям, то эта изолированная замкнутая траектория есть предельный цикл.

Простые примеры позволяют убедиться, что система общего вида (5.1) допускает в качестве траекторий предельные циклы. Например, для системы

$$\begin{cases} \frac{dx}{dt} = y + x[1 - (x^2 + y^2)]; \\ \frac{dy}{dt} = -x + y[1 - (x^2 + y^2)] \end{cases} \quad (5.43)$$

траектория  $x^2 + y^2 = 1$  является предельным циклом. Его параметрические уравнения будут:

$$x = \cos(t - t_1); \quad y = \sin(t - t_1),$$

а уравнения всех других фазовых траекторий запишутся в виде

$$x = \frac{\cos(t - t_0)}{\sqrt{1 + Ce^{-2(t-t_0)}}}; \quad y = \frac{\sin(t - t_0)}{\sqrt{1 + Ce^{-2(t-t_0)}}}.$$

Значениям постоянной интегрирования  $C > 0$  соответствуют фазовые траектории, накручивающиеся на предельный цикл изнутри (при  $t \rightarrow +\infty$ ), а значениям  $-1 < C < 0$  — траектории, накручивающиеся снаружи.

### 5.6.2. Устойчивость предельных циклов

Предельный цикл называется *устойчивым*, если существует область на фазовой плоскости, содержащая этот предельный цикл, —  $\varepsilon$ -окрестность, такая, что все фазовые траектории, начинающиеся в  $\varepsilon$ -окрестности, асимптотически при  $t \rightarrow \infty$  приближаются к предельному циклу.

Если же, наоборот, в любой сколь угодно малой  $\varepsilon$ -окрестности предельного цикла существует по крайней мере одна фазовая траектория, не приближающаяся к предельному циклу при  $t \rightarrow +\infty$ , то такой предельный цикл называется *неустойчивым*. Такие циклы разделяют области влияния (бассейны) разных притягивающих множеств.

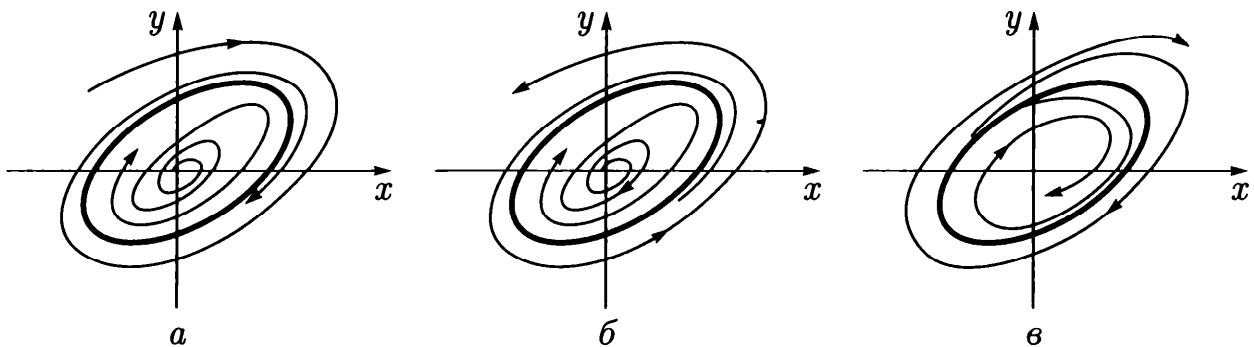


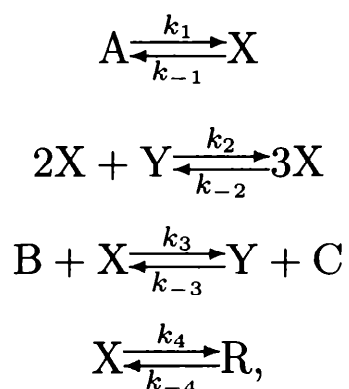
Рис. 5.24. Устойчивый (а) и неустойчивые (б и в) предельные циклы на фазовой плоскости

На рис. 5.24 изображены устойчивый (а) и неустойчивые (б) и (в) предельные циклы. Неустойчивые предельные циклы, подобные изображенному на рис. 5.24, б, такие, что все траектории с одной стороны (например, изнутри) приближаются к ним, а с другой (например, извне) — удаляются от них при  $t \rightarrow \infty$ , называют «полуустойчивыми», или двойными. Последнее название связано с тем, что обычно такие циклы при подходящем изменении параметра системы расщепляются на два, один из которых устойчив, а другой — неустойчив.

### 5.6.3. Брюсселятор

Простейшим классическим примером существования автоколебаний в системе химических реакций является тримолекулярная модель «брюсселятор», предложенная в Брюсселе И. Р. Пригожиным и Р. Лефевром (1965). Основной целью при изучении этой модели было установление качественных типов поведения, совместимых с фундаментальными законами химической и биологической кинетики. В этом смысле брюсселятор играет роль базовой модели, такую же как гармонический осциллятор в физике, или модели Вольтерра в динамике популяций. Рассмотрим свойства брюсселятора как автоколебательной системы.

Брюсселятор представляет собой следующую схему гипотетических химических реакций:



где А, В — исходные вещества; С, R — продукты; X, Y — промежуточные вещества.

Пусть конечные продукты С и R немедленно удаляются из реакционного пространства. Это означает, что обратные константы  $k_{-3} = k_{-4} = 0$ . Если субстрат А находится в избытке,  $k_{-1} = 0$ . Предположим также, что  $k_{-2} = 0$ . Значения остальных констант положим равными единице. Тогда схема рассмотренных реакций (в случае точечной системы) описывается системой уравнений:

$$\begin{cases} \frac{dx}{dt} = A + x^2y - (B + 1)x; \\ \frac{dy}{dt} = Bx - x^2y. \end{cases} \quad (5.44)$$

Модель (5.44) имеет одну особую точку с координатами

$$\bar{x} = A, \quad \bar{y} = \frac{B}{A}. \quad (5.45)$$

Исследуем стационарное решение (5.45) системы (5.44) на устойчивость по методу Ляпунова. Введем переменные, характеризующие отклонения от особой точки:

$$\xi = x - \bar{x}, \eta = y - \bar{y}.$$

Линеаризованная система имеет вид

$$\begin{cases} \frac{d\xi}{dt} = (B - 1)\xi + A^2\eta; \\ \frac{d\eta}{dt} = -B\xi - A^2\eta. \end{cases}$$

Характеристическое уравнение

$$\begin{vmatrix} B - 1 - \lambda & A^2 \\ -B & -A^2 - \lambda \end{vmatrix} = 0,$$

или

$$\lambda^2 + (A^2 + 1 - B)\lambda + A^2 = 0,$$

имеет корни

$$\lambda_{1,2} = -\frac{1}{2}(A^2 + 1 - B) \pm \frac{1}{2}\sqrt{(A^2 + 1 - B)^2 - 4A^2}. \quad (5.46)$$

Напомним, что особая точка является устойчивой, если действительные части корней характеристического уравнения отрицательны. Из выражения (5.46) видно, что при  $B < 1 + A^2$  особая точка (5.45) устойчива. Если  $B > 1 + A^2$ , то особая точка становится неустойчивой, и у системы (5.44) появляется устойчивый предельный цикл. Значение  $B = 1 + A^2$  является бифуркационным. Если оно лишь немного превосходит бифуркационный порог, то автоколебания в системе носят квазигармонический характер. Таким образом, брюсселятор при выполнении условия

$$B > 1 + A^2$$

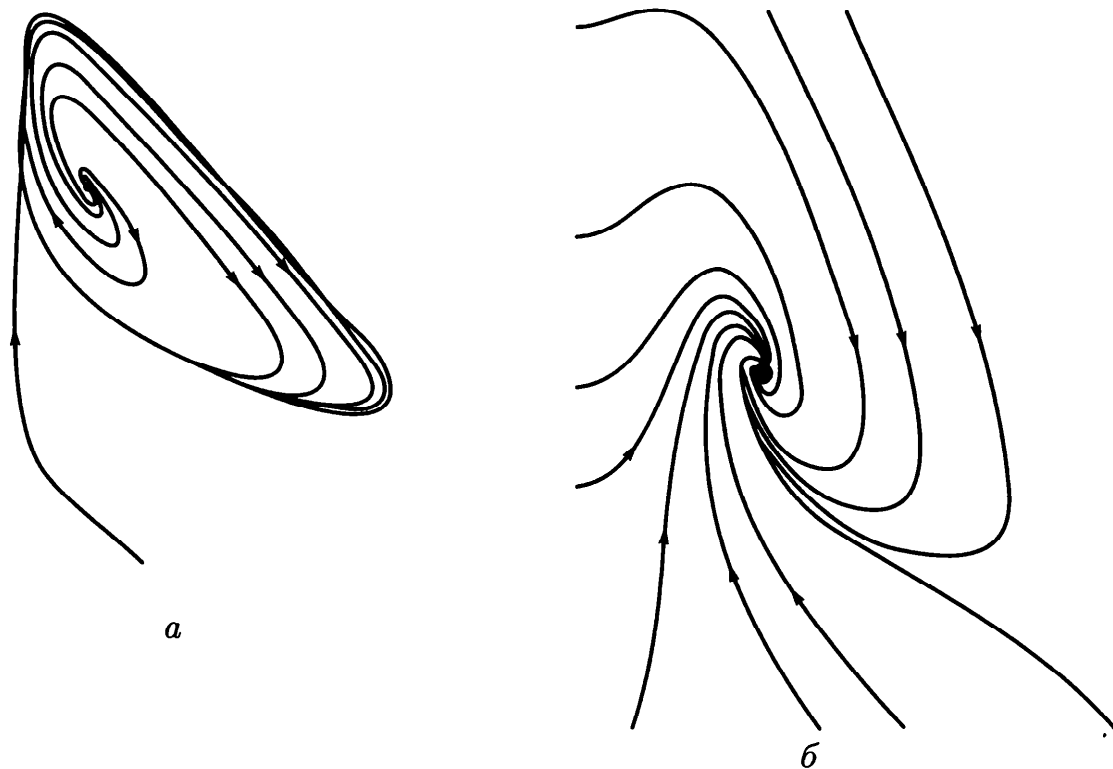


Рис. 5.25. Фазовый портрет брюсселятора:

$$a - B > 1 + A^2; \quad b - B < 1 + A^2$$

является автоколебательной системой. Фазовый портрет брюсселятора при разных значениях параметров изображен на рис. 5.25.

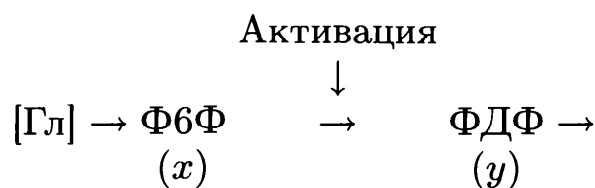
Далее будет рассмотрен «точечный» брюсселятор в качестве локального элемента системы «реакция — диффузия» (см. 6.3). Исследования брюсселятора как распределенной системы легли в основу науки о пространственно-временной самоорганизации химических и биологических систем.

#### 5.6.4. Колебания в гликолизе

Классическим примером колебательной биохимической реакции является гликолиз. В процессе гликолиза осуществляется распад глюкозы и других сахаров, при этом соединения, содержащие шесть молекул углерода, превращаются в трикарбоновые кислоты, включающие три молекулы углерода. За счет избыт-

Рис. 5.26. Упрощенная схема реакций гликолиза:

[Гл] — глюкоза; Ф6Ф — фруктозо-6-фосфат — субстрат ключевой реакции; ФДФ — фруктозо-ди-фосфат — продукт этой реакции, который является субстратом в следующей стадии



ка свободной энергии в процессе гликолиза на одну молекулу шестиуглеродного сахара образуются две молекулы АТФ. Основную роль в генерации наблюдаемых колебаний концентраций компонентов реакции: фруктозо-6-фосфата, фруктозо-1,6-фосфата и восстановленного НАД (*никотинаминадениндинуклеотид*) играет ключевой фермент гликолитического пути — фосфофруктокиназа (ФФК). Упрощенная схема реакций представлена на рис. 5.26.

Обе реакции катализируются ферментами. В безразмерных переменных система описывающих реакции уравнений может быть записана в виде:

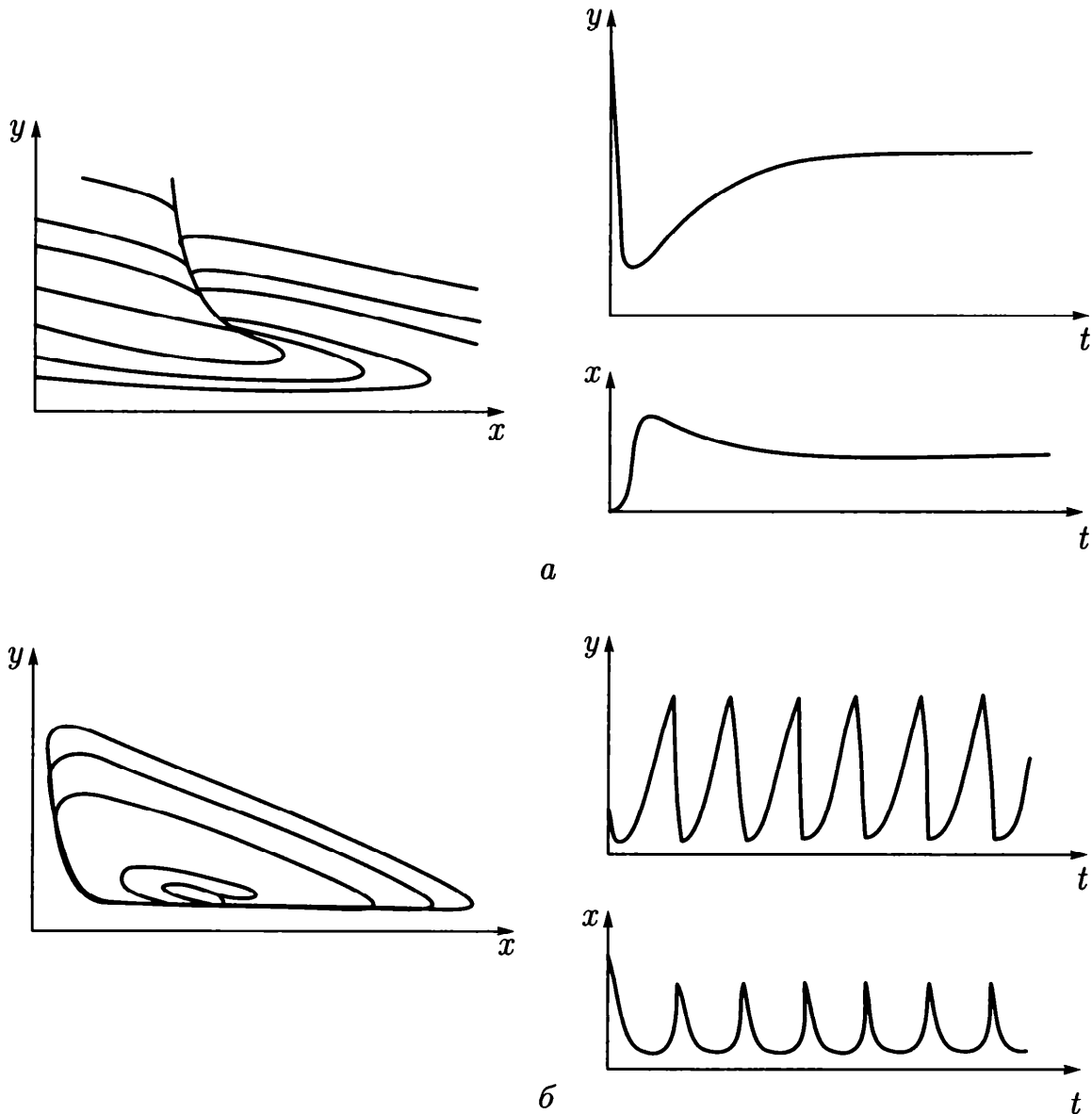


Рис. 5.27. Модель гликолиза (5.48). Кинетика изменений концентраций фруктозо-6-фосфата ( $x$ ) и фруктозо-ди-фосфата ( $y$ ) (справа) и фазовый портрет системы (слева) при разных значениях параметров системы:

$a$  — бесколебательный процесс (узел на фазовой плоскости),  $\alpha = 0,25$ ;  $r = 1$ ;  
 $b$  — колебания с постоянной амплитудой и фазой (предельный цикл на фазовой плоскости),  $\alpha = 8$ ;  $r = 0,5$

$$\begin{cases} \frac{dx}{dt} = k - \chi \frac{x}{(K_{mx} + x)} \frac{y}{(K_{my} + y)}; \\ \frac{dy}{dt} = \chi \frac{x}{(K_{mx} + x)} \frac{y}{(K_{my} + y)} - q \frac{y}{(K'_{my} + y)}. \end{cases} \quad (5.47)$$

Здесь зависимости скоростей реакций от переменных записаны в форме Михаэлиса — Ментен (Моно). Если выполняются условия:  $K_{mx} \gg x$ ,  $K_{my} \gg y$ , можно произвести замену переменных:

$$t' = \frac{t\chi k K'_{my}}{K_{mx} K_{my} (q - k)}, \quad x' = \frac{x\chi K'_{my}}{K_{mx} K_{my} (q - k)}, \quad y' = y \frac{q - k}{k K'_{my}}.$$

Опустив штрихи у новых переменных, перепишем систему (5.47) в безразмерном виде :

$$\begin{cases} \frac{dx}{dt} = 1 - xy; \\ \frac{dy}{dt} = \alpha y \left( x - \frac{1 + r}{1 + ry} \right), \end{cases} \quad (5.48)$$

где

$$\alpha = \frac{(q - k)^2 K_{mx} K_{my}}{(K'_{my})^2 k \chi}, \quad r = \frac{k}{q - k}.$$

Кинетика изменений переменных и фазовые портреты системы (5.48) при разных значениях параметров представлены на рис. 5.27.

Интересно, что колебательные реакции в системе гликолиза были сначала предсказаны на математической модели [J. A. Higgins, 1954] и лишь после этого зарегистрированы экспериментально с помощью метода дифференциальной спектрофотометрии в лаборатории Б. Чанса (1966).

## 5.7. Динамический хаос

Ранее рассматривались модели систем, которые описываются с помощью двух дифференциальных уравнений, их поведение можно наглядно изобразить на фазовой плоскости. Для таких двумерных систем в рамках качественной теории дифференциальных уравнений разработана исчерпывающая теория возможных типов динамического поведения.

Когда встает вопрос описания сложных многокомпонентных систем, например биологических сообществ, необходимо использовать системы большей размерности.



При исследовании систем размерности больше двух используется понятие **фазового пространства**, которое является расширением понятия фазовой плоскости. В фазовом пространстве по осям координат отложены значения переменных системы. Размерность фазового пространства соответствует числу переменных. Для трехмерного пространства говорят о **фазовом объеме**.

Полной классификации типов динамического поведения не существует. Известно, что увеличение размерности позволяет описать качественно новые типы поведения. Так, одно автономное уравнение может описать лишь монотонные изменения переменной. Система двух автономных уравнений может иметь более сложные типы поведения — предельные циклы, множественные стационарные состояния. Во второй половине XX в. стало понятно, что в автономной системе третьего и более высокого порядка возможны квазистохастические режимы.

### 5.7.1. Система Лоренца. Детерминированный хаос

Впервые проблема динамического хаоса была рассмотрена в работах великого французского математика Ж. Пуанкаре в начале XX в. Позднее (во второй половине XX в.) метеоролог Е. Лоренц [Lorenz, 1963] обнаружил, что даже простая система из трех нелинейных дифференциальных уравнений

$$\begin{cases} \dot{x} = \sigma y - \sigma x; \\ \dot{y} = rx - y - xz; \\ \dot{z} = xy - bz \end{cases} \quad (5.49)$$

может привести к хаотическим траекториям (рис. 5.28).

**Хаотическое поведение** затем было обнаружено при расширении размерности в большинстве классических моделей биологических систем, имеющих колебательные решения, в том числе в моделях взаимодействия видов, моделях гликолиза и клеточного цикла, моделях ферментативного катализа и др. Некоторые из этих моделей будут рассмотрены далее.

Причина нерегулярности определяется свойством нелинейных систем экспоненциально быстро разводить первоначально близкие траектории в ограниченной области фазового пространства. Механической системой такого типа является бильярд Синая, у которого стенки выпуклы внутрь, отчего угол отражения шара от стенки приводит к большому (экспоненциальному) разбеганию траекторий при малых отклонениях угла падения.

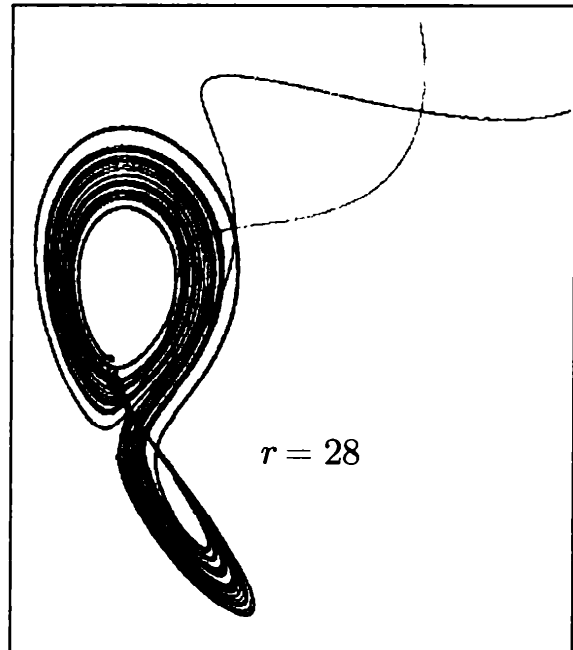
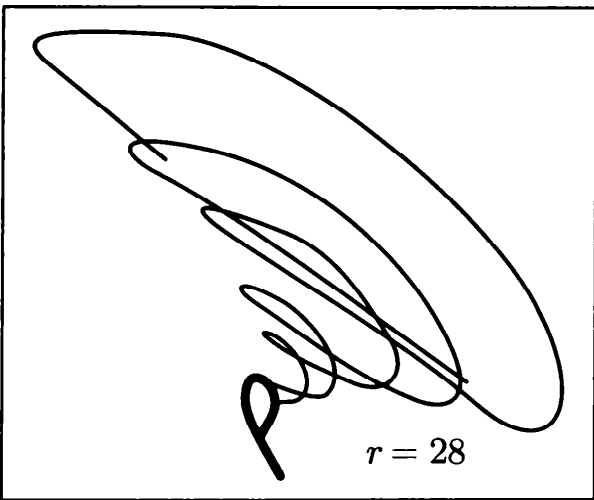
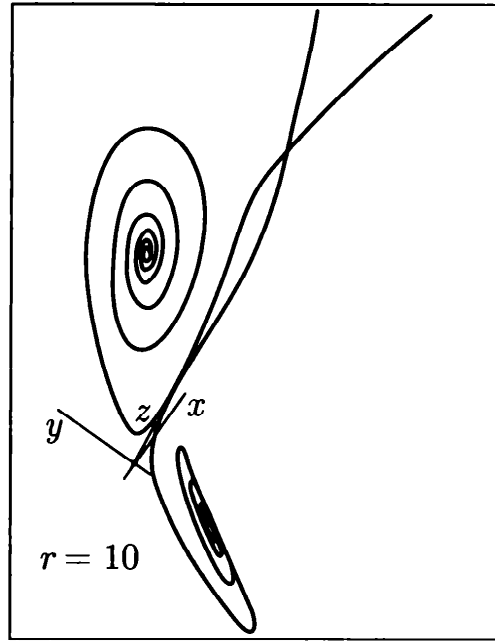
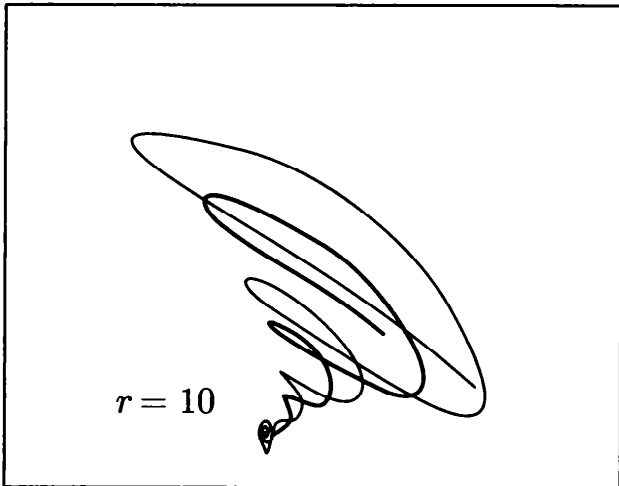
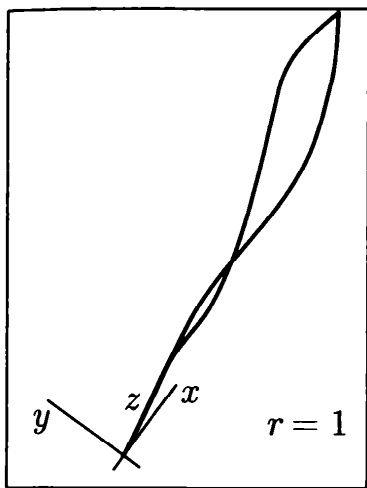


Рис. 5.28. Проекция фазовых траекторий системы Лоренца при разных значениях параметра  $r$  ( $r = 1$  — единственное стационарное состояние;  $r = 10$  — два стационарных состояния — устойчивых фокуса в разных плоскостях;  $r = 28$  — область странного аттрактора. Справа — траектории в окрестности аттракторов в увеличенном масштабе)

То же происходит при рассеивании частиц на круглых шарах. В таких системах траектория частицы становится непредсказуемой на больших временах. К процессам такого типа относятся течение жидкости вблизи порога возникновения турбулентности, некоторые химические реакции, метеорологические процессы, движения горных масс при землетрясениях. К ним относятся и многие биологические процессы в достаточно узкой области значений параметров. Изучение роли *динамического хаоса* в организации биологических процессов — одна из актуальных задач математической биологии.

Понятие «хаотическое поведение» означает неустойчивость фазовых траекторий, рост малого начального возмущения во времени, перемешивание элементов фазового объема и, как следствие, непредсказуемость поведения системы на больших временах.

Важно, что режимы такого типа обнаруживаются в *детерминированных системах*, где однозначно задан закон изменения системы с течением времени. Частный случай рождения динамического хаоса мы наблюдали при изучении дискретного аналога логистического уравнения (см. 4.4). При некоторых значениях параметра эта система демонстрировала квазистохастическое поведение, траектории системы при этом приобретали сложный непериодический характер. Как и в случае истинно хаотического броуновского движения, в каждой новой реализации при тех же начальных условиях (в пределах возможной точности) мы получали другие сложные траектории, сильно отличающиеся друг от друга. На самом деле, если бы начальные значения воспроизводились с абсолютной точностью, сложная траектория также бы повторилась. Но в области детерминированного хаоса траектории являются неустойчивыми по отношению к малым отклонениям. Поэтому даже малейшие отклонения, допускаемые компьютером, приводят к разбеганию.

Этим и объясняется название «*детерминированный хаос*», объединяющее два несовместимых представления — детерминированность (однозначную определенность) и непредсказуемость поведения. В системах с размерностью фазового пространства  $N \geq 3$  возможны установившиеся изменения переменных, не являющиеся ни периодическими, ни квазипериодическими. Таким хаотическим изменениям переменных соответствуют *хаотические аттракторы*, представляющие собой геометрически сложные множества дробной размерности.

### 5.7.2. Линейный анализ устойчивости траекторий

Линейный анализ устойчивости траекторий проводится аналогично линейному анализу устойчивости стационарных состояний. Поскольку анализируется малое возмущение, можно линеаризовать оператор эволюции в окрестности исследуемой траектории и провести линейный анализ ее устойчивости.

Рассмотрим автономную динамическую систему

$$\dot{x} = F(x, \alpha),$$

где  $x$  — вектор переменных;  $\alpha$  — вектор параметров;  $F$  — вектор-функция с компонентами  $f_j$ . Изучим устойчивость решения  $x^0(t)$ . Введем малое возмущение  $y = x(t) - x^0(t)$ . Для него можно записать

$$\dot{y} = F(x^0 + y) - F(x^0).$$

Раскладывая  $F(x^0 + y)$  в ряд в окрестности  $x^0$  и учитывая малость возмущения, получим линеаризованное уравнение относительно  $y$ :

$$\dot{y} = A(t)y,$$

где  $A$  — матрица линеаризации системы с элементами

$$a_{j,k} = \left. \frac{\partial f_j}{\partial x_k} \right|_{x_0}, \quad j, k = 1, 2, \dots, N.$$

Матрица  $A$  характеризуется собственными векторами  $e_i$  и собственными значениями  $\rho_i$ :

$$Ae_i = \rho_i e_i, \quad i = 1, 2, \dots, N.$$

Собственные числа являются корнями характеристического уравнения

$$\det[A - \rho E] = 0,$$

где  $E$  — единичная матрица.

Начальное возмущение, заданное в момент времени  $t^*$  вдоль  $i$ -го собственного вектора, с течением времени будет меняться в соответствии с эволюцией этого вектора:

$$y^i(t) = y^i(t^*) \exp \rho_i(t - t^*). \quad (5.50)$$

Будет отклонение уменьшаться или нарастать, определяется значением действительной части  $\rho_i$ .

Элементы матрицы  $A$  со временем могут меняться. Соответственно меняются ее собственные векторы и собственные значения, а также знак действительной части  $\rho_i$ . Поэтому, вообще говоря, (5.50) выполняется только в пределе при  $t - t^* \rightarrow 0$ . Для общей характеристики устойчивости траектории по отношению к возмущению вдоль  $i$ -го собственного вектора используют величину, называемую *характеристическим показателем Ляпунова*:

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t - t_0} \ln \|y^i(t)\|,$$

где  $y^i(t)$  — возмущение вдоль  $i$ -го собственного вектора в момент времени  $t$ , соответствующее бесконечно малому начальному возмущению  $y^i(t^*)$ .

Для  $N$ -мерной задачи устойчивость траектории характеризуется набором  $N$  ляпуновских характеристических показателей. Они связаны с собственными значениями матрицы линеаризации соотношением

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t - t_0} \int_{t_0}^t \operatorname{Re} \rho_i(t') dt'.$$

Таким образом, ляпуновский показатель — это усредненное вдоль исследуемой траектории значение действительной части собственного значения  $\rho_i$  матрицы линеаризации. Устойчивость траектории по Ляпунову означает, что произвольное начальное возмущение  $y(t^*)$  в среднем вдоль траектории не возрастает. Для этого необходимо и достаточно, чтобы спектр ляпуновских показателей  $\lambda_i$  не содержал положительных значений.

### 5.7.3. Диссипативные системы

В физике системы принято подразделять на консервативные и диссипативные. В консервативных системах энергия сохраняется (маятник без затухания). В диссипативных системах энергия со временем уменьшается (маятник в вязкой среде). Для того чтобы диссипативная система поддерживала непрерывное движение (например, автоколебания), необходимы источники энергии.

Существование аттрактора в диссипативной системе связано со свойством сжатия элемента фазового объема под действием оператора эволюции. Рассмотрим множество точек, заполня-

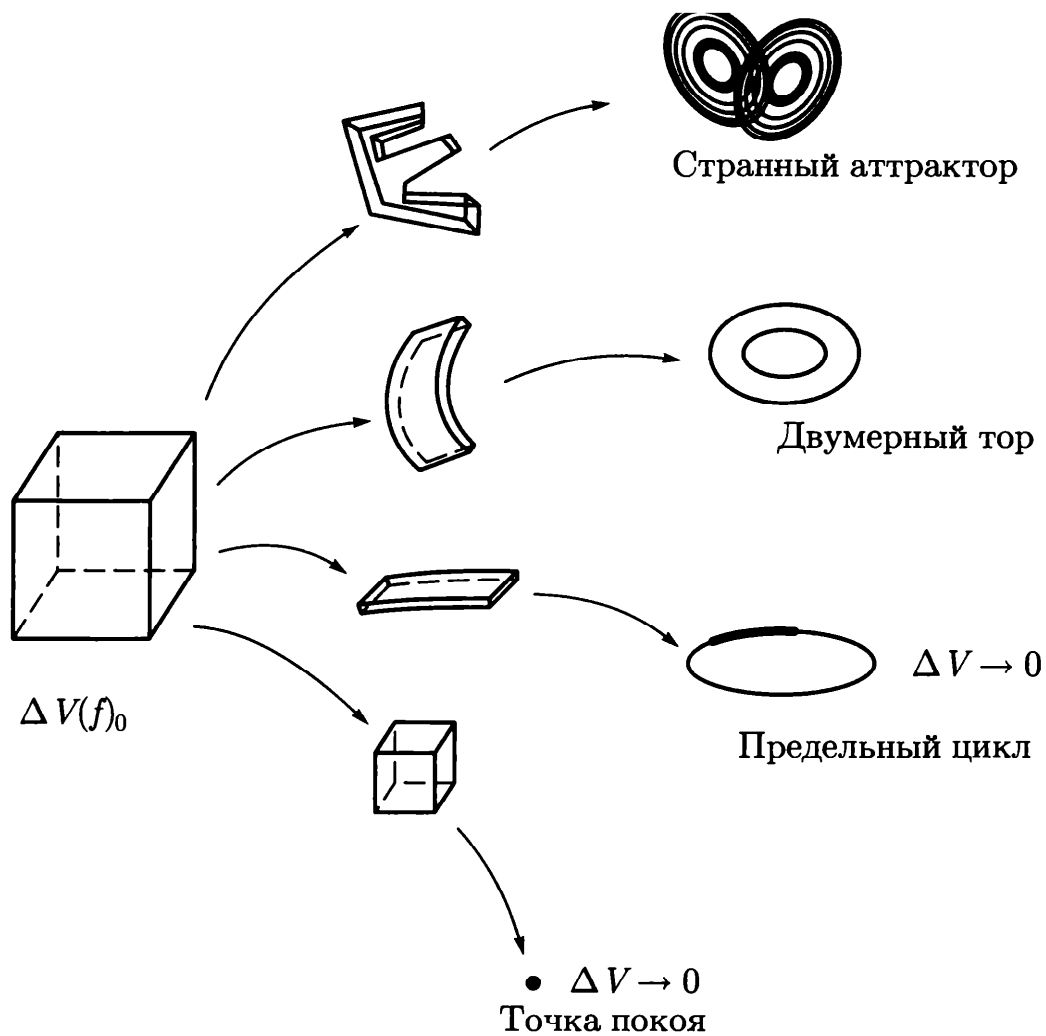


Рис. 5.29. Сжатие элемента фазового объема в разные типы аттракторов

ющих элемент объема  $\Delta V$ , и множество фазовых траекторий, стартующих из этих точек в момент времени  $t_0$ .

С течением времени объем  $\Delta V$  меняется по закону:

$$\Delta V(t) = \Delta V(t_0) \exp[(t - t_0) \overline{\operatorname{div} \mathbf{F}(x(t))}],$$

где  $\mathbf{F}(x(t))$  — поле фазовых скоростей (поток) динамической системы. Черта сверху означает усреднение вдоль фазовой траектории. Если в среднем дивергенция потока отрицательна, а это всегда выполняется для систем с потерями, то элемент фазового объема  $\Delta V$  в пределе при  $t \rightarrow \infty$  стремится к нулю. Это означает, что рассматриваемое множество фазовых траекторий, которые берут свое начало в  $\Delta V$ , стремится попасть на некоторое предельное множество, размерность которого меньше размерности  $N$  фазового пространства системы.

На рис. 5.29 изображены различные типы аттракторов, в которые может перейти элемент фазового пространства размерности 3: точка покоя, предельный цикл, двумерная поверхность — поверхность тора и странный аттрактор.

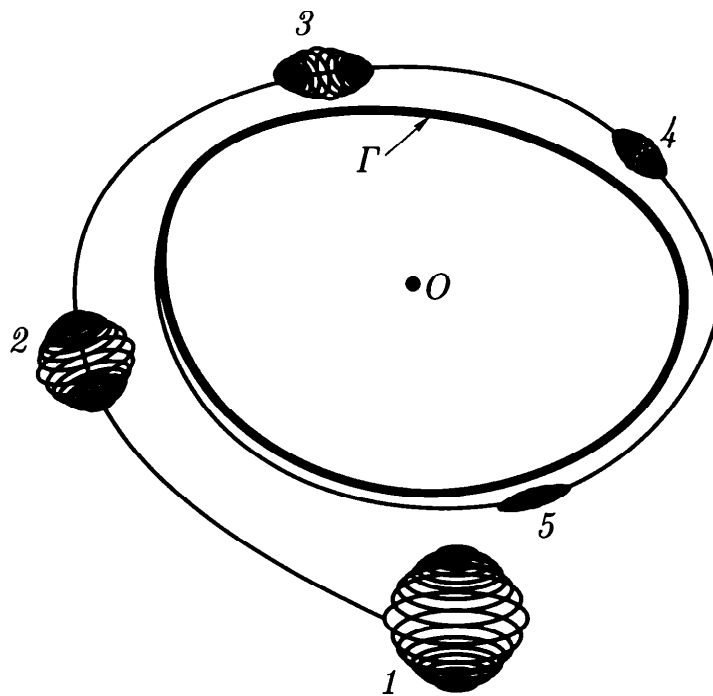


Рис. 5.30. Сжатие элемента фазового пространства радиуса  $\epsilon$  при «наматывании» траектории на устойчивый предельный цикл — траектория  $\Gamma$

Средняя вдоль траектории дивергенция потока и, следовательно, эволюция элемента фазового объема определяются суммой ляпуновских показателей:

$$\sum_{i=1}^N \lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t - t_0} \int_{t_0}^t \operatorname{div} F(t') dt'.$$

Для фазовых траекторий на аттракторе должно иметь место сжатие элемента фазового объема. Соответственно дивергенция потока вдоль траектории отрицательна, а значит, ляпуновские показатели удовлетворяют неравенству

$$\sum_{i=1}^N \lambda_i < 0.$$

Хаотические аттракторы имеют по крайней мере один положительный ляпуновский показатель. Если траектории на аттракторе имеют более чем одно направление неустойчивости, хаос называется *гиперхаосом*.

Непредсказуемость поведения системы в области динамического хаоса связана с неустойчивостью системы по отношению к малым отклонениям начального состояния. Это означает, что необходимо анализировать эволюцию во времени не начальной

точки, а начального объема вокруг этой точки. Рассмотрим малую сферу радиуса  $\varepsilon > 0$ , окружающую начальное состояние  $x_0$ . Любая точка внутри сферы характеризует малое отклонение от начального состояния. Применим оператор эволюции и посмотрим за трансформацией этого малого объема во времени. Если система устойчива, любое малое отклонение со временем будет затухать, шарик радиуса  $\varepsilon$  будет уменьшаться со временем и в пределе при  $t \rightarrow \infty$  его радиус уменьшится до нуля. На рис. 5.30 представлено последовательное сжатие (1—5) первоначальной области неопределенности фазового объема радиуса в случае, когда устойчивое предельное множество представляет собой предельный цикл.

Для неустойчивых режимов процесс происходит сложнее. Неустойчивость режима ведет к росту возмущений. Но если система диссипативна, то независимо от того, устойчива или неустойчива система, происходит уменьшение элемента фазового объема во времени, что связано с потерями энергии. Это значит, что элемент фазового пространства по одним направлениям растягивается (что соответствует положительным показателям Ляпунова), а по другим — сжимается. Причем степень сжатия превалирует над степенью расширения. Пример такой трансформации для системы, описывающей радиотехническое устройство (модифицированный генератор с инерционной нелинейностью), представили В.С. Анищенко с соавторами в книге «Нелинейная динамика хаотических и стохастических систем» (рис. 5.31). Модель генератора описывается системой уравнений:

$$\begin{cases} \dot{x} = mx + y - xz; \\ \dot{y} = -x; \\ \dot{z} = -gz + gI(x)x^2, \end{cases} \quad (5.51)$$

$$I = \begin{cases} 1, x > 0, \\ 0, x \leq 0. \end{cases}$$

При определенных значениях параметров система демонстрирует квазистохастическое поведение.

Рассмотрим, как будет себя вести малый фазовый объем радиуса  $\varepsilon$ , окружающий начальную точку, для такой квазистохастической системы.

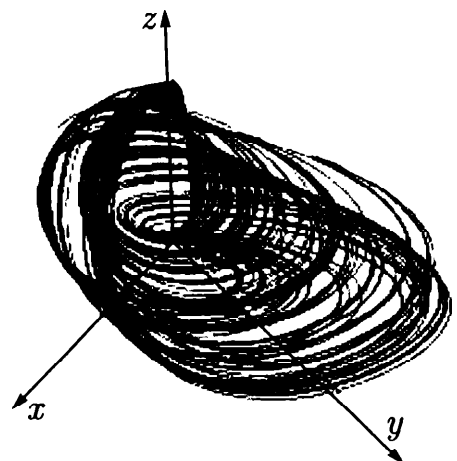


Рис. 5.31. Детерминированный хаос в трехмерной системе (5.51). Фазовая траектория для значений параметров  $m = 1,5$ ;  $g = 0,5$  [В. С. Анищенко и др.], 1999]



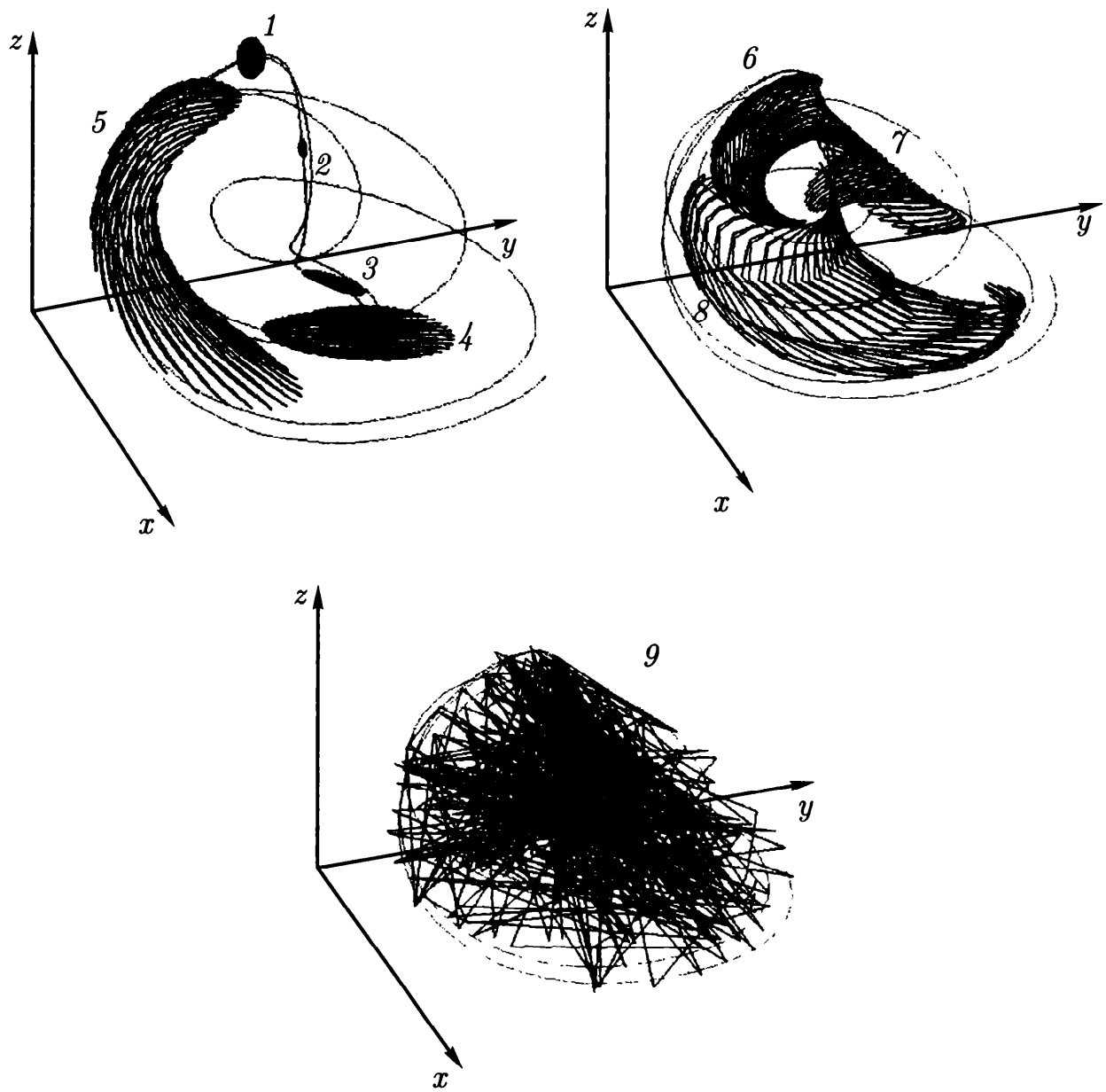


Рис. 5.32. Перемешивание в квазистохастической системе. Эволюция малого первоначального фазового объема во времени в динамической системе (5.51). Наблюдения траектории в различные моменты времени (1—9) [В. С. Анищенко и [др.], 1999]

Результаты компьютерного моделирования представлены на рис. 5.32.

Видно, что со временем имеет место растяжение в одних направлениях и сжатие — в других. Так, точки траекторий, начинающихся в элементе 1, можно обнаружить в любой части фазового пространства, занятого аттрактором.

Процесс перемешивания имеет простую аналогию. Поместим в жидкость, находящуюся в сосуде, капельку чернил и будем жидкость перемешивать. В силу «неустойчивости» капли молекулы чернил под влиянием потоков жидкости скоро «разбегутся» по всему объему. Их траектории будут представлять собой хаотические траектории. Если же в сосуд поместить твердую ча-

стицу, то молекулы вещества будут перемещаться под влиянием потока жидкости тоже по сложной траектории, но не удаляясь друг от друга (траектория устойчива).

Явление детерминированного хаоса было обнаружено практически во всех областях науки. Рассмотрим далее модель квазистохастического поведения биологического сообщества из трех видов.

#### 5.7.4. Динамический хаос в сообществе из трех видов

Для системы из трех видов в случае разветвленной трофической цепи даже исследование автономной локальной системы становится чрезвычайно сложным. Здесь отступление от вольтерровской схемы и учет биологических факторов, влияющих на динамику численности сосуществующих популяций, приводят к большому разнообразию модельных систем.

В системах, состоящих из хищника и двух жертв при наличии внутривидовой конкуренции между жертвами при изменении параметров системы, возможны бифуркационные явления, приводящие к появлению квазистохастических режимов [Е. А. Аполина и [др.], 1982]:

$$\begin{cases} \frac{du_1}{dt} = u_1(\alpha_1 - u_1 - a_1u_2 - b_1u_3); \\ \frac{du_2}{dt} = u_2(\alpha_2 - u_2 - a_2u_1 - b_2u_3); \\ \frac{du_3}{dt} = u_3(-1 + d_1u_1 + d_2u_2 - u_3), \end{cases} \quad (5.52)$$

где  $u_1, u_2$  — безразмерные численности жертв;  $u_3$  — безразмерная численность хищников;  $\alpha_1, \alpha_2$  — параметры, соответствующие скоростям роста численностей жертв.

В такой системе наряду с устойчивым состоянием равновесия возможны колебательные изменения численностей всех трех видов. В некоторой области параметров в системе имеется предельный цикл сложной формы.

При уменьшении  $\alpha_2$  наблюдается серия последовательных удвоений цикла (рис. 5.33, *a — в*). В некотором диапазоне значений  $\alpha_2$  из результатов численного эксперимента видно, что траектория системы полностью заполняет некоторый фазовый объем. При этих значениях параметров поведение системы неотлично от случайного, т. е. является квазистохастическим

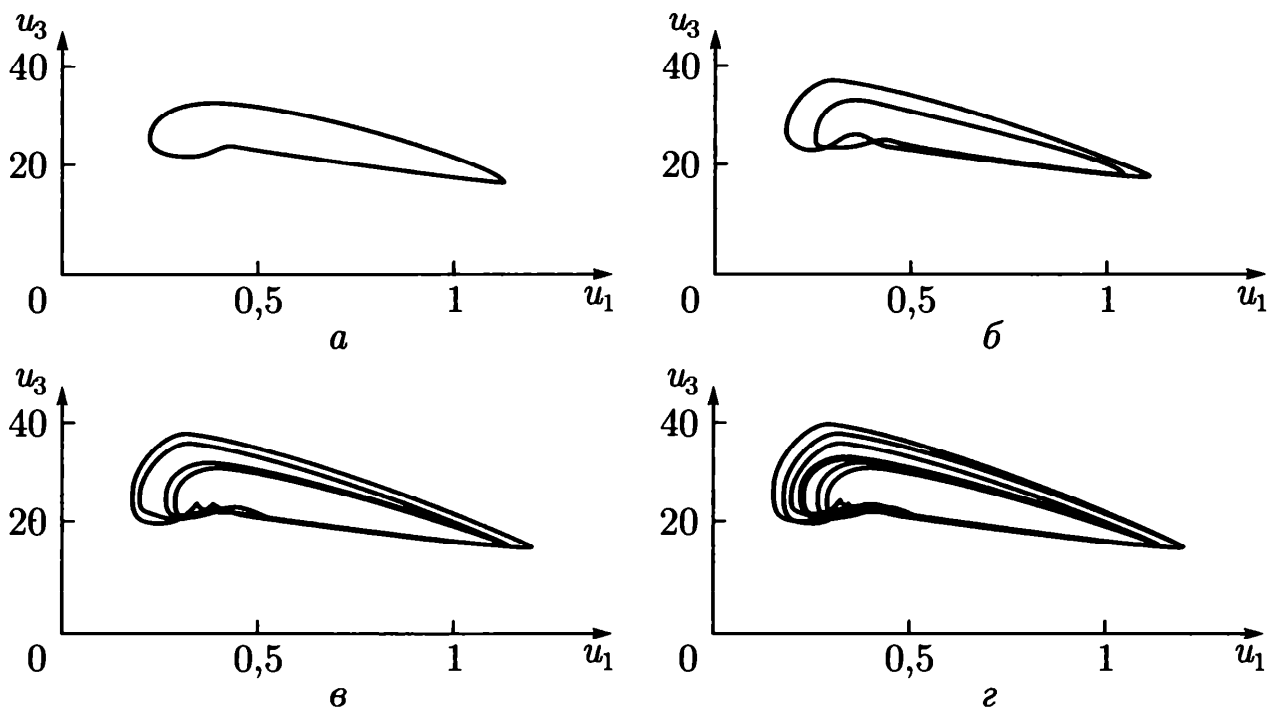


Рис. 5.33. Фазовые портреты системы (5.52), описывающей взаимодействие трех видов: хищник — две жертвы [Е. А. Апонина и [др.], 1982; А. Д. Базыкин, 2003]. Параметры модели:  $a_1 = 6$ ;  $a_2 = 1$ ;  $b_1 = 4$ ;  $b_2 = 10$ ;  $d_1 = 0,25$ ;  $d_2 = 4$ . При уменьшении параметра скорости роста численности второй жертвы происходит усложнение траектории (последовательное удвоение предельного цикла), колебательная динамика переходит в квазистохастическую:

$$\begin{aligned}
 a - \alpha_1 = 2,4; \alpha_2 = 1,7537; \quad б - \alpha_1 = 5,4; \alpha_2 = 1,7533; \quad в - \alpha_1 = 5,4; \alpha_2 = 1,7532; \\
 г - \alpha_1 = 5,4; \alpha_2 = 1,7532
 \end{aligned}$$

(рис. 5.33, г). Притягивающая область в фазовом пространстве является странным аттрактором.

Модельному режиму, изображенному на рис. 5.33, г, соответствуют наблюдаемые в природе нерегулярные квазистохастические изменения численностей животных.

РАСПРЕДЕЛЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ МОДЕЛИ

6.1. Уравнение реакция — диффузия

6.1.1. Активные кинетические среды в живых системах

Все биологические системы, в частности биологически активные мембраны и ткани, сообщества живых организмов, относятся к классу активных распределенных систем, основные свойства которых следующие:

- существует распределенный источник энергии или веществ, богатых энергией;
- каждый элементарный объем среды находится в состоянии, далеком от термодинамического равновесия, т. е. является открытой термодинамической системой, в которой диссипирует часть энергии, поступающей из распределенного источника;
- связь между соседними элементарными объемами осуществляется за счет процессов переноса.

Благодаря этим свойствам биологические системы обладают рядом нетривиальных типов поведения, составляющих сущность живых систем. Исследование простейших моделей активных сред показало, что сложное пространственно-временное поведение систем может быть описано системами нелинейных дифференциальных уравнений в частных производных вида:

$$\frac{\partial x_i}{\partial t} = F_i(X_1, X_2, \dots, X_n) + \frac{\partial}{\partial r} \left( \sum_{j=1}^n D_{ij} \frac{\partial x_j}{\partial r} \right), \quad (6.1)$$

$$i = 1, 2, \dots, n,$$

где  $D_{ii}$  и  $D_{ij}$  ( $i \neq j$ ) — коэффициенты диффузии и взаимной диффузии;  $F_i$  — нелинейные функции, описывающие взаимодействие компонентов.

В зависимости от вида функций  $F_i$  и коэффициентов диффузии  $D_{ij}$  в уравнении (6.1) в системах реакция — диффузия могут возникать следующие нетривиальные типы поведения переменных [Ю. М. Романовский, Н. В. Степанова, Д. С. Чернавский, 1984; 2004]:

- 1) распространяющиеся возмущения в виде бегущего импульса;

2) генерация волн автономными источниками импульсной активности. В качестве источников волн могут выступать либо неоднородности среды, вызванные отклонением значений параметров системы из-за механических или других повреждений, либо локальные кратковременные флуктуации переменных (источники типа «ведущий центр»);

3) стоячие волны;

4) синхронные автоколебания во всем пространстве. Синхронизация происходит с частотой того элемента пространства, который обладает наименьшим периодом колебаний;

5) квазистохастические волны, которые могут быть связаны с динамическим хаосом в локальной системе, но могут и возникать в распределенной системе с устойчивыми локальными элементами;

6) стационарные неоднородные распределения переменных в пространстве — диссипативные структуры.

### 6.1.2. Уравнение диффузии

Существование процессов переноса вещества и энергии является необходимым условием того, что система функционирует в пространстве как единое целое. Перемещение молекул вещества под действием случайных сил — диффузия — имеет место в любом веществе — газе, жидкости, твердом теле. Неоднородная смесь веществ в замкнутом объеме будет со временем благодаря диффузии однородной — вещества перемещаются, и концентрация каждого из них во всем объеме станет одинаковой. В такой системе установится так называемое *однородное (гомогенное) стационарное состояние*. Время установления стационарного состояния, естественно, определяется свойствами вещества, в основном подвижностью его молекул.

В активных кинетических средах, которыми являются биологические системы, кроме процессов диффузии, происходят также взаимодействия между компонентами, описываемые, как правило, нелинейными функциями. Эти нелинейные процессы могут приводить к установлению различных концентраций взаимодействующих компонентов в разных точках пространства, препятствуя таким образом вызванному диффузией выравниванию концентраций. Противоборство этих двух процессов — взаимодействия компонентов в каждой отдельной точке пространства и диффузии — определяет поведение распределенной системы.

Рассмотрение процессов, происходящих в распределенной системе, начнем с простейшего случая — с изучения процесса диффузии в одномерном реакторе — трубке, заполненной раствором некоторого вещества, предполагая, что во всякий момент времени концентрация раствора по сечению трубки одинакова. Тогда процесс диффузии может быть описан функцией  $C(r, t)$ , представляющей концентрацию вещества в сечении с координатой  $r$  в момент времени  $t$ .

Опыт показывает, что диффузионный поток какого-либо компонента, т. е. масса диффундирующего компонента, проходящая в единицу времени через единицу площади, перпендикулярной направлению диффузии, пропорционален градиенту концентрации этого компонента, взятому с обратным знаком (закон Фика):

$$I = -D \frac{\partial C}{\partial r}, \quad (6.2)$$

где  $I$  — диффузионный поток интересующего нас компонента в направлении оси  $r$ . Знак «минус» в правой части (6.2) показывает, что диффузионный поток направлен в сторону убывания концентрации. Коэффициент  $D$  в уравнении (6.2) называется *коэффициентом диффузии*, он численно равен диффузионному потоку при градиенте концентраций, равном 1, и зависит от свойств диффундирующего вещества и свойств остальных компонентов, составляющих смесь. При не слишком больших концентрациях веществ  $D$  мало зависит от концентрации самого вещества и определяется степенью подвижности молекул.

Пусть поперечное сечение трубки  $S$ . Выделим в ней элементарный объем  $\Delta V_r$  с координатами границ  $r$  и  $r + \Delta r$ , тогда  $\Delta V_r = S \Delta r$ . Не нарушая общности, предположим, что диффузия протекает в направлении оси  $r$ .

Масса  $\Delta M_r$  вещества, втекающего за время от  $t$  до  $t + \Delta t$  в рассматриваемый объем через границу  $r$ , согласно закону Фика, равна:

$$\Delta M_r = -D \frac{\partial C(r, t)}{\partial r} S \Delta t.$$

Через другую границу с координатой  $r + \Delta r$  из выделенного объема вытекает вещество массой  $\Delta M_{r+\Delta r}$ :

$$M_{r+\Delta r} = -D \frac{\partial C(r + \Delta r, t)}{\partial r} S \Delta t.$$

Общее изменение массы  $\Delta M$  в объеме  $\Delta V$  составляет

$$\Delta M = \Delta M_r - \Delta M_{r+\Delta r}, \quad \text{или}$$

$$\Delta M = \left[ D \frac{\partial C(r + \Delta r, t)}{\partial r} - D \frac{\partial C(r, t)}{\partial r} \right] S \Delta t.$$

Запишем уравнение для изменения концентрации:

$$\Delta C(r, t) = \frac{\Delta M}{\Delta V} = \frac{\Delta M}{S \Delta r} = \frac{D \frac{\partial C(r + \Delta r, t)}{\partial r} - D \frac{\partial C(r, t)}{\partial r}}{\Delta r} \Delta t.$$

Перейдем к пределу при  $\Delta r \rightarrow 0$  и получим

$$\Delta C(r, t) = \frac{\partial}{\partial r} \left( D \frac{\partial C(r, t)}{\partial r} \right) \Delta t.$$

Разделив левую и правую части этого выражения на  $\Delta t$  и перейдя к пределу при  $\Delta t \rightarrow 0$ , получим уравнение диффузии в дифференциальной форме:

$$\frac{\partial C(r, t)}{\partial t} = \frac{\partial}{\partial r} \left( D \frac{\partial C(r, t)}{\partial r} \right).$$

В случае когда коэффициент диффузии  $D$  постоянен, имеем

$$\frac{\partial C(r, t)}{\partial t} = D \frac{\partial^2 C(r, t)}{\partial r^2} \quad (6.3)$$

или в другой форме записи  $C_t = DC_{rr}$ .

В случае когда диффузия происходит в трехмерном пространстве, причем имеет место **изотропная диффузия**, т. е. перемещение частиц вещества равновероятно по всем направлениям, уравнение (6.3) примет вид

$$\frac{\partial C(r, t)}{\partial t} = D \Delta C(r, t).$$

Здесь  $\Delta$  — оператор Лапласа.

Если в среде диффундируют  $n$  веществ с концентрациями  $C_i$  ( $i = 1, \dots, n$ ), процесс описывается системой  $n$  уравнений:

$$\frac{\partial C_i(r, t)}{\partial t} = D_i \Delta C_i(r, t), \quad i = 1, 2, \dots, n, \quad (6.4)$$

где  $D_i$  — коэффициент диффузии  $i$ -го вещества.

Вообще говоря, коэффициент диффузии определяется не только свойствами самого диффундирующего вещества, но и свойствами остальных компонентов системы. Тогда вместо ко-

эффицента самодиффузии  $D_i$  следует пользоваться коэффициентом взаимной диффузии (кросс-диффузии).

Уравнения (6.3), (6.4) описывают изменение во времени и в пространстве концентраций веществ, когда в системе происходит единственный процесс — диффузия. Однако специфика химических и биологических систем определяется тем, что кроме диффузии в них протекают и другие процессы. Это приводит к возникновению новых членов в правых частях уравнений типа (6.3), (6.4), описывающих изменения концентраций  $C_i$  во времени. Например, в случае одного вещества в одномерном реакторе кроме диффузии возможно наличие в некоторых местах трубки источников или стоков этого вещества, с учетом которых уравнение (6.3) примет вид

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial r^2} + F(r, t),$$

где  $F$  — функция источника.

Если, кроме того, имеется диффузия вещества через стенки трубки, то в правой части уравнения появятся также члены, пропорциональные концентрациям:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial r^2} + F(r, t) + \alpha C.$$

В многокомпонентных системах возможны разнообразные взаимопревращения компонентов, например химические превращения веществ в ходе реакций. Пусть скорость изменения концентрации  $C_i$  за счет химической реакции описывается функцией  $f_i = f_i(C_1, C_2, \dots, C_n)$ , вид которой определяется механизмом реакции. В общем случае функция  $f_i$  зависит как от концентраций реагирующих веществ, так и явно от пространственной координаты  $r$  и времени  $t$ . В дальнейшем, как и в случае точечных систем, будем считать, что вид функции  $f_i$  не зависит явно от времени и координаты пространства.

С учетом химических превращений, происходящих в каждой точке системы, уравнения (6.4) следует переписать в виде:

$$\frac{\partial C_i}{\partial t} = f_i(C_1, C_2, \dots, C_n) + D_i \Delta C_i(r, t).$$

В случае одномерного реактора:

$$\frac{\partial C_i}{\partial t} = f_i(C_1, C_2, \dots, C_n) + D_i \frac{\partial^2 C_i}{\partial r^2}. \quad (6.5)$$



Дифференциальные уравнения с частными производными имеют, вообще говоря, бесчисленное множество решений. Поэтому для однозначной характеристики процесса необходимо к уравнениям добавить некоторые дополнительные условия. Для обыкновенных дифференциальных уравнений необходимо задать начальные значения переменных в момент времени  $t = t_0$ . В случае распределенных систем этого недостаточно.

### 6.1.3. Начальные и граничные (краевые) условия

Кроме начальных условий — значений функций  $C_i(t, r)$  в некоторый начальный момент времени  $t = t_0$  — следует задать также так называемые *граничные (краевые)* условия, т. е. условия на границе области, в которой разворачивается изучаемый процесс.

Начальные условия задаются в виде функции зависимости концентрации от пространственной координаты в начальный момент времени  $t_0$ :

$$C_i(t_0, r) = \varphi_i(r). \quad (6.6)$$

Граничные условия могут быть заданы в различном виде в зависимости от закона изменения концентрации веществ на границе изучаемой области. Рассмотрим основные типы краевых условий на примере одномерного реактора — трубки длиной  $l$ .

1. На краю трубки задана концентрация, которая может определяться, например, концентрацией вещества в резервуаре, с которым трубка находится в контакте:

$$C(0, t) = \mu(t), \quad (6.7)$$

где  $\mu(t)$  — концентрация вещества в резервуаре в момент времени  $t$ .

В частности, если на границе трубки происходит поглощение вещества, то

$$C(0, t) = 0.$$

2. На границе задано значение производной:

$$\frac{\partial C}{\partial r}(0, t) = \nu(t). \quad (6.8)$$

К этому условию мы приходим, если задана величина диффузионного потока  $I$ , протекающего через торцевое сечение трубки:

$$I(0, t) = D \frac{\partial C}{\partial r}(0, t),$$

тогда величина  $v$  в формуле (6.8) — известная функция, которая выражается через заданный поток  $I$  по формуле

$$v(t) = \frac{I(0, t)}{D}.$$

В частности, в случае непроницаемости торца трубки, когда диффузионный поток на границе равен нулю:  $\frac{\partial C}{\partial r}(0, t) = 0$ .

3. На краю трубки задано линейное соотношение между производной и функцией:

$$\frac{\partial C}{\partial r}(0, t) = -\lambda [C(0, t) - \theta(t)]. \quad (6.9)$$

Граничное условие (6.9) соответствует случаю, когда на границе имеется диффузионный поток между трубкой и резервуаром, концентрация вещества в котором известна. Пользуясь двумя выражениями для диффузионного потока, протекающего через сечение  $r = 0$ ,  $I = h(C - \theta)$  и  $I = -D \frac{\partial C}{\partial r}$ , получаем математи-

ческую формулировку краевого условия в виде (6.9), где  $\lambda = \frac{h}{D}$ ;  $\theta$  — заданная функция.

В случае если функции  $C_i(x, t)$  непрерывны в изучаемой области вместе со своими вторыми производными, то система уравнений типа (6.5) вместе с начальными (6.6) и граничными условиями одного из типов (6.7) — (6.9) имеет единственное решение.

Возможны типы задач, в которых краевые или начальные условия не следует учитывать. Рассмотрим процесс диффузии в очень большом объеме. В течение небольшого промежутка времени влияние потока веществ, заданного на границе, в центральной части рассматриваемого объема оказывается весьма слабо, и концентрация веществ определяется, помимо химических реакций, в основном начальным распределением вещества. При анализе поведения переменных на малых временах в таких задачах считают, что реакция происходит в неограниченном объеме, и ищут решение системы уравнений при заданных начальных условиях без учета граничных условий.

Влияние начальных условий с течением времени ослабевает. При этом если концентрация веществ поддерживается постоянной на границе, то в момент времени, достаточно удаленный от начального, концентрация веществ определяется лишь гранич-

ными условиями. В этом случае можно считать, что опыт продолжается бесконечно долго, и начальные условия тем самым отпадают.

## 6.2. Решение уравнения диффузии

### 6.2.1. Общий ход решения

Решение краевых задач для системы нелинейных уравнений типа (6.5) при произвольных функциях  $f_i$  может быть выполнено лишь с помощью компьютера. Методы аналитического решения разработаны только для линейных уравнений вида

$$C_t = DC_{rr} + f(r, t) \quad (6.10)$$

с соответствующими начальными и граничными условиями, не содержащими нелинейные функции относительно концентраций  $C$ . Методы решения таких задач подробно изложены в учебниках и монографиях, посвященных уравнениям математической физики [А. Н. Тихонов, А. А. Самарский, 1999]. Здесь достаточно подробно рассмотрим решение линейной краевой задачи и увидим, что такое решение в пространстве — времени может быть представлено в виде ряда Фурье (бесконечной суммы гармонических функций пространственной координаты с убывающими со временем амплитудами). Отсюда становится понятным, почему в нелинейных системах с диффузией при наличии малых флуктуаций могут возникнуть автоволновые процессы и периодические по пространству, постоянные во времени структуры (диссипативные структуры). Нелинейная система оказывается своеобразным «фильтром», выделяющим некоторые из членов гармонического ряда и поддерживающим их существование, в то время как более высокие гармоники затухают во времени.

Рассмотрим общий путь решения одномерной краевой задачи для уравнения (6.10) с начальным условием

$$C(r, 0) = \varphi(r) \quad (6.11)$$

и граничными условиями первого рода на обеих границах узкой трубки длины  $l$ :

$$C(0, t) = \mu_1(t); C(l, t) = \mu_2(t). \quad (6.12)$$

Задачу решают в три этапа. Сначала ищут решение однородного уравнения

$$C_t = DC_{rr} \quad (6.13)$$

с начальным условием (6.11) и нулевыми краевыми условиями

$$C(0, t) = 0; C(l, t) = 0. \quad (6.14)$$

Затем находят решение неоднородного уравнения (6.10) с нулевыми краевыми условиями (6.14). Наконец, последний этап — решение общей краевой задачи — уравнения (6.10) с начальным условием (6.11) и граничными условиями (6.12).

### 6.2.2. Решение однородного уравнения

Пусть в системе имеет место только один процесс — диффузия. Решим основную вспомогательную задачу: найдем решение уравнения (6.13), не равное тождественно нулю и удовлетворяющее нулевым краевым условиям (6.14). При этом воспользуемся методом разделения переменных, представляя решение в виде

$$C(r, t) = R(r)T(t), \quad (6.15)$$

где  $R(r)$  — функция только пространственной переменной  $r$ ;  $T(t)$  — функция только переменной времени  $t$ .

Подставим решение в форме (6.15) в уравнение (6.13):

$$T'R = DTR''$$

и разделим обе части равенства на  $DRT$ . Получаем

$$\frac{1}{D} \frac{T'}{T} = \frac{R''}{R} = -\lambda,$$

где  $\lambda = \text{const}$ , так как левая часть равенства зависит только от  $t$ , а правая — только от  $r$ . Отсюда получим два самостоятельных уравнения для переменных  $r$  и  $t$ :

$$R''(r) + \lambda R(r) = 0, \quad (6.16)$$

$$T'(t) + D\lambda T(t) = 0. \quad (6.17)$$

Для определения функции  $R(r)$  получено обыкновенное дифференциальное уравнение (6.16), причем вследствие граничных условий (6.14) функция  $R(r)$  должна удовлетворять дополнительным условиям:

$$R(0) = R(l) = 0, \quad (6.18)$$

так как в противном случае было бы:

$$T(t) = 0 \quad \text{и} \quad C(r, t) = 0,$$

в то время как задача состоит в нахождении нетривиального решения. Для функции  $T(t)$  никаких дополнительных условий нет.

Таким образом, чтобы найти функцию  $R(r)$ , необходимо определить значения  $\lambda$ , называемые *собственными значениями*, при которых существует нетривиальное решение задачи (6.16), (6.18), а также сами эти решения. Такая задача называется *задачей о собственных значениях*, или задачей Штурма — Лиувилля. Собственные значения задачи Штурма — Лиувилля обладают следующими свойствами:

- множество собственных значений не пусто и не имеет конечных предельных точек;
- собственные значения вещественны и конечной кратности;
- собственные значения неотрицательны;
- множество собственных значений счетно;
- каждое собственное значение — простое.

Согласно теореме В. А. Стеклова, всякая функция  $f$ , непрерывная и имеющая непрерывные производные не ниже второй, разлагается в регулярно сходящийся ряд Фурье по собственным функциям  $\{X_k\}$  задачи Штурма — Лиувилля:

$$f(x) = \sum_{n=1}^{\infty} (f, X_k) X_k(x).$$

Доказательство этих утверждений можно найти в учебнике [В. С. Владимиров, 1981].

Общее решение уравнения (6.16) имеет вид:

$$R(r) = C_1 e^{\sqrt{-\lambda}r} + C_2 e^{-\sqrt{-\lambda}r}.$$

При  $\lambda \leq 0$  задача не имеет нетривиальных решений. При  $\lambda > 0$  общее решение содержит мнимые показатели и поэтому может быть записано в виде

$$R(r) = D_1 \cos \sqrt{\lambda}r + D_2 \sin \sqrt{\lambda}r.$$

Краевые условия (6.18) дают:

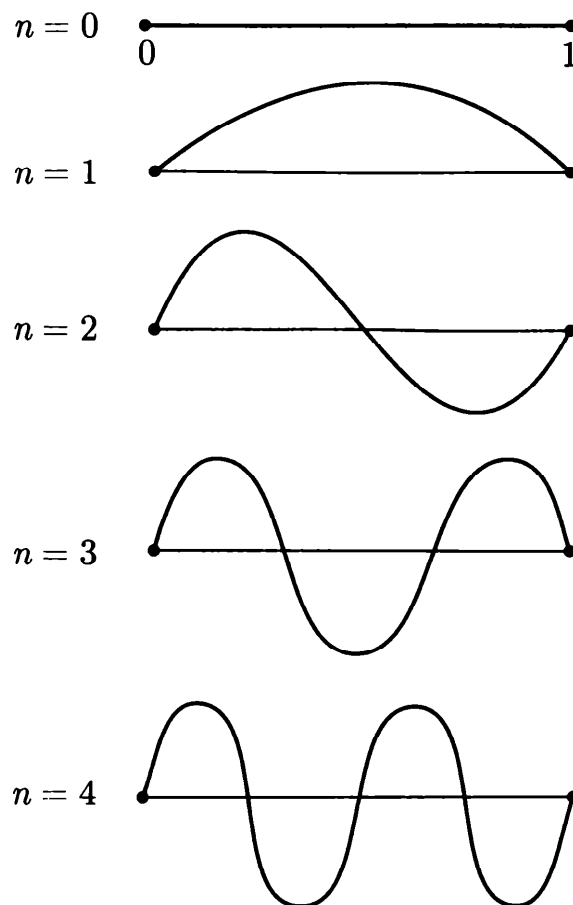
$$R(0) = D_1 = 0,$$

$$R(l) = D_2 \sin \sqrt{\lambda}l = 0.$$

Если  $R(r)$  не равно тождественно нулю, то  $D_2 \neq 0$ , поэтому

$$\sin \sqrt{\lambda}l = 0, \quad \text{или} \quad \sqrt{\lambda} = \frac{\pi n}{l},$$

Рис. 6.1. Собственные функции  $R_n(r)$  для краевой линейной задачи о диффузии в одномерном реакторе длины  $l$  с нулевыми краевыми условиями



где  $n$  — любое целое число. Величину  $\sqrt{\lambda} = \frac{\pi n}{l}$  в литературе, посвященной волновым процессам, обычно называют *волновым числом* и обозначают буквой  $k$ . Таким образом, нетривиальные решения задачи возможны лишь при собственных значениях

$$\lambda_n = \left(\frac{\pi n}{l}\right)^2. \quad (6.19)$$

Этим собственным значениям соответствуют *собственные функции*:

$$R_n(r) = D_n \sin \frac{n\pi}{l} r. \quad (6.20)$$

В дальнейшем произвольный множитель  $D_n$  будем считать равным единице. Вид собственных функций (6.20) для краевой задачи о диффузии в одномерном реакторе длины  $l$  с нулевыми концентрациями на границе при различных значениях  $n$  изображен на рис. 6.1.

Собственным значениям  $\lambda_n$  (6.19) соответствуют решения уравнения (6.17):

$$T'(t) + D\lambda_n T(t) = 0.$$

Это линейное обыкновенное дифференциальное уравнение, его решения для каждого  $n$  представляют собой затухающие со временем экспоненты:

$$T_n(t) = A_n e^{-D\lambda_n t}, \quad (6.21)$$

где  $A_n$  — не определенные пока коэффициенты. Возвращаясь к задаче (6.13), (6.14), видим, что функции концентраций  $C_n(r, t)$ , представляющие собой произведения временного и пространственного сомножителей:

$$C_n(r, t) = R_n(r)T_n(t) = A_n e^{-D\lambda_n t} \sin \frac{\pi n}{l} r, \quad (6.22)$$

являются частными решениями уравнения (6.13), удовлетворяющими нулевым краевым условиям.

Таким образом, частные решения (6.22) представляют собой затухающие со временем синусоидальные распределения концентрации  $C$ . Легко видеть, что выражение, стоящее под знаком  $\sin$ , представляет собой произведение волнового числа  $k_n = \frac{\pi n}{l}$  и координаты  $r$ . Таким образом,  $k_n = \frac{\pi n}{l}$  является «частотой колебания» переменной  $C$  в пространстве или, что то же самое, величина  $\Lambda_n = \frac{2\pi}{kn}$  является «периодом» колебаний концентрации  $C$  по пространству  $r$ . Иначе,  $\Lambda_n = \frac{2l}{n}$  есть длина волны синусоиды, представляющей собой собственное решение  $C_n$  (см. рис. 6.1). Чем больше номер гармоники  $n$ , тем меньше период синусоиды в пространстве и тем больше коэффициент затухания этой синусоиды во времени (за счет множителя  $e^{-(\frac{\pi n}{l})^2 Dt}$ ).

### **Условие нулевых потоков на границах**

При граничных условиях непроницаемости концов одномерного реактора на решения уравнения (6.16) накладываются краевые условия

$$\frac{\partial C}{\partial r}(0, t) = \frac{\partial C}{\partial r}(l, t) = 0, \quad (6.23)$$

которые дают следующее:  $\frac{\partial R}{\partial r}(0) = 0$ ;  $\frac{\partial R}{\partial r}(l) = 0$ .

Продифференцировав выражение для функции  $R(r)$ , получим

$$\frac{\partial R}{\partial r} = -D_1 \sqrt{\lambda} \sin \sqrt{\lambda} r + D_2 \sqrt{\lambda} \cos \sqrt{\lambda} r.$$

Из краевых условий (6.23) следует

$$\left. \frac{\partial R}{\partial r} \right|_{r=0} = D_2 \sqrt{\lambda} = 0, \quad \left. \frac{\partial R}{\partial r} \right|_{l=0} = D_2 \sqrt{\lambda} = 0.$$

Отсюда, как и в случае, когда краевые условия соответствуют нулевым концентрациям на границах, получаются те же величины собственных значений  $\sqrt{\lambda} = \frac{\pi n}{l}$ , но собственными функциями для одномерного реактора с непроницаемыми концами являются функции

$$R_n = D_n \cos \frac{\pi n}{l} r.$$

Частные решения задачи о диффузии в замкнутом одномерном реакторе — концентрации  $C_n(r, t)$ , как и в задаче с нулевыми концентрациями на границе — произведения временного и пространственного сомножителей:

$$C_n(r, t) = R_n(r)T_n(t) = A_n e^{-D\lambda_n t} \cos \frac{\pi n}{l} r. \quad (6.24)$$

Общее решение представляет собой суперпозицию частных решений. Именно благодаря такой структуре решения определенные типы взаимодействия переменных в системах с диффузией могут привести к образованию в системе стационарных периодических структур, а также возникновению режимов типа стоящих и бегущих волн.

### 6.2.3. Зависимость решений от начальных условий

Вернемся к задаче с начальными условиями (6.11). Представим решение в виде ряда:

$$C(r, t) = \sum_{n=1}^{\infty} A_n e^{-\left(\frac{\pi n}{l}\right)^2 D t} \sin \frac{\pi n}{l} r. \quad (6.25)$$

Функция  $C(r, t)$  удовлетворяет граничным условиям, так как им удовлетворяют все члены ряда. Требуя выполнения начальных условий (6.11), получим выражение:

$$\varphi(r) = C(r, 0) = \sum_{n=1}^{\infty} A_n \sin \frac{\pi n}{l} r. \quad (6.26)$$

Таким образом,  $A_n$  представляют собой коэффициенты разложения Фурье, известного из курса математического анализа, функции  $\varphi(r)$  по синусам в интервале  $(0, l)$ :

$$A_n = \varphi_n = \frac{2}{l} \int_0^l \varphi(\xi) \sin \frac{\pi n}{l} \xi d\xi, \quad (6.27)$$

где  $\xi$  — переменная интегрирования.



Учитывая выражения (6.27) для множителей  $A_n$ , формулу (6.25) можно записать в виде

$$C(r, t) = \int_0^l \left[ \frac{2}{l} \sum_{n=1}^{\infty} e^{-\left(\frac{\pi n}{l}\right)^2 Dt} \sin \frac{\pi n}{l} r \sin \frac{\pi n}{l} \xi \right] \varphi(\xi) d\xi. \quad (6.28)$$

Изменение порядков суммирования и интегрирования законно, если ряды сходятся равномерно по  $\xi$  при  $t > 0$ .

Обозначив

$$G(r, \xi, t) = \frac{2}{l} \sum_{n=1}^{\infty} e^{-\left(\frac{\pi n}{l}\right)^2 Dt} \sin \frac{\pi n}{l} r \sin \frac{\pi n}{l} \xi,$$

представим  $C(r, t)$  через  $G(r, \xi, t)$  в виде

$$C(r, t) = \int_0^l G(r, t, \xi) \varphi(\xi) d\xi,$$

где  $G(r, \xi, t)$  называется *функцией мгновенного источника* и характеризует распределение вещества в трубке  $0 \leq r \leq l$  в момент времени  $t$ , если в начальный момент концентрация вещества равна нулю, и в этот момент в точке  $r = \xi$  мгновенно выделяется некоторое количество вещества, а концентрация вещества на концах трубки все время поддерживается нулевой. Как будет показано далее, выражение для функции источника удобно использовать при решении неоднородного уравнения диффузии.

Итак, получено решение однородного уравнения с заданными начальными условиями и нулевыми краевыми условиями.

Решение краевой задачи с нулевыми потоками на границе будет иметь вид

$$C(r, t) = \int_0^l \left[ \frac{2}{l} \sum_{n=1}^{\infty} e^{-\left(\frac{\pi n}{l}\right)^2 Dt} \cos \frac{\pi n}{l} r \cos \frac{\pi n}{l} \xi \right] \varphi(\xi) d\xi,$$

а функция источника соответственно может быть выражена формулой

$$G(r, \xi, t) = \frac{2}{l} \sum_{n=1}^{\infty} e^{-\left(\frac{\pi n}{l}\right)^2 Dt} \cos \frac{\pi n}{l} r \cos \frac{\pi n}{l} \xi.$$

Таким образом, для краевых задач как первого рода (на границах заданы концентрации), так и второго рода (на границах заданы потоки) собственными функциями являются периодические гармонические функции пространственной координаты

$$\sin \frac{\pi n}{l} r, \quad \cos \frac{\pi n}{l} r, \quad n = 1, 2, \dots, \infty.$$

Если реакция происходит в безграничной трубке, то решение однородного уравнения

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial r^2}$$

с начальным условием  $C(0, r) = g(r)$  имеет вид

$$C(r, t) = \frac{1}{2\sqrt{\pi Dt}} \int_{-\infty}^{\infty} g(\xi) e^{-\frac{(x-\xi)^2}{4Dt}} d\xi.$$

Из этой формулы, в частности, следует, что если начальная концентрация была положительна только на конечном отрезке  $0 \leq r \leq l$ , то при любом  $t > 0$  концентрация в момент  $t$  будет положительна всюду на числовой прямой:  $-\infty < r < \infty$ . С помощью диффузии большие концентрации распространяются сравнительно медленно, в то время как малые — за малое время на большие расстояния. Следует иметь в виду, что уравнение (6.10) на очень малых интервалах времени плохо описывает процесс диффузии. Если рассматривается диффузия на конечном отрезке  $[0, l]$  с условием непроницаемости на концах, то любая начальная концентрация с ростом  $t$  стремится к равномерному распределению по отрезку.

#### **6.2.4. Решение неоднородного уравнения диффузии с нулевым начальным условием и нулевыми краевыми условиями**

Присутствие в правой части уравнения (6.10) члена  $f(r, t)$  означает наличие источника (или стока) вещества в данном месте через стенки трубки. Например, при описании процессов диффузии ионов вдоль мембраны возможен трансмембранный перенос ионов.

Для неоднородного уравнения диффузии

$$C_t = DC_{rr} + f(r, t)$$

с нулевым начальным условием

$$C(r, 0) = 0$$

и нулевыми краевыми условиями

$$C(0, t) = 0, C(l, t) = 0$$

решение  $C(r, t)$  также ищут в виде разложения в ряд Фурье по  $\sin \frac{\pi n}{l} r$ :

$$C(r, t) = \sum_{n=1}^{\infty} C_n(t) \sin \frac{\pi n}{l} r.$$

Здесь функции  $C_n(t)$  могут быть получены при подстановке предполагаемой формы решения в исходное уравнение (6.10), где функция  $f(r, t)$  также представляется в виде ряда Фурье:

$$f(r, t) = \sum_{n=1}^{\infty} f_n(t) \sin \frac{\pi n}{l} r.$$

Общее решение поставленной задачи имеет вид

$$C(r, t) = \sum_{n=1}^{\infty} \left[ \int_0^t e^{-\left(\frac{\pi n}{l}\right)^2 D(t-\tau)} f_n(\tau) d\tau \right] \sin \frac{\pi n}{l} r. \quad (6.29)$$

Как и в случае однородного уравнения, оно может быть представлено через функцию источника:

$$C(r, t) = \int_0^t \int_0^l G(r, \xi, t - \tau) f(\xi, \tau) d\xi d\tau, \quad (6.30)$$

$$G(r, \xi, t - \tau) = \frac{2}{l} \sum_{n=1}^{\infty} e^{-\left(\frac{\pi n}{l}\right)^2 D(t-\tau)} \sin \frac{\pi n}{l} r \sin \frac{\pi n}{l} \xi.$$

Легко видеть, что функция (6.30) совпадает с функцией (6.27). Различие состоит в том, что в случае формулы (6.27) изучалось однородное уравнение диффузии и поэтому рассматривался источник вещества, действующий лишь в момент времени  $t = 0$ , согласно начальному условию (6.11). Дальнейшее распределение вещества определялось в этом случае «пассивной»

диффузией по градиенту концентраций. В случае неоднородного уравнения (6.10) функция  $f(r, t)$  задает распределение источников вещества, действующих постоянно. Поэтому в выражении для  $C(r, t)$  через функцию источника необходимо суммировать действие мгновенных точечных источников во все моменты времени от  $t = 0$  до рассматриваемого момента  $t$  (интеграл по  $\tau$ ) и во всех точках одномерного реактора (интеграл по  $\xi$ ). Таким образом, исходя из физического смысла функции источника  $G(r, \xi, t)$  можно было бы сразу написать выражение (6.30) для функции, дающей решение неоднородного уравнения.

### 6.2.5. Общая краевая задача

Решение общей краевой задачи — уравнения (6.10) с начальными (6.11) и краевыми (6.12) условиями — сводится к решению задачи с нулевыми краевыми условиями. Для этого решение  $C(r, t)$  представляют в виде суммы двух функций:

$$C(r, t) = V(r, t) + v(r, t).$$

Здесь  $V(r, t)$  — известная функция:

$$V(r, t) = \mu_1(t) + \frac{r}{l} [\mu_2(t) - \mu_1(t)],$$

а  $v$  — неизвестная функция, которая определяется как решение уравнения

$$v_t = Dv_{rr} + \bar{f}(r, t), \quad \text{где } \bar{f}(r, t) = f(r, t) - [V_t - DV_{rr}].$$

Начальные условия для функции  $v$ :

$$v(r, 0) = \bar{\varphi}(r); \quad \bar{\varphi}(r) = \varphi(r) - V(r, 0),$$

а граничные условия — нулевые:

$$\bar{\mu}_1(t) = 0, \quad \bar{\mu}_2(t) = 0.$$

Метод нахождения функции  $v$  был разобран ранее.

Итак, мы рассмотрели аналитические методы решения однородных (типа (6.13)) и неоднородных (типа (6.10)) уравнений, описывающих диффузию одного вещества в одномерном реакторе. Решение представляется в виде интегралов, причем удобный для аналитического исследования вид решения может быть получен лишь в небольшом числе частных случаев. Еще более

сложной становится ситуация при изучении системы нескольких веществ, способных вступать в химические реакции и диффундировать в трехмерном пространстве, а именно с такими системами мы имеем дело в биологии. Однако некоторые выводы о свойствах решений могут быть сделаны на основании качественного исследования моделей. Одна из проблем, при решении которой оказываются эффективными методы качественного исследования, — изучение устойчивости стационарных состояний распределенных систем.

В любом случае условием возникновения в распределенных системах сложных пространственно-временных режимов является неустойчивость гомогенного стационарного состояния. Границы области параметров, в которой возникает такая неустойчивость, могут быть установлены на основе анализа линеаризованной системы аналогично локальным системам (см. гл. 4, 5).

### 6.2.6. Устойчивость гомогенных стационарных состояний нелинейных систем

Если некоторое стационарное состояние системы неустойчиво, то это означает, что с течением времени в системе устанавливается какой-либо иной режим. Для точечной системы это могут быть другие устойчивые стационарные состояния в мультистационарных системах, автоколебания или динамический хаос (см. гл. 2). В распределенных системах неустойчивость однородных в пространстве (гомогенных) стационарных решений может приводить к возникновению диссипативных структур, автоволновых процессов и квазистохастических режимов.

Стационарные, т. е. неизменные во времени, решения можно найти из условия обращения в нуль производных по времени:

$$\frac{\partial C_i}{\partial t} = 0.$$

Поясним, как ставится задача об устойчивости стационарных решений распределенных систем на примере одного уравнения с одной пространственной переменной. Пусть реакция происходит в тонкой трубке длины  $l$ . Уравнение, описывающее изменение переменной  $C$  в пространстве и во времени, имеет вид

$$\frac{\partial C}{\partial t} = f(C) + D \frac{\partial^2 C}{\partial r^2}. \quad (6.31)$$

Пусть краевые условия соответствуют непроницаемости трубки на торцах:

$$\frac{\partial C}{\partial r}(t, 0) = \frac{\partial C}{\partial r}(t, l) = 0. \quad (6.32)$$

Поскольку речь идет о стационарном решении (поведении переменной  $C$  при  $t \rightarrow \infty$ ), начальные условия не играют роли. Пусть  $C_0(r)$  — стационарное решение уравнения (6.31), т. е. решение задачи:

$$f(C) + D \frac{\partial^2 C}{\partial r^2} = 0$$

с краевым условием непроницаемости границ.

Зададим системе некоторое возмущение  $\delta(r)$ , т. е. выберем в качестве начальной функции в этой задаче функцию, близкую к  $C_0$ :

$$C(0, r) = C_0(r) + \delta(r); \quad |\delta(r)| \ll 1.$$

Пусть  $C_\delta(t, r)$  — решение задачи (6.31), (6.32) с такой начальной функцией.

Стационарное решение  $C_0(r)$  называется *устойчивым*, если для достаточно малых отклонений от стационарного состояния  $|\delta(r)|$  функция  $C_\delta(t, r)$  при всех  $t \geq 0$  мало отличается от  $C_0(r)$ .

Вблизи  $C_0(r)$  нелинейную функцию  $f(C)$  можно приблизить линейной функцией, используя первый член разложения по  $C$  в ряду Тейлора:

$$f(C) = f(C_0) + f'_c(C_0)(C - C_0).$$

Здесь  $C - C_0 = \delta(t, r)$ . При малых  $\delta(r)$  функция  $C_\delta(t, r)$  может быть представлена в виде

$$C_\delta(t, r) \approx C_0(r) + \delta(t, r),$$

где  $\delta(t, r)$  — решение линейной задачи:

$$\frac{\partial \delta(t, r)}{\partial t} = \frac{\partial^2 \delta(t, r)}{\partial r^2} + f'_c(C_0) \delta(t, r); \quad (6.33)$$

$$\delta(0, r) = \delta(r); \quad \frac{\partial \delta(t, 0)}{\partial r} = \frac{\partial \delta(t, l)}{\partial r} = 0. \quad (6.34)$$

Здесь и в дальнейшем считаем для краткости, что коэффициент диффузии  $D = 1$ .

Итак, для исследования устойчивости стационарных состояний распределенных систем нужно изучить поведение решения линейной задачи при  $t \rightarrow \infty$ . Как правило, свойства линейной задачи определяют устойчивость или неустойчивость решения соответствующей нелинейной системы. Исключение составляют

это имело место и при исследовании точечных систем.

Пусть задача (6.33), (6.34) имеет однородные по пространству стационарные решения. Рассмотрим вопрос об устойчивости таких решений. Для однородных стационарных решений  $f'_c(C_0) = A = \text{const}$ . Задача (6.33), (6.34) представляет собой линейное уравнение диффузии, которое отличается от рассмотренного ранее наличием в правой части слагаемого, которое является отклонением от стационарной концентрации  $\delta(t, r)$ , умноженным на число  $A$ , с соответствующими краевыми и начальными условиями.

Ранее было показано, что собственными функциями линейного уравнения диффузии с условиями непроницаемости на концах отрезка являются функции

$$\cos \frac{k\pi r}{l}, \quad k = 0, 1, \dots$$

Решение  $\delta(t, r)$  задачи (6.33), (6.34) с начальной функцией  $\delta(r)$  также можно представить в виде

$$\delta(r, t) = \sum_{k=0}^{\infty} a_k(t) \cos \frac{k\pi r}{l}.$$

Подставив это выражение в (6.33), получим следующее уравнение для нахождения  $a_k$ :

$$\frac{\partial a_k(t)}{\partial t} = \left( -\frac{k^2\pi^2}{l^2} + A \right) a_k(t), \quad a_k(0) = 1,$$

откуда

$$a_k(t) = \exp \left\{ \left( -\frac{k^2\pi^2}{l^2} + A \right) t \right\}. \quad (6.35)$$

Величины  $a_k$  задают временной характер нарастания или затухания соответствующей гармоники возмущения  $\delta(t, r)$ , в то время как множители  $\cos \frac{k\pi r}{l}$  определяют распределение начального отклонения вдоль пространственной координаты.

Если в формулах (6.35)  $A < 0$ , то при любом  $k = 1, 2, \dots$  функция  $\delta(t, r) \rightarrow 0$  при  $t \rightarrow \infty$ , какова бы ни была начальная функция  $\delta(r)$ . Таким образом, в этом случае любое малое возмущение однородного по пространству стационарного решения со временем затухает.

бых  $k$ , кроме  $k = 0$ . В такой системе будут затухать все гармоники  $\cos \frac{k\pi r}{l}$ ,  $k = 1, 2, \dots$  (относительно нулевой гармоники линейное приближение не дает ответа).

Если  $A > 0$ , то существует конечное число гармоник вида  $\cos \frac{k\pi r}{l}$ , которое приводит к развитию возмущений стационарного однородного решения. А именно, это только те гармоники, для которых  $\frac{k^2\pi^2}{l^2} < A = f'(C_0)$ . Если начальное возмущение не содержит этих гармоник, то оно со временем будет исчезать.

Вспомним, что в начале анализа было принято  $D = 1$ . Если учесть коэффициент диффузии в системе, где  $f'(C_0) > 0$ , возможны незатухающие начальные возмущения, номера незатухающих гармоник  $k^*$  в соответствии с (6.35) можно определить по формуле

$$k^* < \sqrt{\frac{f'(C_0)l^2}{D\pi^2}}.$$

Таким образом, номер наивысшей незатухающей гармоники тем больше, чем длиннее реактор, и тем меньше, чем выше значение коэффициента диффузии. Незатухающие гармоники, развиваясь, могут приводить систему к установлению пространственно неоднородных диссипативных структур или автоволновых режимов.

### 6.2.7. Распределенная модель популяции с наименьшей критической численностью

В качестве примера исследуем устойчивость однородных стационарных решений одного уравнения с одной пространственной переменной

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial r^2} + c(a - c)(c - b). \quad (6.36)$$

Соответствующее точечное уравнение

$$\frac{dc}{dt} = c(a - c)(c - b)$$

имеет три стационарных решения:  $\bar{c}_1 = 0$ ,  $\bar{c}_2 = a$ ;  $\bar{c}_3 = b$ .

Для  $0 < a < b$  вид функции  $f(c)$  в этом уравнении соответствует графику, изображенному на рис. 6.2.



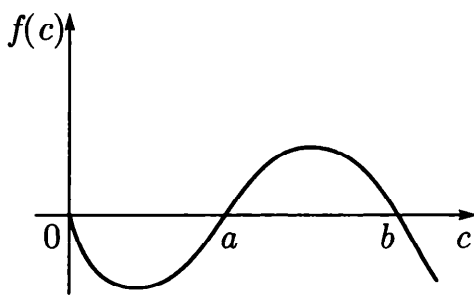


Рис. 6.2. Функция  $f(c)$  для уравнения (6.36)

Предположим, что процесс, описываемый уравнением (6.36), происходит в трубке длины  $l$  ( $0 \leq r \leq l$ ) с непроницаемыми концами. Это накладывает граничные условия:

$$\frac{\partial c}{\partial r}(t, 0) = \frac{\partial c}{\partial r}(t, l) = 0.$$

Исследуем теперь стационарные решения уравнения (6.36). Для их определения имеем обыкновенное дифференциальное уравнение, независимой переменной которого является пространственная координата  $r$ :

$$D \frac{d^2 \bar{c}}{dr^2} + \bar{c}(\bar{c} - a)(b - \bar{c}) = 0, \quad 0 \leq r \leq l, \quad (6.37)$$

причем  $\frac{d\bar{c}}{dr}(0) = 0$ ;  $\frac{d\bar{c}}{dr}(l) = 0$ .

Исследуем устойчивость однородных стационарных состояний системы, характеризуемой уравнением (6.37). В соответствии с процедурой, описанной ранее, вблизи стационарного решения  $\bar{c}(r)$  аппроксимируем нелинейную функцию

$$f(c) = c(a - c)(c - b)$$

линейной функцией  $f(c) = f(\bar{c}) + f'_c(\bar{c})(c - \bar{c})$ . Прежде всего имеем три стационарных решения, которые являются положениями равновесия соответствующей точечной системы:  $\bar{c}_1 = 0$ ,  $\bar{c}_2 = a$ ,  $\bar{c}_3 = b$ . Рассмотрим знак производной функции  $f'_c(\bar{c})$  в стационарных точках. Как видно из рис. 6.2, при  $0 < a < b$  значения  $f'(0)$  и  $f'(b)$  отрицательны и, следовательно, стационарные решения  $c = 0$ ,  $c = b$  устойчивы. Решение  $c = a$  неустойчиво, так как  $f'(a) > 0$ . Заметим, что существует только конечное число гармоник вида  $\cos \frac{k\pi r}{l}$ , которые приводят к развитию возмущений стационарного решения  $c = a$ . Это те гармоники, для которых  $\frac{k^2 \pi^2}{l^2} < f'(a)$ . Если начальное возмущение не содержит этих гармоник, то оно со временем будет затухать.

При не очень малых длинах реактора  $l$  имеются еще и неоднородные по пространству решения. Чтобы найти их, обозначим через  $F(c)$  функцию, для которой  $F'(c) = c(c - a)(b - c)$ . Умножив уравнение (6.37) на  $\frac{d\bar{c}}{dr}$  и интегрируя по  $r$ , получим

$$\frac{D}{2} \left( \frac{d\bar{c}}{dr} \right)^2 + F(\bar{c}) = \frac{A}{2} = \text{const.}$$

Здесь  $A$  — произвольная постоянная. Решив полученное урав-

нение, имеем соотношение:  $\pm \int_{c_0}^{\bar{c}} \frac{dc}{\sqrt{\frac{1}{D}(A - 2F(c))}} = r$ , которое ха-

рактеризует все решения уравнения (6.37). Постоянные  $c_0$  и  $A$  определяются из двух граничных условий:  $\frac{d\bar{c}}{dr}(0, t) = 0$ ;  $\frac{d\bar{c}}{dr}(l, t) = 0$ . Можно доказать, что при  $l$ , бóльших некоторого  $l_0 > 0$ , эти граничные условия можно удовлетворить и, следовательно, существует пространственно неоднородное решение задачи. С увеличением длины реактора число различных стационарных решений возрастает.

### 6.3. Система двух уравнений реакция — диффузия

Поведение распределенных систем из двух уравнений может быть чрезвычайно разнообразным: возмущения в виде бегущего импульса, генерация волн автономными источниками импульсной активности, стоячие волны, синхронные автоколебания во всем пространстве, квазистохастические волны и диссипативные структуры — стационарные неоднородные распределения переменных в пространстве. Как и в случае точечных систем, важное место в изучении распределенных систем занимает исследование устойчивости стационарного состояния.

#### 6.3.1. Устойчивость однородных стационарных решений

Рассмотрим распределенную систему, в которой имеется два вещества, т. е. две кинетические переменные  $x$  и  $y$ , которые претерпевают химические превращения и, кроме того, могут диффундировать в реакционном объеме. В случае одномерного реактора такая система может быть описана системой уравнений:

$$\begin{cases} \frac{\partial x}{\partial t} = P(x, y, r) + D_x \frac{\partial^2 x}{\partial r^2}; \\ \frac{\partial y}{\partial t} = Q(x, y, r) + D_y \frac{\partial^2 y}{\partial r^2}, \end{cases} \quad (6.38)$$

где  $r$  — пространственная переменная.

Пусть краевыми условиями являются условия непроницаемости торцов одномерного реактора:

$$\frac{\partial x}{\partial r} \Big|_{r=0} = \frac{\partial x}{\partial r} \Big|_{r=l} = \frac{\partial y}{\partial r} \Big|_{r=0} = \frac{\partial y}{\partial r} \Big|_{r=l} = 0.$$

Характер поведения такой системы со временем может быть различным. В простейшем случае при  $t \rightarrow \infty$  во всех точках реактора установятся одинаковые концентрации  $\bar{x}$  и  $\bar{y}$ , т. е. система придет к своему устойчивому однородному в пространстве (гомогенному) стационарному состоянию. Если же однородное стационарное состояние неустойчиво, при  $t \rightarrow \infty$  могут реализоваться другие режимы — либо иные однородные стационарные состояния, либо неоднородные стационарные (диссипативные) структуры. Возможно также, что в каждой точке пространства переменные с течением времени не стремятся к определенным значениям, а в системе устанавливается автоколебательный или квазистохастический режим. Таким образом, как и в случае точечных моделей, первым необходимым этапом изучения модели распределенной системы является исследование устойчивости ее однородного стационарного состояния.

Рассмотрим пространственно однородное решение системы (6.38):

$$x = \bar{x}_k = \text{const}; \quad y = \bar{y}_k = \text{const},$$

где  $\bar{x}$  и  $\bar{y}$  являются корнями алгебраической системы уравнений

$$\begin{cases} P(x, y) = 0; \\ Q(x, y) = 0 \end{cases}$$

и, следовательно, особыми точками точечной системы.

В подразд. 6.2 была рассмотрена устойчивость гомогенного стационарного состояния в случае одного уравнения. Такое состояние устойчиво, если малые возмущения сил (в том числе и распределенных в пространстве), действующих на систему, вызывают малые возмущения ее решений. Предполагается, что эти возмущения остаются малыми при любом  $t \rightarrow \infty$ .

Как и в случае одного уравнения, исследование устойчивости будем проводить на основе анализа линеаризованной системы уравнений. Пусть  $\xi_k(t, r)$  и  $\eta_k(t, r)$  — малые отклонения от пространственно однородных решений  $\bar{x}_k$  и  $\bar{y}_k$ . Тогда для  $\xi_k$  и  $\eta_k$  можно записать распределенную линеаризованную систему (нижние индексы  $k$  для краткости опускаем):

$$\begin{cases} \frac{\partial \xi}{\partial t} = a\xi + b\eta + D_\xi \frac{\partial^2 \xi}{\partial r^2}; \\ \frac{\partial \eta}{\partial t} = c\xi + d\eta + D_\eta \frac{\partial^2 \eta}{\partial r^2}. \end{cases} \quad (6.39)$$

Коэффициенты  $a, b, c, d$  в линеаризованных уравнениях (6.39) определяются, как и в случае обыкновенных дифференциальных уравнений (см. 4.3):

$$\begin{aligned} a &= \frac{\partial P(\bar{x}, \bar{y})}{\partial x}; & b &= \frac{\partial P(\bar{x}, \bar{y})}{\partial y}; \\ c &= \frac{\partial Q(\bar{x}, \bar{y})}{\partial x}; & d &= \frac{\partial P(\bar{x}, \bar{y})}{\partial y}. \end{aligned}$$

Коэффициенты диффузии:  $D_x = D_\xi$ ,  $D_y = D_\eta$ .

Решение будем искать в виде

$$\xi(t, r) = Ae^{pt}e^{ikr}; \quad \eta(t, r) = Be^{pt}e^{ikr}. \quad (6.40)$$

Здесь множитель  $e^{ikr}$  характеризует отклонение величин переменных от однородного стационарного состояния в точке с координатой  $r$  для собственных функций, соответствующих волновому числу  $k$ . Для трубки длиной  $l$  (см. подразд. 6.2) волновое число принимает дискретные значения  $k = k_n = \frac{\pi n}{l}$ . Для бесконечного одномерного реактора волновые числа  $k$  меняются непрерывно от 0 до  $\infty$ . Множитель  $e^{pt}$  характеризует поведение отклонения от стационарного состояния во времени. После подстановки выражений (6.40) в (6.39) и сокращения на  $e^{pt}e^{ikr}$  получим

$$Ap = aA + bB - D_\xi k^2 A;$$

$$Bp = cA + dB - D_\eta k^2 B,$$

или

$$A(p - a + D_\xi k^2) - bB = 0;$$

$$cA - B(p - d + D_\eta k^2) = 0.$$

Величины  $A, B$  тождественно не равны нулю только в том случае, если определитель системы равен нулю:

$$(p - a + k^2 D_\xi)(p - d + k^2 D_\eta) - bc = 0.$$

Полученное уравнение называется *дисперсионным уравнением*. Его решение

$$p_{1,2} = \frac{a+d - (D_{\xi} - D_{\eta})k^2 \pm \sqrt{[a-d - k^2(D_{\xi} - D_{\eta})]^2 + 4bc}}{2} \quad (6.41)$$

определяет поведение системы во времени через параметры системы и волновое число  $k$ . Как и в случае точечных систем, знак действительной части  $p_{1,2}$  показывает, устойчивым или неустойчивым будет исследуемое однородное стационарное решение. В области параметров, где оба  $\operatorname{Re} p_{1,2} < 0$ , решение устойчиво.

К устойчивому стационарному состоянию переменные могут приближаться колебательным или бесколебательным образом, в зависимости от знака подкоренного выражения в формуле (6.41). Если подкоренное выражение отрицательно, то корни  $p_{1,2}$  комплексно-сопряженные и в каждой точке пространства имеют место затухающие колебания переменных вокруг стационарных значений. Если же подкоренное выражение положительно, то корни  $p_{1,2}$  действительные отрицательные и имеет место бесколебательное стремление переменных к стационарным значениям. В случае когда действительные части  $p_{1,2}$  положительны или имеют разные знаки, однородное стационарное состояние является неустойчивым. Здесь также возможны различные типы неустойчивостей: если  $p_{1,2}$  действительны и положительны — неустойчивость типа «узел», если  $p_{1,2}$  — комплексные и  $\operatorname{Re} p_{1,2} > 0$  — неустойчивость типа «фокус», и наконец, если  $p_1$  и  $p_2$  имеют разные знаки — неустойчивость типа «седло».

На рис 6.3,  $a - e$  изображены возможные виды зависимости действительной части  $p_{1,2}$  от волнового числа  $k$ . На графиках выделены три области: I — оба корня имеют положительную действительную часть ( $\operatorname{Re} p_{1,2} > 0$ ); II — один корень имеет положительную, а другой — отрицательную действительную часть ( $\operatorname{Re} p_1 > 0, \operatorname{Re} p_2 < 0$ ); III — оба корня имеют отрицательную действительную часть: ( $\operatorname{Re} p_{1,2} < 0$ ).

На рис. 6.3  $a, б, e$  оба характеристических числа  $p_{1,2}$  действительны для любых волновых чисел  $k$ , а потому  $\operatorname{Re} p_1$  и  $\operatorname{Re} p_2$  различны во всей области изменения параметров. На рис. 6.3,  $в - д$  существуют две области — область, в которой характеристические числа  $p_{1,2}$  комплексно-сопряженные и потому их действительные части равны, и область, где оба числа  $p_{1,2}$  действительны и различны.

На рисунках указаны значения волнового числа, которые соответствуют изменению типа устойчивости системы. Рассмотрим, какие изменения могут происходить при увеличении числа  $k$ .

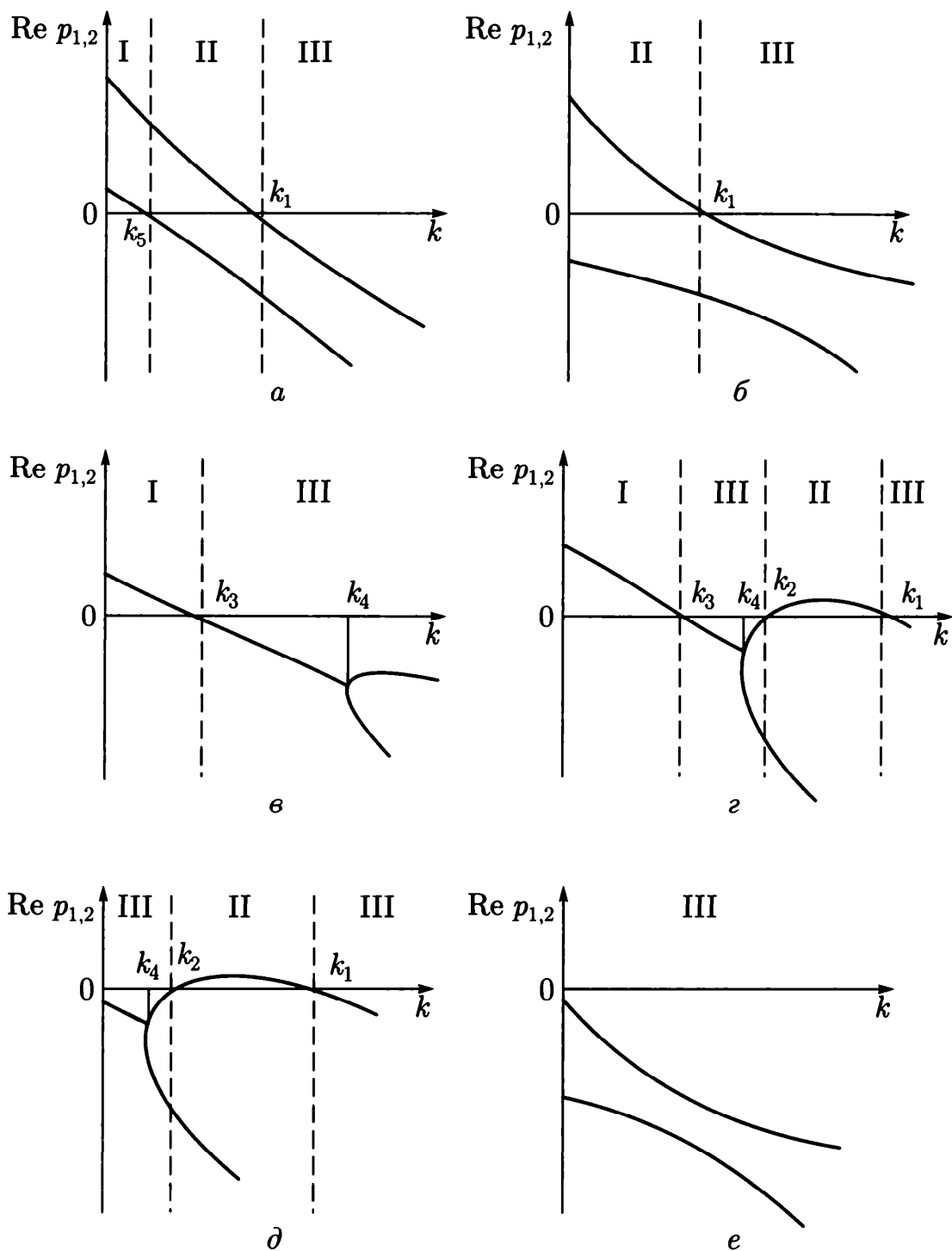


Рис. 6.3. Различные типы зависимости действительной части корней дисперсионного уравнения для системы (6.39) от волнового числа  $k$  [Ю. М. Романовский и [др.], 2004]

Волновое число  $k_1$  соответствует значению, при котором один из вещественных корней (большой) переходит из положительной в отрицательную область, это соответствует переходу из области седловой неустойчивости (два положительных действительных корня разных знаков) II в область устойчивого узла III (см. рис. 6.3, а, б, г, д). Волновое число  $k_2$ , наоборот, соответствует переходу из области устойчивого узла III в область седловой неустойчивости II (см. рис. 6.3, г, д). Волновое число  $k_3$  характеризует переход из области колебательной неустойчи-

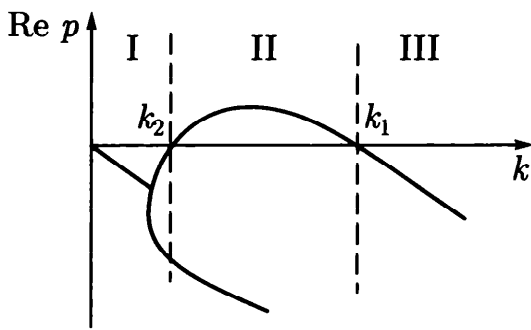


Рис. 6.4. Зависимость действительной части корней дисперсионного уравнения  $\text{Re } p$  от волнового числа  $k$  в случае неустойчивости Тьюринга

вости I ( $\text{Re } p_1 = \text{Re } p_2 > 0$ ) в область устойчивых колебаний:  $\text{Re } p_1 = \text{Re } p_2 < 0$  (см. рис. 6.3, в, г). Волновое число  $k_4$  определяет превращение колебательной системы в бесколебательную — рождение двух разных значений  $\text{Re } p_{1,2}$  из одного, соответствующего двум комплексно-сопряженным числам  $p_{1,2}$  (см. рис. 6.3, в, г, д). Волновое число  $k_5$  характеризует переход из области неустойчивого узла в область седловой неустойчивости II (см. рис. 6.3, а).

Возмущения с длиной волны из области I (см. рис. 6.3, а, в, г) в нелинейной распределенной системе могут приводить к возникновению бегущих волн конечной амплитуды, стоячих волн, ведущих центров. В системе с двумя переменными появление таких режимов возможно лишь в случае, когда точечная система ( $D_\xi = 0, D_\eta = 0$ ) является автоколебательной.

На рис. 6.3, г, д имеется область II изменения параметра  $k$ :  $k_2 < k < k_1$ , в которой одно из характеристических чисел положительно, а другое — отрицательно. Для значений волновых чисел  $k < k_1$  и  $k > k_2$  такая система имеет устойчивое гомогенное стационарное состояние, а в интервале  $k_2 < k < k_1$  демонстрирует неустойчивость. Область неустойчивости такого рода называется областью **неустойчивости Тьюринга**.

Зависимость  $\text{Re } p_{1,2}$  от волнового числа в случае неустойчивости Тьюринга изображена на рис. 6.4. Границы области неустойчивости на оси абсцисс определяются значениями  $k_1, k_2$ , для которых одна из действительных частей  $\text{Re } p_{1,2}$  обращается в нуль:

$$k_{1,2} = \left[ (aD_\eta + dD_\xi) \pm \sqrt{(aD_\eta + dD_\xi)^2 - 4D_\xi D_\eta (ad - bc)} \right] \frac{1}{2D_\xi D_\eta}. \quad (6.42)$$

Флуктуации от однородного стационарного состояния в области неустойчивости Тьюринга могут привести к нарушению го-

могенности системы и возникновению стационарной неоднородной структуры.

При анализе устойчивости гомогенных стационарных состояний систем более высокого порядка

$$\frac{\partial x_i}{\partial t} = F_i(x_1, x_2, \dots, x_n) + D \frac{\partial^2 x_i}{\partial r^2} \quad (6.43)$$

поступают таким же образом, как и при исследовании систем второго порядка. Пусть координаты особой точки системы:

$$\{\bar{x}_{1m}, \bar{x}_{2m}, \dots, \bar{x}_{nm}\}, \quad m = 1, 2, \dots, M,$$

где  $m$  — номер особой точки. Возмущение представляется в виде суперпозиции волн вида

$$x'_{im} = x_i - \bar{x}_{im} = \alpha_{im} \exp \left( p_{mk} t + i \frac{\pi k}{L} r \right) \quad (6.44)$$

( $i$  — мнимая единица),

где  $k$  — волновое число, определяющее длину волны  $\lambda_{mk} = \frac{2L}{k}$ .

Подставив (6.44) в линеаризованную систему уравнений, записанную в системе координат с началом в  $m$ -й особой точке, с учетом условия существования нетривиальных решений такой системы получим дисперсионное уравнение, связывающее комплексные частоты  $p_{mk} = \delta_{mk} \pm i\omega_{mk}$ , длины волн  $\lambda_{mk}$  (волновые числа  $k$ ) и коэффициенты системы (6.43):

$$p_{mk}^n + q_{n-1}(k^2)p_{mk}^{n-1} + \dots + q_0(k^2) = 0.$$

Если исследуемое стационарное состояние неустойчиво, то имеется хотя бы одно значение комплексной частоты  $p_{mk}$ , для которой  $\text{Re } p_{mk} > 0$ .

Количество корней дисперсионного уравнения с положительной действительной частью определяет тип неустойчивости системы. Если имеется четное число корней  $p_{mk}$  с  $\delta > 0$ , неустойчивость называют колебательной. Нечетному числу таких корней соответствует неустойчивость Тьюринга, приводящая к образованию стационарных неоднородных структур, названных диссипативными структурами (ДС). Этот термин подчеркивает термодинамический аспект проблемы — ДС рождаются и существуют в термодинамически открытых системах за счет диссипативных процессов утилизации энергии и энтропии [Г. Николис, И. Р. Пригожин, 1979].



### 6.3.2. Распределенный брюсселятор

В гл. 5 (см. 5.6.3) была изучена простейшая модельная автоколебательная система «брюсселятор», описывающая химическое взаимодействие двух веществ по схеме, включающей реализацию кубической нелинейности:



Рассмотрим, каким может быть пространственно-временное поведение системы, если в каждой точке локальное взаимодействие описывается уравнениями типа «брюсселятор». Пусть реакции протекают в узкой длинной трубке (одномерном реакторе), вдоль которой вещества могут диффундировать. Коэффициенты диффузии  $D_x$ ,  $D_y$  будем считать постоянными параметрами системы. Концы трубки, так же как и ее стенки, непроницаемы для веществ, участвующих в реакции. Уравнения, описывающие распределенный брюсселятор, имеют вид

$$\begin{aligned} \frac{\partial X}{\partial t} &= A + X^2Y - (B + 1)X + D_x \frac{\partial^2 X}{\partial r^2}; \\ \frac{\partial Y}{\partial t} &= BX - X^2Y + D_y \frac{\partial^2 Y}{\partial r^2}, \end{aligned} \quad (6.45)$$

где  $r$  — пространственная координата.

Напомним, что для точечной системы имеется одно стационарное состояние, которое характеризуется значениями концентраций:

$$\bar{X} = A; \quad \bar{Y} = \frac{B}{A}.$$

Таковыми будут концентрации во всех точках реактора, если однородное стационарное состояние системы устойчиво.

Для исследования условий потери устойчивости однородного по пространству решения введем переменные, характеризующие малые отклонения системы от однородного решения:

$$\xi = X - A; \quad \eta = Y - \frac{B}{A}.$$

Линеаризуем систему (6.45), и решение полученной системы будем искать в виде

$$\begin{aligned} \xi(t, r) &= C_1 e^{pt+ikr}, \\ \eta(t, r) &= C_2 e^{pt+ikr}. \end{aligned}$$

Величины  $p$  и  $k$  связаны дисперсионным уравнением, которое позволяет определить характер устойчивости исследуемого гомогенного решения. Дисперсионное уравнение для брюсселятора имеет вид

$$(p - B - 1 + k^2 D_\xi)(p + A^2 + k^2 D_\eta) + BA^2 = 0. \quad (6.46)$$

Если уравнение (6.46) имеет два действительных корня, причем один из них  $p_1 < 0$ , а второй  $p_2 > 0$ , то система в области гомогенного стационарного решения имеет неустойчивость седлового типа (неустойчивость Тьюринга). Условия существования такой неустойчивости выполняются при

$$B \geq 1, D_x \leq \frac{D_y(1 - \sqrt{B})^2}{A^2}.$$

Границы области волновых чисел  $k$ , в которой реализуется неустойчивость Тьюринга, задаются выражением

$$k_{1,2}^2 = \frac{1}{2D_x D_y} \left\{ [(B - 1)D_y - D_x A^2] \pm \sqrt{[(B - 1)D_y - D_x A^2]^2 - 4A^2 D_x D_y} \right\}.$$

Именно в этой области система (6.45) образует диссипативные структуры.

Аналитическое исследование устойчивости неоднородных стационарных решений представляет значительные трудности, и в основном для этой цели используют асимптотические методы. Так, устойчивость диссипативных структур в брюсселяторе исследовали методом малых возмущений. Предполагая, что в системе существует диссипативная структура, которая носит квазигармонический характер [В. А. Васильев, Ю. М. Романовский, 1976], стационарные решения представляли в виде

$$\bar{x}(r) = a_1 + a_2 \cos kr; \quad \bar{y}(r) = b_1 + b_2 \cos kr + c_2 \sin kr,$$

а малые возмущения описывали формулами

$$\xi(t, r) = \sum_{j=0}^n \xi_j(t) \cos ikr; \quad \eta(t, r) = \sum_{j=0}^n \eta_j(t) \cos ikr.$$

В результате этих и ряда других исследований было показано, что при  $k_{\min} < k < k_{\max}$  наблюдаются устойчивые структуры, для брюсселятора

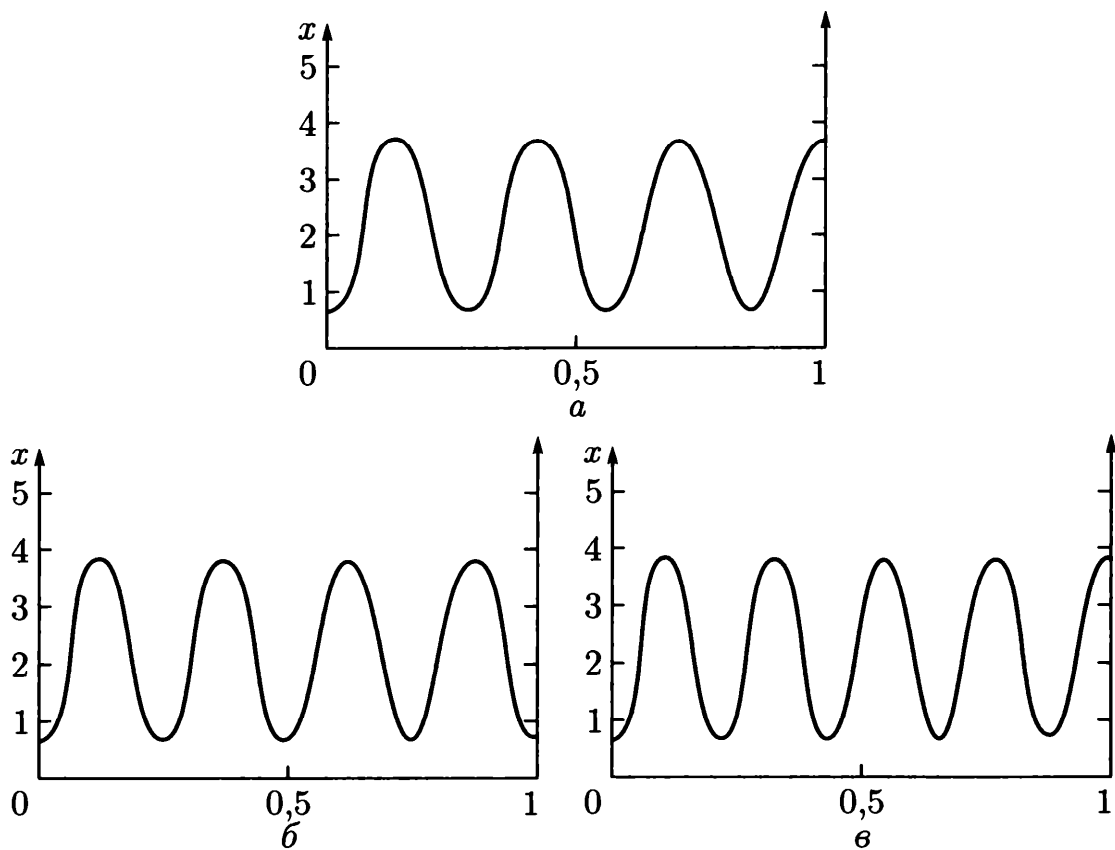


Рис. 6.5. Стационарные диссипативные структуры, полученные при одинаковых значениях параметров путем наложения на однородное стационарное состояние локализованного возмущения [Г. Николис, И. Р. Пригожин, 1979]

$$k_{\min} = \frac{B - 1}{D_x}; \quad k_{\max} = \frac{A^2}{D_y(B - 1)}.$$

Компьютерные эксперименты показали, что в отсутствие потоков на границах в системе может возникать несколько диссипативных структур в зависимости от локализации возмущений однородного состояния. Стационарные профили переменной  $X$  для различных возмущений представлены на рис. 6.5.

Отрезок  $\{0, 1\}$  разбивали на 101 одинаковых интервала, после чего возмущение одного знака и одинаковой амплитуды налагались в точках: 9, 21, 48, 72 (рис. 6.5, а); 9, 17, 34, 43 (рис. 6.5, б) и 9, 55, 70 (рис. 6.5, в)

### 6.3.3. Локализованные диссипативные структуры

Описанные диссипативные структуры распространяются на всю систему. Этот факт является следствием предположения о том, что концентрации исходных веществ реакций  $A$  и  $B$  поддерживаются постоянными во всех точках пространства. Такая ситуация является идеализированной. В реальном экспери-

менте реагенты вводятся в реакционный объем через границы. Это означает, что вещества  $A$  и  $B$  будут диффундировать в среде, что вызовет установление их концентрационных профилей. Неоднородное пространственное распределение исходных веществ приводит к локализации диссипативных структур внутри определенных границ.

В случае когда вещество  $B$  равномерно распределено по объему, а концентрация вещества  $A$  поддерживается постоянной на границе, систему уравнений (6.45) следует дополнить уравнением, описывающим потребление и диффузию вещества  $A$ . Получим систему из трех уравнений:

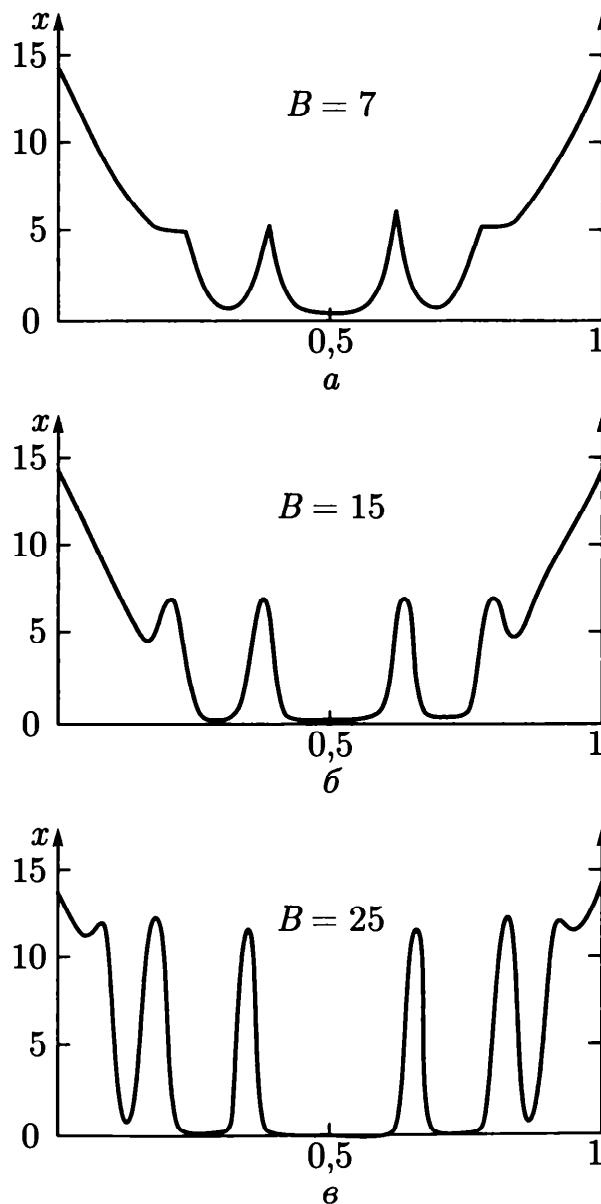


Рис. 6.6. Локализованная диссипативная структура, полученная при численном решении системы (6.47) для последовательно увеличивающихся значений параметра  $B$ . Параметры системы:  $D_A = 0,026$ ;  $D_x = 1,052 \cdot 10^{-3}$ ;  $D_y = 5,26 \cdot 10^{-3}$ ;  $l = 1$  [Г. Николис, И. Р. Пригожин, 1979]

$$\begin{cases} \frac{\partial X}{\partial t} = A + X^2 Y - (B + 1)X + D_x \frac{\partial^2 X}{\partial r^2}; \\ \frac{\partial Y}{\partial t} = BX - X^2 Y + D_y \frac{\partial^2 Y}{\partial r^2}; \\ \frac{\partial A}{\partial t} = -A + D \frac{\partial^2 A}{\partial r^2} \quad (0 \leq r \leq l) \end{cases} \quad (6.47)$$

с граничными условиями

$$A(0) = A(l) = \bar{A}.$$

Концентрационные профили вещества  $X$ , полученные в соответствии с системой (6.47) для разных значений параметра  $B$ , представлены на рис. 6.6. На границах поддерживались значения переменных, соответствующие гомогенному стационарному состоянию.

## 6.4. Распространение фронтов, импульсов и волн

### 6.4.1. Распространение волны в системах с диффузией

В популяционной генетике к задаче распространения волны приводит рассмотрение области, занятой особями — носителями доминантного гена. Подобные задачи встречаются в экологии при изучении распространения вида. Эффекты, возникающие при распространении волн в активной кинетической среде, играют особую роль в процессах передачи информации и управления в биологических системах.

Рассмотрим дифференциальное уравнение

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial r^2} + f(C) \quad (6.48)$$

с начальным условием

$$C(0, r) = g(r) = \begin{cases} 1, & r < 0; \\ 0, & r > 0. \end{cases}$$

Начальное условие такого вида означает, например, что обширная территория (в одномерном случае — полупрямая  $r < 0$ ) занята доминантным геном, концентрация которого близка к единице. В начальный момент времени область, где  $C = 1$ , имеет резкую границу, и при всех  $r > 0$  концентрация  $C = 0$ . При

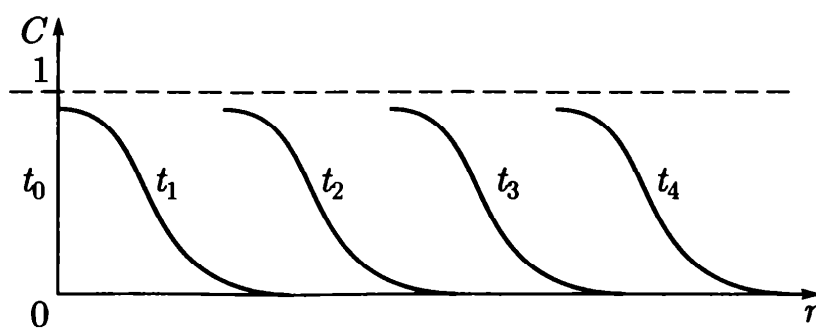


Рис. 6.7. Профиль плотности  $C$  в зависимости от координаты  $r$  в последовательные моменты времени:  $t_4 > t_3 > t_2 > t_1 > t_0 = 0$

$t = 0$  начинается распространение «волны» ненулевых концентраций доминантного гена в область  $r > 0$ , которое является следствием взаимодействия двух процессов: случайного перемещения особей (диффузии частиц) и размножения, описываемого функцией  $f(C)$ . Эти процессы ведут к перемещению области плотностей, близких к единице, с возрастанием  $t$  слева направо. Примерная картина профиля плотности  $C$  в зависимости от координаты  $r$  в разные моменты времени изображена на рис. 6.7.

В начальный момент времени  $t_0 = 0$  кривая имеет вид ступеньки. С течением времени фронт волн перемещается вправо, причем его форма приближается к определенной предельной кривой. А. Н. Колмогоров, И. Г. Петровский и Н. С. Пискунов [А. Н. Колмогоров и [др.], 1975] решили задачу (1937) о предельной скорости перемещения фронта волны и определили предельную форму фронта. В качестве функции  $f(C)$  в этой работе рассматривали функцию, равную нулю при  $C = 0$  и  $C = 1$  и положительную в промежуточных точках. Вид такой функции показан на рис. 6.8, а.

Сделанные относительно  $f(C)$  предположения означают, что при малых  $C$  концентрация резко нарастает за счет функции размножения  $f(C)$ . При дальнейшем возрастании  $C$  функция  $f(C)$  — скорость роста концентрации  $C$  — начинает убывать, и при  $C$ , близких к единице, наступает насыщение. Изменение вида функции  $C(t)$  в зависимости от времени в фиксированных точках пространства  $r_i^*$  изображено на рис. 6.9.

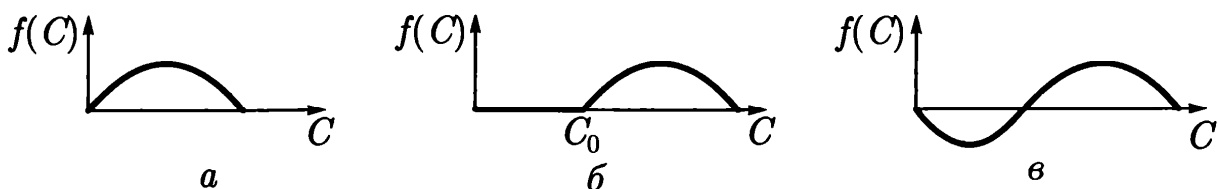


Рис. 6.8. Возможные типы функции  $f(C)$  для уравнения (6.48)

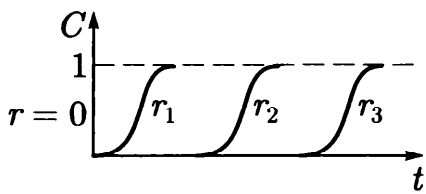


Рис. 6.9. Вид функции  $C(t)$  в фиксированных точках пространства  $r_i$

Продолжительность лаг-периода (времени запаздывания роста) функции  $C(r, t)$  зависит от координаты  $r$ . Как только благодаря диффузии малые, но конечные концентрации достигли точки  $r^*$ , начинается бурный рост  $C$  до значений, близких к единице, в соответствии с функцией  $f(C)$ .

Таким образом, малые концентрации, которые распространяются за счет диффузии, увеличиваются за счет нелинейных свойств локальной системы. Взаимодействие этих двух процессов приводит к тому, что волна концентрации, близкой к единице, движется слева направо (см. рис. 6.9). В работе [А. Н. Колмогоров и [др.], 1975] установлено, что предельная скорость распространения фронта волны для функции, изображенной на рис. 6.8, *a*, равна:

$$\lambda_0 = 2\sqrt{Df'(0)}. \quad (6.49)$$

Чтобы найти предельную форму фронта распространения концентраций, вводят так называемую *автоволновую переменную*:  $z = r - \lambda t$  и ищут решение уравнения (6.48) в виде

$$C(t, r) \approx V(r - \lambda t). \quad (6.50)$$

Предельная форма кривой плотности задается решением уравнения

$$D \frac{d^2 V}{dz^2} + \lambda_0 \frac{dV}{dz} + f(V) = 0, \quad (6.51)$$

которое обращается в нуль при  $z = +\infty$  и в единицу при  $z = -\infty$ . Такое решение  $V(z)$  всегда существует и единственно с точностью до преобразования  $z' = z + A$  ( $A$  — произвольная постоянная), не меняющего форму кривой.

Автоволновое решение (6.50) обладает тем свойством, что при изменении  $t$  форма кривой, изображающей зависимость  $V(t)$ , не меняется, а сама эта кривая перемещается слева направо со скоростью  $\lambda$ . Можно показать, что уравнение (6.50) имеет решение, удовлетворяющее граничным условиям при  $z \rightarrow \pm\infty$ , лишь при  $\lambda = \lambda_0$  (формула (6.49)). За счет множителя  $\sqrt{f'(0)}$  скорость распространения волны (6.49) может быть существенно больше, чем скорость, с которой распространяются не очень малые концентрации за счет диффузии. Это увеличение скорости распространения происходит вследствие действия «размножи-

теля», функционирование которого описывается точечной (локальной) системой

$$\frac{\partial C}{\partial t} = f(C).$$

В ряде моделей возникают функции иного вида, чем изображенные на рис. 6.8, *а*. Например функция на рис. 6.8, *б* описывает случай, когда размножение частиц начинается не при любой сколь угодно малой концентрации, а только при достижении некоторой достаточно большой концентрации  $C_0$ . Такая функция встречается в некоторых моделях химической кинетики и моделях передачи сигналов в биологических системах, когда реакция начинается только в том случае, когда концентрация достигает порогового значения. Эта функция также применяется в теории горения, пороговое значение переменной величины здесь имеет смысл температуры возгорания. В случае когда начальное условие задано «ступенькой» (см. рис. 6.8, *б*), в распределенной системе также распространяется концентрационная волна.

Случай, изображенный на рис. 6.8, *в*, встречается в ряде задач популяционной генетики и экологии. Отрицательность функции  $f(C)$  при малых  $C$  описывает, например, эффект, связанный с тем, что при малой концентрации скорость размножения мала, так как мала частота встреч особей разного пола. Поэтому за счет смертности при малых  $C$  скорость изменения численности отрицательна. Именно такой вид функции скорости роста был рассмотрен в гл. 4 (см. рис. 4.10, *б*) при изучении популяций с наименьшей критической численностью. В случае, представленном на рис. 6.8, *в*, для распространения концентрационной волны постоянной амплитуды нужно еще потребовать,

чтобы величина  $\int_0^1 f(C) dC$  была положительна, в противном случае будет распространяться область малых значений концентраций.

В важном частном случае  $f(C) = C(C - 1)(a - C)$  скорость волны, которая устанавливается при больших  $t$ , можно вычислить явно. Именно этот тип функции используется в уравнении Фитцхью — Нагумо для описания распространения нервного импульса. Модель этого чрезвычайно важного для биологии явления будет рассмотрена в 6.4.2. В этом случае скорость распространения волны оказывается равной  $\sqrt{\frac{D}{2}} (1 - 2a)$ .



### 6.4.2. Задача распространения нервного импульса

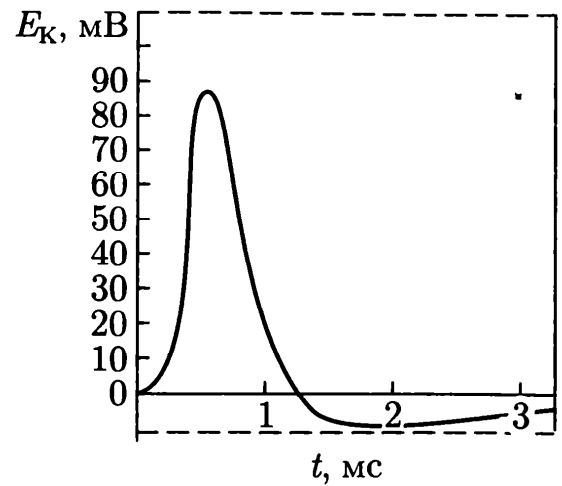
Основным свойством нервных клеток является способность к проведению нервного импульса — распространяющейся по нервному волокну электрической волны. Свойство проведения обусловлено тем, что элементарный объем мембраны представляет собой «локальный активный элемент», обладающий способностью усиливать проходящий электрический сигнал.

Основополагающая работа по изучению механизмов принадлежит А. Ходчкину и А. Хаксли [A. L. Hodgkin, A. F. Huxley, 1952], которые провели эксперименты и разработали модель, описывающую распространение нервного импульса. Их исследование, удостоенное Нобелевской премии (1963), было выполнено на гигантском аксоне кальмара. Аксон представляет собой цилиндрическую трубку диаметром 50—70 Å, вдоль внешней поверхности которой распространяется электрический импульс. В настоящее время известно, что приложение внешнего напряжения вызывает конформационные изменения в белковых комплексах — каналах, осуществляющих трансмембранный перенос ионов: Na, K, Ca, Cl и др. Каналы начинают «работать» — переносить ионы через мембрану, при этом происходит локальное изменение потенциала, которое носит обратимый характер. В свою очередь это приводит к распространению фронта нарастания и следующего за ним фронта спада потенциала на мембране. Таким образом, вдоль нервного волокна «бежит» импульс потенциала.

В терминах кинетических моделей трансмембранные ионные каналы работают как «нелинейные функции» (типа функции — множителя  $f(C)$  в уравнении Петровского — Колмогорова — Пискунова (6.48)), обеспечивающие распространение вдоль мембраны импульса потенциала. В ответ на ступенчатое изменение напряжения (начальные условия) вдоль мембраны проходит одиночный нервный импульс (рис. 6.10) — потенциал действия, который длится примерно 1 мс и распространяется со скоростью от 1 до 100 м/с.

Опыты с использованием фармакологических агентов позволили разделить протекающий поперек мембраны ток на составляющие, соответствующие  $\text{Na}^+$ - и  $\text{K}^+$ -токам ( $I_{\text{Na}}$  и  $I_{\text{K}}$ ), которые играют основную роль в распространении импульса. Роль других ионов, в частности ионов  $\text{Ca}^{2+}$ , в модели Ходчкина — Хаксли охарактеризована током «утечки» ( $I_{\text{L}}$ ). Общий ток  $I(t)$  представляет собой сумму токов отдельных ионов через мембрану и

Рис. 6.10. Форма нервного импульса во времени. Потенциал отсчитывается от уровня потенциала покоя



емкостного тока, вызванного вариациями трансмембранного потенциала. Таким образом,

$$I(t) = C \frac{dV}{dt} + I_i,$$

где  $C$  — емкость мембраны;  $I_i = I_{Na} + I_K + I_L$ .

На основе экспериментальных данных о зависимости проводимости мембраны для различных ионов от потенциала на мембране А. Ходчкин и А. Хаксли описали величины ионных токов с помощью функций, зависящих от «концентраций» некоторых гипотетических частиц, перемещающихся в мембране под действием электрического поля. Моделью Ходчкина — Хаксли называется следующая система нелинейных дифференциальных уравнений:

$$\begin{cases} \frac{a}{2R} \frac{\partial^2 V}{\partial x^2} = C \frac{\partial V}{\partial t} + (V - E_K) \bar{g}_K n^4 + (V - E_{Na}) \bar{g}_{Na} m^3 h + \\ + (V - E_0) \bar{g}_0; \\ \frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n; \\ \frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m; \\ \frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h, \end{cases} \quad (6.52)$$

где  $n$ ,  $m$  и  $h$  — функции потенциала и времени, определяющие поведение натриевого и калиевого токов;  $a$  — радиус волокна;  $R$  — удельное сопротивление протоплазмы;  $C$  — удельная емкость мембраны;  $E_K$ ,  $E_{Na}$ ,  $E_0$ ,  $\bar{g}_K$ ,  $\bar{g}_{Na}$ ,  $\bar{g}_0$  — постоянные;  $\alpha_n$ ,  $\alpha_m$ ,  $\alpha_h$ ,  $\beta_n$ ,  $\beta_m$ ,  $\beta_h$  — заданные функции потенциала [А. Hodgkin, А. Huxley, 1952]. Модель позволяет описать основные свойства проводящего нервного волокна.

Система (6.52) слишком сложна для аналитического исследования. Были предложены упрощенные модели, способные описать те же свойства. Наиболее популярная из них — предложенная независимо Р. Фитцхью [R. Fitzhugh, 1961] и Дж. Нагумо [J. S. Nagumo, S. Arimoto [et al.], 1962]. Система для локальных процессов здесь сводится к системе из двух уравнений, которую можно представить в безразмерном виде:

$$\begin{cases} \frac{du}{dt} = f(u) - v + I_a; \\ \frac{dv}{dt} = bu - \gamma v; \\ f(u) = u(a - u)(u - 1). \end{cases} \quad (6.53)$$

Здесь переменная  $u$  выполняет роль мембранного потенциала  $V$ , а  $v$  — роль всех трех переменных:  $m, n, h$  в уравнениях (6.52);  $I_a$  — составляющая тока, обусловленная наличием собственного постоянного потенциала (потенциала покоя) на мембране.

В случае когда  $I_a = 0$ , расположение главных изоклин на плоскости  $u, v$  представлено на рис. 6.11, а, б. Графики на рис. 6.11, а и б отличаются значением отношения  $b/\gamma$ . На рис. 6.11, а система имеет одно устойчивое решение, соответствующее нулевому значению потенциала, на рис. 6.11, б — три стационарных решения, два из которых устойчивы ( $0$  и  $S_2$ ), а третье  $S_1$  — неустойчивое.

С функцией вида  $f(u)$  мы неоднократно встречались, ее график изображен на рис. 6.8, в. В задаче о распространении кон-

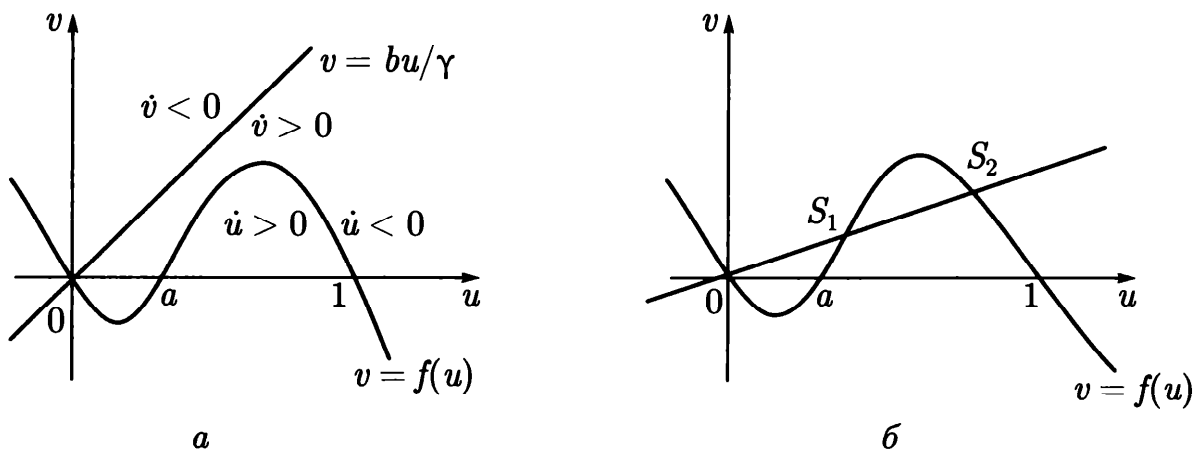
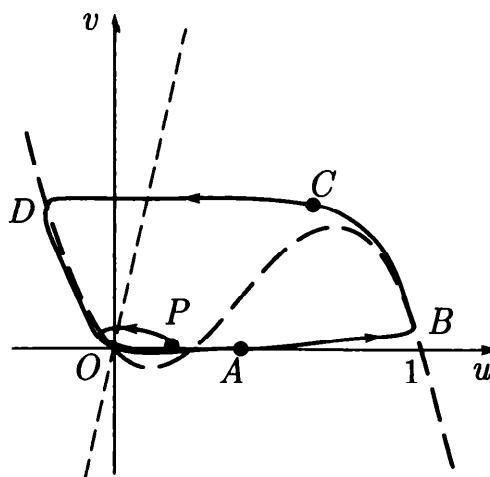


Рис. 6.11. Расположение главных изоклин системы (6.53) в случае  $I_a = 0$  на фазовой плоскости  $u, v$  для разных значений отношения  $b/\gamma$ : а — одно устойчивое стационарное состояние  $(0, 0)$ ; б — два устойчивых стационарных состояния  $(0$  и  $S_2)$  и одно неустойчивое — седло  $S_1$

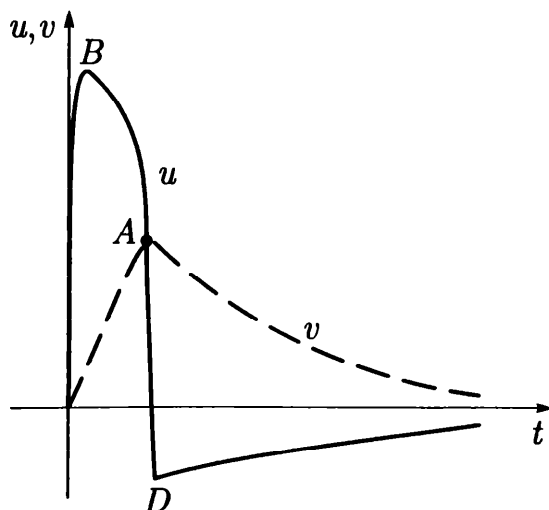
Рис. 6.12. Фазовый портрет системы (6.53) при  $I_a = 0$ ,  $a = 0,25$ ;  $b = \gamma = 0,002$  для малых подпороговых (точка  $P$ ) и больших надпороговых (точка  $A$ ) первоначальных отклонений потенциала  $u$  от стационарного нулевого значения. Штриховыми линиями обозначены изоклины горизонтальных (прямая) и вертикальных (имеет  $N$ -образный характер) касательных [J. Rinzel, 1981]



центрационной волны такой вид локальной функции источника означал, что «размножение» и «разгорание» процесса в данной точке пространства начинаются, если величина локальной концентрации превысит некоторое пороговое значение. Уравнения (6.53) также отражают тот факт, что бурный рост потенциала начинается лишь в случае, когда значение потенциала превысит некоторый порог (величину  $a$ ). В противном случае потенциал затухает. Поясним, как в такой системе возникает «импульс».

На рис. 6.12 штриховыми линиями изображены главные изоклины системы. Зададим некоторое начальное отклонение потенциала  $u$ . Если это отклонение переменной  $u$  мало, а именно, меньше величины  $a$  в третьем уравнении системы (6.53) — точка  $P$ , то по короткой фазовой траектории система возвратится в состояние покоя (подпороговое возбуждение). Если же отклонение переменной  $u$  велико (больше  $a$ ) — точка  $A$ , то фазовая траектория имеет иной характер. Изображающая точка движется по траектории  $ABCD$ , достигает изоклины вертикальных касательных ( $AB$ ), потом движется вдоль нее ( $BC$ ), затем довольно быстро переходит в область отрицательных значений ( $CD$ ), пересекая изоклину горизонтальных касательных, и после этого

Рис. 6.13. Поведение переменных  $u$  (сплошная линия) и  $v$  (штриховая линия) во времени в случае, когда фазовая траектория начинается в точке  $A$  [J. Rinzel, 1981]



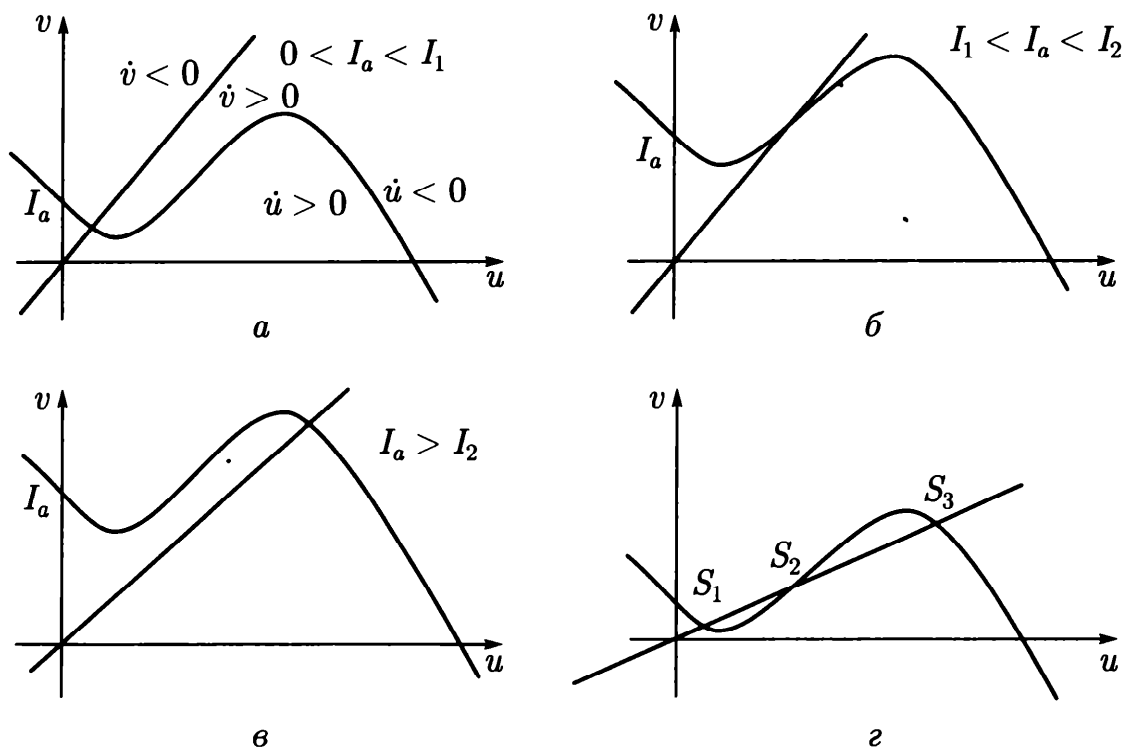


Рис. 6.14. Главные изоклины в модели Фитцхью — Нагумо (6.53) для разных значений приложенного напряжения:

*а, в* — стационарное состояние локально устойчиво, но «возбудимо» в том смысле, что при достаточно больших отклонениях изображающая точка описывает траекторию, соответствующую одиночному «импульсу» (см. рис. 6.12); *б* — стационарное состояние неустойчиво, возможно существование предельного цикла и автоколебаний состояния; *г* — состояние  $S_2$  неустойчиво; состояния  $S_1$  и  $S_3$  — устойчивы, между ними возможны переключения

вдоль изоклины вертикальных касательных ( $DO$ ) возвращается в точку покоя. На рис. 6.13 изображено поведение переменных  $u$  и  $v$  во времени. Локальный элемент системы, обладающий таким типом поведения, называется *возбудимым элементом*.

В случае когда равновесное значение потенциала на мембране не равно нулю, изоклина вертикальных касательных будет пересекать ось  $v$  не в нуле (см. рис. 6.12), а в точке  $v = I_a$ . На рис. 6.14, *а—г* представлены возможные варианты расположения главных изоклин в системе (6.53) при  $I_a \neq 0$ .

В распределенных системах, где имеет место диффузия, или в случае проводящего нервного волокна, когда ток распространяется вдоль пространственной координаты, такое локальное поведение приводит к автоволновым явлениям. Возможно распространение импульса (см. рис. 6.14, *а* и *в* (возбудимая среда)), стоячие и бегущие волны и возникновение устойчивых негомогенных распределений концентраций (см. рис. 6.14, *б* (колебательная неустойчивость) и *г* (седловая неустойчивость)).

Рассмотрим пространственно распределенную систему Фитцхью — Нагумо, которая отличается от системы (6.53) тем, что в

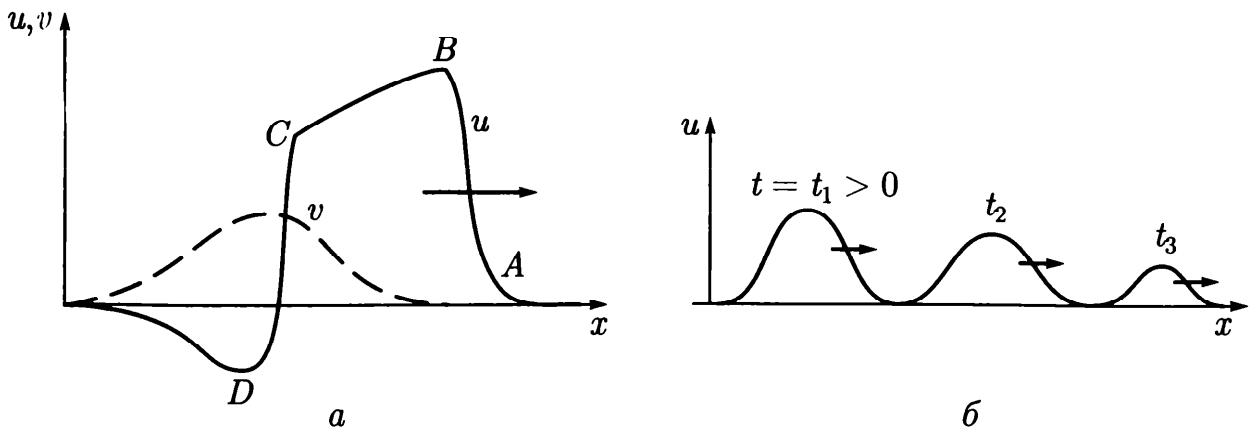


Рис. 6.15. Распространение импульса в случае надпорогового (а) и подпорогового (б) воздействия:

а — при начальном отклонении потенциала больше пороговой величины  $u = a$  в системе возникает бегущий импульс, который распространяется вдоль нерва без изменения формы (уединенная волна). Точки  $A, B, C, D$  соответствуют точкам на фазовом портрете (см. рис. 6.12); б — при начальном отклонении потенциала в точке возбуждения, меньшем пороговой величины  $u = a$  (точка  $P$  на рис. 6.12), бегущий импульс быстро затухает со временем ( $t_1 < t_2 < t_3$ )

ней учтено распространение потенциала вдоль кабеля (нервного волокна):

$$\begin{cases} \frac{\partial u}{\partial t} = f(u) - v + I_a + D \frac{\partial^2 u}{\partial x^2}; \\ \frac{\partial v}{\partial t} = bu - \gamma v; \\ f(u) = u(a - u)(u - 1). \end{cases} \quad (6.54)$$

Процесс распространения бегущего импульса в такой системе в случае надпорогового и подпорогового возбуждения изображен на рис. 6.15, а, б. Изучим распространение переднего фронта импульса. Для упрощения качественного рассмотрения будем считать  $b$  и  $\gamma$  малыми:  $b = \epsilon L$ ,  $\gamma = \epsilon M$ ,  $0 < \epsilon < 1$ .

Это предположение соответствует распространению весьма длительных (с характерным временем  $\epsilon^{-1}$ ) импульсов, однако более подробный анализ и вычислительные эксперименты показывают, что вывод относительно скорости распространения переднего фронта импульса справедлив и для относительно коротких импульсов.

При  $I_a = 0$  система (6.54) принимает вид:

$$\begin{cases} \frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + f(u) - v; \\ \frac{\partial v}{\partial t} = \epsilon(Lu - Mv). \end{cases} \quad (6.55)$$

Рассмотрим передний фронт импульса, изображенного на рис. 6.15. При  $\varepsilon \rightarrow 0$  из (6.55) следует, что  $v = \text{const}$ . Предположим, что локальная система имеет фазовый портрет, представленный на рис. 6.12, и единственное стационарное состояние  $(0, 0)$ . Приходим к одному уравнению для потенциала  $u$ :

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + f(u),$$

представляющему собой уравнение Петровского — Колмогорова — Пискунова (6.48) с функцией  $f(u) = u(a - u)(u - 1)$ .

Распространение фронта в модели Петровского — Колмогорова — Пискунова было рассмотрено ранее, где отмечалось также, что в такой системе скорость распространения переднего фронта импульса постоянна и равна  $\sqrt{\frac{D}{2}}(1 - 2a)$ . Учет медленных изменений величины  $v$  на участке траектории  $BC$  (см. рис. 6.12 и 6.15,  $a$ ) приводит к следующему выражению для скорости распространения одиночного импульса [J. D. Murray, 1993]:

$$c = \sqrt{\frac{D}{2}}(u_C - 2u_P + u_D),$$

где  $u_C, u_P, u_D$  — значения переменной  $u$  в точках  $C, P, D$  (см. рис. 6.15).

Добавление к модели (6.54) уравнения, описывающего трансмембранный перенос ионов  $\text{Ca}^{++}$ , позволяет описать периодическое возникновение пачек импульсов, бегущих вдоль нервного волокна [R. E. Plant, 1981].

Модель Фитцхью — Нагумо представляет собой в некотором смысле упрощенную «модель» системы Ходчкина — Хаксли, позволяющую увидеть на фазовых портретах те свойства системы, которые определяют ее качественное поведение, в частности существование порога возбуждения.

В настоящее время получены рентгеноструктурные данные о молекулярном строении каналов, обеспечивающих «воротные» токи и осуществляющих перенос ионов  $\text{Na}^+$ ,  $\text{K}^+$  и  $\text{Ca}^{2+}$  через возбудимую мембрану нервной клетки. С точки зрения этих данных модель Ходчкина — Хаксли, зарекомендовавшая себя как очень хорошая эмпирическая модель, требует новой биологической интерпретации.

## Часть I

*Айвазян С. А.* Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М. : Финансы и статистика, 1985. — 487 с.

*Айвазян С. А.* Классификация и снижение размерностей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М. : Финансы и статистика, 1989. — 607 с.

*Аффифи А.* Статистический анализ. Подход с использованием ЭВМ / А. Аффифи, С. Эйзен. — М. : Мир, 1982. — 488 с.

*Благовещенский Ю. Н.* Непараметрические методы в почвенных исследованиях / Ю. Н. Благовещенский, В. П. Самсонова, Е. А. Дмитриев. — М. : Наука, 1987. — 98 с.

*Бокс Дж.* Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. — М.: Мир, Вып. 1, 1974. — 405 с.

*Большев Л. Н.* Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. — М. : Наука, 1983. — 416 с.

*Браунли К. А.* Статистическая теория и методология в науке и технике. — М. : Наука, 1977. — 407 с.

*Боровков А. А.* Математическая статистика. — Новосибирск : Наука, 1997. — 772 с.

*Глотов Н. В.* Биометрия / Н. В. Глотов и [др.]. — Л. : Изд-во ЛГУ, 1982. — 264 с.

*Гмурман В. Е.* Теория вероятностей и математическая статистика. — М. : Юрайт, 2005. — 479 с.

*Гнеденко Б. В.* Элементарное введение в теорию вероятностей / Б. В. Гнеденко, А. Я. Хинчин. — М. : Наука, 1982.

*Дрейпер Н.* Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. — М. : Диалектика, 2007. — 912 с.

*Ивашов-Мусатов О. С.* Теория вероятностей и математическая статистика. — М. : Фима, 2003. — 224 с.

*Кендалл М. Дж.* Теория распределений / М. Дж. Кендалл, А. Стьюарт. — М. : Наука, 1966. — 586 с.

*Кендалл М. Дж.* Статистические выводы и связи / М. Дж. Кендалл, А. Стьюарт. — М. : Наука, 1973. — 892 с.

*Кендалл М. Дж.* Многомерный статистический анализ и временные ряды / М. Дж. Кендалл, А. Стьюарт. — М. : Наука, 1976. — 734 с.

*Колмогоров А. Н.* Основные понятия теории вероятностей. — М. : Наука, 1974. — 119 с.



*Компьютерная биометрика* / под ред. В. Н. Носова. — М. : Изд-во Моск. ун-та, 1990. — 232 с.

*Коршунов Д. А.* Сборник задач и упражнений по математической статистике / Д. А. Коршунов, Н. И. Чернова. — Новосибирск : Изд-во Ин-та математики, 2004. — 128 с.

*Крамер Г.* Математические методы статистики. — М. : Мир, 1975. — 648 с.

*Макаров А. А.* Анализ данных на компьютере / А. А. Макаров, Ю. Н. Тюрин. — М. : ИНФРА-М, 2003. — 544 с.

*Манита А. Д.* Теория вероятностей и математическая статистика. — М. : Изд-во Моск. ун-та, 2001. — 120 с.

*Мешалкин Л. Д.* Сборник задач по теории вероятностей. — М. : Изд-во Моск. ун-та, 1963. — 154 с.

*Мостеллер Ф.* Пятьдесят занимательных вероятностных задач с решениями. — М. : Наука, 1971. — 103 с.

*Налимов В. В.* Логические основания планирования эксперимента / В. В. Налимов, Т. И. Голикова. — М. : Metallургия, 1981. — 151 с.

*Тутубалин В. Н.* Теория вероятностей и случайных процессов. — М. : Изд-во Моск. ун-та, 1992. — 400 с.

*Тюрин Ю. Н.* Непараметрические методы статистики. — М. : Знание, 1978. — 64 с.

*Феллер В.* Введение в теорию вероятностей и ее приложения: в 2 т. Т. 1. — М. : Мир, 1984. — 528 с.

*Фишер Р. А.* Статистические методы для исследователей. — М. : Госстатиздат, 1958. — 267 с.

*Хан Г.* Статистические модели в инженерных задачах / Г. Хан, С. Шапиро. — М. : Мир, 1969. — 396 с.

*Харман Г.* Современный факторный анализ. — М. : Статистика, 1972. — 486 с.

*Холлендер М.* Непараметрические методы статистики / М. Холлендер, Д. А. Вульф. — М. : Финансы и статистика, 1983. — 518 с.

*Чжун К. Л.* Элементарный курс теории вероятностей / К. Л. Чжун, Ф. Айтсахлиа. — М. : Бином, 2007. — 455 с.

*Шеффе Г.* Дисперсионный анализ. — М. : Наука, 1980. — 512 с.

*Lilliefors H. W.* The Kolmogorov—Smirnov Test for Normality with Mean and Variance Unknown // Journal of the American Statistical Association, 1967. — V. 62. — P. 399—402.

*Sokal R. R.* 1995. Biometry / R. R. Sokal, F. J. Rohlf. — N.Y. : W. H. Freeman Co. — 887 p.

*Teriokhin A. T.* On the power of some binomial modifications of the Bonferroni multiple test / A. T. Teriokhin, T. de Meeus, J.-F. Guegan // Journal of General Biology, 2007. — V. 68. P. 332—340.

*Zar J. H.* Biostatistical Analysis. Upper Saddle River, N.J. : Prentice-Hall, Inc., 1999. — 998 p.

Учебные пособия и монографии

*Анищенко В. С.* Нелинейная динамика хаотических и стохастических систем / В. С. Анищенко, Т. Е. Вадивасова, В. В. Астахов. — Саратов : Изд-во Саратовского гос. ун-та, 1999. — 368 с.

*Базыкин А. Д.* Нелинейная динамика взаимодействующих популяций. — М.-Ижевск : Изд-во ин-та компьютерных исследований, 2003. — 368 с.

*Бигон М.* Экология. Особи, популяции и сообщества / М. Бигон, Дж. Харпер, К. Таусенд. — М. : Мир, 1989. — Т. 1. — 668 с., Т. 2. — 478 с.

*Братусь А. С.* Математические модели экологии и динамические системы с непрерывным временем / А. С. Братусь, А. С. Новожилов. — М. : Изд-во Моск. ун-та, 2004. — 242 с.

*Васильев В. А.* Автоволновые процессы / В. А. Васильев, Ю. М. Романовский, В. А. Яхно. — М. : Наука, 1987. — 238 с.

*Владимиров В. С.* Уравнения математической физики. — М. : Наука, 1981.

*Иваницкий Г. Р.* Математическая биофизика клетки / Г. Р. Иваницкий, В. И. Кринский, Е. Е. Сельков. — М. : Наука, 1978. — 310 с.

*Лоскутов А. Ю.* Введение в синергетику / А. Ю. Лоскутов, А. С. Михайлов. — М. : Наука, 1990. — 272 с.

*Николис Г.* Самоорганизация в неравновесных системах. От диссипативных структур к упорядоченности через флуктуации / Г. Николис, И. Р. Пригожин. — М. : Мир, 1979. — 512 с.

*Ризниченко Г. Ю.* Биофизическая динамика продукционных процессов / Г. Ю. Ризниченко, А. Б. Рубин. — М.-Ижевск : Изд-во института компьютерных исследований, 2004. — 464 с.

*Ризниченко Г. Ю.* Лекции по математическим моделям в биологии. — М.-Ижевск : Изд-во РХД, 2002. — 232 с.

*Ризниченко Г. Ю.* Математические модели биологических продукционных процессов / Г. Ю. Ризниченко, А. Б. Рубин. — М. : Изд-во Моск. ун-та, 1993. — 302 с.

*Романовский Ю. М.* Математическая биофизика / Ю. М. Романовский, Н. В. Степанова, Д. С. Чернавский. — М. : Наука, 1984. — 344 с.

*Романовский Ю. М.* Математическое моделирование в биофизике. Введение в теоретическую биофизику / Ю. М. Романовский, Н. В. Степанова, Д. С. Чернавский. — М.-Ижевск : Изд-во института компьютерных исследований, 2004. — 472 с.

*Рубин А. Б.* Биофизика. — М. : Изд-во «Книжный Дом Университет», 1999. — 448 с.

*Рубин А. Б.* Кинетика биологических процессов / А. Б. Рубин, Н. Ф. Пытьева, Г. Ю. Ризниченко. — М. : Изд-во Моск. ун-та, 1988. — 304 с.

*Сви́режев Ю. М.* Устойчивость биологических сообществ / Ю. М. Сви́режев, Д. О. Логофет. — М. : Наука, 1978. — 352 с.

*Тихо́нов А. Н.* Уравнения математической физики / А. Н. Тихо́нов, А. А. Самарский. — М. : Наука, 1999. — 768 с.

*Чернавский Д. С.* Синергетика и информация. Динамическая теория информации. — М. : УРСС, 2004. — 288 с.

*Шноль С. Э.* Физико-химические факторы биологической эволюции. — М. : Наука, 1979. — 262 с.

*Lotka A. J.* Elements of Physical biology. — Baltimore: Williams and Wilkins, 1925.

*Malthus T. R.* An essay on the principal of Population, 1798; Penguin Books, 1970.

*Murray J. D.* Mathematical biology. — I: An intriduction. — Berlin: Springer, 2002. — 568 p.

*Volterra V.* Lecons sur la Theorie Mathematique de la Lutte pour la Vie. — Paris, 1931.

### Оригинальные статьи

*Апони́на Е. А.* Анализ сложного динамического поведения в модели хищник — две жертвы / Е. А. Апони́на, Ю. М. Апони́н, А. Д. Базыкин // Проблемы экологического мониторинга и моделирования экосистем. — Ленинград : Гидрометеиздат, 1982. — Т. 5. — С. 163 — 180.

*Колмогоров А. Н.* Качественное изучение математических моделей динамики популяций // Вопросы кибернетики. — М., 1972. — № 55.

*Колмогоров А. Н.* Исследование уравнения диффузии, соединенной с возрастанием вещества, и его применение к одной биологической проблеме / А. Н. Колмогоров, И. Г. Петровский, Н. С. Пискунов // Бюлл. МГУ. — Сер. А. Математика и механика, 1937 (Вопросы кибернетики). — М., 1975. — Вып. 12. — С. 3 — 30.

*Тихо́нов А. Н.* Системы дифференциальных уравнений, содержащие малые параметры при производных // Математический сборник. — М., 1952. — № 3. — Т. 32.

*Fitzhugh R.* Impulses and physiological states in theoretical model of nerve membrane // Biophys J., 1961. — V. 1. — P. 445 — 466.

*Higgins J. A.* A chemical mechanism for oscillations in glicolitic intermediates in yeast cells // Proc. Nat. Acad. Sci. USA, 1954. — V. 51.

*Higgins J. A.* The theory of oscillating reactions // Ing. Chem., 1967. — № 5. — V. 9.

*Hodgkin A. L.* A quantitative description of membrane current and its application to conduction and excitation in nurve / A. L. Hodgkin, A. F. Huxley // J. Physiol. — London, 1952. — V. 117. — P. 500 — 544.

*Lorenz E. N.* Deterministic non-periodic flow // J. Atmos. Sci., 1963. — V. 20. — P. 131 — 141.

*May R. M.* When two and two make four: nonlinear phenomena in ecology // Proc. R. Soc. — London, 1986. — № 1252. — V. B228. — P. 241 — 268.

- Michaelis L.* Die Kinetik der Investinwirkung / L. Michaelis, M. I. Menten // Biochem. Z., 1913. — V. 49. — P. 333 — 369.
- Moran H. F. H.* Some remarks on animal population dynamics // Biometrics, 1950. — V. 6. — P. 250 — 258.
- Nagumo J. S.* An active pulse transmission line simulating nerve axon / J. S. Nagumo, S. Arimoto, S. Yoshizava // Proc. IRE, 1962. — V. 50. — P. 2061 — 2071.
- Plant R. E.* Bifurcation and resonance in a model for bursting nerve cells // J. Math. Biol., 1981. — V. 11. — P. 15 — 32.
- Prigogine I. R.* Symmetry breaking instabilities in dissipative systems / I. R. Prigogine, R. Lefever // J. Chem. Phys., 1968. — N. 48. — P. 1665 — 1700.
- Ricker W. E.* Stock and recruitment // J. Fish. Res. Bd. Can., 1954. — V. 11. — P. 559 — 623.
- Rinzel J.* Models in neurobiology // In: Nonlinear Phenomena in Physics and Biology. N.Y. Plenum Press, 1981. — P. 345 — 367.
- Turing A. M.* The chemical basis of the morphogenesis // Proc. Roy. Soc. B, 1952. — V. 237. — P. 37 — 71.
- Verhulst P. F.* Notice sur la loi que la population suit dans son accroissement // Corr. Math. et Phys., 1838. — V. 10. — P. 113 — 121.
- Volterra V.* Variazione e fluttuazione del numero d'individui in specie animali conviventi // Mem. Acad. Lincei., 1926. — V. 2. — P. 31 — 113.

# ОГЛАВЛЕНИЕ

Предисловие . . . . .	3
Введение . . . . .	4

## Часть I

### Статистический анализ биологических данных

<b>Глава 1. Основы теории вероятностей . . . . .</b>	<b>6</b>
1.1. Случайные события и их вероятности . . . . .	7
1.1.1. Пространство элементарных событий . . . . .	7
1.1.2. Операции над событиями . . . . .	8
1.1.3. Вероятности событий . . . . .	9
1.1.4. Модель равновероятных элементарных событий . . . . .	11
1.1.5. Геометрическая модель . . . . .	13
1.1.6. Условная вероятность и независимость . . . . .	14
1.1.7. Формула полной вероятности и формула Байеса . . . . .	16
1.1.8. Формула Бернулли . . . . .	19
1.2. Случайные величины и их распределения . . . . .	20
1.2.1. Функция распределения случайной величины . . . . .	21
1.2.2. Дискретные случайные величины . . . . .	21
1.2.3. Непрерывные случайные величины . . . . .	25
1.2.4. Нормальное и связанные с ним распределения . . . . .	28
1.2.5. Характеристики распределений . . . . .	32
1.3. Многомерные случайные величины . . . . .	41
1.3.1. Дискретные многомерные случайные величины . . . . .	42
1.3.2. Непрерывные многомерные случайные величины . . . . .	43
<b>Глава 2. Основы математической статистики . . . . .</b>	<b>48</b>
2.1. Случайная выборка и ее описание . . . . .	49
2.2. Статистическое оценивание . . . . .	56
2.2.1. Точечное статистическое оценивание . . . . .	56
2.2.2. Интервальное статистическое оценивание . . . . .	62
2.3. Статистическая проверка гипотез . . . . .	71
2.3.1. Логика проверки статистических гипотез . . . . .	71
2.3.2. Проверка гипотез о математических ожиданиях . . . . .	73
2.3.3. Проверка гипотез о дисперсиях . . . . .	81

2.3.4. Сравнение параметров двух биномиальных распределений . . . . .	83
2.3.5. Сравнение параметров двух пуассоновских распределений . . . . .	83
2.3.6. Проверка гипотезы о равенстве заданному числу коэффициента корреляции . . . . .	84
2.3.7. Критерии согласия . . . . .	85
2.3.8. Непараметрические критерии . . . . .	92
<b>Глава 3. Анализ многомерных данных . . . . .</b>	<b>100</b>
3.1. Классификация методов анализа многомерных данных . . . . .	100
3.2. Матричная алгебра . . . . .	104
3.3. Регрессионный анализ . . . . .	108
3.3.1. Множественная линейная регрессия . . . . .	110
3.3.2. Множественный и частный коэффициенты корреляции . . . . .	118
3.3.3. Пошаговая регрессия . . . . .	120
3.3.4. Нелинейная регрессия . . . . .	122
3.3.5. Логистическая регрессия . . . . .	123
3.4. Дисперсионный анализ . . . . .	124
3.4.1. Однофакторный дисперсионный анализ . . . . .	126
3.4.2. Двухфакторный дисперсионный анализ . . . . .	134
3.4.3. Смешанные модели дисперсионного анализа . . . . .	139
3.4.4. Непараметрические методы анализа связи между количественной и качественными переменными . . . . .	146
3.5. Кластерный анализ . . . . .	151
3.6. Факторный анализ . . . . .	157
3.7. Дискриминантный анализ . . . . .	167

## Часть II

### Моделирование динамики биологических процессов

<b>Глава 4. Модели, описываемые одним уравнением . . . . .</b>	<b>174</b>
4.1. Основные понятия . . . . .	174
4.2. Исследование уравнения в окрестности стационарного состояния . . . . .	177
4.2.1. Исследование устойчивости стационарного состояния методом линеаризации . . . . .	177
4.2.2. Модель роста колонии микроорганизмов . . . . .	179
4.2.3. Модель перехода вещества в раствор . . . . .	180
4.3. Непрерывные модели популяций . . . . .	181
4.3.1. Уравнение экспоненциального роста . . . . .	181
4.3.2. Ограниченный рост . . . . .	183
4.3.3. Модель популяции с наименьшей критической численностью . . . . .	187

4.4. Дискретные модели популяций . . . . .	190
4.4.1. Модели популяций с неперекрывающимися поколениями . . . . .	191
4.4.2. Диаграмма Ламерея . . . . .	192
4.4.3. Понятия равновесия и его устойчивости для разностных уравнений . . . . .	193
4.4.4. Циклы и динамический хаос . . . . .	195
4.5. Вероятностные модели популяций . . . . .	196

## Глава 5. Модели, описываемые системой дифференциальных уравнений . . . . . 199

5.1. Основные понятия . . . . .	199
5.1.1. Фазовая плоскость и фазовый портрет . . . . .	199
5.1.2. Метод изоклин . . . . .	200
5.1.3. Устойчивость стационарного состояния . . . . .	202
5.2. Исследование систем двух линейных уравнений . . . . .	203
5.2.1. Характеристическое уравнение . . . . .	203
5.2.2. Корни $\lambda_1, \lambda_2$ действительны и одного знака . . . . .	205
5.2.3. Корни $\lambda_1, \lambda_2$ действительны и разных знаков . . . . .	206
5.2.4. Корни $\lambda_1, \lambda_2$ комплексные сопряженные . . . . .	208
5.2.5. Бифуркационная диаграмма . . . . .	211
5.3. Системы двух нелинейных дифференциальных уравнений . . . . .	213
5.3.1. Метод Ляпунова линеаризации системы в окрестности стационарного состояния . . . . .	213
5.3.2. Кинетические уравнения Лотки . . . . .	216
5.3.3. Модель Вольтерра . . . . .	219
5.4. Проблема быстрых и медленных переменных . . . . .	225
5.4.1. Иерархия времен. Метод квазистационарных концентраций . . . . .	225
5.4.2. Теорема Тихонова . . . . .	227
5.4.3. Фермент-субстратная реакция Михаэлиса — Ментен . . . . .	230
5.5. Мультистационарность . . . . .	235
5.5.1. Фазовый портрет мультистационарной системы . . . . .	235
5.5.2. Типы переключения триггера . . . . .	237
5.5.3. Отбор одного из равноправных видов . . . . .	239
5.6. Колебания в биологических системах . . . . .	244
5.6.1. Автоколебания и предельные циклы . . . . .	244
5.6.2. Устойчивость предельных циклов . . . . .	246
5.6.3. Брюсселятор . . . . .	247
5.6.4. Колебания в гликолизе . . . . .	249
5.7. Динамический хаос . . . . .	251
5.7.1. Система Лоренца. Детерминированный хаос . . . . .	252
5.7.2. Линейный анализ устойчивости траекторий . . . . .	255
5.7.3. Диссипативные системы . . . . .	256
5.7.4. Динамический хаос в сообществе из трех видов . . . . .	261

<b>Глава 6. Распределенные дифференциальные модели</b>	<b>263</b>
6.1. Уравнение реакция — диффузия	263
6.1.1. Активные кинетические среды в живых системах	263
6.1.2. Уравнение диффузии	264
6.1.3. Начальные и граничные (краевые) условия	268
6.2. Решение уравнения диффузии	270
6.2.1. Общий ход решения	270
6.2.2. Решение однородного уравнения	271
6.2.3. Зависимость решений от начальных условий	275
6.2.4. Решение неоднородного уравнения диффузии с нулевым начальным условием и нулевыми краевыми условиями	277
6.2.5. Общая краевая задача	279
6.2.6. Устойчивость гомогенных стационарных состояний нелинейных систем	280
6.2.7. Распределенная модель популяции с наименьшей критической численностью	283
6.3. Система двух уравнений реакция — диффузия	285
6.3.1. Устойчивость однородных стационарных решений	285
6.3.2. Распределенный брюсселятор	292
6.3.3. Локализованные диссипативные структуры	294
6.4. Распространение фронтов, импульсов и волн	296
6.4.1. Распространение волны в системах с диффузией	296
6.4.2. Задача распространения нервного импульса	300
Список литературы	307