

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Вятский государственный университет»**

Биологический факультет

Кафедра микробиологии

А.В. Крючков

Математические методы и модели в биологии

Курс лекций

Киров
2012

УДК 57.087.1(07)

К856

Допущено к изданию методическим советом биологического факультета ФГБОУ ВПО «ВятГУ» в качестве учебного пособия для студентов направления 020400 «Биология» (бакалавриат и магистратура) всех профилей подготовки, всех форм обучения

Рецензент

доктор технических наук, главный научный сотрудник ФГУ «48 ЦНИИ» Минобороны России А.С. Кучеренко

Крючков, А.В.

К 856 Математические методы и модели в биологии. Курс лекций /А.В.

Крючков. – Киров: ФГБОУ ВПО «ВятГУ», 2012. – 172 с.

УДК 856

В учебном пособии рассмотрены основные понятия теории вероятностей и математической статистики, начала биологической статистики, закономерности в массовых биологических объектах и методы их изучения. Одна из важных задач пособия – привить интерес к изучению объектов живой природы с использованием методов биологической статистики, научить правилам обработки статистических данных.

Учебное пособие «Биометрия» составлено в соответствии с программой курса «Математические методы и модели в биологии» и предназначено для студентов университета, обучающихся по направлению 020400 «Биология».

© ФГБОУ ВПО «ВятГУ», 2012

СОДЕРЖАНИЕ

Лекция 1. МЕТОДЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ В БИОЛОГИИ	4
Лекция 2. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ	11
Лекция 3. ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА	23
Лекция 4. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ	34
Лекция 5. КРИТЕРИИ СОГЛАСИЯ	40
Лекция 6. КРИТЕРИЙ СТЬЮДЕНТА	48
Лекция 7. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ	53
Лекция 8. УРАВНЕНИЯ РЕГРЕССИИ	60
Лекция 9. КРИВАЯ «ДОЗА-ЭФФЕКТ»	64
Лекция 10. МЕТОДЫ ОЦЕНОК ПАРАМЕТРОВ КРИВОЙ «ДОЗА-ЭФФЕКТ»	72
Лекция 11. МЕТОДЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТОВ	81
Лекция 12. МЕТОДЫ ОПТИМИЗАЦИИ	96
Лекция 13. МОДЕЛИ БИОЛОГИЧЕСКИХ СИСТЕМ, ОПИСЫВАЕМЫЕ ОДНИМ ДИФФЕРЕНЦИАЛЬНЫМ УРАВНЕНИЕМ ПЕРВОГО ПОРЯДКА	105
Лекция 14. МОДЕЛИ БИОЛОГИЧЕСКИХ СИСТЕМ, ОПИСЫВАЕМЫХ СИСТЕМАМИ ДВУХ АВТОНОМНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ	114
Лекция 15. МУЛЬТИСТАЦИОНАРНЫЕ СИСТЕМЫ	127
Лекция 16. КОЛЕБАНИЯ В БИОЛОГИЧЕСКИХ СИСТЕМАХ	138
Лекция 17. МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ДВУХ ВИДОВ	147
Лекция 18. МОДЕЛИРОВАНИЕ МИКРОБНЫХ ПОПУЛЯЦИЙ	159
Учебно-методическое обеспечение дисциплины	172

Лекция 1. МЕТОДЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ В БИОЛОГИИ

1. Дискретные (прерывные) случайные величины.
2. Математическое ожидание и дисперсия дискретной случайной величины и их свойства
3. Структурные средние.
4. Биномиальное распределение и распределение редких событий (распределение Пуассона). Математическое ожидание, дисперсия для этих распределений.

1. Дискретные (прерывные) случайные величины.

Случайной называют величину, которая в результате испытания примет одно и только одно возможное значение, наперед не известное и зависящее от случайных причин, которые заранее не могут быть учтены.

Будем далее обозначать случайные величины прописными буквами X, Y, Z , а их возможные значения – соответствующими строчными буквами x, y, z ... с индексами.

Целесообразно различать случайные величины, принимающие лишь отдельные, изолированные значения, и случайные величины, возможные значения которых сплошь заполняют некоторый промежуток.

Их и делят по этому признаку на прерывные (дискретные) и непрерывные.

Дискретной (прерывной) называют случайную величину, которая принимает отдельные, изолированные возможные значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или бесконечным.

Непрерывной называют случайную величину, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Очевидно, число возможных значений непрерывной случайной величины бесконечно.

Для задания дискретной случайной величины недостаточно перечислить все возможные ее значения, нужно еще указать их вероятности.

Законом распределения дискретной случайной величины называют соответствие между возможными значениями и их вероятностями; его можно задать таблично, аналитически (в виде формулы) и графически.

При табличном задании закона распределения дискретной случайной величины первая строка таблицы содержит ее возможные значения, а вторая – их вероятности:

$$\begin{array}{l} X \quad x_1, x_2, \dots, x_n \\ P \quad p_1, p_2, \dots, p_n \end{array} \quad (1)$$

Приняв во внимание, что в одном испытании случайная величина принимает одно и только одно возможное значение, заключаем, что события

$X = x_1, X = x_2, \dots, X = x_n$ образуют полную группу; следовательно, сумма вероятностей этих событий, т. е. сумма вероятностей этих событий равна единице:

$$p_1 + p_2 + \dots + p_n = 1. \quad (2)$$

Если множество возможных значений X бесконечно (счетно), то ряд $p_1 + p_2 + \dots$ сходится и его сумма равна единице.

Для наглядности закон распределения дискретной случайной величины можно изобразить и графически, для чего в прямоугольной системе координат строят точки (x_i, p_i) , а затем соединяют их отрезками прямых. Полученную фигуру называют *многоугольник распределения*.

Закон распределения полностью характеризует случайную величину. Однако часто закон распределения неизвестен и приходится ограничиваться меньшими сведениями. Иногда даже выгоднее пользоваться числами, которые описывают случайную величину суммарно; такие числа называют числовыми характеристиками случайной величины. К числу важнейших числовых характеристик относится *математическое ожидание и дисперсия* случайной величины.

2. Математическое ожидание и дисперсия дискретной случайной величины и их свойства

Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности:

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n \quad (\text{для конечного множества значений}) \quad (3)$$

$$\text{или} \quad M(X) = x_1 p_1 + x_2 p_2 + \dots = \sum_{i=1}^{\infty} x_i p_i \quad (4)$$

(для счетного множества значений),
причем математическое ожидание существует, если ряд в правой части сходится абсолютно.

Вероятностный смысл математического ожидания таков: математ. ожидание приближенно равно (тем точнее, чем больше число испытаний) среднему арифметическому наблюдаемых значений случайной величины.

Свойства математического ожидания:

Свойство 1. Математическое ожидание постоянной величины равно самой постоянной:

$$M(C) = C. \quad (5)$$

Свойство 2. Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX) = C \cdot M(X). \quad (6)$$

Свойство 3. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY) = M(X) \cdot M(Y). \quad (7)$$

Следствие. Математическое ожидание произведения нескольких взаимно независимых случайных величин равно произведению их математических ожиданий.

Свойство 4. Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

$$M(X+Y) = M(X) + M(Y). \quad (8)$$

Следствие. Математическое ожидание суммы нескольких случайных величин равно сумме их математических ожиданий.

Можно указать такие случайные величины, которые имеют одинаковые математические ожидания, но различные возможные значения.

Например,

X	−0,7	0,7	Y	−100	100
p	0,5	0,5	p	0,5	0,5

Оба эти распределения имеют одинаковые математические ожидания (равные нулю). Однако первое распределение имеет возможные значения, расположенные ближе к математическому ожиданию, чем второе.

Таким образом, математическое ожидание полностью случайную величину не характеризует.

Поэтому наряду с математическим ожиданием вводят и другие числовые характеристики. Так, например, для того, чтобы оценить, как рассеяны возможные значения случайной величины вокруг ее математического ожидания, пользуются, в частности, числовой характеристикой, которую называют дисперсией.

Дисперсией (рассеянием) дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M[(X - M(X))^2] \quad (9)$$

или

$$D(X) = M[(X - M(X))^2] = [x_1 - M(X)]^2 p_1 + [x_2 - M(X)]^2 p_2 + \dots + [x_n - M(X)]^2 p_n$$

(для конечного множества значений).

Таким образом, чтобы найти дисперсию достаточно вычислить сумму произведений возможных квадратов отклонения на их вероятности.

Для вычисления дисперсии часто удобно пользоваться следующей теоремой

Теорема. Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания:

$$D(X) = M[X^2] - [M(X)]^2. \quad (10)$$

Свойства дисперсии.

Свойство 1. Дисперсия постоянной величины C равна нулю: $D(C) = 0$

Свойство 2. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(CX) = C^2 D(X) \quad (11)$$

Свойство 3. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:

$$D(X+Y) = D(X) + D(Y) \quad (12)$$

Следствие 1. Дисперсия суммы нескольких взаимно независимых случайных величин равна сумме дисперсий этих величин.

Следствие 2. Дисперсия суммы постоянной величины и случайной равна дисперсии случайной величины:

$$D(C+X) = D(X). \quad (13)$$

Свойство 4. Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин:

$$D(X-Y) = D(X) + D(Y). \quad (14)$$

Средним квадратическим отклонением (или стандартным отклонением) случайной величины называют квадратный корень из дисперсии:

$$\sigma(x) = \sqrt{D(x)}. \quad (15)$$

4. Биномиальное распределение и распределение редких событий (распределение Пуассона). Математическое ожидание, дисперсия для этих распределений.

Пусть производится n независимых испытаний, в каждом из которых событие A может появиться либо не появиться. Вероятность наступления события во всех испытаниях постоянна и равна p (следовательно, вероятность не появления $q = 1 - p$). Рассмотрим в качестве дискретной случайной величины X число появлений события A в этих испытаниях.

Найдем закон распределения величины X . Для ее решения требуется определить возможные значения X и их вероятности. Очевидно, событие A в n испытаниях может не появиться, либо появиться 1 раз, либо 2 раза, ..., либо n раз. Таким образом, возможные значения X таковы: $x_1 = 0, x_2 = 1, \dots, x_{n+1} = n$. Найдем вероятности этих возможных значений. Для этого достаточно воспользоваться формулой Бернулли:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad \text{где } k = 0, 1, 2, \dots, n. \quad (16)$$

Формула (3.28) и является аналитическим выражением искомого закона распределения.

Биномиальным называют распределение вероятностей, определяемое формулой Бернулли. Закон назван биномиальным потому, что правую часть

равенства (3.28) можно рассматривать как общий член разложения бинома Ньютона:

$$(p+q)^n = C_n^n p^n + C_n^{n-1} p^{n-1} q + C_n^{n-2} p^{n-2} q^2 + \dots + C_n^k p^k q^{n-k} + \dots + C_n^0 q^n \quad (17)$$

Таким образом, первый член разложения определяет вероятность наступления рассматриваемого события n раз в n независимых испытаниях, второй член $np^{n-1}q$ определяет вероятность наступления события $n - 1$ раз, ..., последний член q^n определяет вероятность того, что событие не появится ни разу.

Биномиальное распределение имеет среднее $n \cdot p$ и дисперсию $n \cdot p \cdot q$.

На рисунке приведен график закона биномиального распределения с параметрами $n = 10, p = 0,3$.

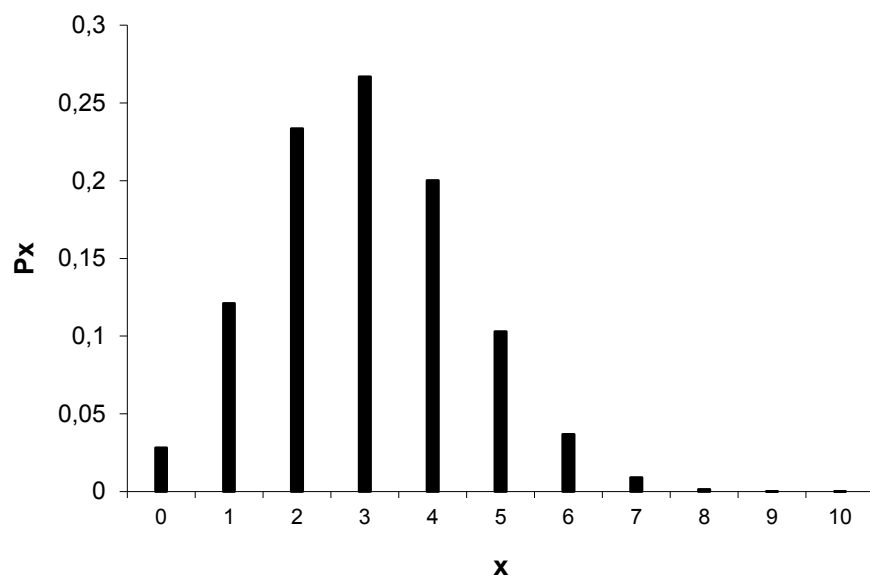


Рисунок 1 – Биномиальное распределение с параметрами $n = 10, p = 0,3$

Случайная величина X , принимающая целые неотрицательные значения, распределена по закону Пуассона, если

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}, \quad (18)$$

где $r = 0, 1, 2, \dots$ и λ – положительная константа.

Для распределения Пуассона среднее и дисперсия равны λ .

Графики распределения Пуассона со средним 0,9 и 5,0 показаны на рисунках 2 и 3.

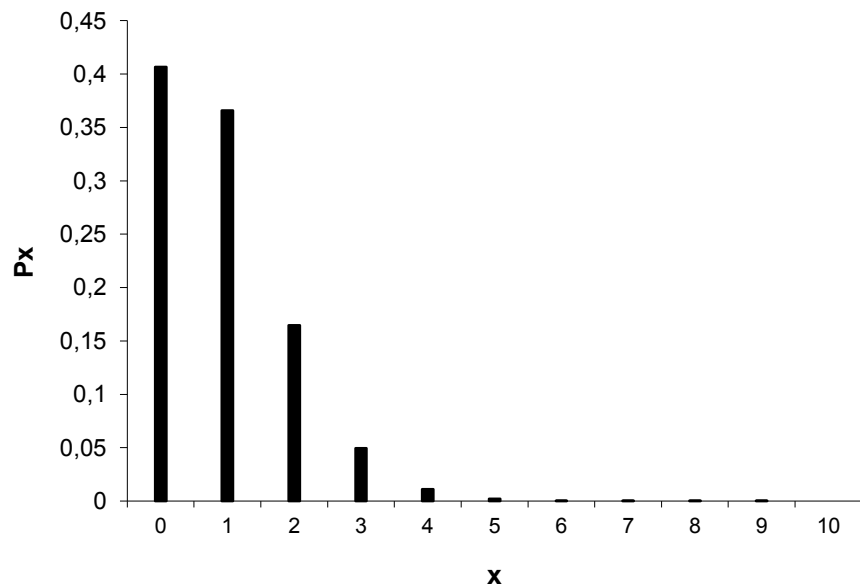


Рисунок 2 – Распределение Пуассона со средним $\lambda = 0,9$

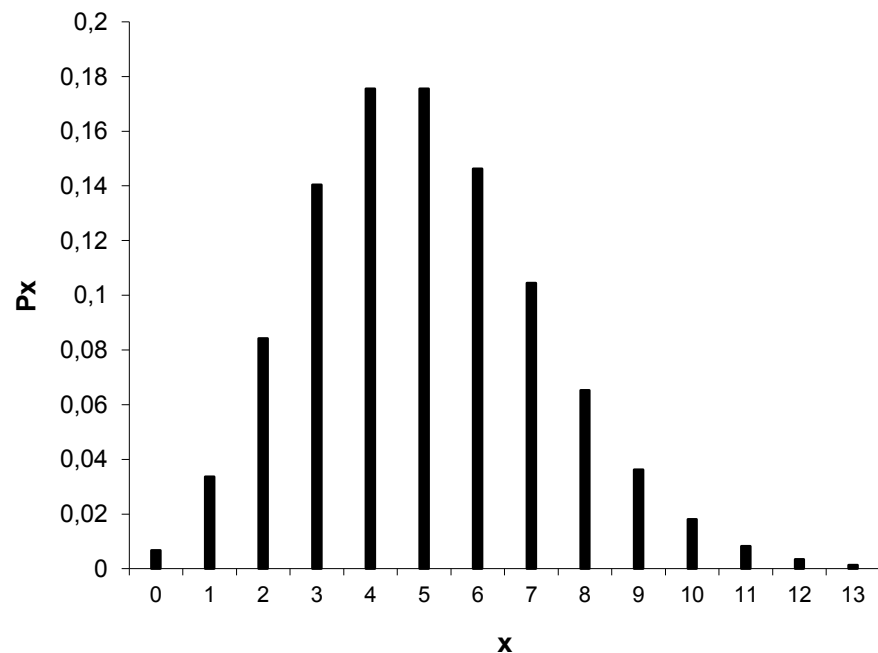


Рисунок 3 – Распределение Пуассона со средним $\lambda = 5,0$

Получим формулу (18) из биномиального распределения при допущении: произведение np сохраняет постоянное значение, а именно равно λ

$$np = \lambda \quad (19)$$

Это означает (это будет следовать из дальнейшего), что среднее число появлений события в различных сериях испытаний, т.е. при различных значениях n , остается неизменным.

Воспользуемся формулой Бернулли для вычисления интересующей нас вероятности:

$$P_n(k) = \frac{n(n-1)(n-2)\dots[n-(k-1)]}{k!} p^k (1-p)^{n-k}$$

Так как $np = \lambda$, то $p = \lambda/n$ и, следовательно,

$$P_n(k) = \frac{n(n-1)(n-2)\dots[n-(k-1)]}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Так как n имеет очень большое значение, вместо $P_n(k)$ найдем предел $P_n(k)$ при n стремящимся к бесконечности. При этом будет найдено лишь приближенное значение отыскиваемой величины: n хотя и велико, но конечно. Заметим, что при $n \rightarrow \infty$ вероятность $p \rightarrow 0$ (т.к. λ постоянно).

Итак,

$$\begin{aligned} P_n(k) &\approx \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots[n-(k-1)]}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left[1 - \frac{1}{n}\right] \left[1 - \frac{2}{n}\right] \dots \left[1 - \frac{k-1}{n}\right] \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \end{aligned}$$

Таким образом,

$$P_n(k) = \lambda^k e^{-\lambda} / k!$$

Эта формула и выражает закон распределения Пуассона вероятностей массовых (n велико) и редких (p мало) событий.

Легко проверяется, что (3.30) действительно выражает закон распределения: сумма вероятностей $P_n(r)$ по r от нуля до бесконечности равна 1 (одно из свойств распределения вероятностей).

Легко показывается прямым вычислением, что математическое ожидание случайной величины, распределенной по закону Пуассона, и дисперсия равны λ .

Распределение Пуассона является однопараметрическим распределением, т.к. оно зависит только от одного параметра λ .

Лекция 2. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

1. Функция распределения непрерывной случайной величины. Плотность вероятности непрерывной случайной величины.
2. Математическое ожидание и дисперсия непрерывной случайной величины.
3. Нормально распределенные случайные величины, их параметры. Основные свойства нормального распределения.
4. Распределение «хи-квадрат»; t-распределение Стьюдента; распределение F (распределение Фишера-Снедекора).
5. Логарифмически нормальное распределение случайных величин

1. Функция распределения непрерывной случайной величины. Плотность вероятности непрерывной случайной величины.

Если дискретная случайная величина может быть задана перечнем всех ее возможных значений и их вероятностей, то для других случайных величин, так называемых непрерывных, такой способ не применим.

Действительно, рассмотрим случайную величину X , возможные значения которой сплошь заполняют интервал (a, b) . Для нее нельзя составить перечень всех возможных значений с указанием их вероятностей, т.к. вероятность любого конкретного значения X будет равна нулю. Это указывает на целесообразность дать общий способ задания любых типов случайных величин. С этой целью и вводят функции распределения вероятностей случайной величины.

Функцией распределения (ФР) называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x , т.е.

$$F(x) = P(X < x). \quad (1)$$

Геометрически это можно истолковать так: $F(x)$ есть вероятность того, что случайная величина примет значение, которое изображается на числовой оси точкой, лежащей левее точки x .

Иногда вместо термина «функция распределения» используют термин «интегральная функция».

Теперь можно дать более точное определение непрерывной случайной величины: случайную величину называют непрерывной, если ее функция

распределения есть непрерывная, кусочно-дифференцируемая функция с непрерывной производной.

Свойства функции распределения:

Свойство 1. Значения ФР принадлежат отрезку $[0,1]$, т.е.

$$0 \leq F(x) \leq 1$$

Свойство 2. $F(x)$ – неубывающая функция, т.е.

$$F(x_2) \geq F(x_1), \text{ если } x_2 > x_1.$$

Следствие 1. Вероятность того, что случайная величина примет значение, заключенное в интервале (a, b) , равна приращению функции на этом интервале:

$$P(a \leq X < b) = F(b) - F(a) \quad (2)$$

Следствие 2. Вероятность того, что случайная величина X примет определенное значение равно нулю.

Таким образом, не представляет интереса говорить о вероятности того, что непрерывная случайная величина примет одно определенное значение, но имеет смысл рассматривать вероятность попадания ее в интервал, пусть даже сколь угодно малый.

Заметим, что было бы неправильно думать, что равенство нулю вероятности $P(X=x_1)$ означает, что событие $X=x_1$ невозможно. Действительно, в результате испытания случайная величина обязательно примет одно из возможных значений; в частности, это значение может оказаться равным x_1 .

Свойство 3. Если возможные значения случайной величины принадлежат интервалу (a, b) , то: 1) $F(x) = 0$ при $x \leq a$; 2) $F(x) = 1$ при $x \geq b$

Следствие. Если возможные значения непрерывной случайной величины расположены на всей оси x , то справедливы следующие предельные соотношения:

$$F(x) \rightarrow 0 \text{ при } x \rightarrow -\infty \text{ и } F(x) \rightarrow 1 \text{ при } x \rightarrow \infty. \quad (3)$$

Плотность вероятности непрерывной случайной величины.

Непрерывную случайную величину можно также задать, используя другую функцию, которую называют *плотностью распределения* или *плотностью вероятности* (иногда называют *дифференциальной функцией*).

Плотностью вероятности непрерывной случайной величины X называют функцию $f(x)$ – первую производную от функции распределения $F(x)$:

$$f(x) = F'(x). \quad (4)$$

Из этого определения следует, что ФР является первообразной для плотности вероятности.

Замечу, что для описания распределения вероятностей дискретной случайной величины плотность вероятностей неприменима.

Зная плотность вероятности $f(x)$ можно найти функцию распределения $F(x)$ по формуле

$$F(x) = \int_{-\infty}^x f(x) dx. \quad (5)$$

Таким образом, зная плотность вероятности можно найти ФР. И наоборот: зная ФР можно найти плотность вероятности (см. формулу (4)).

График плотности вероятности называют *кривой распределения*.

Свойства плотности вероятности

Свойство 1. Плотность вероятности – неотрицательная функция:

$$f(x) \geq 0 \quad (6)$$

Свойство 2. Несобственный интеграл от плотности вероятности в пределах от $-\infty$ до ∞ равен единице:

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (7)$$

В частности, если все возможные значения случайной величины принадлежат интервалу (a, b) то

$$\int_a^b f(x) dx = 1. \quad (8)$$

Из дифференциального исчисления известно, что приращение функции приближенно равно дифференциалу функции, т. е.

$$F(x+\Delta x) - F(x) \approx dF = F'(x) dx = f(x) dx. \quad (9)$$

Смысл этого равенства такой: вероятность того, что случайная величина примет значение, принадлежащее интервалу $(x, x + \Delta x)$ приближенно равна произведению плотности вероятности в точке x на длину интервала Δx .

2. Математическое ожидание и дисперсия непрерывной случайной величины

Дадим определения математического ожидания и дисперсии непрерывной случайной величины.

Математическим ожиданием непрерывной случайной величины X называют интеграл

$$M(x) = \int_{-\infty}^{\infty} xf(x)dx. \quad (10)$$

Предполагается, что интеграл сходится абсолютно.

Дисперсией непрерывной случайной величины называют математическое ожидание квадрата ее отклонения:

$$D(x) = \int_{-\infty}^{\infty} (x - M(x))^2 f(x)dx. \quad (11)$$

Среднее квадратическое отклонение непрерывной случайной величины определяется, как и для дискретной (прерывной) случайной величины, равенством:

$$\sigma(x) = \sqrt{D(x)}. \quad (12)$$

Моменты случайных величин. Асимметрия и эксцесс

Кроме характеристик положения – средних (математического ожидания, медианы, моды), типичных значений случайной величины, – употребляются еще ряд характеристик, каждая из которых описывает то или иное свойство распределения. В качестве таких характеристик чаще всего применяются так называемые моменты.

(Замечание: Это название попало в теорию вероятностей из механики).

Чаще всего на практике применяются моменты двух видов: начальные и центральные.

Начальным моментом порядка k случайной величины X называют математическое ожидание величины X^k :

$$\nu_k = M(X^k). \quad (13)$$

В частности,

$$\nu_1 = M(X), \quad (14)$$

т. е. начальный момент первого порядка равен математическому ожиданию случайной величины.

Для прерывной (дискретной) случайной величины начальный момент порядка k выражается суммой

$$v_k = \sum_{i=1}^n x_i^k p_i, \quad (15)$$

а для непрерывной случайной величины – интегралом

$$v_k = \int_{-\infty}^{\infty} x^k f(x) dx. \quad (16)$$

Кроме начальных моментов случайной величины X рассматривают и моменты отклонения величины X от математического ожидания $M(X)$, которые называют центральными моментами.

Таким образом, центральным моментом порядка k случайной величины X называют математическое ожидание величины $(X - M(X))^k$:

$$\mu_k = M[(X - M(X))^k]. \quad (17)$$

В частности,

$$\mu_1 = M[X - M(X)] = 0, \quad \mu_2 = M[(X - M(X))^2] = D(X) \text{ (дисперсия)}.$$

Для прерывной (дискретной) случайной величины центральный момент порядка k выражается суммой

$$\mu_k = \sum_{i=1}^n (x_i - v_1)^k p_i, \quad (18)$$

а для непрерывной случайной величины – интегралом

$$\mu_k = \int_{-\infty}^{\infty} (x - v_1)^k f(x) dx. \quad (19)$$

Легко выводятся соотношения, связывающие начальные и центральные моменты.

Например, известно, что дисперсию можно вычислить по следующему соотношению:

$$D(X) = M(X^2) - [M(X)]^2,$$

или в обозначениях начальных моментов:

$$\mu_2 = v_2 - v_1^2.$$

Нетрудно, исходя из определения центрального момента и пользуясь свойствами математического ожидания, получить формулы:

$$\mu_3 = v_3 - 3 v_2 v_1 + 2 v_1^3, \quad \mu_4 = v_4 - 4 v_3 v_1 + 6 v_2 v_1^2 - 3 v_1^4.$$

Моменты более высоких порядков применяются редко.

Замечание: Моменты, рассмотренные здесь, называются теоретическими.

Первый начальный момент и второй центральный момент – соответственно математическое ожидание и дисперсия, – наиболее часто применяемые характеристики случайной величины. Они характеризуют наиболее важные черты распределения – его положение и степень разбросанности. Для более подробного описания распределения применяются моменты высших порядков (главным образом центральные моменты третьего и четвертого порядков).

Третий центральный момент служит для характеристики *асимметрии* (или «скошенности») распределения. Если распределение симметрично, то все моменты нечетного порядка равны нулю. Естественно поэтому в качестве характеристики асимметрии распределения выбрать какой-то из нечетных моментов. Простейший из них есть третий центральный момент. Он имеет размерность куба случайной величины. Чтобы получить безразмерную характеристику, третий момент μ_3 делят на куб среднего квадратического отклонения. Полученная величина называется «коэффициент асимметрии» или просто «асимметрия»:

$$Sk = \mu_3 / \sigma^3 . \quad (20)$$

Если кривая функции распределения скошена влево, то «асимметрия» имеет положительное значение, если вправо – то отрицательное.

Четвертый центральный момент служит для характеристики т.н. «крутости», т.е. островершинности или плосковершинности распределения. Эти свойства описываются с помощью так называемого *эксцесса*.

Эксцессом случайной величины X называется величина

$$Ex = \mu_4 / \sigma^4 - 3 . \quad (21)$$

Число 3 вычитается из отношения μ_4 / σ^4 потому, что для весьма важного и широко распространенного нормального распределения $\mu_4 / \sigma^4 = 3$. Таким образом, для нормального распределения эксцесс равен нулю; кривые, более островершинные по сравнению с нормальной, имеют положительный эксцесс, а кривые более плосковершинные – отрицательный эксцесс.

Кроме начальных и центральных моментов, на практике иногда применяют и т.н. абсолютные моменты (начальные и центральные). Из абсолютных моментов наиболее часто применяют первый абсолютный центральный момент, называемый *средним арифметическим отклонением*:

$$M(|X - m|), \quad (22)$$

Наряду с дисперсией и средним квадратическим отклонением *среднее арифметическое отклонение* иногда применяется как характеристика рассеяния.

3. Нормально распределенные случайные величины, их параметры. Основные свойства нормального распределения.

Нормальный закон распределения (часто называемый законом Гаусса или гауссовским распределением) исключительно важен в теории вероятности и в статистике и занимает среди других законов распределения особое положение. Это – наиболее часто встречающийся на практике закон распределения. Главная особенность этого закона состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях.

В ТВ доказывается, что сумма достаточно большого числа независимых (или слабо зависимых) случайных величин приближенно подчиняется нормальному закону, и это выполняется тем точнее, чем большее количество случайных величин суммируется. Основное ограничение, налагаемое на суммируемые случайные величины, состоит в том, чтобы они все равномерно играли в общей сумме относительно малую роль. Если это условие не выполняется и, например, одна из случайных величин окажется по своему влиянию на сумму преобладающей над всеми другими, то закон распределения этой преобладающей случайной величины определит в основных чертах закон распределения суммарной случайной величины.

Нормальный закон распределения характеризуется плотностью вероятности вида:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (23)$$

Легко показывается, что действительно соотношение (3.54) есть плотность распределения. Для этого доказывается (путем прямого вычисления), что интеграл от функции (3.54) в пределах от $-\infty$ до $+\infty$ равен единице.

Кривая плотности нормального распределения имеет симметричный холмообразный вид.

Максимальная ордината кривой, равная $\frac{1}{\sigma\sqrt{2\pi}}$, соответствует точке $x = m$; по мере удаления от точки m плотность распределения падает, и при $x \rightarrow \pm\infty$ кривая асимптотически приближается к оси абсцисс.

Легко доказать, что величина m есть не что иное, как математическое ожидание, а величина σ – среднее квадратическое отклонение (или другими словами, стандартное отклонение) величины X .

Для этого вычисляют сначала следующий интеграл:

$$\int_{-\infty}^{\infty} x \cdot f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx, \quad (24)$$

и показывается, что он равен m ; а потом показывают, что дисперсия:

$$\int_{-\infty}^{\infty} (x-m)^2 \cdot f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx, \quad (25)$$

равна σ^2 , т.е. показывается, что параметр σ есть не что иное, как среднее квадратическое отклонение (стандартное отклонение).

При выводе используют тот факт, что известный интеграл Эйлера–Пуассона:

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}. \quad (26)$$

Параметр m , таким образом, характеризует положение распределения на оси абсцисс.

Параметр σ характеризует не положение, а форму кривой распределения. Наибольшая ордината кривой распределения обратно пропорциональна σ ; при увеличении σ максимальная ордината уменьшается. Так как площадь под кривой распределения всегда остается равной единице, то при увеличении σ кривая распределения становится более плоской. Напротив, при уменьшении σ кривая распределения вытягивается вверх.

Размерность параметра σ , естественно совпадает с размерностью случайной величины X .

Свойство 1. Для нормального распределения математическое ожидание случайной величины совпадает с медианой и модой. Равенство этих показателей указывает на то, что случайная величина распределена по

нормальному закону и используется при выяснении того, подчиняется ли закон распределения случайной величины нормальному закону.

Свойство 2. Вероятность случайной величины отклониться в ту или другую сторону от средней m на σ , 2σ , 3σ как следует из таблиц для значений интеграла вероятностей равна приблизительно соответственно 0,683; 0,955; 0,997.

Это значит, например, что при распределении совокупности наблюдений по нормальному закону 68,3% вариант от объема совокупности окажутся в интервале от $m - \sigma$ до $m + \sigma$, 95,5% в интервале от $m - 2\sigma$ до $m + 2\sigma$ и 99,7% – в интервале от $m - 3\sigma$ до $m + 3\sigma$. Последний вывод известен в биометрии как правило «плюс–минус трех сигм».

Свойство 3. В точках при $x = m - \sigma$, $x = m + \sigma$ нормальная кривая имеет точки перегиба.

В справочниках приводятся таблицы для так называемой *стандартной нормальной функции распределения*, под которой понимают функции распределения вида (1) с $m = 0$ и $\sigma = 1$.

Можно показать, что вероятность того, что X примет значение, принадлежащее интервалу (α, β) такова:

$$P(\alpha < X < \beta) =$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^{(\beta-m)/\sigma} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_0^{(\alpha-m)/\sigma} e^{-\frac{x^2}{2}} dx =$$

$$= \Phi((\beta - m)/\sigma) - \Phi((\alpha - m)/\sigma), \quad (27)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ есть функция Лапласа.

4. Распределение «хи-квадрат»; t-распределение Стьюдента; распределение F (распределение Фишера-Снедекора).

Пусть X_i ($i = 1, 2, \dots, n$) – нормальные независимые случайные величины, причем математическое ожидание каждой из них равно нулю, а среднее квадратическое отклонение – единице. Тогда сумма квадратов этих величин

$$\chi^2 = \sum_{i=1}^n X_i^2 \quad (28)$$

распределена по закону χ^2 («хи-квадрат») с n степенями свободы. Если эти величины связаны одним линейным соотношением, например, $\sum_{i=1}^n X_i = n\bar{X}$, то число степеней свободы равно $k = n - 1$.

Плотность этого распределения $f(x)$ равна нулю при $x \leq 0$ и $\frac{1}{2^{k/2} \Gamma(k/2)} e^{-\frac{x}{2}} x^{(k/2)-1}$ при $x > 0$.

Отсюда видно, что распределение «хи-квадрат» определяется одним параметром – числом степеней свободы k . С увеличением числа степеней свободы распределение медленно приближается к нормальному.

Распределение «хи-квадрат» расположено правее оси ординат и является асимметричным распределением.

Математическое ожидание равно $M(\chi^2) = n$, а дисперсия равна $D(\chi^2) = 2n$.

Если $n > 30$, то распределение величины $z = \sqrt{2\chi^2} - \sqrt{2n-1}$ является приближенно нормальным с нулевым средним и единичной дисперсией.

Замечание. Распределение «хи-квадрат» играет важную роль при исследовании выборочной дисперсии s^2 для выборки, взятой из нормально распределенной совокупности.

t – распределение Стьюдента

Пусть Z – нормальная случайная величина, причем $M(Z) = 0$, $D(Z) = 1$, а V – независимая от Z величина, которая распределена по закону «хи-квадрат» с k степенями свободы. Тогда величина

$$T = \frac{Z}{\sqrt{V/k}}, \quad (29)$$

имеет распределение, которое называется t -распределением или распределением Стьюдента с k степенями свободы.

С возрастанием числа степеней свободы распределение Стьюдента быстро приближается к нормальному.

Плотность распределения Стьюдента имеет следующий вид:

$$S(x, k) = B_k [1 + x^2/k]^{-(k+1)/2}, \quad \text{где } B_k = \frac{\Gamma((k+1)/2)}{\sqrt{\pi(k)} \cdot \Gamma(k/2)}.$$

Распределение Стьюдента симметрично относительно оси ординат с нулевым математическим ожиданием. Его дисперсия равна $k/(k-2)$.

Замечание. Это распределение играет важную роль в тех случаях, когда рассматриваются выборочные средние и неизвестна дисперсия генеральной совокупности.

Распределение F (Фишера – Снедекора)

Если U и V – независимые случайные величины, распределенные по закону «хи- квадрат» со степенями свободы m и n , то величина

$$F = \frac{U / m}{V / n} \quad (30)$$

имеет распределение, которое называют распределением F Фишера-Снедекора со степенями свободы m и n (иногда его обозначают через V^2).

Плотность этого распределения

$$f(x) = 0 \text{ при } x \leq 0 \text{ и } f(x) = C_0 x^{(m-2)/2} / (n+mx)^{(n+m)/2} \text{ при } 0 \leq x < \infty,$$

$$\text{где нормирующий множитель } C_0 = \frac{\Gamma(\frac{n+m}{2}) m^{m/2} n^{n/2}}{\Gamma(m/2) \Gamma(n/2)}.$$

Распределение F определяется двумя параметрами – числами степеней свободы.

Распределение Фишера–Снедекора асимметрично и расположено правее оси ординат.

Математическое ожидание F -распределения равно $n/(n - 2)$, а дисперсия равна

$$2n^2 (m + n - 2) / (m (n - 2)^2 (n - 4)) \quad (n > 4).$$

В пределе при $n \rightarrow \infty$ математическое ожидание становится равным 1, а дисперсия равной $2/m$.

Замечание. F -распределение приобретает особую важность, когда сравниваются выборочные дисперсии из нормально распределенных совокупностей. Оно также используется в регрессионном и дисперсионном анализе.

5. Логарифмически нормальное распределение случайных величин

Говорят, что случайная величина X имеет логарифмически нормальное распределение, если существует такое число Θ , что величина $z = \ln(X - \Theta)$ распределена нормально.

Так как логарифмическая функция определена только для положительных чисел, значения X должны быть больше Θ . Распределение случайной величины X включает три параметра: Θ , μ , σ . Последние являются математическим ожиданием и дисперсией нормального распределения.

Плотность вероятности величины X имеет вид:

$$f(x) = \frac{1}{(x - \Theta)\sigma\sqrt{2\pi}} \exp(-0,5 [\ln(x - \Theta) - \mu]^2 / \sigma^2) \quad (31)$$

Отметим, что математическое ожидание и дисперсия X равны:

$$M(X) = \Theta + \exp(\mu + 0,5\sigma^2) \quad (32)$$

$$D(X) = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1) \quad (33)$$

Распределение имеет одну моду и положительный коэффициент асимметрии.

Когда σ^2 мало, то распределение по форме близко к нормальному.

Известно, что во многих приложениях Θ равно нулю, и тогда X имеет двухпараметрическое логарифмически нормальное распределение.

Лекция 3. ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

1. Генеральная совокупность и выборка, ее репрезентативность.
2. Группировка результатов наблюдений. Понятие о вариационных рядах.
3. Эмпирическая функция распределения.
4. Средняя ряда. Показатели вариации. Дисперсия, среднее квадратичное отклонение (стандартное отклонение).
5. Средняя геометрическая величина.
6. Коэффициент корреляции (Пирсона). Основные свойства коэффициента корреляции.

1. Генеральная совокупность и выборка, ее репрезентативность.

Пусть требуется изучить совокупность однородных объектов относительно некоторого *качественного* или *количественного признака*, характеризующего эти объекты.

Иногда проводят *сплошное* обследование. На практике, однако, сплошное обследование применяют редко. Чаще случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению.

Выборочной совокупностью или просто выборкой называют совокупность случайно отобранных объектов.

Генеральной совокупностью называют совокупность объектов, из которых производится выборка.

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности.

Различают повторную (с возвращением) и бесповторную (при которой отобранный объект в генеральную совокупность не возвращается) выборки.

Для того, чтобы по данным выборки можно было судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно ее представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование коротко формулируют так: выборка должна быть *репрезентативной (представительной)*.

Для этого необходимо, чтобы выборка была:

случайной;

объем выборки должен быть достаточно велик.

Способы отбора:

- Отбор не требующий расчленения генеральной совокупности на части: а) простой случайный бесповторный отбор; б) простой случайный повторный отбор.

- Отбор, при котором генеральная совокупность разбивается на части: а) типический отбор; б) механический отбор; в) серийный отбор.

Чтобы избежать грубых ошибок в работе и получить сопоставимые результаты, необходимо соблюдать признанные правила записи и округления приближенных чисел. Очень важно, чтобы числа, фиксируемые в документах учета, соответствовали *точности, принятой при измерении варьирующих объектов*. Так, если измерения проводят с точностью до одного десятичного знака, то результаты измерений нельзя записывать, например, в таком виде: 4,2; 6; 6,69; 5,082 и т.д. Правильная запись этих чисел будет такова: 4,2; 6,0; 6,7; 5,1.

Числа округляют следующим образом: если за последней сохраняемой цифрой следуют цифры 0, 1, 2, 3, 4, они отбрасываются (округление с недостатком); если же за последней сохраняемой цифрой следуют цифры 5, 6, 7, 8 и 9, то последняя цифра сохраняемая цифра увеличивается на единицу (округление с избытком).

2. Группировка результатов наблюдений. Понятие о вариационных рядах.

Обработка начинается с упорядочения или систематизации собранных данных. Процесс систематизации результатов массовых наблюдений, объединения их в относительно однородные группы по некоторому признаку называется *группировкой*.

Наиболее распространенной формой группировки являются *статистические таблицы*; они бывают простыми и сложными. К простым относятся, например, четырехпольные таблицы, применяемые при альтернативной группировке, когда одна группа вариант противопоставляется другой.

К сложным относятся многопольные таблицы.

Особую форму группировки представляют так называемые статистические ряды.

Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалась n_1 раз, x_2 – n_2 раз, ... x_k – n_k раз и $\sum n_k = n$ – объем выборки.

Наблюдаемые значения x_i называют *вариантами*, а последовательность вариант, записанных в возрастающем порядке, – вариационным рядом. Числа наблюдений (n_i) называют *частотами*, а их отношения к объему выборки

$n_i/n = W_i$ *относительными частотами.* (Вариационный ряд – ряд ранжированных значений признака, в котором указана частота их встречаемости в данной совокупности.)

Статистическим распределением выборки называют перечень вариантов и соответствующих им частот или относительных частот.

Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот попавших в этот интервал).

Вариационные ряды бывают:

- безынтервальные;
- интервальные (неравноинтервальные или равноинтервальные).

Число классов в интервальном вариационном ряду определяют:

- по таблицам (например, при объеме выборки $n=25\ldots 40$ число классов принимают 5..6; при $n=40\ldots 60$ – 6-8;
- по формуле, например, Стерджеса: $K = 1 + 3,32 \cdot \lg n$,
- или по формуле Брукса и Карузерс: $K = 5 \cdot \lg n$ при $n > 100$.

Графическое изображение вариационных рядов.

Для наглядного представления закономерностей варьирования количественных признаков, вариационные ряды изображают в виде графиков. При построении графика безынтервального вариационного ряда по оси абсцисс откладывают срединные значения классов, по оси ординат – частоты. Высота перпендикуляров, восставляемых по оси абсцисс, соответствует частотам классов. Соединяя вершины перпендикуляров прямыми линиями, получают геометрическую фигуру в виде многоугольника, называемую *полигоном распределения частот*.

При построении графика интервального вариационного ряда по оси абсцисс откладывают границы классовых интервалов, по оси ординат – частоты интервалов. В результате получается *гистограмма распределения частот*.

3. Эмпирическая функция распределения.

Пусть n_x – число наблюдений, при которых наблюдалось значение признака, меньшее x ; n – объем выборки. Ясно, что относительная частота события $X < x$ равна n_x/n .

Эмпирической функцией распределения (функцией распределения выборки или выборочной функцией распределения) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$.

Итак, по определению,

$$F^*(x) = n_x/n, \quad (1)$$

где n_x – число вариантов, меньших x ;

n – объем выборки.

В отличие от эмпирической функции распределения выборки функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*.

При больших n $F^*(x)$ и $F(x)$ мало отличаются одно от другого. Отсюда следует целесообразность использования эмпирической функции распределения выборки для приближенного представления теоретической (интегральной) Функции распределения генеральной совокупности.

Следует заметить, что $F^*(x)$ обладает всеми свойствами $F(x)$ (принадлежит $[0,1]$; неубывающая; $F^*(x) = 0$ при $x \leq x_1$; $F^*(x) = 1$ при $x > x_i$).

4. Средняя ряда. Показатели вариации. Дисперсия, среднее квадратичное отклонение (стандартное отклонение).

Средняя ряда (выборочная средняя) определяется по формуле:

а) Если все значения $x_1, x_2, \dots, x_n$ признака вариационного ряда (выборки) объема n различны, то

$$\bar{x}_B = (x_1 + x_2 + \dots + x_n) / n \quad (2)$$

б) Если же значения признака $x_1, x_2, \dots, x_k$ имеют соответственно частоты $n_1, n_2, \dots, n_k$, причем $n_1 + n_2 + \dots + n_k = n$, то

$$\bar{x}_B = (x_1 n_1 + x_2 n_2 + \dots + x_k n_k) / n \quad \text{или}$$

$$\bar{x}_B = \left(\sum_{i=1}^k n_i x_i \right) / n \quad (2^*)$$

т.е. средняя вариационного ряда (выборочная средняя) есть средняя взвешенная значений признака с весами, равными соответствующим частотам.

Выборочную среднюю можно рассматривать как случайную величину, а следовательно, можно говорить о распределениях (теоретическом и

эмпирическом) выборочной средней и о числовых характеристиках этого распределения (его называют выборочным), в частности о математическом ожидании и дисперсии выборочного распределения.

Медианой называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно, т.е. если $n=2k+1$, то $m_e = x_{k+1}$; при четном $n = 2k$, медиана $m_e = (x_k + x_{k+1})/2$.

Модой M_0 называют варианту, которая имеет наибольшую частоту.

Квантили отсекают в пределах ряда определенную часть его вариант (квартили отсекают $1/4$ ряда, децили - отсекают $1/10$ ряда, перцентили или проценти - $1/100$ ряда). В практике обычно используют перцентили: $P_3, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}, P_{97}$.

Медиану и моду, а также квантили называют структурными средними.

Квантили, как и медиана (средняя) – ценные характеристики варьирующих объектов, особенно в тех случаях, когда обнаруживается значительная или резкая асимметрия в распределении частот по классам вариационного ряда.

Показатели вариации

Для того чтобы охарактеризовать рассеяние наблюдаемых значений количественного признака выборки около своего среднего значения $\bar{x}_B$, вводят характеристику - выборочную дисперсию.

Выборочной дисперсией D_B называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{x}_B$.

Если все значения $x_1, x_2, \dots, x_n$ признака вариационного ряда (выборки) объема n различны, то

$$D_B = (\sum_{i=1}^k (x_i - \bar{x}_B)^2) / n \quad (3)$$

Если же значения признака $x_1, x_2, \dots, x_k$ имеют соответственно частоты $n_1, n_2, \dots, n_k$, причем $n_1 + n_2 + \dots + n_k = n$, то

$$D_B = (\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2) / n. \quad (3^*)$$

То есть, выборочная средняя есть средняя взвешенная квадратов отклонений с весами, равными соответствующим частотам.

Кроме дисперсии для характеристики рассеяния значений признака выборочной совокупности около своего выборочного среднего пользуются и *выборочным средним квадратическим отклонением*.

Выборочным средним квадратическим отклонением (стандартом) называют квадратный корень из выборочной дисперсии:

$$\sigma_B = \sqrt{D_B} \quad (4)$$

Вычисление дисперсии можно упростить, используя следующую формулу:

$$D_B = \overline{x^2} - (\bar{x})^2. \quad (5)$$

(Дисперсия равна среднему квадратов значений минус квадрат выборочной средней).

Оказывается, что если в качестве оценки генеральной дисперсии принять выборочную дисперсию, то эта оценка будет приводить к систематическим ошибкам, давая заниженное значение генеральной дисперсии. Легко исправить выборочную дисперсию так, чтобы ее математическое ожидание было равно генеральной дисперсии. Достаточно для этого умножить D_B на дробь $n/(n - 1)$. Сделав это получим так называемую исправленную дисперсию, которую обычно обозначают через s^2 :

$$s^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n - 1} \quad (6)$$

Замечание: Исправленная дисперсия является несмещенной оценкой генеральной дисперсии, т.е.

$$M[s^2] = D_{\Gamma}. \quad (7)$$

Итак, в качестве оценки генеральной дисперсии принимают исправленную дисперсию (соотношение (6)).

Для оценки же среднего квадратического отклонения используют «исправленное» среднее квадратическое отклонение, которое равно квадратному корню из исправленной дисперсии:

$$s = \sqrt{\frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n - 1}} \quad (8)$$

Замечание: s не является несмещенной оценкой.

Кроме выборочной средней и выборочной дисперсии применяются и другие характеристики вариационного ряда:

Размахом варьирования (размахом вариации) R называют разность между наибольшей и наименьшей вариантами:

$$R = x_{\max} - x_{\min}. \quad (9)$$

Коэффициентом вариации V называют выраженное в процентах отношение выборочного (иногда «исправленного») среднего квадратического отклонения к выборочной средней: $V = \sigma_B / \bar{x}_B \cdot 100\%$.

Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого коэффициент вариации больше. Коэффициент вариации безразмерная величина, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность.

Показатель точности определения средней: $Cs = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100\%$,

(отношение ошибки средней к средней), где $s_{\bar{x}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$.

Показатель точности можно вычислить и по следующей формуле:

$$Cs = V / \sqrt{n}$$

где V – показатель вариации, а n – объем выборки.

5. Средняя геометрическая величина.

Доверительный интервал средней геометрической

В тех случаях, когда распределение данных экспериментов приближается к нормальному распределению при оперировании не непосредственно полученными числовыми результатами, а их логарифмами, следует применять среднюю геометрическую.

Средняя геометрическая ($\bar{X}_{geom}$) вычисляется по формуле:

$$\bar{X}_{geom} = \sqrt[n]{\prod_{i=1}^n x_i}, \quad (10)$$

где $\prod$ – знак произведения;

n – объем выборки.

Практически удобнее рассчитывать среднюю геометрическую не непосредственно по формуле (4.6), а так:

вычислить логарифмы всех величин x_i ;

рассчитать среднюю арифметическую логарифмов всех величин x_i ;

определить ее антилогарифм (пропотенцировать).

Последовательность расчета доверительного интервала:

рассчитать доверительный интервал, как для средней арифметической для логарифмов;

взять антилогарифм от средней логарифма и от концов доверительного интервала для логарифмов и получить доверительный интервал для средней геометрической в виде:

$$(\bar{X}_{геом} : K, \bar{X}_{геом} \times K) \quad (11)$$

где K – антилогарифм доверительного интервала для средней арифметической логарифмов величин.

6. Коэффициент корреляции (Пирсона). Основные свойства коэффициента корреляции.

Для описания системы двух случайных величин кроме математических ожиданий и дисперсий используют и другие характеристики; к их числу относят корреляционный момент и коэффициент корреляции (Пирсона).

Корреляционным моментом μ_{xy} (ковариацией) случайных величин X и Y называют математическое ожидание произведения отклонений этих величин:

$$\mu_{xy} = M\{[X - M(X)][Y - M(Y)]\} \quad (12)$$

Корреляционный момент служит для характеристики связи между величинами X и Y . Корреляционный момент равен нулю, если X и Y независимы:

Теорема 1. Корреляционный момент двух независимых случайных величин равен нулю.

Док-во. Т.к. X и Y – независимые случайные величины, то и их отклонения $X - M(X)$ и $Y - M(Y)$ также независимы. Пользуясь свойством математического ожидания: математ. ожидание произведения независимых случайных величин равно произведению математических ожиданий сомножителей, и отклонения (математическое ожидание отклонения равно нулю), получим:

$$\mu_{xy} = M\{[X - M(X)][Y - M(Y)]\} = M[X - M(X)] \cdot M[Y - M(Y)] = 0 . \blacktriangle$$

Величина корреляционного момента зависит от единиц измерения случайных величин. Это является недостатком этой числовой

характеристики. Для того чтобы устранить этот недостаток, вводят новую числовую характеристику – *коэффициент корреляции*.

Коэффициентом корреляции r_{xy} случайных величин X и Y называют отношение корреляционного момента к произведению средних квадратичных отклонений этих величин. Часто индекс y коэффициента корреляции Пирсона опускают и пишут r .

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}. \quad (13)$$

r – безразмерная величина и не зависит от выбора единиц измерения случайных величин.

Очевидно, коэффициент корреляции независимых случайных величин равен нулю (т.к. $\mu_{xy} = 0$).

Абсолютная величина коэффициента корреляции не превышает единицы:

$$|r| \leq 1.$$

Две случайные величины X и Y называют коррелированными, если корреляционный момент (или, что то же, коэффициент корреляции) отличен от нуля;

X и Y называют некоррелированными, если корреляционный момент (или, что то же, коэффициент корреляции) равен нулю.

Две коррелированные величины также и зависимы (иначе было бы $\mu_{xy} = 0$, а это противоречит условию, т.к. для коррелированных величин $\mu_{xy} \neq 0$).

Обратное не всегда верно, т.е. если две величины зависимы, то они могут быть как коррелированными, так и некоррелированными. Иначе говоря, корреляционный момент двух зависимых величин может быть не равен нулю, но может быть и равняться нулю.

Итак, из коррелированности двух случайных величин следует их зависимость, но из зависимости еще не вытекает коррелированность. Из независимости двух величин следует их некоррелированность, но из некоррелированности еще нельзя заключить о независимости этих величин.

Однако для нормально распределенных случайных величин из их некоррелированности вытекает их независимость (т.е. для нормально распределенных величин понятия независимости и некоррелированности равносильны).

Коэффициент корреляции характеризует не всякую зависимость, а только *линейную* зависимость. Линейная вероятностная зависимость случайных величин заключается в том, что при возрастании одной случайной

величины другая имеет тенденцию возрастать (или же убывать) по линейному закону. Эта тенденция к линейной зависимости может быть более или менее ярко выраженной, более или менее приближаться к функциональной, т.е. самой тесной линейной зависимости. Коэффициент корреляции характеризует степень тесноты линейной зависимости между случайными величинами. Если случайные величины X и Y связаны линейной функциональной зависимостью:

$$Y = aX + b$$

то $r = \pm 1$, причем, если a положителен то $+1$, если a отрицателен, то -1 .

В случае $r > 0$ говорят о *положительной корреляции*, в случае $r < 0$ – об *отрицательной корреляции*.

Если корреляция равна нулю, то из этого следует отсутствие какой-либо *линейной* зависимости между X и Y .

Считается, что: при $|r| < 0,3$ – корреляция отсутствует, при $0,3 \leq |r| < 0,5$ – корреляция слабая, при $0,5 \leq |r| < 0,7$ – корреляция средняя, при $0,7 \leq |r| < 0,9$ – корреляция сильная и при $0,9 \leq |r| \leq 1,0$ – корреляция значительная.

При обработке малочисленных выборок (при $n < 30$) расчет коэффициента корреляции по формуле (9.2) дает заниженные оценки генерального коэффициента корреляции. В таких случаях лучшую оценку коэффициента корреляции генеральной совокупности получают при использовании поправки, на которую умножают эмпирический коэффициент корреляции, т.е.

$$r^* = r \cdot \left[1 + \frac{1 - r^2}{2(n - 3)} \right]. \quad (14)$$

Для оценки статистической значимости корреляции удобно воспользоваться формулой:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} \quad (15)$$

Число степеней свободы здесь $\nu = n - 2$.

Если при выбранном уровне значимости (например, 0,05) вычисленное значение t меньше критического, то корреляция статистически не значима.

Выборочный коэффициент корреляции не будет точной оценкой генерального параметра, если он вычислен на малочисленной выборке и его абсолютное значение превышает 0,5. Р. Фишер нашел более точный способ оценки генерального параметра по значению выборочного коэффициента корреляции. Этот способ сводится к замене r преобразованной величиной z , которая связана с выборочным коэффициентом корреляции следующим образом:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}. \quad (16)$$

Величина z меняет свое значение от $-\infty$ до $+\infty$, а ее распределение быстро приближается к нормальному со средним значением

$$\bar{z} = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} \right) \text{ и дисперсией } \sigma_z^2 = \frac{1}{n-3}.$$

Критерием достоверности показателя z является следующее отношение:

$$t_z = \frac{z}{s_z} = z \cdot \sqrt{n-3}. \quad (17)$$

Этот критерий используют в тех случаях, когда вместо коэффициента корреляции берут соответствующее ему число z . Нулевая гипотеза отвергается на принятом уровне значимости α и числе степеней свободы $k = n - 2$. Значения критических точек приведены в таблице П.1 Приложения 1.

Лекция 4. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ

1. Доверительный интервал средней арифметической; определение необходимого объема выборки; величины доверительных вероятностей, принятых в биологических экспериментах.

2. Правила исключения выскакивающих величин.

3. Анализ качественных признаков. Доверительный интервал для долей.

4. Доверительный интервал средней геометрической.

5. Определение необходимого объема выборки.

1. Доверительный интервал средней арифметической; величины доверительных интервалов, принятых в биологических экспериментах.

Точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные на предыдущих занятиях – точечные. При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам. По этой причине при небольшом объеме выборки следует пользоваться *интервальными оценками*.

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Точностью оценки называется величина

$$|\theta - \theta^*|,$$

где θ^* – оценка неизвестного параметра θ .

Надежностью (доверительной вероятностью) оценки θ по θ^* называют вероятность γ , с которой осуществляется неравенство

$$|\theta - \theta^*| < \delta. \quad (1)$$

Обычно надежность (доверительная вероятность) задается наперед. Наиболее часто задают надежность γ , равную 0,95; 0,99 и 0,999.

Из (1) следует:

$$\theta^* - \delta < \theta < \theta^* + \delta. \quad (2)$$

Интервал $(\theta^* - \delta, \theta^* + \delta)$ называют доверительным. Он покрывает неизвестный параметр (параметр θ) с заданной надежностью (доверительной вероятностью) γ . Величину α , равную $(1 - \gamma)$ называют уровнем значимости.

Доверительные интервалы для оценки математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении σ

Пусть количественный признак X генеральной совокупности распределен нормально, но среднее квадратическое отклонение σ неизвестно. Требуется оценить неизвестное математическое ожидание с помощью метода доверительных интервалов.

Оказывается, что по данным выборки можно построить случайную величину (ее возможные значения будем обозначать через t):

$$T = \frac{\bar{X} - a}{S / \sqrt{n}}, \quad (3)$$

которая имеет распределение Стьюдента с $k=n-1$ степенями свободы. Здесь $\bar{X}$ – выборочная средняя, S – «исправленное» среднее квадратическое отклонение, n – объем выборки.

Плотность распределения Стьюдента

$$S(t, n) = B_n [1 + t^2 / (n - 1)]^{-n/2}, \quad (4)$$

где $B_n = \Gamma(n/2) / \sqrt{\pi(n-1)} \Gamma((n-1)/2)$

Видим, что распределение Стьюдента определяется параметром n – объемом выборки (или, что то же, числом степеней свободы $k = n - 1$) и не зависит от неизвестных параметров a и σ , что является его большим достоинством.

Вероятность осуществления неравенства

$$\left| \frac{\bar{X} - a}{S / \sqrt{n}} \right| < t_\gamma,$$

определяется так:

$$P\left(\left| \frac{\bar{X} - a}{S / \sqrt{n}} \right| < t_\gamma\right) = 2 \int_0^{t_\gamma} S(t, n) dt = \gamma \quad (5)$$

Заменив неравенство в круглых скобках равносильным ему двойным неравенством, получим

$$P(\bar{X} - t_\gamma S / \sqrt{n} < a < \bar{X} + t_\gamma S / \sqrt{n}) = \gamma \quad (5^*)$$

Итак, пользуясь распределением Стьюдента, нашли доверительный интервал

$$(\bar{X} - t_\gamma S / \sqrt{n}, \bar{X} + t_\gamma S / \sqrt{n}), \quad (6)$$

покрывающий неизвестный параметр a с надежностью (доверительной вероятностью) γ . Здесь величины $\bar{x}$ и s найдены по выборке.

t_γ можно найти по таблице критических точек распределения Стьюдента при заданных n и α .

Определение необходимого объема выборки. Если требуется определить среднее значение с наперед заданной точностью δ и надежностью γ , то минимальный объем выборки, который обеспечит эту точность, можно найти по формуле $n = t_\gamma^2 S^2 / \delta^2$, которая следует из равенства $\delta = t_\gamma S / \sqrt{n}$.

В биологических экспериментах приняты следующие величины доверительных вероятностей: 0,95 (наиболее часто используемый); 0,99 и 0,999.

2. Правила исключения выскакивающих значений.

Используется большое количество различных способов для отсева грубых погрешностей наблюдения (выскакивающих значений, аномальных значений). Если выборка небольшого объема ($n \leq 25$), то можно воспользоваться методом максимального относительного отклонения. Для выделения аномального значения вычисляют

$$\tau = \frac{|x_i - \bar{x}|}{\sqrt{(n-1)/n} \cdot S}$$

и полученный результат сравнивают с критическим значением, взятым из следующей таблицы при соответствующем n и p .

Таблица. Квантили распределения величины τ_p

n	$p = 0,90$	$p = 0,95$	$p = 0,99$
3	1,41	1,41	1,41
4	1,64	1,69	1,72
5	1,79	1,87	1,96
6	1,89	2,00	2,13
7	1,97	2,09	2,26
8	2,04	2,17	2,37
9	2,10	2,24	2,46
10	2,15	2,29	2,54

Если соблюдается неравенство $\tau \leq \tau_p$, то наблюдение не отсеивают, если не соблюдается, то наблюдение исключают. После исключения того или иного наблюдения или нескольких наблюдений характеристики эмпирического распределения должны быть пересчитаны по данным сокращенной выборки.

Отсев грубых погрешностей можно провести и для больших выборок. Для практических целей лучше всего использовать таблицы распределения Стьюдента. Этот метод исключения аномальных значений отличается простотой, а таблицы распределения Стьюдента имеются практически в любой книге по математической статистике.

3. Анализ качественных признаков. Доверительный интервал для доли.

Многие признаки невозможно измерить численно. Здесь мы имеем дело с качественными признаками. Эти признаки не связаны между собой никакими арифметическими соотношениями, упорядочить их также нельзя. Единственный способ описания качественных признаков состоит в том, чтобы подсчитать число объектов, имеющих один и тот же качественный признак. Кроме того, можно подсчитать, какая доля от общего числа объектов приходится на тот или иной признак. Расчет генеральной доли из общей генеральной совокупности $N = 200$ и выборочных долей из 25 выборок объемом 10 иллюстрирует рисунок 1.

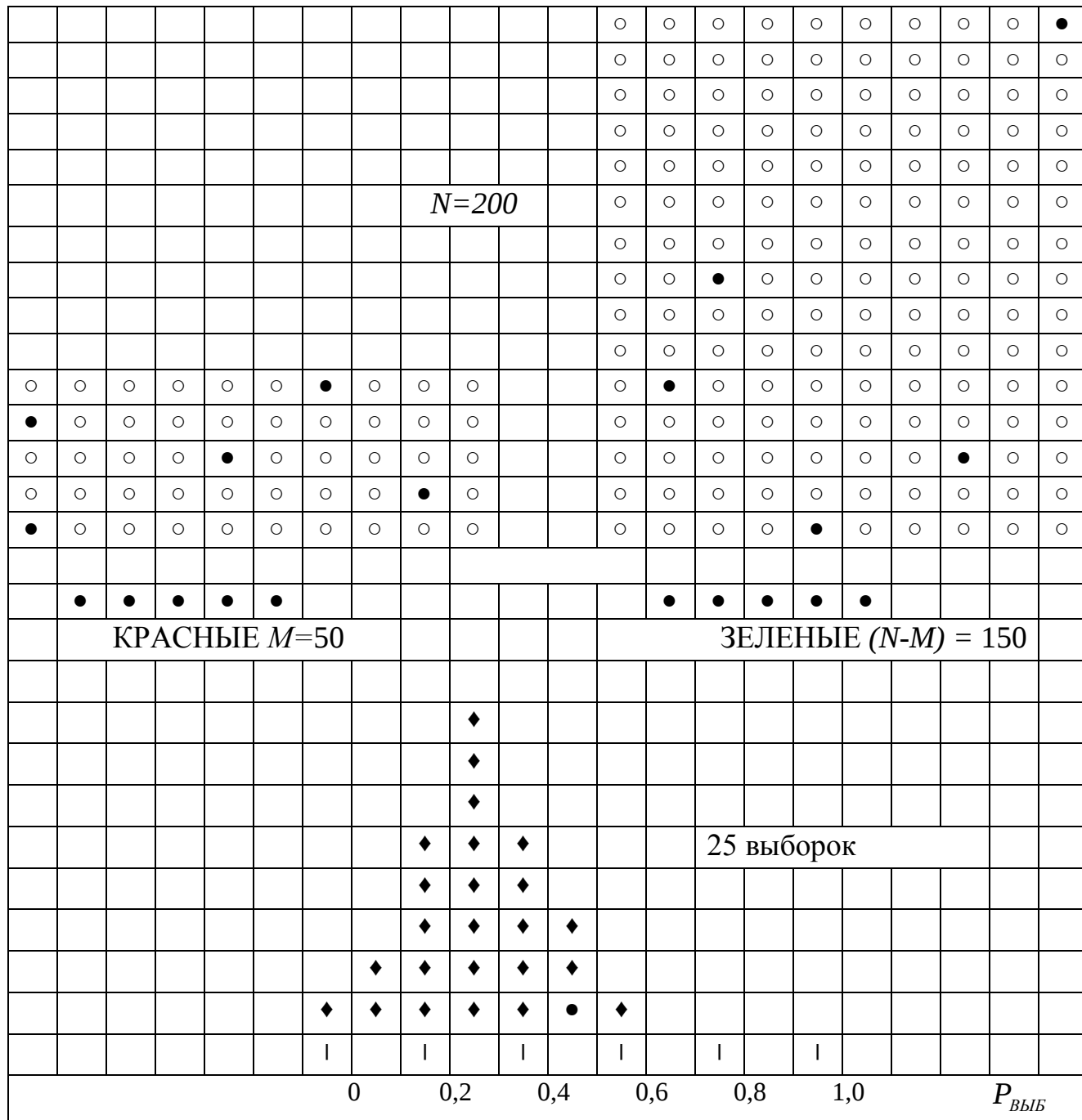


Рисунок 1 – Анализ качественных признаков – цвета (красного и зеленого)

Среднее выборочных долей $\bar{P} = 0,3$

Среднее значение признака и стандартное отклонение равно

$$\mu = \frac{\sum X}{N} = \frac{M \times 1 + (N - M) \times 0}{N} = \frac{M}{N} = p ; \quad \mu = p = 0,25$$

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum (X - \mu)^2}{N}} = \sqrt{\frac{(1-p)^2 + \dots + (1-p)^2 + (0-p)^2 + \dots + (0-p)^2}{N}} = \\ &= \sqrt{\frac{M(1-p)^2 + (N-M)p^2}{N}} = \sqrt{\frac{M}{N}(1-p)^2 + (1-\frac{M}{N})p^2} = \\ &= \sqrt{p(1-p)^2 + (1-p)p^2} = \sqrt{[p(1-p) + p^2](1-p)} = \sqrt{p(1-p)}; \end{aligned}$$

$$\text{Стандартная ошибка доли: } \sigma_{\bar{P}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}} \Rightarrow$$

$$\text{Оценка стандартной ошибки доли: } S_{\bar{P}} = \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$$

Нормальное распределение служит хорошим приближением для выборочной оценки доли, если $n\bar{P}$ и $n(1-\bar{P})$ больше 5. В этом случае:

$$\bar{P} - t \cdot S_{\bar{P}} < p < \bar{P} + t \cdot S_{\bar{P}}. \quad (7)$$

Если $\bar{P} < 0,25$ или $\bar{P} > 0,75$, то доверительный интервал определяют с помощью величины $\bar{\varphi} = 2 \arcsin \sqrt{\bar{P}}$. Ее распределение близко к нормальному, а выборочная ошибка равна $\frac{1}{\sqrt{n}}$.

В этом случае определяют нижнюю и верхнюю границы интервала для φ :

$$\varphi_{\text{НИЖ}} = \bar{\varphi} - t_{\text{КРИТ}} \cdot \frac{1}{\sqrt{n}}; \quad \varphi_{\text{ВЕР}} = \bar{\varphi} + t_{\text{КРИТ}} \cdot \frac{1}{\sqrt{n}}$$

и рассчитывают доверительный интервал для P :

$$\left(\sin\left(\frac{\varphi_{\text{НИЖ}}}{2}\right)\right)^2 < p < \left(\sin\left(\frac{\varphi_{\text{ВЕР}}}{2}\right)\right)^2. \quad (8)$$

4. Доверительный интервал средней геометрической.

В тех случаях, когда распределение данных экспериментов приближается к нормальному распределению при оперировании не непосредственно полученными числовыми результатами, а их логарифмами, следует применять среднюю геометрическую.

Средняя геометрическая ($\bar{X}_{geom}$) вычисляется по формуле:

$$\bar{X}_{geom} = \sqrt[n]{\prod_{i=1}^n x_i}, \quad (9)$$

где Π – знак произведения;

n – объем выборки.

Практически удобнее рассчитывать среднюю геометрическую не непосредственно по формуле (9), а так:

- вычислить логарифмы всех величин x_i ;
- рассчитать среднюю арифметическую логарифмов всех величин x_i ;
- определить ее антилогарифм (пропотенцировать).

Доверительный интервал средней геометрической рассчитывается в следующей последовательности:

- рассчитать доверительный интервал, как для средней арифметической, для логарифмов;
- взять антилогарифм (т.е. пропотенцировать) от средней логарифма и от концов доверительного интервала для логарифмов и получить доверительный интервал для средней геометрической в виде:

$$(\bar{X}_{geom} : K, \bar{X}_{geom} \times K) \quad (10)$$

где K – антилогарифм доверительного интервала для средней арифметической логарифмов величин.

Лекция 5. КРИТЕРИИ СОГЛАСИЯ

1. Критерий согласия хи-квадрат (Пирсона).
2. Критерий согласия Колмогорова-Смирнова.
3. Однофакторный дисперсионный анализ.

Часто необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить, что он имеет определенный вид (назовем его A), выдвигают гипотезу: генеральная совокупность распределена по закону A . Таким образом, в этой гипотезе речь идет о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр Θ равен определенному значению Θ_0 , выдвигают гипотезу: $\Theta = \Theta_0$. Таким образом, в этой гипотезе речь идет о предполагаемой величине параметра одного известного распределения.

Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений, о независимости выборок и другие.

Статистической называют гипотезу о виде неизвестного распределения, или о параметрах известных распределений.

Например, статистическими являются гипотезы:

- 1) генеральная совокупность распределена по закону Пуассона;
- 2) дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух неизвестных распределений.

Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу H_0 .

Конкурирующей (альтернативной) называют гипотезу, которая противоречит нулевой.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание a нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что $a \neq 10$. Коротко это записывают так: $H_0: a = 10$; $H_1: a \neq 10$.

Различают гипотезы, которые содержат только одно и более одного предположений.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, т.е. могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза.

Ошибка второго рода состоит в том, что будет принята неправильная гипотеза.

Правильное решение может быть принято также в двух случаях:

- 1) гипотеза принимается и в действительности она правильная;
- 2) гипотеза отвергается, причем и в действительности она неверна.

Вероятность совершить ошибку первого рода принято обозначать через α ; ее называют уровнем значимости. Наиболее часто его принимают равным 0,05 или 0,01. Вероятность совершить ошибку второго рода принято обозначать через β , при этом $\beta = 1 - \alpha$.

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближенное распределение которой известно. Эту величину обозначают через U или Z , если она распределена нормально, F или v^2 по закону Фишера-Снедекора, T – по закону Стьюдента, χ^2 – по закону «хи-квадрат» и т.д. обозначим эту величину в целях общности через K .

Статистическим критерием (или просто критерием) называют случайную величину K , которая служит для проверки нулевой гипотезы.

Для проверки гипотезы по данным выборок вычисляют частные значения входящих в критерий величин и таким образом получают частное (наблюдаемое) значение критерия.

Наблюдаемым значением $K_{\text{набл}}$ называют значение критерия, вычисленное по выборке.

После выбора определенного критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая – при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Критическими точками (границами) называют точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Имеется несколько критериев согласия: χ^2 («хи-квадрат») К. Пирсона; Колмогорова – Смирнова и др.

Ограничимся описанием применения критерия Пирсона («хи-квадрат») к проверке гипотезы о нормальном распределении генеральной совокупности (критерий применяется и для других распределений, в этом состоит его достоинство) и описанием критерия Колмогорова–Смирнова.

1. Критерий согласия хи-квадрат (Пирсона)

При применении критерия Пирсона будем сравнивать эмпирические (наблюдаемые) и теоретические (вычисленные в предположении нормального распределения) частоты.

Пусть по выборке объема n получено эмпирическое распределение:

Варианты	x_1	x_2		x_s
Эмпирические частоты	n_1	n_2		n_s

Допустим, что в предположении нормального распределения генеральной совокупности вычислены теоретические частоты n'_i . При уровне значимости α требуется проверить нулевую гипотезу: генеральная совокупность распределена нормально.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$\chi^2 = \sum (n_i - n'_i)^2 / n'_i \quad (1)$$

Эта величина случайная, т.к. в различных опытах она принимает различные, заранее не известные значения. Ясно, что чем меньше различаются эмпирические и теоретические частоты, тем меньше величина

критерия (1) и, следовательно, он характеризует близость эмпирического и теоретического распределений.

Доказано, что при $n \rightarrow \infty$ закон распределения случайной величины (1) независимо от того, какому закону распределения подчинена генеральная совокупность, стремится к закону распределения χ^2 с k степенями свободы. Поэтому случайная величина (1) обозначена через χ^2 , а сам критерий называют критерием согласия «хи–квадрат».

Число степеней свободы находят по равенству $k = s - 1 - r$, где s – число групп (частичных интервалов) выборки; r – число параметров предполагаемого распределения, которые оценены по данным выборки. В частности, если предполагаемое распределение – нормальное, то $r = 2$ и $k = s - 3$.

Если, например, предполагают, что генеральная совокупность распределена по закону Пуассона, то оценивают один параметр и поэтому $k = s - 2$.

Обозначим значение критерия, вычисленное по данным наблюдений, через $\chi^2_{\text{набл}}$ и сформулируем правило проверки нулевой гипотезы.

Правило. Для того, чтобы при заданном уровне значимости проверить нулевую гипотезу H_0 : генеральная совокупность распределена нормально, надо сначала вычислить теоретические частоты, а затем наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \sum (n_i - n'_i)^2 / n'_i \quad (2)$$

и по таблице критических точек распределения χ^2 , по заданному уровню значимости α и числу степеней свободы $k = s - 3$ найти критическую точку $\chi^2_{\text{кр}}(\alpha; k)$ (см. Приложение 1, таблица П.2). Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}(\alpha; k)$ – нет оснований отвергнуть нулевую гипотезу. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(\alpha; k)$ – нулевую гипотезу отвергают.

Замечание: Для применения критерия согласия «хи–квадрат» объем выборки должен быть достаточно велик (не менее 50). Каждая группа должна содержать не менее 5–8 вариантов.

2. Критерий согласия Колмогорова–Смирнова

При использовании критерия согласия Колмогорова вычисляют

$$D = \frac{\max |F_{\text{нак}} - F_{\text{теор}}|}{n} \quad \text{или} \quad D = \frac{\max |F_{\text{нак}} - F_{\text{теор}}|}{\sqrt{n}} \quad (3)$$

где $F_{\text{нак}}$ – накопленная наблюдаемая частота;

$F_{\text{теор}}$ – накопленная ожидаемая (теоретическая) частота.

Полученное значение сравнивают с критическим $D_{кр}(n;\alpha)$, которое берут из таблиц. Если $D \ll D_{кр}(n;\alpha)$ то нулевая гипотеза принимается, если же $D \geq D_{кр}(n;\alpha)$ то нулевая гипотеза отвергается.

Замечание: в том случае, когда значение статистики лишь немного меньше номинального (табличного) нужно быть осторожным в принятии определенного решения.

Критерий Колмогорова своей простотой выгодно отличается от критерия «хи-квадрат», поэтому его охотно применяют на практике. Следует оговорить, что этот критерий можно применять только в том случае, когда гипотетическое распределение $F(x)$ полностью известно заранее из каких-либо теоретических соображений, т.е. известен не только вид функции распределения $F(x)$, но и все входящие в нее параметры. Такой случай сравнительно редко встречается на практике. Обычно из теоретических соображений известен только общий вид функции, а входящие в нее числовые параметры определяются по выборке. При применении критерия «хи-квадрат» это обстоятельство учитывается соответствующим уменьшением числа степеней свободы. Критерий Колмогорова такого согласования не предусматривает. Если все же применять этот критерий в тех случаях, когда параметры теоретического распределения оцениваются по выборке, критерий дает заведомо завышенные значения вероятности, поэтому есть риск принять как правильную гипотезу, в действительности плохо согласующуюся с опытными данными (выборкой).

3. Однофакторный дисперсионный анализ.

Пусть генеральные совокупности $X_1, X_2, \dots, X_p$ распределены нормально и имеют одинаковую, хотя и неизвестную дисперсию; математические ожидания также неизвестны, но могут быть различными. Требуется при выбранном уровне значимости по выборочным средним проверить нулевую гипотезу $H_0: M(X_1) = M(X_2) = \dots = M(X_p)$ о равенстве всех математических ожиданий. Другими словами, проверить значимо или незначимо различаются выборочные средние. Казалось бы, для сравнения нескольких средних ($p > 2$) можно сравнить их попарно. Однако с возрастанием числа средних возрастает и наибольшее различие между ними: среднее новой выборки может оказаться больше наибольшего или меньше наименьшего из средних, полученных до нового опыта. По этой причине для сравнения нескольких средних пользуются другим методом, который основан на сравнении

дисперсий и поэтому назван дисперсионным анализом (в основном развит англ. статистиком Р. Фишером).

На практике дисперсионный анализ применяют, чтобы установить, оказывает ли существенное влияние некоторый качественный фактор, который имеет p уровней $1, 2, \dots, p$ на изучаемую величину X .

Основная идея дисперсионного анализа состоит в сравнении так называемой «факторной дисперсии», порождаемой воздействием фактора, и «остаточной дисперсии», обусловленной случайными причинами. Если различие между этими дисперсиями значимо, то фактор оказывает существенное влияние на X ; в этом случае средние наблюдаемых значений на каждом уровне (групповые средние) различаются также значимо.

Если уже установлено, что фактор существенно влияет на X , а требуется выяснить, какой из уровней оказывает наибольшее воздействие, то дополнительно производят попарное сравнение средних.

Иногда дисперсионный анализ применяется, чтобы установить *однородность* нескольких совокупностей (дисперсии этих совокупностей одинаковы по предположению; если дисперсионный анализ покажет, что и мат. ожидания одинаковы, то в этом смысле совокупности однородны). Однородные же совокупности можно объединить в одну и тем самым получить о ней более полную информацию, следовательно, и более надежные выводы.

Ограничимся простейшим случаем однофакторного анализа, когда на нормально распределенный признак X воздействует только один фактор, который имеет p постоянных уровней.

Будем предполагать, что число наблюдений (испытаний) на каждом уровне одинаково и равно q .

Пусть наблюдалось $n = pq$ значений x_{ij} признака X , где i – номер испытания ($i = 1, 2, \dots, q$), j – номер уровня фактора ($j = 1, 2, \dots, p$). Результаты наблюдений приведены в нижеследующей таблице.

Номер испытания	Уровни фактора			
	1	2	...	p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
...	...	...	...	...
q	x_{q1}	x_{q2}	...	x_{qp}
Групповая	$\bar{x}_{гр1}$	$\bar{x}_{гр2}$	...	$\bar{x}_{грp}$

средняя				
---------	--	--	--	--

Введем, по определению,

$$S_{\text{общ}} = \sum_{i=1}^p \sum_{j=1}^q (x_{ij} - \bar{x})^2, \quad (4)$$

(общая сумма квадратов отклонений наблюдаемых значений от общей средней $\bar{x}$),

$$S_{\text{факт}} = q \sum_{i=1}^p (\bar{x}_{ip} - \bar{x})^2, \quad (5)$$

(факторная сумма квадратов отклонений групповых средних от общей средней, которая характеризует рассеяние «между группами»)

$$S_{\text{ост}} = \sum_{i=1}^q (x_{i1} - \bar{x}_{ip1})^2 + \dots + \sum_{i=1}^q (x_{ip} - \bar{x}_{ipp})^2, \quad (6)$$

(остаточная сумма квадратов отклонений наблюдаемых значений группы от своей групповой средней, которая характеризует рассеяние «внутри групп»).

Разделив суммы квадратов отклонений на соответствующее число степеней свободы, получим общую, факторную и остаточную дисперсии:

$$s_{\text{общ}}^2 = \frac{S_{\text{общ}}}{pq-1}, \quad s_{\text{факт}}^2 = \frac{S_{\text{факт}}}{p-1}, \quad s_{\text{ост}}^2 = \frac{S_{\text{ост}}}{p(q-1)} \quad (7)$$

Практически остаточную дисперсию находят по равенству:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}. \quad (8)$$

Вернемся к поставленной задаче: проверить нулевую гипотезу о равенстве нескольких средних нормальных совокупностей с неизвестными, но равными дисперсиями. Покажем, что решение этой задачи сводится к сравнению факторной и остаточной дисперсий по критерию Фишера–Снедекора.

1. Пусть нулевая гипотеза о равенстве нескольких средних (групповых) правильна. В этом случае факторная и остаточная дисперсии являются несмещенными оценками неизвестной генеральной дисперсии и, следовательно, различаются незначимо. Если сравнить эти оценки по критерию F (Фишера–Снедекора), то критерий укажет, что нулевую гипотезу о равенстве факторной и остаточной дисперсий следует принять. Таким образом, если гипотеза о равенстве групповых средних правильна, то верна и гипотеза о равенстве факторной и остаточной дисперсий.

2. Пусть нулевая гипотеза о равенстве групповых средних ложна. В этом случае с возрастанием расхождения между групповыми средними увеличивается факторная дисперсия, а вместе с ней и отношение $F_{\text{набл}} = s^2_{\text{факт}}/s^2_{\text{ост}}$. В итоге $F_{\text{набл}}$ окажется больше $F_{\text{кр}}$ и, следовательно, гипотеза о равенстве дисперсий будет отвергнута.

Таким образом, если гипотеза о равенстве групповых средних ложна, то ложна и гипотеза о равенстве факторной и остаточной дисперсий.

Легко доказывается от противного и справедливость обратных утверждений: из правильности (ложности) гипотезы о дисперсиях следует правильность (ложность) гипотезы о средних.

Итак, для того чтобы проверить нулевую гипотезу о равенстве групповых средних нормальных совокупностей с одинаковыми дисперсиями, достаточно проверить по критерию F нулевую гипотезу о равенстве факторной и остаточной дисперсий. В этом и состоит метод дисперсионного анализа.

Следует отметить, что имеется обобщение вышеприведенных соотношений на случай разного числа испытаний на различных уровнях.

Итак, если верна нулевая гипотеза, то и факторная (внутригрупповая), и остаточная (межгрупповая) дисперсии служат оценками дисперсии совокупности и должны быть приближенно равны.

Если $F_{\text{набл}}$ значительно превышает $F_{\text{кр } \alpha(m-1, m(n-1))}$, то нулевую гипотезу отвергают (здесь m – число групп, n – численность каждой группы).

Методом однофакторного дисперсионного анализа можно пользоваться, если верны следующие предположения:

- Каждая выборка независима от остальных выборок;
- Каждая выборка случайным образом извлечена из исследуемой совокупности;
- Совокупность нормальна распределена;
- Дисперсии выборок равны.

Лекция 6. КРИТЕРИЙ СТЬЮДЕНТА

1. Сравнение выборочных средних (сравнение двух групп) (критерий Стьюдента).
 2. Анализ парных наблюдений (парный критерий Стьюдента).
 3. Критерий Стьюдента для множественных сравнений (поправка Бонферрони). Множественные сравнения с контрольной группой.
1. Сравнение выборочных средних (сравнение двух групп) (критерий Стьюдента)

Сравним средние арифметические независимых выборок, взятых из нормально распределенных совокупностей с параметрами $\mu_1, \sigma_1^2; \mu_2, \sigma_2^2$. Исходим из предположения, что разница $\bar{x}_1 - \bar{x}_2$ между ними возникла случайно.

Переменная величина

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}} \quad (1)$$

следует распределению Стьюдента с числом степеней свободы $n_1 + n_2 - 2$. В этой формуле $s_{\bar{x}_1 - \bar{x}_2}$ – ошибка разности между выборочными средними.

$s_{\bar{x}_1 - \bar{x}_2}$ определяется по следующей формуле (при равновеликой выборке $n_1 = n_2 = n$):

$$s_d = s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2}{n_1(n_1 - 1)} + \frac{\sum (x_i - \bar{x}_2)^2}{n_2(n_2 - 1)}} = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{n(n - 1)}}, \quad (2)$$

а при неравновеликих малочисленных выборках, когда $n_1 \neq n_2$:

$$s_d = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{n_1 + n_2 - 2} \left(\frac{n_1 + n_2}{n_1 n_2} \right)}. \quad (3)$$

Так как согласно нулевой гипотезе $\mu_1 - \mu_2 = 0$, то критерий достоверности различий, наблюдаемых между выборочными средними, выражается в виде:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_{\bar{x}_1 - \bar{x}_2}}. \quad (4)$$

Нулевая гипотеза отвергается, если $t_{\phi} \geq t_{st}$ для принятого уровня значимости и числа степеней свободы $k = n_1 + n_2 - 2$.

Критерием Стьюдента можно пользоваться, если верны те же предположения, что и при пользованием методом однофакторного дисперсионного анализа:

- Каждая выборка независима от остальных выборок;
- Каждая выборка случайным образом извлечена из исследуемой совокупности;
- Совокупность нормальна распределена;
- Дисперсии выборок равны.

2. Анализ парных наблюдений (парный критерий Стьюдента).

Выявить изменение, располагая парами наблюдений, позволяет парный критерий Стьюдента.

Выбор одной группы объектов наблюдения и измерение у каждого объекта в группе значение признака до и после воздействия какого-либо фактора позволяет более точно улавливать различия, вызванные воздействием фактора, нежели сравнение двух независимых групп, “зашумленное” разбросом значений у разных объектов.

Такой подход повышает чувствительность критерия. Это иллюстрирует следующий пример. На рисунке 4, в левой его части, представлены результаты наблюдения за двумя независимыми группами: первый столбец образуют данные о признаке объектов, на которые фактор не воздействовал, а второй – данные о признаке объектов, на которые фактор воздействовал.

Правые два столбца относятся к одним и тем же объектам, левый столбец содержит данные о признаке до воздействия фактора, правый – после воздействия фактора. Отрезками соединены пары точек, относящиеся к одному объекту.

Из анализа данных, представленных на левой половине рисунка, не следует, что фактор оказывает существенное влияние на значение признака.

По наклону отрезков, соединяющих значения признака до и после воздействия фактора на один и тот же объект, можно сделать вывод, что фактор увеличил значение признака.

	*		*
	*		*
	*		*
*		*	*
*	*	*	*
*		*	*
	*		*
*		*	*
*		*	*
	*		*
*	*	*	*
*		*	*
*		*	*
	*		*
*		*	*
	*		*
*		*	*
Группа 1	Группа 2	Измерение «До»	Измерение «После»

Рисунок 1 – Измерение показателей в двух независимых группах и в последовательных измерениях

Оценить статистическую значимость изменения позволяет *парный критерий Стьюдента*. Нулевая гипотеза состоит в том, что среднее изменение равно нулю.

Если δ – истинное среднее изменение признака, а d – наблюдаемое (выборочное) среднее изменение признака, то выборочное стандартное отклонение изменения признака:

$$S_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}} ;$$

а стандартная ошибка изменения признака: $S_{\bar{d}} = \frac{S_d}{\sqrt{n}}$.

Таким образом, парный критерий Стьюдента принимает вид:

$$t = \frac{\bar{d} - \delta}{S_{\bar{d}}}. \quad \text{При } H_0 : \delta = 0 \quad t = \frac{\bar{d}}{S_{\bar{d}}} \quad (5)$$

и это значение следует сравнить с $t_{\text{крит}}$ для выбранного уровня значимости и числа степеней свободы $\nu = n - 1$.

Если обычный критерий Стьюдента требует нормального распределения самих данных, то парный критерий Стьюдента требует нормального распределения их *изменений*.

3. Критерий Стьюдента для множественных сравнений (поправка Бонферрони) Множественные сравнения с контрольной группой.

Критерий Стьюдента предназначен для сравнения двух групп.

Однако на практике он широко и неправильно используется для оценки различий большего числа групп посредством попарного их сравнения. При этом вступает в силу *эффект множественных сравнений*.

В общем случае вероятность ошибиться *хотя бы в одном* из k сравнений составит не $1-P$, а значительно больше, а именно:

$$P' = 1 - P^k$$

где k - число сравнений.

При небольшом числе сравнений можно использовать приближенную формулу

$$P' = (1 - P)k$$

то есть вероятность ошибиться хотя бы в одном из сравнений примерно равна вероятности ошибиться в одном, помноженной на число сравнений.

Если критерий Стьюдента был использован для проверки различий между несколькими группами, то истинный уровень значимости можно получить, умножив уровень значимости на число возможных сравнений.

Простейший из методов множественного сравнения - введение *поправки Бонферрони*. При трехкратном применении критерия Стьюдента, с 5% уровнем значимости, вероятность обнаружить различия там, где их нет, составляет не 5%, а почти $3 \times 5 = 15\%$. Этот результат является частным случаем *неравенства Бонферрони*, если k раз применить критерий с уровнем значимости $1-P$, то вероятность хотя бы в одном случае найти различие там, где его нет не превышает произведения k на $1-P$. Неравенство Бонферрони выглядит так:

$$\alpha' < k\alpha$$

где α' - вероятность хотя бы один раз ошибочно выявить различия. α' собственно и является истинным уровнем значимости многократно примененного критерия. Из неравенства Бонферрони следует, что если мы хотим обеспечить вероятность ошибки α' , то в каждом из сравнений мы должны принять уровень значимости α'/k - это и есть поправка Бонферрони. Например, при трехкратном сравнении уровень значимости должен быть $0,05/3 = 1,7\%$.

Поправка Бонферрони хорошо работает, если число сравнений невелико. Если оно превышает 8, метод становится слишком строгим и даже весьма большие различия приходится признавать статистически незначимыми. Существуют не столь жесткие методы множественного сравнения, например критерий Ньюмена-Кейлса.

Иногда задача заключается в том, чтобы сравнить несколько групп с единственной - контрольной. Применить поправку Бонферрони к сравнению нескольких групп с одной контрольной очень просто. Ход вычислений такой же что и при применении поправки Бонферрони в общем случае. Надо только учесть, что число сравнений k составляет теперь $m - 1$ (здесь m - число групп, включая контрольную) и соответственно рассчитать уровень требуемый значимости в каждом из сравнений $\alpha = \alpha/k$.

Лекция 7. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

1. Критерий Крускала-Уоллиса.
2. Х-критерий Ван-дер-Вардена.
3. Критерий знаков Z.
4. U-критерий Манна-Уитни для двух независимых выборок (U-критерий Уилкоксона).
5. T-критерий Манна-Уитни.
6. Критерий Уилкоксона для парных выборочных наблюдений.
7. Ранговый коэффициент корреляции Спирмена.

Правильное применение параметрических критериев для проверки статистических гипотез основано на предположении о нормальном распределении совокупностей, из которых взяты сравниваемые выборки. Однако не всегда имеет место, т.к. не все биологические признаки распределяются по нормальному закону. Немаловажным является и то обстоятельство, что исследователю приходится иметь дело не только с количественными, но и с качественными признаками, многие из которых выражаются порядковыми номерами, индексами и другими условными знаками. В таких случаях необходимо использовать *непараметрические критерии*. Видное место среди них занимают ранговые критерии, применение которых основано на ранжировании членов сравниваемых групп. При этом сравниваются не сами по себе члены ранжированных рядов, а их порядковые номера, или ранги. Ниже рассмотрены некоторые непараметрические критерии, применяемые для проверки нулевой гипотезы при сравнении как независимых, так и зависимых выборочных групп.

1. Критерий Крускала-Уоллиса

Если рассматривается задача сравнения нескольких выборок и нет уверенности, что данные, полученные для каждой группы, подчиняются нормальному распределению и дисперсии по всем группам примерно одинаковы, то можно воспользоваться критерием Крускала-Уоллиса.

Сначала все значения, независимо от того, какой выборке они принадлежат, упорядочивают по возрастанию. Каждому значению присваивается ранг – номер его места в упорядоченном ряду (Совпадающим значениям присваивается общий ранг, равный среднему их мест). Затем вычисляют суммы рангов, относящихся к каждой группе, и для каждой группы определяют средний ранг. Величина

$$H = \frac{12}{N(N+1)} \sum \frac{(\sum R_i)^2}{n_i} - 3(N+1) , \quad (1)$$

где N – общее число наблюдений, объем комплекса;

n_i – численность вариант в каждой группе;

R_i – ранги вариант, ранжированных в общем порядке, является критерием Крускала-Уоллиса. Суммирование в приведенной формуле производится по всем группам.

Условием для отвергания нулевой гипотезы на принятом уровне значимости α будет $H \geq H_{st}$, где H_{st} находят по таблице (см. [2]). При числе групп больше 3 или $n \geq 5$ $H \rightarrow \chi^2$, поэтому величину H можно сравнить с табличным значением χ^2_{st} . H_0 -гипотезу отвергают, если $H \geq \chi^2_{st}$ для принятого уровня значимости α и числа степеней свободы $k = m - 1$, где m – число групп.

2. X-критерий Ван-дер-Вардена

Этот критерий применяют для проверки нулевой гипотезы при сравнении друг с другом независимых выборок. Техника расчетов X-критерия сводится к следующему.

- Сравнимые выборки ранжируют в один общий ряд по возрастающим значениям признака. Каждому члену ряда присваивают порядковый номер.
- По порядковым номерам одной из выборок, обычно меньшей по объему, находят отношение $R/(N+1)$, где $N + 1 = n_1 + n_2 + 1$, т.е. сумма всех объемов выборок, увеличенная на единицу, а R – порядковый номер вариантов выборки в общем ранжированном ряду, их ранг.
- Находят (по специальной таблице, например, [3]) значения функции $\psi[R/(N+1)]$ для каждого значения $R/(N+1)$;
- Суммируют результаты для выборки (обязательно с учетом знаков), получают величину $X_\phi = \sum \psi[R/(N+1)]$, которую сравнивают с критической точкой этого критерия X_{st} (см.[3]) для принятого уровня значимости α и общего числа членов сравниваемых выборок, т.е. $N = n_1 + n_2$ (с учетом разности $n_1 - n_2$);
- Если $X_\phi \geq X_{st}$ то предположение, что сравниваемые выборки извлечены из генеральных совокупностей с одинаковыми функциями распределения отвергают на принятом уровне значимости.

3. Критерий знаков z

В тех случаях, когда результаты наблюдений выражаются не числами, а знаками плюс (+) и минус (–), различия между попарно связанными членами выборок оценивают с помощью критерия знаков z .

Применение и расчет этого критерия демонстрирует следующий пример.

Пример. Изучали влияние туберкулина на состав периферической крови низших обезьян. Результаты приведены в таблице.

Номер подопытных животных	Эозинофилия в крови		Эффект воздействия
	до введения туберкулина	после введения туберкулина	
1	++	++	0
2	+++	++	+
3	++	+	+
4	+++	++	+
5	+++	++	+
6	++	+++	—
7	++	+	+
8	+	++	—
9	+++	++	+
10	++	+	+
11	++	++	0
12	+++	+	+
13	++	+	+
14	++	+	+

Их 14 наблюдений в двух результат воздействия оказался нулевым и эти наблюдения не учитываются в дальнейшем, так что $n = 14 - 2 = 12$. Из этого числа положительных разностей насчитывается 10. Следовательно, $z_{\Phi} = 10$. Из специальной таблицы (например, [3], Приложение XII) для $n = 12$ и $\alpha = 5\%$ находим $z_{st} = 10$. Равенство $z_{\Phi} = z_{st}$ дает основание отвергнуть гипотезу H_0 на 5 % уровне значимости. Следовательно, с вероятностью 95 % можно утверждать, что введение туберкулина (реакция Манту) вызывает заметное снижение эозинофилов в периферической крови.

4. U - Критерий Манна-Уитни для двух независимых выборок (U -критерий Уилкоксона)

Порядок расчета этого критерия, используемого при сравнении двух выборок, следующий.

- Расположить значения сравниваемых выборок в возрастающем порядке в один общий ряд и пронумеровать члены ряда от 1 до $N = n_1 + n_2$ (Эти номера и будут рангами членов ряда);

Выборка 1 (объема n_1)		Выборка 2 (объема n_2)	
наблюдение	ранг	наблюдение	ранг
x_{11}	r_{11}	x_{12}	r_{12}
x_{21}	r_{21}	x_{22}	r_{22}
.	.	.	.
.	.	.	.
$x_{n1\ 1}$	$r_{n1\ 1}$	$x_{n2\ 2}$	$r_{n2\ 2}$
Сумма	R_1	Сумма	R_2

Величины $\{x_{ij}\}$ – реальные наблюдения. Определим $n = n_1 + n_2$. Числа $\{r_{ij}\}$ – ранги этих наблюдений. Они меняются от 1 до n .

- Вычисляют значения критериальной статистики:

1) В случае *малых выборок*:

$$U_1 = n_1 n_2 + n_1(n_1+1)/2 - R_1; U_2 = n_1 n_2 + n_2(n_2 + 1)/2 - R_2; U = \max(U_1, U_2) \quad (2)$$

и сравнивают U с критическим значением для малых выборок.

2) В случае *больших выборок*:

$$T = (U - 0,5 n_1 n_2) / \{ n_1 n_2 (n+1)/12 \}^{0,5} \quad (3)$$

и сравнивают T с критическим значением для стандартного нормального распределения.

Замечания: 1. Не имеет значения, как ранжируются наблюдения: от наименьшего до наибольшего или наоборот.

2. Если два или более наблюдения имеют в точности одинаковые значения, они называются совпадающими и в этом случае каждому из них следует приписать значения ранга, равное среднему из рангов, которые были бы им приписаны при отсутствии совпадений.

3. $U_1 + U_2 = n \cdot n_2$

4. Не следует применять критерий к парным наблюдениям.

5. Если обе выборки извлечены из совокупностей, имеющих нормальные распределения с равными дисперсиями, то следует использовать более мощный критерий Стьюдента.

Вариант U - критерия Манна-Уитни (U – критерия Уилкоксона)

Порядок расчета этого критерия следующий.

– Отдельно для каждой выборки найти суммы рангов R и определить величины

$$U_1 = R_1 - n_1(n_1 + 1)/2 \quad \text{и} \quad U_2 = R_2 - n_2(n_2 + 1)/2 ; \quad (4)$$

– в качестве U -критерия использовать величину U_{Φ} , равную меньшей из U_1 и U_2 , которую сравнить с табличным значением U_{st} (см. [2]). Условием для сохранения принятой H_0 -гипотезы служит неравенство $U_{\Phi} > U_{st}$. Критические точки U -критерия U_{st} для n_1 , n_2 и принимаемого уровня значимости α содержатся в соответствующей таблице.

5. Т-критерий Манна-Уитни

Порядок вычисления критерия:

- Данные обеих групп объединяют и упорядочивают по возрастанию.
- Ранг 1 присваивают наименьшему из всех значений, ранг 2 - следующему и т.д. Если значения совпадают, им присваивают один и тот же средний ранг.
- Для меньшей группы вычисляют T – сумму рангов ее членов. Если численность групп одинакова, T можно вычислить для любой из них.
- Полученное значение T сравнивают с критическими значениями (см., например, [2]). Если T меньше или равно первому из них либо больше или равно второму, то H_0 -гипотеза отвергается (различия статистически значимы).

Следует отметить, что при $n_2 \geq n_1 > 8$ $z_T = \frac{|T - \mu_T| - 0,5}{\sigma_T}$ имеет стандартное нормальное распределение ($\mu_T = 0,5n_1(n_1 + n_2 + 1)$; $\sigma_T = \sqrt{n_1n_2(n_1 + n_2 + 1)/12}$).

6. Критерий Уилкоксона для парных выборочных наблюдений

Порядок вычисления значения критериальной статистики проводят по одному из следующих двух вариантов.

Вариант 1.

а) В случае *малых выборок*.

- Ранжируют абсолютные величины разностей в возрастающем порядке и приписывают им соответствующие ранги от 1 до n .
- Каждому значению ранга приписывают знак его разности.
- Вычисляют сумму значений положительных рангов и обозначают ее через N .
- Проверяют, принадлежит ли N критической области (по таблице значений критерия Уилкоксона).

б) В случае *больших выборок*.

Вычисляют N как в п. а) и полагают

$$T = \{N - n \cdot (n+1) / 4\} / \{n \cdot (n+1) \cdot (2n+1) / 24\}^{1/2} \quad (5)$$

и проверяют по таблице стандартного нормального распределения.

Вариант 2.

- Вычисляют величины изменений наблюдаемого признака. Отбрасывают пары наблюдений, которым соответствует нулевое изменение.
- Упорядочивают изменения по возрастанию их абсолютной величины и присваивают соответствующие ранги. Рангами одинаковых величин назначают средние тех мест, которые они делят в упорядоченном ряду.
- Присваивают каждому рангу знак в соответствии с изменением: если значение увеличилось – « + », если уменьшилось – « - ».
- Вычисляют величину W , равную сумме знаковых рангов.
- Сравнивают W с критическим значением. Если она больше критического значения, изменение показателя статистически значимо.

Замечание:

Если число пар наблюдений больше 20, то распределение величины

$$z_W = \frac{|W| - 0,5}{\sqrt{n(n+1)(2n+1)/12}} \quad \text{близко к стандартному нормальному.}$$

(здесь n – число пар наблюдений, т. е. численность группы).

7. Ранговый коэффициент корреляции Спирмена

Расчет коэффициента корреляции (Пирсона) возможен при линейности связи переменных и распределения их по нормальному закону. Эти условия выполняются не всегда. Кроме того, в исследованиях часто имеют дело с порядковыми признаками, к которым расчет коэффициента корреляции Пирсона не применим.

В этих случаях следует воспользоваться коэффициентом ранговой корреляции Спирмена. Это непараметрический метод – он не требует нормальности распределения, линейной зависимости, его можно применять как к количественным, так и к порядковым признакам.

Рассчитывают коэффициент ранговой корреляции Спирмена r_s следующим образом. Упорядочивают данные по возрастанию и заменяют реальные значения их рангами. Рангом значения называется его номер в упорядоченном ряду. Беря вместо самих значений их ранги, рассчитывают коэффициент корреляции Пирсона. Это и будет коэффициент ранговой корреляции Спирмена. Его можно рассчитать и по следующей формуле:

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}, \quad (6)$$

где d – разность рангов для каждого члена выборки.

Если в ряду встречаются одинаковые значения, то им следует присвоить один и тот же ранг, равный среднему занимаемых ими мест.

Значимость коэффициента ранговой корреляции Спирмена при объеме выборки меньше 50 и для выбранного уровня значимости определяют, сравнивая его с критическими значениями (см., таблица П.4 в Приложении 1). Если объем выборки больше 50, нужно применить критерий Стьюдента

$$t = \frac{r_s}{\sqrt{\frac{1 - r_s^2}{n - 2}}} \text{ с числом степеней свободы } \nu = n - 2.$$

Лекция 8. УРАВНЕНИЯ РЕГРЕССИИ

1. Общие принципы выбора уравнений регрессии.
 2. Прямая регрессия. Метод наименьших квадратов. Оценка параметров уравнения регрессии по выборке.
 3. Оценка линейности регрессии.
 4. Нелинейная парная связь между признаками.
1. Общие принципы выбора уравнений регрессии.

Важной задачей в области регрессионного анализа является выбор уравнения, которое бы наилучшим образом описывало исследуемую закономерность. Обычно эту задачу решают следующим образом. Эмпирический ряд регрессии или динамики, для которого подыскивают наилучшее уравнение, изображают в виде точечного графика в системе прямоугольных координат. Если точки располагаются на одной прямой или могут быть аппроксимированы прямой линией, зависимость между переменными величинами описывают уравнением линейной регрессии. Труднее выбрать наилучшее уравнение регрессии при наличии нелинейной связи между переменными величинами. В таких случаях подходящее уравнение подбирают на основании сравнения эмпирического графика с известными образцами кривых. Немаловажное значение при выборе уравнения регрессии имеют опыт и профессиональные знания исследователя.

Графический анализ не гарантирует от возможных ошибок, особенно в тех случаях, когда главное направление регрессии или динамики сильно затушевывается колебаниями членов ряда. Поэтому наряду с графическим анализом применяют и аналитические способы проверки правильности выбора уравнений. Одним из них является применение принципов дисперсионного анализа.

Необходимо отметить: уравнения регрессии позволяют прогнозировать возможные значения зависимой переменной на основании известных величин аргумента. При этом, однако, не следует экстраполировать регрессию за пределы проведенных опытов, т.к. она может потерять свое направление. Область применения уравнений регрессии лучше ограничить теми областями данных, на которых получены эмпирические уравнения. Это предостережет исследователя от возможных ошибок. Не следует также при отыскании уравнений регрессии допускать значительные сокращения приближенных чисел. Следует округлять лишь «конечные» числа – параметры регрессионных уравнений.

2. Прямая регрессии. Метод наименьших квадратов. Оценка параметров уравнения регрессии по выборке.

Рассмотрим процедуру *линейного парного регрессионного анализа* (метода наименьших квадратов на плоскости).

Парными называются зависимости $y = f(x)$.

Пусть имеется n пар наблюдений (y_i, x_i) . Такие результаты наблюдений могут быть получены в любой экспериментальной работе. Нас будет интересовать вопрос, что произойдет с y (функция отклика) если изменить независимую переменную (фактор x)?

Задача линейного регрессионного анализа (метода наименьших квадратов) состоит в том, чтобы, зная положения точек (y_i, x_i) на плоскости, так провести линию регрессии, чтобы сумма квадратов отклонений вдоль оси Оу (ординаты) этих точек от проведенной прямой была минимальной.

Для проведения вычислений по классическому методу наименьших квадратов (МНК) к форме уравнения регрессии предъявляется требование: это уравнение должно быть линейно по параметрам или допускать возможность линеаризации.

Для простоты предположим, что при проведении парного линейного регрессионного анализа имеем дело только с уравнением прямой линии.

Уравнение прямой на плоскости в декартовых координатах:

$$y = a + bx, \quad (1)$$

где a и b – постоянные числа.

Задачу МНК можно выразить следующим образом:

$$U = \sum_{i=1}^n [y_i - (a + bx_i)]^2 = \min \quad (2)$$

Для решения задачи, поставленной в формуле (2) необходимо в каждом конкретном случае вычислить значения коэффициентов a и b , минимизирующие сумму отклонений. Для этого необходимо вычислить частные производные функции U по коэффициентам a и b и приравнять их к нулю:

$$\frac{\partial U}{\partial a} = -2 \sum_{i=1}^n [y_i - (a + bx_i)] = 0 \quad (3)$$

$$\frac{\partial U}{\partial b} = -2 \sum_{i=1}^n [y_i - (a + bx_i)] x_i = 0 \quad (4)$$

Преобразуем полученную систему так называемых, нормальных уравнений:

$$an + b \sum x_i = \sum y_i \quad (5)$$

$$a \sum x_i + b \sum x_i^2 = \sum (y_i x_i) \quad (6)$$

Решая систему уравнений (10.5– 10.6) (например, по методу Крамера), получим:

$$a = (\Sigma y \Sigma x^2 - \Sigma xy \Sigma x) / (n \Sigma x^2 - (\Sigma x)^2) \quad (7)$$

$$b = (n \Sigma xy - \Sigma x \Sigma y) / (n \Sigma x^2 - (\Sigma x)^2) . \quad (8)$$

Коэффициент a геометрически представляет собой расстояние от начала координат до точки пересечения линии регрессии с ординатой.

Коэффициент b представляет собой тангенс угла наклона линии регрессии к оси абсцисс.

Для y в литературе можно встретить следующие наименования: функция отклика, зависимая переменная, предиктор; x называют входной переменной, независимой переменной, фактором, регрессором.

Если переменные y и x представляют собой двумерную нормально распределенную случайную величину, то существуют две регрессии. Одна определяет зависимость y от x , а другая x от y . Прямые регрессии пересекаются в точках $(\bar{x}, \bar{y})$ и образуют «ножницы». Чем «уже» ножницы, тем ближе стохастическая связь к функциональной. При функциональной связи обе прямые сливаются.

3. Оценка линейности регрессии

Оценка линейности регрессии может быть выполнена в том случае, если общее число значений y больше, чем число k значений x . Каждому значению x_i соответствует n_i значений y_{ij} . В этом случае подсчитываются: отклонения групповых средних $\bar{y}_i$ от прямой регрессии; отклонения значений y_{ij} от групповых средних.

Далее рассчитывают статистику (Фишера–Снедекора):

$$F = \frac{\frac{1}{k-2} \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}, \quad (9)$$

т.е. отношение суммы квадратов отклонений групповых средних от прямой регрессии к сумме квадратов отклонений значений y от групповых средних со степенями свободы в числителе $k - 2$ и в знаменателе $n - k$. Если статистика F достигает или превосходит границу значимости, то гипотезу о линейности нужно отбросить.

В том случае, когда гипотеза линейности может быть отброшена или когда при графическом изображении точек нелинейность явно просматривается, по экспериментальным данным получают нелинейную (квадратичную или высших порядков) формулу парной зависимости. (Например, парная квадратическая регрессия: $y = a + bx + cx^2$).

При этом можно рассчитывать, что нелинейная формула даст меньшую остаточную дисперсию, т.е. лучше предскажет результаты опытов. Следует помнить, что зависимость нелинейна по независимой переменной x . По параметрам зависимость *остаётся линейной*.

4. Нелинейная парная связь между признаками

Используя метод наименьших квадратов, можно построить практически любые формы нелинейной парной связи. Для этого используют линеаризующие преобразования, т.к. только линейные по параметрам функции восстанавливаются методом наименьших квадратов. В нижеследующей таблице приведены часто встречающиеся парные зависимости, допускающие линеаризацию.

Таблица – Аппроксимирующие функции, допускающие линеаризацию

№ п/п	Функция
1	$y = a + b/x$
2	$y = 1/(a + bx)$
3	$y = x/(a + bx)$
4	$y = a \cdot b^x$
5	$y = a \cdot e^{bx}$
6	$y = 1/(a + b \cdot e^{-x})$
7	$y = a \cdot x^b$
8	$y = a + b \cdot \ln(x)$
9	$y = a/(b + x)$
10	$y = ax/(b + x)$
11	$y = a \cdot e^{b/x}$
12	$y = a + b^x$

Для нахождения коэффициентов в уравнении нелинейной по независимой переменной регрессии проводят ту же последовательность действий, что и для линейной: записывают функцию, подобную (2); находят частные производные по коэффициентам a , b , c ,... и, приравнявая их к нулю, получают систему линейных алгебраических уравнений для коэффициентов; решают систему и определяют соотношения для расчета коэффициентов.

Качество предсказания проверяют с помощью линейного уравнения $\hat{y}' = a' + b' \cdot x'$. После вычисления коэффициентов a и b выполняют обратные преобразования, т.е. по a' и b' определяют a и b).

Лекция 9. КРИВАЯ «ДОЗА-ЭФФЕКТ»

1. Принципы (методы) количественной оценки иммуногенности вакцин.
2. Количественный метод определения иммуногенности вакцин, основанный на испытании их постоянным уровнем иммунитета.
3. Количественные методы определения иммуногенности вакцин, основанные на испытании постоянной дозой антигенов.

1. Принципы (методы) количественной оценки иммуногенности вакцин.

Оценка иммуногенности профилактических вакцин производится в опытах на экспериментальных животных. При этом иммуногенность прививочных препаратов определяют или по напряженности иммунитета к соответствующему токсическому или инфекционному агенту или по уровню определяемых у животных гуморальных факторов иммунитета (антитоксинов, агглютининов, комплемент-связывающих и других антител).

В производстве вакцин для текущего контроля продукции обычно применяют специально установленные в опыте минимальные стандарты иммуногенности. При разработке прививочных препаратов, а также в случае изучения закономерностей иммунизации применение этих стандартов оказывается часто недостаточным, так как не дает точного количественного представления об уровне иммунитета и не позволяет «дозировать» иммуногенность препаратов, т.е. с известной точностью количественно выражать защитные свойства препаратов, выявляемые в опытах на лабораторных животных.

Для этих целей существуют специально разработанные методы оценки иммуногенности вакцин, основанные или на определении минимальной дозы вакцины, защищающей какую-то часть (обычно половину) животных от определенной дозы инфекционного агента или специфического токсина, или на установлении среднего уровня иммунологического ответа для группы животных, вакцинированных какой-либо определенной дозой антигена. В первом случае метод может быть вполне закономерно назван методом «постоянного уровня иммунитета», а во втором случае - методом «постоянной дозы антигена». Для выявления относительной иммуногенности отдельных серий вакцин методом постоянного уровня иммунитета сравнивают дозы вакцин, обеспечивающие равноценный иммунологический эффект у вакцинированных животных; при применении метода постоянной дозы антигена сравнительная иммуногенность различных

серий вакцины выражается отношением достигнутых при вакцинации уровней иммунитета.

Возможно ли с известной точностью определить уровень иммунологического ответа у животных и на каких принципах основано измерение иммуногенности профилактических вакцин?

Теория измерения биологически активных веществ, в том числе и вакцин, исходит из предположения, что животные определенного вида, на которых испытывается вакцина, на одинаковое раздражение отвечают качественно однотипной реакцией. Однако выраженность (сила, степень) этой реакции зависит не только от условий проведения испытания, но и от различной индивидуальной чувствительности подопытных животных. Вследствие этого иммунологический ответ у отдельных опытных животных не может служить показателем активности прививочного препарата. Очевидно, что иммуногенность вакцины может быть выявлена лишь при испытании ее на достаточно большой группе животных. Применение достаточно большой группы животных для определения иммуногенности вакцины позволяет вычислять средний уровень иммунологического ответа, выраженный или количеством смертельных доз инфекционного или токсического материала, которому противостоит 50% животных, или средним содержанием антител в их крови.

Вычисление среднего уровня иммунитета при прививке данной дозой антигена или средней эффективной дозы антигена, защищающей 50% животных, основано на статистических соображениях, аналогичных тем, которые используются при описании методов определения биологической активности микроорганизмов и токсинов. Зависимость процента заболеваемости или летальности животных от величины дозы какого-либо биоагента выражается S-образными, обычно асимметричными, кривыми. Аналогичные кривые можно получить, если процент защищенных животных рассматривать как функцию дозы антигена. В этом случае кривые называют кривыми активности данного антигена.

Возможность значительной ошибки измерения активности вакцины по такой кривой обусловлена тем, что в опыте кривую приходится проводить лишь по нескольким полученным в эксперименте точкам, что не совсем точно отражает ход зависимости «процент выживших животных - доза вакцины».

Асимметричные кривые активности антигена, как и в случаях кривых «доза-эффект», оказалось возможным превратить в кривые симметричной формы путем выражения наносимых на оси абсцисс величин прививочных доз вакцины в виде их логарифмов. Полученные при этом S-образные

симметричные кривые являются ни чем иным как кумулятивными кривыми нормального распределения; эти кривые тем больше приближаются по своей форме к теоретическим кривым, чем больше животных участвовало в опыте. В основе симметричности кривой активности антигена лежит нормальное распределение животных по их чувствительности к антигену, если на оси абсцисс откладывать логарифмы доз последнего.

Нормальные S-образные кривые активности антигена могут быть превращены затем в прямые линии с помощью системы пробитов - путем нанесения против логарифма прививочных доз антигена процента летальности или выживаемости, выраженного в специальных условных единицах, - так называемых пробитах.

Построив на основании экспериментальных данных прямую активности изучаемого антигена, можно затем найти с ее помощью дозу вакцины, обеспечивающую защиту 50% животных, и ее доверительный интервал. (Детальное описание техники применения метода пробитов обсудим далее в лекции 10). Соответствующий развернутый пример расчета для вакцинных препаратов будет приведен ниже.

Прежде чем перейти к изложению конкретных методов количественного измерения иммуногенности вакцин, необходимо кратко остановиться на некоторых общих моментах, играющих существенную роль при оценке иммуногенности профилактических препаратов в опытах на животных:

а) Для получения воспроизводимых результатов и уменьшения ошибки количественного определения иммуногенности каждая серия вакцины должна проверяться на достаточно большой группе животных.

б) Испытание вакцины должно проводиться путем однократной прививки, поскольку при кратных прививках уровень достигнутого иммунитета зависит не столько от качества препарата, сколько от иммунологической реактивности животных. Исключение могут составлять те случаи, когда однократная прививка не обеспечивает создания необходимого для измерения уровня иммунитета.

в) Состояние невосприимчивости у вакцинированных животных лучше оценивать не по гуморальным факторам иммунитета (антитоксины, агглютинины, комплементсвязывающие и др. антитела), а по устойчивости целостного организма к специфическому токсину или инфекционному агенту, поскольку между уровнем антител в крови и иммунитетом к токсину или инфекции не существует прямой пропорциональности. Известное исключение могут составлять антитоксины (противостолбнячные, противодифтерийные, противоботулинические).

г) Определение иммуногенности должно проводиться на максимуме иммунитета, достигаемого при иммунизации данной вакциной, поскольку динамика развития иммунитета и его максимальная выраженность зависят от вида вакцинного препарата. Например, при первичной однократной иммунизации сорбированными анатоксинами максимум иммунитета достигается в более поздние сроки, чем при такой же прививке нативными анатоксинами. Кроме того, определением иммуногенности вакцины на максимуме иммунитета можно избежать выраженных колебаний в невосприимчивости у отдельных животных за счет различного увеличения их в весе в течение иммунизационного периода.

д) Дозы антигенов для иммунизации должны избираться с таким расчетом, чтобы испытание напряженности иммунитета "проводилось введением не менее 10 DLM токсона. В таком случае на распределение животных по степени их специфической невосприимчивости не будет оказывать влияния индивидуальная чувствительность неиммунных животных к токсину. Применение чрезмерно больших доз антигенов может привести к гипериммунизации животных, в результате чего может нивелироваться иммуногенность сравниваемых препаратов.

е) В качестве стандартного антигена должны применяться препараты с постоянной и средней иммуногенностью. Для этой цели применимы лиофильновысушенные вакцины, длительно сохраняющие свою иммуногенность без изменения. Нельзя, например, применять в качестве стандартных сорбированные анатоксины в жидком виде, поскольку они повышают иммуногенность при хранении.

ж) При сравнении иммуногенности вакцинных препаратов по методу Пригге параллелизм прямых активности имеет место лишь для антигенов одного вида. Вследствие этого рекомендуемый далее «метод трех точек» применим лишь при определении иммуногенности препаратов одного вида. Например, нельзя сравнивать между собой очищенный и концентрированный дифтерийный анатоксин и этот же препарат, но сорбированный гидроокисью алюминия, поскольку кривые активности этих двух препаратов не параллельны. По-видимому, это связано с различной иммунологической реактивностью животных к очищенному и к депонированному анатоксинам, которая обусловлена неодинаковым механизмом действия этих антигенов.

Учитывая сказанное выше, стандартный вакцинный препарат должен быть того же вида, что и испытуемый.

з) Для испытания иммуногенности вакцины должны применяться по возможности животные одной линии, поскольку доказано, что такие животные отличаются меньшими индивидуальными колебаниями

чувствительности к антигену. Так, по данным Пригге, в группе морских свинок чистой линии колебания в иммунизиремости, выраженной отношением величин доз антигена, необходимых для создания равноценного иммунитета, составляли 1 : 25, а в смешанной группе морских свинок - 1 : 32000! Меньшая вариабельность животных по индивидуальной чувствительности к антигену обеспечивает меньшую ошибку определения иммуногенности вакцины.

и) На абсолютную величину иммуногенности вакцины оказывает влияние сезонность определения. Так, например, различие в иммунизиремости морских свинок дифтерийным анатоксином по данным Пригге в зимний и летний периоды достигает 2,5 раз.

к) Иммуногенность вакцинного препарата должна выражаться в условных единицах, являющихся количественной мерой иммуногенности. Этот способ выражения активности вакцин имеет большие преимущества перед общепринятыми способами дозировать препараты в антигенных единицах, определяемых до прямого испытания иммуногенности вакцины (единицы связывания, флокуляционные единицы, число микробных клеток, вирусов и т. д.).

В отличие от последнего способа выражения активности препарата применение «единиц иммунизирующей активности» характеризует качество препарата по признаку иммуногенности. Пригге ввел понятие о «защитной единице» (SE - Schutzeinheit). За 1 SE было принято количество антигена (дифтерийного или столбнячного анатоксинов) в миллилитрах или в миллиграммах сухого препарата, защищающее 50% иммунизированных животных от избранной для испытания дозы токсина. Например, в некоторых странах до 1955 г. за 1 SE было принято 1,9 мг/мл сухого стандартного дифтерийного анатоксина, а с октября 1955 г. была введена международная стандартная защитная единица (IE) равная 0,75 мг лиофильновысушенного сорбированного дифтерийного анатоксина серии F-236. В качестве 1 SE столбнячного анатоксина с 1951 г. принято 0,03 мг лиофильновысушенного стандартного препарата. Применение «защитных единиц» в качестве меры иммуногенности препаратов позволяет точно дозировать иммуногенность антигенов. Так, в некоторых странах (Австрия, Дания) препараты дифтерийного и столбнячного анатоксинов отвечают требованиям по иммуногенности, если содержат в 1 мл не менее 30 SE.

л) При определении иммуногенности вакцин необходимо вычисление пределов возможных колебаний результата, которые зависят от количества животных, взятых в опыт, от их однородности в отношении

иммунологической реактивности, от условий проведения испытания, от вида сравниваемых антигенов.

2. Количественный метод определения иммуногенности вакцин, основанный на испытании их постоянным уровнем иммунитета.

Одним из методов количественного определения иммуногенности вакцин является метод постоянного уровня иммунитета. его принципиальная основа состоит в сравнении доз антигенов, обеспечивающих одинаковый уровень иммунитета. наиболее целесообразно и удобно сравнение эффективных доз, защищающих 50 % иммунизированных животных от избранной дозы токсина или микроорганизмов (ED_{50}). Различные методы определения ED_{50} будут представлены в лекции 10.

В табл. 1 представлены данные по сравнению иммуногенности двух препаратов столбнячного анатоксина.

Таблица 1 – Сравнение иммуногенности двух препаратов столбнячного анатоксина в опытах на морских свинках методом «трех точек (данные Пригге)

Препарат	Иммунизирующая доза, мг	Количество защищенных животных по отношению к числу взятых в опыт	ED_{50} и возможные пределы ее колебаний
А (стандартный)	4,5	47/98	4,8 (3,9 – 5,5)
	1,5	11/99	
В (испытуемый)	1,7	15/25	1,4 (0,8 – 2,2)

Морские свинки были иммунизированы дозами 1,5 и 4,5 мг стандартного анатоксина (по 98-99 животных на дозу) и одной дозой – 1,7 мг испытуемого анатоксина В (25 животных на дозу). Через 28 суток все животные были испытаны на напряженность иммунитета введением одинаковой дозы (примерно 20 DLM) столбнячного токсина. На основании полученного процента выживаемости морских свинок вычислены значения ED_{50} и их колебания для двух сравниваемых препаратов. Как видно из данных табл.1, для стандартного анатоксина А ED_{50} составляла 4,8 мг с предельными колебаниями от 3,9 до 5,5 мг, а для испытуемого анатоксина В - $ED_{50} = 1,4$ мг с колебаниями от 0,8 до 2,2 мг (доверительная вероятность – 0,95). Для нахождения относительной иммуногенности испытуемого и стандартного препарата берут отношение их ED_{50} (R):

$$R = \frac{(ED_{50})_{\text{станд.}}}{(ED_{50})_{\text{исп.}}}$$

В примере $R = \frac{4.8}{1.4} = 3.4$, т.е. испытуемый вакцинный препарат в 3,4 раза

иммуногеннее, чем стандартный.

Существенным для получения достоверных результатов является вычисление пределов доверительного интервала отношения R . В примере с вероятностью 0,95 действительное отношение иммуногенности двух препаратов столбнячного анатоксина находится в пределах от 2,0 до 5,8.

3. Количественные методы определения иммуногенности вакцин, основанные на испытании постоянной дозой антигенов.

Вышеизложенный метод постоянного уровня иммунитета для оценки иммуногенности вакцин, основанный на сравнении эффективных доз антигенов, защищающих 50% иммунизированных животных от какой-либо избранной дозы токсического или инфекционного агента, не позволяет характеризовать уровень иммунитета, достигаемый при введении различных доз препарата, что весьма важно для всесторонней оценки его эффективности. Кроме того, в тех случаях, когда показателем иммунологического ответа является наличие тех или иных гуморальных факторов иммунитета (агглютининов, антитоксинов, комплементсвязывающих антител, преципитинов), метод постоянного уровня иммунитета вообще не применим для сопоставления иммуногенностей вакцинных препаратов, поскольку весьма трудно подбором прививочных дозировок антигена создать постоянный уровень антител в крови иммунизированных животных.

Часто трудно бывает также пользоваться этим методом при оценке низкоиммуногенных препаратов, обеспечивающих в максимальных иммунизирующих дозах минимальный иммунологический ответ. Поэтому для сравнения иммуногенности вакцин весьма полезен может быть и метод постоянной дозы антигена. Последний, как уже указывалось выше, сводится к нахождению среднего уровня иммунитета в группах животных, привитых равными дозами (в антигенных, объемных или весовых дозировках) сравниваемых препаратов. Средний уровень иммунологического ответа выражают или в ED_{50} (LD_{50} или ID_{50} , т.е. количеством токсина или инфекционного агента, которому противостоит 50% животных), или средней (средней геометрической - поскольку уровень антител сам по себе чаще всего оценивают в опытах на животных) титра специфических антител в крови группы животных, подвергшихся иммунизации.

Следовательно, для сравнения иммуногенности двух вакцинных препаратов необходимо найти ED_{50} соответствующего токсина или микроорганизма или же средний титр антител (AE_{cp}) в крови для групп животных, привитых одинаковыми дозами сравниваемых препаратов.

Вычисление ED_{50} (LD_{50} , ID_{50}) производится описанными далее в лекции 10 методами. Выбор метода вычисления зависит при этом от цели исследования, количества животных в опыте и методики его постановки.

Иммуногенность препаратов с использованием данного метода оценивается по отношению ED_{50} (LD_{50} , ID_{50}). Например, по отношению LD_{50} :

$R = \frac{(LD_{50})_1}{(LD_{50})_2}$ для двух препаратов «1» и «2». Если $R > 1$, то препарат «1» более иммуногенен, чем препарат «2» в R раз.

Лекция 10. МЕТОДЫ ОЦЕНОК ПАРАМЕТРОВ КРИВОЙ «ДОЗА-ЭФФЕКТ»

1. Метод Рида-Менча.
2. Метод Кербера.
3. Прóбит-метод.
4. Количественные закономерности связи между уровнем иммунитета и дозой антигена.

Доля p тест-объектов, дающих положительный ответ, может рассматриваться как оценка вероятности реагирования при данной дозе (интенсивности воздействия). Предметом исследования может быть связь между вероятностью реагирования, оцениваемой долей положительных ответов p , и интенсивностью воздействия (доза D).

Обычно доля p возрастает с увеличением дозы D . Зависимость p от D изображается S-образной кривой, называемой кривой эффекта или кривой «доза-эффект». Как правило, значения p изменяются от 0 до 1, т.е. при очень малых дозах ни один тест-объект не реагирует, а при достаточно больших дозах реагируют все тест-объекты.

Чаше всего кривая эффекта несимметрична, но при замене доз их логарифмами она становится симметричной. Поэтому далее мы будем считать, что аргументом являются значения $l = \lg D$. Для краткости будем называть величину l просто дозой. Одной из характеристик кривой может служить та доза, которая вызывает эффект у 50% тест-объектов; ее называют 50%-ной эффективной дозой и обозначают D_{50} . Ниже будут описаны несколько методов оценки D_{50} . При любом методе расчета величина D_{50} оценивается по выборочным данным, поэтому должны строиться и доверительные интервалы для нее.

1. Метод Рида - Менча

Пусть при испытании на пяти дозах число тест-объектов, давших положительное и отрицательное реагирование, такое, как представлено в нижеследующей таблице (столбцы 1–4).

Логарифм дозы	Число тест-объектов в опыте	Частота эффекта		Положительные ответы, %	Накопленная частота эффекта				Пробит
		есть	нет		есть	нет	сумма	%	
1	2	3	4	5	6	7	8	9	10
2,4	6	0	6	0,0	0	17	17	0,0	3,16
2,8	7	1	6	14,3	1	11	12	8,2	3,93
3,2	7	3	4	42,9	4	5	9	44,5	4,82
3,6	6	5	1	83,3	9	1	10	90,0	5,97
4,0	6	6	0	100,0	15	0	15	100,0	6,84

Считая, что в средней части графика «доза-эффект» зависимость может рассматриваться как линейная, естественно применить для интервала доз от 3,2 до 3,6 линейную интерполяцию и записать:

$$l_{50} = \lg D_{50} = 3,2 + (3,6 - 3,2) \frac{50,0 - 42,9}{83,3 - 42,9} = 3,27, \text{ так что } D_{50} = 1,86 \cdot 10^3.$$

Однако при таком подсчете используются только два опыта – для двух доз, близких к D_{50} . Данные опытов, относящиеся к другим дозам, могли бы уточнить результат.

Метод Рида и Менча исходит из того, что если некоторый тест-объект дал положительный ответ при какой-либо дозе, он дал бы такой же ответ и при более высоких дозах; и наоборот, если тест-объект дал отрицательный ответ при определенной дозе, то он дал бы также отрицательный ответ и при всех меньших дозах.

Скорректированный, с учетом результатов опытов с другими дозами, процент положительных ответов при данной дозе дается в графе 9. Эти проценты и употребляются для нахождения D_{50} , тоже с применением линейной интерполяции. Если ввести обозначения: A – скорректированный процент положительных ответов, ближайший к 50% снизу (в примере $A = 44,5\%$), B – скорректированный процент положительных ответов, ближайший к 50% сверху (в примере $B = 90,0\%$), а l_A и l_B – соответствующие дозы (в примере

$l_A = 3,2$ и $l_B = 3,6$), то

$$\lg D_{50} = l_{50} = l_A + (l_B - l_A) \frac{50 - A}{B - A}. \quad (1)$$

Для примера получаем:

$$\lg D_{50} = 3,2 + (3,6 - 3,2) \frac{50 - 44,5}{90,0 - 44,5} = 3,25 \text{ и } D_{50} = 1,78 \cdot 10^3.$$

Стандартную ошибку величины $l_{50} = \lg D_{50}$, найденную методом Риды и Менча, оценивают по формуле

$$S_{l_{50}} = \sqrt{\frac{0,79kh(l_{75} - l_{25})}{N}}, \quad (2)$$

где k – число доз, h – интервал между двумя соседними дозами (в этом методе дозы должны быть равноотстоящими), N – общее число тест-объектов, l_{25} и l_{75} находятся по тому же правилу, что и l_{50} , т.е. линейной интерполяцией.

Если число тест-объектов одно и то же ($n = \frac{N}{k}$), то, очевидно,

$$S_{l_{50}} = \sqrt{\frac{0,79h(l_{75} - l_{25})}{n}}. \quad (3)$$

В нашем примере $h=0,4$ и

$$l_{25} = 2,8 + 0,4 \frac{25,0 - 8,2}{44,5 - 8,2} = 2,99 \text{ и } l_{75} = 3,2 + 0,4 \frac{75,0 - 44,5}{90,0 - 44,5} = 3,47,$$

так что $S_{l_{50}} = \sqrt{\frac{0,79 \cdot 5 \cdot (3,47 - 2,99)}{32}} = 0,154.$

Для границ доверительного интервала получаем:

$$l_{50} \pm u_p \cdot S_{l_{50}}, u_p - \text{аргумент интеграла вероятностей для дов. вероятности } P, \\ u_{0,95} = 1,96, u_{0,99} = 2,58.$$

В примере 95%-ный доверительный интервал для l_{50} будет:

$$3,25 - 1,96 \cdot 0,154 = 2,95, \quad 3,25 + 1,96 \cdot 0,154 = 3,55, \text{ что дает для } D_{50} \\ \text{доверительный интервал } 0,89 \cdot 10^3 \dots 3,54 \cdot 10^3.$$

Метод Риды и Менча можно применять, если значения $l = \lg D$ являются равноотстоящими, а численности групп для разных доз одинаковы (или почти одинаковы). Крайне желательно, чтобы эти численности не были меньше 4.

2. Метод Кербера

Метод Риды и Менча имеет ряд ограничений: значения $l = \lg D$ должны быть равноотстоящими, численности групп для всех доз должны быть одинаковы. Кроме того, применение линейной интерполяции в решающем пункте вычисления l_{50} может внести значительную ошибку в результат, если дозы l_A и l_B не очень близки к l_{50} , так как на участке от l_A до l_B зависимость «доза – эффект» заметно криволинейна.

Метод Кербера свободен от этих недостатков и сохраняет основную идею метода Рида и Менча: если тест-объект дал положительный ответ при дозе l , то он дал бы также положительный ответ и при более высокой дозе, а если он дал отрицательный ответ при этой дозе, то такой же отрицательный ответ он дал бы и при меньшей дозе. В таком случае l_{50} есть такая доза, что половина тест-объектов дает положительный ответ при этой и всех меньших дозах, а другая половина тест-объектов дает отрицательный ответ при этой и всех больших дозах.

Пусть l_0 обозначает дозу, при которой вероятность положительного ответа есть $p_0 \approx 0$, а l_{100} обозначает дозу, при которой эта вероятность есть $p_{100} \approx 1$.

Тогда вычисление l_{50} по методу Кербера сводятся к формуле

$$l_{50} \approx \frac{1}{2}(l_{(K)} + l_{(K+1)}) - \frac{1}{2} \sum_{i=1}^K p_i (l_{(i+1)} - l_{(i-1)}). \quad (4)$$

Здесь дозы могут быть не равноотстоящими, а численности n_i групп тест-объектов при разных дозах могут быть различны. Даже лучше, если эти численности для средних доз больше, чем для крайних.

При оценке стандартной ошибки величины l_{50} исходят из того, что выборочная l_{50} согласно формуле есть сумма одного постоянного слагаемого $\frac{1}{2}(l_{(K)} + l_{(K+1)})$, дисперсия которого равна нулю, и k случайных слагаемых вида $\frac{1}{2} p_i (l_{(i+1)} - l_{(i-1)})$, причем в каждом из них случайным является p_i , а $\frac{1}{2}(l_{(i+1)} - l_{(i-1)})$ есть постоянный множитель. Поэтому на основании формул имеем:

$\sigma_{l_{50}}^2 = \sum_{i=1}^k \frac{\bar{p}_i(1 - \bar{p}_i)}{4n_i} (l_{(i+1)} - l_{(i-1)})^2$. Заменяя неизвестные $\bar{p}_i$ эмпирическими p_i , а числа n_i величинами $n_i - 1$, получаем оценку:

$$S_{l_{50}}^2 = \sum_{i=1}^k \frac{p_i(1 - p_i)}{4(n_i - 1)} (l_{(i+1)} - l_{(i-1)})^2 \quad (5)$$

Формулы упрощаются, если дозы взяты равноотстоящими: $l_{(i+1)} - l_{(i)} = h$ для всех i . Тогда:

$$\frac{1}{2}(l_{(K)} + l_{(K+1)}) = \frac{1}{2}(l_{(K+1)} - h + l_{(K+1)}) = l_{(K+1)} - \frac{h}{2} = l_{100} - \frac{h}{2}.$$

$$\frac{1}{2}(l_{(i+1)} - l_{(i-1)}) = h, \text{ так что}$$

$$l_{50} = l_{100} - h \left(\sum_{i=1}^k p_i + \frac{1}{2} \right), \quad S_{l_{50}}^2 = h^2 \sum_{i=1}^k \frac{p_i(1-p_i)}{n_i - 1} \quad (6)$$

Для данных из таблицы получаем:

$$l_{50} = 4,00 - 0,4 (0,143 + 0,429 + 0,833 + 0,500) = 3,24,$$

$$S_{l_{50}} = 0,4 \sqrt{\frac{0,143 \cdot 0,857}{6} + \frac{0,429 \cdot 0,571}{6} + \frac{0,833 \cdot 0,167}{5}} = 0,119$$

95%-ный доверительный интервал для $\bar{l}_{50}$:

$$3,24 - 1,96 \cdot 0,119 = 3,01, \quad 3,24 + 1,96 \cdot 0,119 = 3,47$$

или доверительный интервал для D_{50} : $1,02 \cdot 10^3 \dots 2,95 \cdot 10^3$.

Следует отметить ограничения и недостатки метода Кербера:

- диапазон доз должен включать дозы с $p = 0$ и $p = 1$.
- замена площади, ограниченной кривой, на сумму площадей трапеций сказывается на точности результата.
- значение D_{50} зависит от значения l_{100} , которое определяется экспериментально не слишком точно.

3. Пробит-метод

От ограничений и недостатков метода Рида-Менча и метода Кербера можно избавиться, если принять, что в генеральной совокупности кривая «доза-эффект» (после преобразования $l = \lg D$) совпадает с графиком функции нормального распределения. Тогда доли положительных ответов можно приравнять к накопленным частостям z нормального распределения, для которых: $z = \Phi\left(\frac{x - \mu}{\sigma}\right)$, где Φ – интеграл вероятностей, μ и σ –

математическое ожидание и стандартное отклонение распределения. Заменяя

z на p , x на l и μ на l_{50} , имеем: $p = \Phi\left(\frac{l - l_{50}}{\sigma}\right)$ Если построить график,

откладывая по оси ординат не значения p , а значения: $y' = \Psi(p)$, где Ψ – функция, обратная к интегралу вероятностей Φ (Последнее означает, что $p = \Phi(y')$), то тогда $y' = (l - l_{50}) / \sigma = l / \sigma - l_{50} / \sigma$, т.е. уравнение прямой линии. Следовательно, если откладывать по оси абсцисс значения l , а по оси ординат – значения $y' = \Psi(p)$, то точки расположатся вдоль прямой линии (имеющей наклон $1 / \sigma$).

При $p < 0,5$ значения y' отрицательны, что представляет известное неудобство. Во избежание этого заменяют величины y' величинами $y = y' + a$, где a – некоторое положительное число. Его выбирают так, чтобы оно

превышало по абсолютной величине все отрицательные значения $\Psi(p)$, которые могут встретиться на практике. Принимают $a = 5$, так что значению $\Psi(p) = +5$ соответствует очень малое значение $p \approx 0,0000003$. Величины $y = \Psi(p) + 5$ называют пробитами.

Имеется специально разработанная таблица, в которой пробит для данного опыта находится по общему числу тест-объектов в опыте и числу положительных ответов.

Значение $p = 0$ заменяют на $p^* = \frac{1}{5n_i}$, а $p = 1$ на $p^* = 1 - \frac{1}{5n_i}$, после

чего вычисляют пробиты из равенства $y' = \Psi(p^*)$.

Чтобы определить l_{50} (а затем и D_{50}), нужно найти абсциссу той точки прямой, ордината которой равна $y = 5$ (это соответствует $y' = 0$ и $p = 0,5$). Поэтому проводят прямую, используя экспериментальные точки l, y .

Пример. Для данных из таблицы применение пробит-метода дает $l_{50} = 3,21$ и $D_{50} = 1,62 \cdot 10^3$.

Для оценки стандартной ошибки $\sigma_{l_{50}}$ используют формулу:

$$S_{l_{50}} = \frac{\sigma}{\sqrt{N'/2}}, \text{ где } N' - \text{число тест-объектов, для которых значения}$$

пробитов находятся в пределах от 3,5 до 6,5 (в примере $N' = 20$). Величину σ оценивают по графику: точки прямой, имеющие ординаты 4,0 и 6,0, соответствуют значениям l , равным $l_{50} - \sigma$ и $l_{50} + \sigma$, так что $\sigma = l_{y=5,0} - l_{y=4,0} = l_{y=6,0} - l_{y=5,0}$.

Более точную оценку σ дает выражение: $\sigma = 0,5 (l_{y=6,0} - l_{y=4,0})$.

В примере $l_{y=4,0} = 2,79$, $l_{y=6,0} = 3,63$, так что $\sigma = 0,5 (3,63 - 2,79) = 0,42$.

$S_{l_{50}} = \frac{0,42}{\sqrt{10}} = 0,133$, поэтому 95%-ные довер. интервалы для $\bar{l}_{50}$ и $\bar{D}_{50}$ будут:

$$\bar{l}_{50} : 3,21 \pm 1,96 \cdot 0,133 = 2,95 \dots 3,47 ; \quad \bar{D}_{50} : 0,89 \cdot 10^3 \dots 2,96 \cdot 10^3.$$

Замечания:

- применение пробит-метода не требует ни равноотстоящих доз, ни одинаковой численности тест-объектов в группе.
- дозы с $p = 0$ и $p = 1$ могут как присутствовать, так и отсутствовать.
- точность метода по сравнению с методом Рида-Менча и Кербера выше.

Недостаток пробит-метода – использование допущения о нормальности кривой «доза-эффект».

4. Количественные закономерности связи между уровнем иммунитета и дозой антигена.

Весьма важным для характеристики вакцинного препарата является выявление связи между его прививочной дозой и уровнем достигнутого иммунитета. Известно, что связь между уровнем иммунитета и дозой антигена зависит как от качества антигена, так и от иммунологической реактивности к нему организма.

Первые попытки найти количественную взаимосвязь между уровнем антител в крови иммунизированных животных и дозой антигена имелись в работах Смита и Брукса, а также Истрати, Кикш и Пригге. Смит и Брукс, определяя уровень агглютининов в крови у кроликов, иммунизированных различными дозами брюшнотифозной вакцины, высказали предположение, что зависимость между титром агглютининов в крови и дозой антигена напоминает уравнение изотермы адсорбции Фрейндлиха. По данным Пригге, между уровнем дифтерийных антитоксинов в крови морских свинок (I) и иммунизирующей дозой (A) дифтерийного анатоксина имеется следующая зависимость

$$I = cA, \quad (7)$$

где c - эмпирический коэффициент.

В работах А.В. Марковича и А.А. Воробьева в результате изучения закономерностей развития иммунитета к столбнячному токсину при однократной иммунизации белых мышей сорбированным столбнячным анатоксином была установлена линейная зависимость между логарифмом дозы антигена ($\lg D$) и логарифмом напряженности иммунитета (LD_{50}):

$$\lg(LD_{50}) = \alpha + K \lg D, \quad (8)$$

где LD_{50} - доза токсина в DLM, вызывающая гибель 50% иммунизированных животных (показатель напряженности иммунитета); D - доза антигена (в мл, мг, Lf , ЕС и т. д.); α и K - эмпирические коэффициенты. Коэффициент K является тангенсом угла наклона прямых $\lg LD_{50}$ - $\lg D$, коэффициент α представляет отрезок на оси ординат и показывает величину напряженности иммунитета при иммунизации единицей антигена ($\alpha = \lg LD_{50}$ при $D = 1$).

Для двух иммунизирующих доз антигена D_0 и D_1 (за D_0 принимается минимальная иммунизирующая доза антигена в опыте) и им

соответствующих напряженностей иммунитета $(LD_{50})_0$ и $(LD_{50})_1$ из формулы (8) получаем:

$$\lg \frac{(LD_{50})_1}{(LD_{50})_0} = K \lg \frac{D_1}{D_0} \quad (9)$$

или

$$\frac{(LD_{50})_1}{(LD_{50})_0} = \left(\frac{D_1}{D_0} \right)^K. \quad (10)$$

Уравнение (9) показывает, что логарифм относительного повышения напряженности иммунитета прямо пропорционален логарифму отношения соответствующих доз антигена.

На основании собственных экспериментальных данных, а также обработки экспериментальных данных других авторов А.В. Маркович и А.А. Воробьев показали применимость формулы (8) для оценки эффективности иммунизации различными антигенами (столбнячный, дифтерийный и перфрингенс анатоксины, брюшнотифозная и дизентерийная вакцины) при разных способах их введения (подкожный, внутривентриальный), в динамике развития иммунитета, а также для оценки уровней разных факторов иммунитета (гуморальный иммунитет, устойчивость к токсинам и к инфекции).

Зависимость, выражаемая формулой (8), была подтверждена впоследствии на большом экспериментальном материале в работах Холта и Берне и особенно Стивенса. Холт и Бернс назвали эту формулу *уравнением антигенности*.

Уравнение антигенности, устанавливающее количественную связь между уровнем иммунитета и дозой антигена, указывает лишь на принципиальную основу этой связи, зависящую от разнообразных факторов общего и специального характера. Вполне очевидно, что уравнение антигенности справедливо только в области «средних» прививочных доз антигенов и напряженностей иммунитета. Диапазон доз антигенов и уровней иммунитета, в которых применимо уравнение, определяется природой антигена и чувствительностью к нему животного организма. Например, по нашим данным, при испытании на белых мышках сорбированных столбнячного и перфрингенс анатоксинов диапазон прививочных доз, в которых зависимость между напряженностью иммунитета к токсину и дозой антигена подчинялась уравнению антигенности, составлял для столбнячного анатоксина от 0,005 до 0,1 мл (т.е. дозы отличались в 20 раз), а для анатоксина перфрингенс - от 0,005 до 0,01 мл (т.е. дозы анатоксинов отличались в 2 раза).

Ограниченность применения уравнения антигенности может быть показана на примере изучения взаимосвязи между уровнем столбнячного антитоксина и дозой столбнячного анатоксина; в этом случае повышение уровня антитоксина в крови белых мышей, согласно уравнению, происходило в интервале доз антигена от 0,005 до 0,025 мл (т.е. при увеличении доз в 5 раз), тогда как напряженность иммунитета возрастала по тому же уравнению при увеличении доз антигена - в 20 раз (от 0,005 мл до 0,1 мл).

Изучение коэффициентов α и K , в уравнении антигенности показало, что величина коэффициента K : а) постоянна для данной группы иммунизированных животных, независимо от иммуногенности примененного препарата одного вида; б) зависит от вида антигена и не зависит от препарата антигена; в) изменяется вместе с изменением иммунологической реактивности организма во времени; г) зависит от уровня напряженности иммунитета; д) возрастает при кратных прививках; е) не зависит от метода введения антигенов (при прочих равных условиях).

Коэффициент α : а) возрастает с увеличением иммуногенности препаратов одного и того же антигена при их испытании на группах животных, находящихся в равноценных условиях (при равенстве коэффициента K); б) изменяется вместе с коэффициентом K в динамике развития иммунитета; в) зависит от метода введения антигена (при равном K) и схемы иммунизации (однократная, двукратная).

Анализ причин, влияющих на изменение коэффициентов α и K в уравнении антигенности, позволяет сделать предположение, что коэффициент K зависит от иммунологической реактивности организма к антигену, а коэффициент α - от иммуногенности препаратов антигенов.

Наличие математической зависимости между дозой антигена и уровнем иммунологического ответа позволяет при применении умеренных доз антигенов сравнивать иммуногенность препаратов в одном и том же опыте двумя методами: постоянной дозы антигенов и постоянного уровня иммунитета.

Лекция 11. МЕТОДЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТОВ

1. Основные понятия и определения.
2. Параметр оптимизации и требования к нему.
3. Факторы, требования к ним.
4. Полный факторный эксперимент.
5. Дробный факторный эксперимент.

1. Основные понятия и определения.

Под экспериментом будем понимать совокупность операций совершаемых над объектом исследования с целью получения информации о его свойствах. Эксперимент, в котором исследователь по своему усмотрению может изменять условия его проведения, называется активным экспериментом. Если исследователь не может самостоятельно изменять условия его проведения, а лишь регистрирует их, то это пассивный эксперимент.

Важнейшей задачей методов обработки полученной в ходе эксперимента информации является задача построения математической модели изучаемого явления, процесса, объекта. Ее можно использовать и при анализе процессов и при проектировании объектов. Можно получить хорошо аппроксимирующую математическую модель, если целенаправленно применяется активный эксперимент. Другой задачей обработки полученной в ходе эксперимента информации является задача оптимизации, т.е. нахождения такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.

Опыт – это отдельная экспериментальная часть.

План эксперимента – совокупность данных определяющих число, условия и порядок проведения опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям, совокупность действий направленных на разработку стратегии экспериментирования (от получения априорной информации до получения работоспособной математической модели или определения оптимальных условий). Это целенаправленное управление экспериментом, реализуемое в условиях неполного знания механизма изучаемого явления.

В процессе измерений, последующей обработки данных, а также формализации результатов в виде математической модели, возникают погрешности и теряется часть информации, содержащейся в исходных данных. Применение методов планирования эксперимента позволяет определить погрешность математической модели и судить о ее адекватности. Если точность модели оказывается недостаточной, то применение методов планирования эксперимента позволяет модернизировать математическую модель с проведением дополнительных опытов без потери предыдущей информации и с минимальными затратами.

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов при которых удастся получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.

Пусть интересующее нас свойство (Y) объекта зависит от нескольких (n) независимых переменных ($X_1, X_2, \dots, X_n$) и мы хотим выяснить характер этой зависимости - $Y=F(X_1, X_2, \dots, X_n)$, о которой мы имеем лишь общее представление. Величина Y – называется “отклик”, а сама зависимость $Y=F(X_1, X_2, \dots, X_n)$ – “функция отклика”.

Отклик должен быть определен количественно. Однако могут встречаться и качественные признаки Y . В этом случае возможно применение рангового подхода. Пример рангового подхода - оценка на экзамене, когда одним числом оценивается сложный комплекс полученных сведений о знаниях студента.

Независимые переменные $X_1, X_2, \dots, X_n$ – иначе факторы, также должны иметь количественную оценку. Если используются качественные факторы, то каждому их уровню должно быть присвоено какое-либо число. Важно выбирать в качестве факторов лишь независимые переменные, т.е. только те которые можно изменять, не затрагивая другие факторы. Факторы должны быть однозначными. Для построения эффективной математической модели целесообразно провести предварительный анализ значимости факторов (степени влияния на функцию), их ранжирование и исключить малозначащие факторы.

Диапазоны изменения факторов задают область определения Y . Если принять, что каждому фактору соответствует координатная ось, то полученное пространство называется факторным пространством. При $n=2$ область определения Y представляется собой прямоугольник, при $n=3$ – куб, при $n > 3$ - гиперкуб.

При выборе диапазонов изменения факторов нужно учитывать их совместимость, т.е. контролировать, чтобы в этих диапазонах любые сочетания факторов были бы реализуемы в опытах и не приводили бы к абсурду. Для каждого из факторов указывают граничные значения

$$X_{\min} \leq X_i \leq X_{\max},$$

где $i=1, \dots, n$.

Регрессионный анализ функции отклика предназначен для получения ее математической модели в виде уравнения регрессии

$$Y = F(X_1, X_2, \dots, X_n; B_1, B_2, \dots, B_m) + e,$$

где $B_1, \dots, B_m$ – некоторые коэффициенты; e – погрешность.

Среди основных методов планирования, применяемых на разных этапах исследования, используют:

- планирование отсеивающего эксперимента, основное значение которого выделение из всей совокупности факторов группы существенных факторов, подлежащих дальнейшему детальному изучению;

- планирование эксперимента для дисперсионного анализа, т.е. составление планов для объектов с качественными факторами;

- планирование регрессионного эксперимента, позволяющего получать регрессионные модели (полиномиальные и иные);

- планирование экстремального эксперимента, в котором главная задача – экспериментальная оптимизация объекта исследования;

- планирование при изучении динамических процессов и т.д.

Инициатором применения планирования эксперимента является Рональд А. Фишер, другой автор известных первых работ – Френк Йетс. Далее идеи планирования эксперимента формировались в трудах Дж. Бокса, Дж. Кифера. В нашей стране - в трудах Г.К. Круга, Е.В. Маркова и др.

В настоящее время методы планирования эксперимента заложены в специализированных пакетах, широко представленных на рынке программных продуктов, например: StatGrafics, Statistica, SPSS, SYSTAT и др.

Представление результатов экспериментов

При использовании методов планирования эксперимента необходимо найти ответы на 4 вопроса:

- Какие сочетания факторов и сколько таких сочетаний необходимо взять для определения функции отклика?

- Как найти коэффициенты $B_0, B_1, \dots, B_m$?

- Как оценить точность представления функции отклика?

Как использовать полученное представление для поиска оптимальных значений Y ?

Геометрическое представление функции отклика в факторном пространстве $X_1, X_2, \dots, X_n$ называется поверхностью отклика (рис. 1).

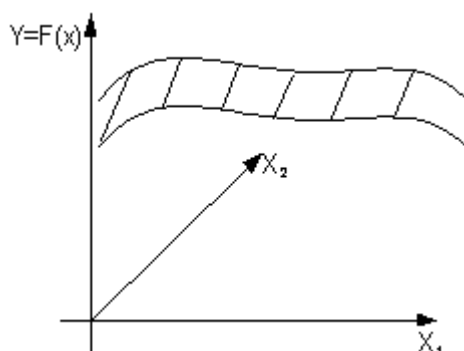


Рис. 1. Поверхность отклика

Если исследуется влияние на Y лишь одного фактора X_1 , то нахождение функции отклика - достаточно простая задача. Задавшись несколькими значениями этого фактора, в результате опытов получаем соответствующие значения Y и график $Y = F(X)$ (рис. 2).

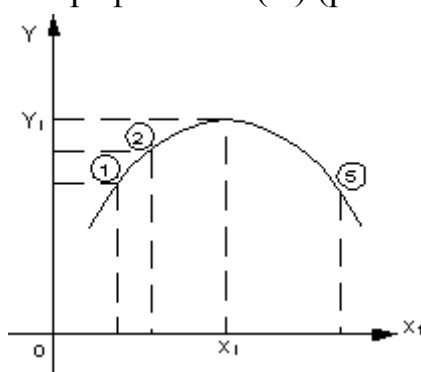


Рис. 2. Построение функции отклика одной переменной по опытным данным

По его виду можно подобрать математическое выражение функции отклика. Если мы не уверены, что опыты хорошо воспроизводятся, то обычно опыты повторяют несколько раз и получают зависимость с учетом разброса опытных данных.

Если факторов два, то необходимо провести опыты при разных соотношениях этих факторов. Полученную функцию отклика в 3х-мерном пространстве (рис. 1) можно анализировать, проводя ряд сечений с фиксированными значениями одного из факторов (рис. 3). Вычлененные графики сечений можно аппроксимировать совокупностью математических выражений.

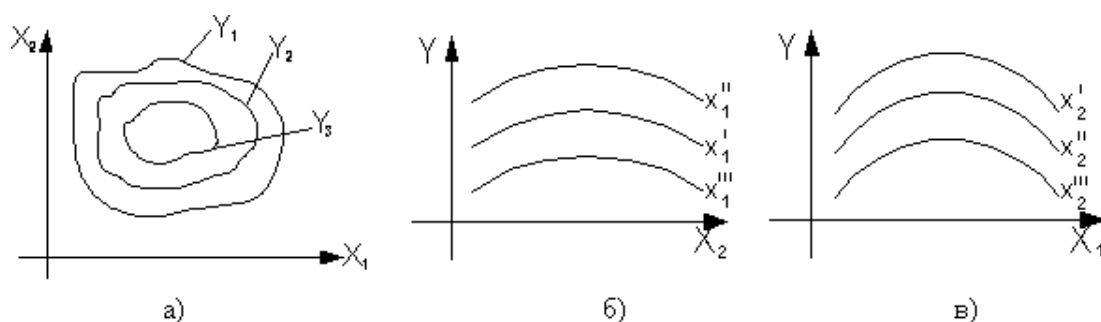


Рис. 3. Сечения поверхности отклика при фиксированных откликах (а) и переменных (б,в).

При трех и более факторах задача становится практически неразрешимой. Если и будут найдены решения, то использовать совокупность выражений достаточно трудно, а часто и не реально.

Например, пусть необходимо исследовать влияние U , f и R_r на M_p и P_2 асинхронного двигателя (АД) (рис. 4).

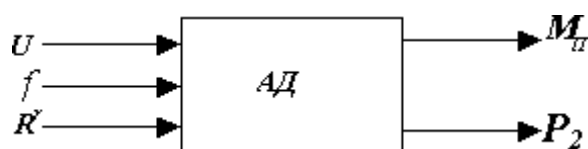


Рис. 4. Исследование влияния U , f и R_r на M_p и P_2 АД

Если в диапазоне изменения каждого фактора взять хотя бы по пять точек

$U, В$	170	180	190	200	210	220
$f, Гц$	40	45	50	55	60	65
$R_r, Ом$	0,1	0,2	0,3	0,4	0,5	0,6

то для того чтобы выполнить опыты при всех возможных сочетаниях значений факторов (их три) необходимо выполнить $5^3=125$ опытов и сформировать по $5^2=25$ кривых для каждой из двух функций отклика. Если мы хотим хотя бы продублировать опыты чтобы снизить погрешность, то число опытов пропорционально возрастает, поэтому произвольное выполнение опытов при числе факторов более двух и использование их результатов - практически нереально.

Эксперимент занимает центральное место в науке. Однако возникает вопрос, насколько эффективно он используется. Джон Бернал, например, отмечал, что научные исследования организуются и проводятся настолько хаотично, что их коэффициент полезного действия может быть оценен величиной порядка 2%. Для того чтобы повысить эффективность исследований, требуется нечто совершенно новое. Одним из возможных

путей является применение математических методов, построение математической теории планирования эксперимента.

Планирование эксперимента - это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью. При этом существенно следующее: стремление к минимизации общего числа опытов; одновременное варьирование всеми переменными, определяющими процесс, по специальным правилам - алгоритмам; использование математического аппарата, формализующего многие действия экспериментатора; выбор четкой стратегии, позволяющей принимать обоснованные решения после каждой серии экспериментов.

Задачи, для решения которых может использоваться планирование эксперимента, чрезвычайно разнообразны. Поиск оптимальных условий, построение интерполяционных формул, выбор существенных факторов, оценка и уточнение констант теоретических моделей (например, кинетических), выбор наиболее приемлемых из некоторого множества гипотез о механизме явлений, исследование диаграмм состав-свойство - вот примеры задач, при решении которых применяется планирование эксперимента. Можно сказать, что там, где есть эксперимент, имеет место и наука о его проведении - планирование эксперимента.

Поиск оптимальных условий является одной из наиболее распространенных научно-технических задач. Они возникают в тот момент, когда установлена возможность проведения процесса и необходимо найти наилучшие (оптимальные в некотором смысле) условия его реализации. Этим задачам и посвящены ближайшие две лекции.

Эксперимент, который ставится для решения задач оптимизации, называется экстремальным. Это название связано с глубокой аналогией между оптимизацией и поиском экстремума некоторой функции.

Определим еще ряд важных понятий, первое из которых - «объект исследования». Для описания объекта исследования удобно пользоваться представлением о кибернетической системе. Иногда такую кибернетическую систему называют «черным ящиком». Стрелки справа на схеме этой системы изображают численные характеристики целей исследования. Мы обозначаем их буквой Y и называем параметрами оптимизации. В литературе можно встретить и другие названия: критерий оптимизации, целевая функция, выход «черного ящика» и т.д.

Для проведения эксперимента необходимо иметь возможность воздействовать на поведение «черного ящика». Все способы такого

воздействия мы обозначаем буквой x и называем факторами. Их называют также входами «черного ящика».

При решении задачи будем использовать математические модели объекта исследования. Под математической моделью мы понимаем уравнение, связывающее параметр оптимизации с факторами. Это уравнение в общем виде можно записать в виде $y = f(x)$. Такая функция называется функцией отклика.

Далее мы рассмотрим вопрос о том, как эту функцию можно выбрать и построить. А сейчас важно понять, как получаются условия проведения опытов в том эксперименте, который собираются провести.

Каждый фактор может принимать в опыте одно из нескольких значений. Такие значения называют уровнями. Может оказаться, что фактор способен принимать бесконечно много значений (непрерывный ряд). Однако на практике точность, с которой устанавливается некоторое значение, не беспредельна. Поэтому можно считать, что всякий фактор имеет определенное число дискретных уровней. Это соглашение существенно облегчает построение «черного ящика» и эксперимента, а также упрощает оценку их сложности.

Фиксированный набор уровней факторов (т. е. установление каждого фактора на некоторый уровень) определяет одно из возможных состояний «черного ящика». Одновременно это есть условия проведения одного из возможных опытов. Если перебрать все возможные наборы состояний, то мы получим полное множество различных состояний данного «ящика». Одновременно это будет число возможных различных опытов. Чтобы узнать число различных состояний, достаточно число уровней факторов (если оно для всех факторов одинаково) возвести в степень числа факторов k : p^k , где p - число уровней.

Реальные объекты, с которыми мы сталкиваемся ежедневно, обладают огромной сложностью. Так, на первый взгляд простая система с пятью факторами на пяти уровнях имеет 3125 состояний, а для десяти факторов на четырех уровнях их уже свыше миллиона!

В этих условиях мы просто вынуждены отказаться от таких экспериментов, которые включают все возможные опыты: перебор слишком велик. Тогда возникает вопрос: сколько и каких опытов надо включить в эксперимент, чтобы решить поставленную задачу? Здесь-то и приходит на помощь планирование эксперимента.

Однако нужно иметь в виду, что при планировании эксперимента не безразлично, какими свойствами обладает объект исследования. Укажем два основных требования, с которыми приходится считаться. Прежде всего

существенно, воспроизводятся ли на объекте результаты эксперимента. Выберем некоторые уровни для всех факторов и в этих условиях проведем эксперимент. Затем повторим его несколько раз через неравные промежутки времени и сравним значения параметра оптимизации. Разброс этих значений характеризует воспроизводимость результатов. Если он не превышает некоторой заранее заданной величины (наших требований к точности эксперимента), то объект удовлетворяет требованию воспроизводимости результатов, а если превышает, то не удовлетворяет этому требованию. Мы будем рассматривать только такие объекты, для которых требование воспроизводимости выполняется.

Планирование эксперимента предполагает активное вмешательство в процесс и возможность выбора в каждом опыте тех уровней факторов, которые представляют интерес. Поэтому такой эксперимент называется активным. Объект, на котором возможен активный эксперимент, называется управляемым. Это и есть второе требование к объекту исследования.

На практике нет абсолютно управляемых объектов. На реальный объект обычно действуют как управляемые, так и неуправляемые факторы. Неуправляемые факторы влияют на воспроизводимость эксперимента и являются причиной ее нарушения. Если требования воспроизводимости не выполняются, приходится обращаться к активно-пассивному эксперименту. Возможно, плохая воспроизводимость объясняется действием фактора, систематически изменяющегося (дрейфующего) во времени. Тогда нужно обращаться к специальным методам планирования. Наконец, возможно, что все факторы неуправляемы. В этом случае возникает задача установления связи между параметром оптимизации и факторами по результатам наблюдений за поведением объекта, или, как говорят, по результатам пассивного эксперимента. Эти случаи мы не будем рассматривать. Наша цель - изложение методов планирования экстремального эксперимента для воспроизводимых управляемых статических объектов.

Планирование экстремального эксперимента - это метод выбора количества и условий проведения опытов, минимально необходимых для отыскания оптимальных условий, т.е. для решения поставленной задачи.

Приступая к знакомству с планированием экстремального эксперимента, надо иметь в виду, что при оптимизации распространен так называемый детерминированный подход. При этом предполагается построение физической модели процесса на основании тщательного изучения механизма явлений (например, кинетики процесса, что позволяет получить математическую модель объекта в виде системы дифференциальных уравнений. Несомненно, что детерминированный и

статистический (связанный с планированием эксперимента) подходы должны разумно дополнять друг друга, а не противопоставляться, как это иногда делается.

Теперь можно считать, что основные определения введены, и мы готовы перейти к детальному рассмотрению нашей задачи.

Итак, мы познакомились с основными определениями, которые используются в теории планирования экстремального эксперимента. Прежде чем приступить к эксперименту, необходимо однозначно и непротиворечиво сформулировать его цель и выбрать подходящую количественную характеристику этой цели, которую мы назвали параметром оптимизации.

Понятие «объект исследования» требует точного формального определения. Для такого определения удалось приспособить кибернетическое понятие «черный ящик» - модель объекта. Экспериментатор, вставший на путь применения методов планирования эксперимента, должен уметь формулировать свою задачу в терминах «черного ящика».

Входы «черного ящика» называются факторами. Каждый фактор может принимать некоторое определенное число различных значений, называемых уровнями. Сочетание определенных уровней всех факторов определяет возможное состояние «черного ящика» и условия одного из возможных опытов. Совокупность всех различных возможных состояний определяет сложность «черного ящика» и общее число возможных опытов.

Результаты эксперимента используются для получения математической модели объекта исследования, которая представляет собой уравнение, связывающее параметр оптимизации и факторы.

Такое уравнение называется функцией отклика.

Использование для получения модели всех возможных опытов приводит к абсурдно большому количеству экспериментов. Задача выбора необходимых для эксперимента опытов, методов математической обработки их результатов и принятия решений - это и есть задача планирования эксперимента. Частный случай этой задачи - планирование "экстремального эксперимента, т.е. эксперимента, поставленного с целью поиска оптимальных условий функционирования объекта. Планирование экстремального эксперимента - метод выбора минимального количества опытов, необходимых для отыскания оптимальных условий.

2. Параметр оптимизации и требования к нему

При планировании экстремального эксперимента очень важно определить параметр, который нужно оптимизировать. Сделать это совсем не так просто, как кажется на первый взгляд. Цель исследования должна быть сформулирована очень четко и допускать количественную оценку. Будем называть характеристику цели, заданную количественно, параметром оптимизации. Параметр оптимизации является реакцией (откликом) на воздействие факторов, которые определяют поведение выбранной вами системы.

Реакция объекта многогранна, многоаспектна. Выбор того аспекта, который представляет наибольший интерес, как раз и задается целью исследования.

При традиционном нематематическом подходе исследователь стремится как-то учесть разные аспекты, взвесить их и принять согласованное решение о том, какой опыт лучше. Однако разные экспериментаторы проведут сравнение опытов неодинаково. Различия, если хотите, одно из проявлений таланта исследователя или его бездарности.

Сформулируем требования к параметрам оптимизации и рекомендации по их выбору.

Параметр оптимизации - это реакция (отклик) на воздействия факторов, которые определяют поведение изучаемой системы.

Параметры оптимизации бывают экономическими, технико-экономическими, технико-технологическими, статистическими, психологическими и т.д.

Параметр оптимизации должен быть:

- эффективным с точки зрения достижения цели;
- универсальным;
- количественным и выражаться одним числом;
- статистически эффективным;
- имеющим физический смысл, простым и легко вычисляемым;
- существующим для всех различных состояний.

В тех случаях, когда возникают трудности с количественной оценкой параметров оптимизации, приходится обращаться к ранговому подходу. В ходе исследования могут меняться априорные представления об объекте исследования, что приводит к последовательному подходу при выборе параметра оптимизации.

Из многих параметров, характеризующих объект исследования, только один, часто обобщенный, может служить параметром оптимизации. Остальные рассматриваются как ограничения.

3. Факторы, требования к ним

Способы воздействия на системы были названы факторами.

После того как выбран объект исследования и параметр оптимизации, нужно включить в рассмотрение все существенные факторы, которые могут влиять на процесс. Если какой-либо существенный фактор окажется неучтенным, то это может привести к неприятным последствиям. Так, если неучтенный фактор произвольно флуктуировал - принимал случайные значения, которые экспериментатор не контролировал, - это значительно увеличит ошибку. При поддержании фактора на некотором фиксированном уровне может быть получено ложное представление об оптимуме, так как нет гарантии, что фиксированный уровень является оптимальным.

Фактором называется измеряемая переменная величина, принимающая в некоторый момент времени определенное значение.

Факторы соответствуют способам воздействия на объект исследования.

Так же, как и параметр оптимизации, каждый фактор имеет область определения. Мы будем считать фактор заданным, если вместе с его названием указана область его определения.

Под областью определения понимается совокупность всех значений, которые в принципе может принимать данный фактор. Ясно, что совокупность значений фактора, которая используется в эксперименте, является подмножеством из множества значений, образующих область определения.

Область определения может быть непрерывной и дискретной. Однако в тех задачах планирования эксперимента, которые мы собираемся рассматривать, всегда используются дискретные области определения. Так, для факторов с непрерывной областью определения, таких, как температура, время, количество вещества и т.п., всегда выбираются дискретные множества уровней.

В практических задачах области определения факторов, как правило, ограничены. Ограничения могут носить принципиальный либо технический характер.

Факторы разделяются на количественные и качественные. Качественные факторы - это разные вещества, разные технологические способы, аппараты, исполнители и т.д. Хотя качественным факторам не соответствует числовая шкала в том смысле, как это понимается для количественных факторов, однако можно построить условную порядковую шкалу, которая ставит в соответствие уровням качественного фактора числа

натурального ряда, т.е. производит кодирование. Порядок уровней может быть произволен, но после кодирования он фиксируется.

В ряде случаев граница между понятием качественного и количественного фактора весьма условна.

Итак, факторы - это переменные величины, соответствующие способам воздействия внешней среды на объект. Они определяют как сам объект, так и его состояние.

Требования к факторам: управляемость и однозначность.

Управлять фактором - это значит установить нужное значение и поддерживать его постоянным в течение опыта или менять по заданной программе. В этом состоит особенность «активного» эксперимента. Планировать эксперимент можно только в том случае, если уровни факторов подчиняются воле экспериментатора.

Факторы должны непосредственно воздействовать на объект исследования. Трудно управлять фактором, если он является функцией других переменных, но в планировании эксперимента могут участвовать сложные факторы, такие, как логарифмы, соотношения и т.д.

Факторы должны быть определены операционально.

Требования к совокупности факторов: совместимость и отсутствие линейной корреляции. Выбранное множество факторов должно быть достаточно полным. Если какой-либо существенный фактор пропущен, это приведет к неправильному определению оптимальных условий или к большой ошибке опыта.

Факторы могут быть количественными и качественными.

Точность фиксации факторов должна быть высока. Степень точности определяется диапазоном изменения факторов.

Выбор факторов - очень ответственный этап при подготовке к планированию эксперимента. От удачного выбора зависит успех оптимизации.

4. Полный факторный эксперимент

Первой серии опытов предшествует этап неформализованных решений, направленных на выбор локальной области факторного пространства. При этом оцениваются границы областей определения факторов, задаваемые либо принципиальными ограничениями, либо технико-экономическими соображениями, либо конкретными условиями проведения процесса. Установление области связано с тщательным анализом априорной

информации об изменении параметра оптимизации и о кривизне поверхности отклика.

Локальная область проведения эксперимента выбирается в два этапа: определение основного уровня и интервалов варьирования. Основной (нулевой) уровень - многомерная точка в факторном пространстве, задаваемая комбинацией уровней факторов.

Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно основного уровня.

При установлении основного уровня приходится рассматривать различные ситуации. Ситуации задаются информацией о наилучших точках и определяют решения.

Следующий этап - выбор интервалов варьирования факторов. Для каждого фактора определяются два уровня, на которых он варьируется в эксперименте. Уровни факторов изображаются двумя точками на координатной оси, симметричными относительно основного уровня. Один из уровней - верхний, другой - нижний. Интервалом варьирования факторов называется некоторое число (свое для каждого фактора), прибавление которого к основному уровню дает верхний, а вычитание - нижний уровень.

Для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям задают так, чтобы верхний уровень соответствовал $+1$, нижний -1 , основной - нулю.

На выбор интервалов варьирования накладываются ограничения снизу (он не может быть меньше ошибки фиксирования уровня фактора) и сверху (верхний или нижний уровни не должны выходить за область определения).

В задачах оптимизации выбирают подобласть, которая давала бы возможность реализовать шаговую процедуру движения к оптимуму. В задачах интерполяции интервал варьирования охватывает всю описываемую область.

При определении интервала варьирования используется информация о точности, с которой фиксируются значения факторов, о кривизне поверхности отклика и о диапазоне изменения параметра оптимизации. Для принятых градаций этих признаков существует 27 различных ситуаций. Низкая точность фиксирования факторов определяет типичное решение - широкий интервал варьирования. Для средней точности характерен выбор среднего интервала. Высокая точность обычно приводит либо к узкому, либо к среднему интервалам.

Эксперимент, в котором реализуются все возможные сочетания уровней-, называется полным факторным экспериментом.

Если число уровней равно двум, то это полный факторный эксперимент типа 2^k . Условия эксперимента представляют в виде таблицы - матрицы планирования, где строки соответствуют различным опытам, а столбцы - значениям факторов. Геометрическая интерпретация полных факторных планов: план 2^2 задается координатами вершин квадрата, план 2^3 - координатами вершин куба, при $k > 3$ — координатами вершин гиперкуба.

Полный факторный эксперимент типа 2^k обладает свойствами симметричности, нормировки, ортогональности, ротатабельности (для линейной модели).

Коэффициенты, вычисленные по результатам эксперимента, указывают на силу влияния факторов. Эффект фактора численно равен удвоенному коэффициенту. В тех случаях, когда эффект одного фактора зависит от уровня, на котором находится другой фактор, говорят о наличии эффекта взаимодействия двух факторов. Для его количественной оценки получают столбец произведений этих факторов и обращаются с ним как с вектор-столбцом любого фактора.

Из полного факторного эксперимента нельзя извлечь информацию о квадратичных членах. Вектор-столбцы для квадратичных членов совпадают друг с другом и со столбцом x_0 . Величина свободного члена b_0 включает вклады квадратичных членов, получается смешанная оценка. Оценки остальных коэффициентов не смешаны.

В полном факторном эксперименте разность между числом опытов и числом коэффициентов велика. Возникает проблема уменьшения числа опытов. Этому вопросу посвящен следующий раздел лекции.

5. Дробный факторный эксперимент

Количество опытов в полном факторном эксперименте значительно превосходит число определяемых коэффициентов линейной модели. Другими словами, полный факторный эксперимент обладает большой избыточностью опытов. Было бы заманчивым сократить их число за счет той информации, которая не очень существенна при построении линейных моделей. При этом нужно стремиться к тому, чтобы матрица планирования не лишилась своих оптимальных свойств. Сделать это не так просто, но все же возможно.

Дробные факторные эксперименты (дробные реплики) находят широкое применение при получении линейных моделей. Целесообразность их применения возрастает с ростом количества факторов. Показано, что при исследовании влияния 15 факторов можно в 2048 раз сократить число

опытов, применяя реплику большой дробности – 16 опытов вместо 32768). Эффективность применения дробных реплик зависит от удачного выбора системы смешивания линейных эффектов с эффектами взаимодействия, а также от умелой стратегии экспериментирования в случае значимости некоторых взаимодействий. Априорные сведения о взаимодействиях могут оказать большую услугу экспериментатору.

При построении дробных реплик используют следующее правило: для того чтобы сократить число опытов при введении в планирование нового фактора, нужно поместить этот фактор в вектор- столбец матрицы, принадлежащий взаимодействию, которым можно пренебречь.

Реплики, которые используются для сокращения опытов в 2^k раз, где $k = 1, 2, 3, 4, \dots$, называются регулярными. Они пользуются большой популярностью, так как позволяют производить расчет коэффициентов уравнения так же просто, как и в случае полного факторного эксперимента.

С ростом числа факторов быстро увеличивается число реплик различной дробности.

Лекция 12. МЕТОДЫ ОПТИМИЗАЦИИ

1. Метод крутого восхождения или метод Бокса-Уилсона.
2. Симплексный метод оптимизации.

Существует большое число разнообразных методов многомерного поиска (см. рис. 1)



Рисунок 1- Методы поиска оптимума

Методы поиска делятся на две большие группы: градиентные и неградиентные. В качестве градиентных методов поиска можно привести пример «метода крутого восхождения» (МКВ).

Неградиентные методы поиска отличаются большим разнообразием идей, положенных в их основу.

1. Метод крутого восхождения.

Планирование эксперимента при поиске оптимальных условий методом крутого восхождения.

Общие положения

Для достижения прогресса в любом производстве необходимо оптимизировать производственные процессы. Важное значение оптимизация имеет в технологических процессах, т.к. уровень технологии определяется не только производительностью, уровнем брака и другими показателями, непосредственно на стадии производства, но и показателями, реализующиеся при эксплуатации продукции. Например, способность конструкции выдерживать электрические, механические и другие нагрузки, обеспечивать показатели надежности и долговечности изделий.

Оптимизация технологических процессов связана с установлением наилучших режимов работы оборудования, с оптимальным выбором применяемых материалов или условий проведения технологических процессов, которые бы в совокупности обеспечивали наилучшие показатели качества при возможно меньших затратах.

Для нахождения оптимальных условий, как правило, применяются два подхода.

1. Определение оптимальных условий с помощью модели объекта.

Предполагается, что имеется статистическая (экспериментальная) модель процесса, т.е. полученная с помощью эксперимента методом наименьших квадратов, планирования экспериментов или другими методами математического описания объекта (процесса).

2. Нахождение оптимальных условий непосредственно на объекте (в эксперименте) без использования математического описания процесса.

Этот метод предполагает возможность проводить на объекте эксперименты.

Оба метода используют (первый явно, второй неявно) понятие *функции отклика*

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon$$

где y - параметр оптимизации (функция отклика); x_j – факторы (влияющие переменные).

Выбором значений факторов необходимо обеспечить экстремальное значение целевой функции y .

Если имеется один фактор, и в пределах изменения этого фактора функция унимодальная, то используются такие методы поиска экстремума, как *метод деления пополам*, *метод золотого сечения* и др.

Методы оптимизации при числе факторов больше единицы.

Биологические процессы отличаются разнообразием всевозможных связей. Причем в большинстве случаев не удастся процессы описать с помощью дифференциальных уравнений, т.к. не всегда понятна сущность явлений и в связи с этим взаимосвязь между отдельными явлениями.

Поэтому наиболее удобными для математического описания процессов являются полиномиальные модели, дающие возможность учитывать много факторов.

Напомним, что при планировании экспериментов целью исследований является нахождение связи между выходными параметрами (параметрами оптимизации) и множеством входных параметров (факторов), отвлекаясь от механизма явлений, протекающих в объекте.

При этом предполагается, что механизм явлений можно описать дифференциальными уравнениями, но практически из-за сложности процесса это сделать очень трудно. Вследствие этого зависимость между параметрами оптимизации и факторами представляется в виде полинома. Коэффициенты полинома интерпретируются как коэффициенты ряда Тейлора, т.е. как значения частных производных в точке, вокруг которой производится разложение найденной функции, задающей решение неизвестных дифференциальных уравнений.

С точки зрения познания механизма явлений, происходящих в исследуемом объекте, полиномиальная модель представляет незначительный интерес, т.к. по коэффициентам ряда Тейлора невозможно восстановить исходные дифференциальные уравнения, описывающие механизм процесса. С точки зрения оптимизации процессов полиномиальная модель является весьма удобной.

В случае, когда не проводится описание процесса во всей возможной области факторного пространства, для поиска оптимума проводится локальное изучение функции отклика по результатам ряда опытов, специально поставленных около исходной точки. При этом используются эффективные методы оптимизации однофакторных процессов и методы факторного планирования экспериментов.

Движение к экстремуму в многомерном пространстве независимых переменных осуществляется обычно не непрерывно, а шагами.

Анализируя результаты текущих экспериментов и сравнивая их с результатами предыдущих, исследователь принимает решение о дальнейших действиях по поиску оптимума. Экстремальное значение параметра оптимизации определяется путем многократного последовательного продвижения в факторном пространстве.

Существует несколько экспериментальных методов оптимизации, различающихся способом определения направления движения и организацией самого движения. Рассмотрим метод крутого восхождения.

1. Метод крутого восхождения или метод Бокса – Уилсона.

В обычной практике экспериментатор часто использует метод поочередной оптимизации по факторам, известный в специальной литературе как метод Гаусса-Зайделя. В такой классической схеме оптимизации последовательно находят частные экстремумы функции отклика по каждому фактору, когда остальные факторы фиксированы.

Очевидна громоздкость такого метода оптимизации, особенно при большом числе факторов, когда движение к оптимуму сильно усложняется и удлиняется. При сложных формах поверхности отклика такой метод может привести к ошибочным решениям – к застреванию в точках, не соответствующих истинному оптимуму функции.

В настоящее время в теории эксперимента разработаны методы оптимизации, которые свободны от ряда недостатков, свойственных методу Гаусса-Зайделя, и часто требуют меньшего числа опытов. Основными из них, получившими широкое распространение в практике, являются метод крутого восхождения (Бокса- Уилсона) и симплексный метод оптимизации.

Шаговое движение при оптимизации методом крутого восхождения (МКВ) осуществляется в направлении наибольшего изменения функции или в направлении градиента, но в отличие от градиентного метода корректировка направления движения производится не после каждого шага, а после достижения частного экстремума функции отклика (рис. 1)

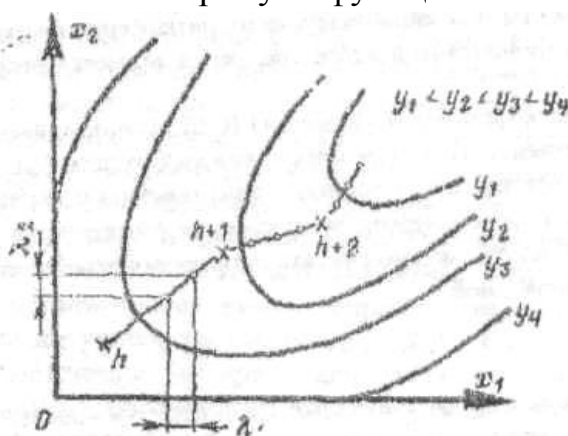


Рис. 1. Схема метода крутого восхождения

Используется модель линейной регрессии. Большими буквами будем обозначать размерные (не кодированные) факторы и коэффициенты модели.

План поиска оптимума следующий:

1. Вблизи исходной точки проводится эксперимент для определения градиента функции отклика. Результаты подвергаются статистическому анализу и определяются коэффициенты уравнения линейной регрессии.

Исходя из этого, можно получить размерный вид уравнения регрессии. Данные коэффициенты совпадают с координатами вектора градиента.

2. Для базового фактора выбирается шаг движения по направлению к градиенту функции отклика. Далее определяются размеры шагов крутого восхождения по остальным переменным. Так как при движении к оптимуму по градиенту все исследуемые факторы должны изменяться пропорционально коэффициентам наклона поверхности отклика.

3. Производятся вычисления значений функции отклика в точках факторного пространства, лежащих на пути к экстремуму от исходной точки, т.е. осуществляется мысленное движение по градиенту к оптимуму.

4. Некоторые расчетные значения функции отклика по уравнению проверяются экспериментально (обычно через 3-5 значения) с целью проверки соответствия аппроксимации процесса найденной зависимостью. Наблюдаемые экспериментальные значения сравниваются с предсказываемыми. Точка, где в реальном опыте получено наиболее благоприятное значение параметра оптимизации, принимается за новую начальную точку следующей серии опытов.

5. Каждый цикл крутого восхождения приближает к экстремуму, где крутизна поверхности отклика больше, поэтому рекомендуется шаг для каждой следующей серии опытов выбирать равным или меньшим, чем предыдущий.

Эксперимент прекращается, когда все или почти все коэффициенты уравнения становятся незначимыми или равными нулю, это говорит о выходе в область оптимума.

2. Симплексный метод оптимизации.

Совокупность $(n+1)$ точек задает в n -мерном факторном пространстве вершины геометрической фигуры, называемой *n -мерным симплексом*. Планы, построенные с помощью подобного рода геометрических фигур, часто называют *симплекс-планами*.

Симплекс представляет собой простейшую фигуру в евклидовом пространстве. В частности, одномерный симплекс - это отрезок прямой (рис. 2а), двумерный - треугольник (рис. 2б), трехмерный - пирамида (тетраэдр) (рис. 2в), n -мерный - выпуклый многогранник с $(n+1)$ вершиной. Рассмотрим

некоторые особые виды симплексов, которые применяются в планировании эксперимента. Симплекс называется *правильным*, если для него расстояние между двумя любыми вершинами есть величина постоянная.

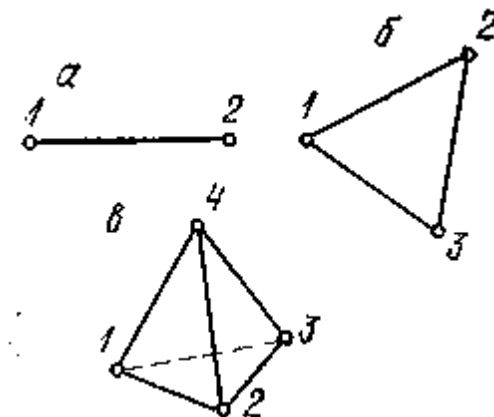


Рис. 2. Симплексы в одномерном, двумерном и трехмерном пространстве

Будем рассматривать простой *симплексный метод поиска*. Он основан на использовании симплекс-планов, а движение к оптимуму осуществляется в n -мерном пространстве последовательным отражением вершин симплекса.

Из любого симплекса можно получить новый, если отбросить одну из его вершин и добавить к оставшимся одну точку. Данное свойство успешно используется для организации шагового перемещения симплекса к оптимуму. С этой целью следует провести измерения отклика Y во всех вершинах симплекса и сопоставить получившиеся решения, выделив среди них наименьшее y . Шаг поиска осуществляется переходом от данного симплекса к новому путем исключения той вершины, где отклик минимален, и определения новой вершины, представляющей собой зеркальное отражение вершины, в которой y минимально, относительно грани, общей обоим симплексам. Многократное отражение вершин с худшими значениями отклика приводит к постепенному перемещению центра симплекса к экстремуму по некоторой ломаной линии. После проведения эксперимента в вершинах исходного симплекса на каждом шаге поиска, очевидно, требуется реализовать всего один дополнительный опыт.

Сформулируем основные пункты, составляющие алгоритм симплексного метода поиска.

1. Построение исходного симплекса в окрестностях начальной точки поиска. Выбор шагов варьирования определяет размер используемого симплекса и в конечном итоге скорость его перемещения в факторном пространстве, поэтому должен производиться с учетом компромисса между необходимой точностью локализации точки экстремума и обеспечением достаточно быстрого его достижения.

2. Реализация опытов в вершинах исходного симплекса и получение соответствующих значений отклика.

3. Сравнение значений отклика между собой и определение вершины с наихудшим значением отклика.

4. Осуществление движения путем перехода от реализованного симплекса к новому. В старом симплексе отбрасывается вершина в которой значение отклика минимально, вместо нее строится точка, зеркальная к ней относительно $(n-1)$ – мерной оставшейся грани симплекса, образованной остальными n точками реализованного симплекса.

Эта точка и оставшаяся грань прежнего симплекса, образуют новый симплекс, сдвинутый в сторону экстремума.

5. Проведение опыта в новой вершине симплекса и получение соответствующего значения отклика, после чего повторяются действия, начиная с пункта 3.

При этом могут возникнуть сложные ситуации для выхода из которых необходимы дополнительные правила.

Прежде всего одно и то же наихудшее значение отклика может наблюдаться сразу в нескольких вершинах симплекса. Тогда вопрос о том, какую из них следует отбросить, лучше всего решить некоторым случайным образом, например с помощью таблицы равномерно распределенных случайных чисел.

Если наблюдение в зеркальной точке нового симплекса снова окажется наихудшим, с наименьшим среди значений отклика во всех вершинах этого симплекса, то формальное применение правила, сформулированного в п. 4, вновь приведет в отброшенную вершину предыдущего симплекса. Тогда вместо движения к оптимуму возникают колебания симплекса относительно одной и той же грани. Чтобы избежать их, следует, вернувшись к старому симплексу, отбросить в нем другую вершину со вторым наименьшим значением отклика, т.е. худшую из оставшихся n вершин - без учета ранее использовавшейся с наименьшим Y . Это правило следует применять многократно путем постепенного перебора вершин данного симплекса до тех пор, пока симплекс перестанет колебаться и вновь начнется его перемещение в факторном пространстве. Если подобного не происходит, нужно попытаться применить названное правило к одному из новых ранее отброшенных как неудачный вариант симплексов.

Критерием останова, с помощью которого обнаруживается достижение экстремальной области, служат факт прекращения поступательного движения симплекса и переход к вращению вокруг определенной вершины. Обычно считают, что симплекс начал вращаться, когда в течение k

последовательных шагов сохраняется неотброшенной хотя бы одна из вершин симплекса. В частности, для $n = 2$ имеем $k = 4$ и, следовательно, если в четырех последовательных положениях симплекса сохраняется одна общая точка, можно говорить о вращательном движении симплекса вокруг этой точки.

В такой ситуации необходимо прежде всего повторить опыт, дающий завышенный результат на предмет выяснения, не обусловлен ли он влиянием случайной помехи. Желательно повторить опыты и во всех соседних с точкой вращения вершинах симплексов, а также продолжить движение, построив в соответствии с обычной процедурой еще несколько новых симплексов. Получение в точке вращения наибольшего значения отклика будет свидетельствовать о достижении экстремума, и процедура поиска должна быть завершена.

Существуют многочисленные модификации данного алгоритма. Важнейшие из них связаны с изменением размера симплекса – с уменьшением по мере продвижения к экстремуму, а также с отказом от требования правильности симплекса. Тогда в зависимости от результатов поиска симплекс деформируется с тем, чтобы обеспечить максимальную скорость движения к экстремуму с учетом конкретных особенностей функции отклика.

Симплексный метод поиска прост в реализации и не требует сложных расчетов. Он достаточно помехоустойчив и эффективен. В принципе данный метод позволяет не только находить значения оптимума, но и отслеживать его положение, если объект оптимизации нестационарен и экстремум медленно смещается. Все это обусловило весьма широкое применение данного метода поиска, причем не только в лабораторных, но и в промышленных условиях.

Основные *достоинства* симплексного метода оптимизации:

1. Метод полностью формализован и дает четкое правило, когда и как надо изменять условия процесса.

2. Использование метода не требует проведения статистического анализа результатов, как при крутом восхождении, и не предъявляет строгих требований к адекватности линейного приближения исследуемой области поверхности отклика.

3. В процессе оптимизации вычисления крайне просты и сводятся на каждом шаге к определению координат одной новой точки.

4. Для принятия решений о постановке нового опыта не требуется точно измерять значения выходного показателя, достаточно их ранжировать, чтобы различить худшие значения.

5. На любом этапе оптимизации легко добавить еще один фактор путем дополнительного введения в симплекс еще одной точке, которая вместе с другими точками образует симплекс, размерностью на единицу больше.

6. В процессе оптимизации движение к оптимуму осуществляется после каждого опыта, а не после серии опытов, что имеет большое психологическое значение, т.к. здесь экспериментатору реже приходится ставить опыты, которые по запланированной программе ожидаются плохими.

7. Метод хорошо приспособлен для оптимизации дрейфующих объектов (например, объектов биотехнологии). В этом случае симплекс будет отслеживать дрейфующий оптимум.

8. Постановка опытов по симплекс-методу не предъявляет строгих требований к регулярности симплекса, т.е. к строгому поддержанию факторов на заданных уровнях.

Недостатки симплексного метода оптимизации:

1. Метод дает ограниченную информацию о поверхности отклика и не оценивает взаимодействия факторов.

2. Все факторы должны быть количественными.

3. Эффективность метода сильно падает с ростом ошибки эксперимента. при этом совершается больше ошибочных движений на пути к оптимуму с заикливанием в отдельных точках.

Для устранения последнего недостатка используют модификации симплексного метода, в котором размер симплекса регулируется в зависимости от уровня шумов – ошибки эксперимента.

Лекция 13. МОДЕЛИ БИОЛОГИЧЕСКИХ СИСТЕМ, ОПИСЫВАЕМЫЕ ОДНИМ ДИФФЕРЕНЦИАЛЬНЫМ УРАВНЕНИЕМ ПЕРВОГО ПОРЯДКА

1. Автономное дифференциальное уравнение первого порядка.
2. Исследование устойчивости стационарных точек дифференциального уравнения.

1. Автономное дифференциальное уравнение первого порядка.

Изучение математических моделей биологических систем начнем с систем первого порядка, которым соответствует одно дифференциальное уравнение первого порядка:

$$\frac{dx}{dt} = f(x, t)$$

Если система автономная, то правая часть уравнений не зависит явно от времени уравнение и имеет вид:

$$\frac{dx}{dt} = f(x) \quad (1)$$

Состояние таких систем в каждый момент времени характеризуется одной единственной величиной - значением переменной x в данный момент времени t .

Рассмотрим плоскость t, x . Решениями уравнения (1) $x(t)$ являются кривые на плоскости t, x , называемые интегральными кривыми (рисунок 1).

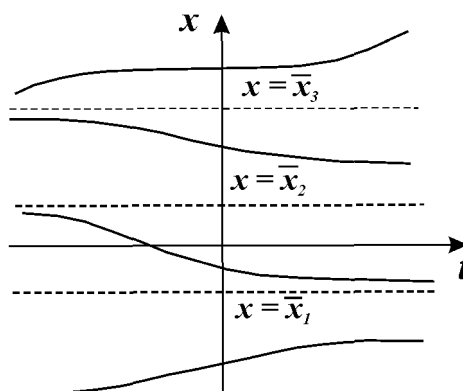


Рис.1 - Интегральные кривые t, x ; $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ - решения уравнения $f(x)=0$

Пусть заданы начальные условия $x = x_0$ при $t = 0$ или, иначе, пусть на плоскости t, x задана точка с координатами (t_0, x_0) . Если для уравнения (1) выполнены условия *теоремы Коши*, то имеется единственное решение уравнения (1), удовлетворяющее этим начальным условиям, и через точку (t_0, x_0) проходит одна единственная интегральная кривая $x(t)$.

Интегральные кривые уравнения (1) не могут пересекаться. Решения уравнения (1) не могут быть периодическими, они монотонны.

Поведение интегральных кривых на плоскости t, x можно установить, не решая в явном виде дифференциального уравнения (1), если известен характер движения изображающей точки на фазовой прямой.

Рассмотрим плоскость t, x , причем, фазовую прямую совместим с осью x . Построим на плоскости t, x точку с абсциссой t и с ординатой, равной смещению изображающей точки по оси x в данный момент времени t . С течением времени в соответствии с уравнением (1) изображающая точка будет двигаться по фазовой прямой (рис. 2), а на плоскости t, x описывать некую кривую. Это будет интегральная кривая уравнения (1).

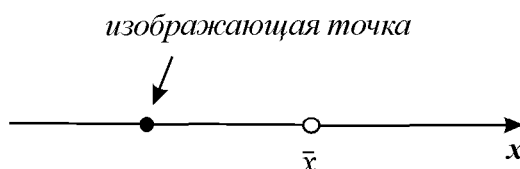


Рис. 2 - Фазовая прямая

Решения одного автономного дифференциального уравнения либо уходят в бесконечность (чего не бывает в реальных системах), либо асимптотически приближаются к стационарному состоянию.

Стационарное состояние (точка покоя, особая точка, состояние равновесия)

В стационарном состоянии значения переменных в системе не меняются со временем. На языке дифференциальных уравнений это означает:

$$\frac{dx}{dt} = 0 \quad (2)$$

Если левая часть уравнения равна нулю, значит равна нулю и его правая часть:

$$f(x) = 0. \quad (3)$$

Корни алгебраического уравнения (3) $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ суть стационарные состояния дифференциального уравнения (1). На плоскости (t, x) прямые $x = \bar{x}_i$ - асимптоты, к которым приближаются интегральные кривые. На фазовой прямой (рис. 2) стационарное состояние x_i - точка, к которой стремится величина x .

Реальные биологические системы испытывают многочисленные флуктуации, переменные при малых отклонениях возвращаются к своим стационарным значениям. Поэтому при построении модели важно знать, устойчивы ли стационарные состояния модели.

Устойчивость состояния равновесия

Каждый имеет интуитивное представление об устойчивости. На рис. 3 в обоих положениях (а и б) шарик находится в равновесии, т.к. сумма сил, действующих на него, равна нулю.

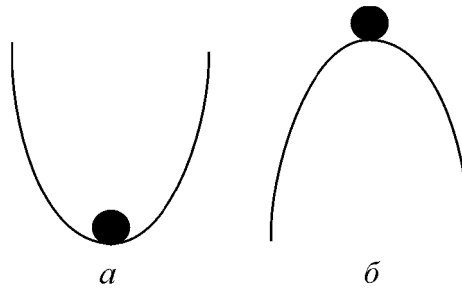


Рис. 3 - К понятию устойчивости состояния равновесия

Чтобы ответить на вопрос: "Какое из этих состояний равновесия устойчиво"? необходимо дать шарiku малое отклонение от состояния равновесия. В случае (а) шарик вернется в начальное положение, а в случае (б) покинет состояние равновесия навсегда.

Устойчивое состояние равновесия можно определить так: если при достаточно малом отклонении от положения равновесия система никогда не уйдет далеко от особой точки, то особая точка будет *устойчивым состоянием равновесия*, что соответствует *устойчивому режиму* функционирования системы.

Строгое математическое определение устойчивости состояния равновесия уравнения $dx/dt = f(x)$ выглядит следующим образом:

Состояние равновесия *устойчиво по Ляпунову*, если задав сколь угодно малое положительное ε , всегда можно найти такое δ , что

$$|x(t) - \bar{x}| < \varepsilon \text{ для } t_0 \leq t < +\infty, \text{ если } |x(t_0) - \bar{x}| < \delta.$$

Иначе говоря, для устойчивого состояния равновесия справедливо утверждение: если и момент времени t_0 отклонение от состояния равновесия мало ($|x(t_0) - \bar{x}| < \delta$), то в любой последующий момент времени $t > t_0$ отклонение решения системы от состояния равновесия будет также мало: $|x(t) - \bar{x}| < \varepsilon$

Другими словами: *стационарное состояние* называется устойчивым, если малые отклонения не выводят систему слишком далеко из окрестности этого стационарного состояния. Пример - шарик в ямке (с трением или без трения).

Стационарное состояние называется асимптотически устойчивым, если малые отклонения от него со временем затухают. Пример - шарик в ямке в вязкой среде.

Стационарное состояние называется неустойчивым, если малые отклонения со временем увеличиваются. Пример: шарик на горке.

Устойчивое стационарное состояние представляет собой простейший тип аттрактора.

Аттрактором называется множество, к которому стремится изображающая точка системы с течением времени (притягивающее множество).

Мы рассмотрим следующие типы аттракторов:

- устойчивая точка покоя;
- предельный цикл - режим колебаний с постоянными периодом и амплитудой (начиная с размерности системы 2).

2. Исследование устойчивости стационарных точек дифференциального уравнения

Метод Ляпунова приложим к широкому классу систем различной размерности, точечным системам, которые описываются обыкновенными дифференциальными уравнениями, и распределенным системам, описываемым уравнениями в частных производных, непрерывным и дискретным.

Рассмотрим метод линеаризации Ляпунова для одного автономного дифференциального уравнения первого порядка. Пусть $\bar{x}$ - стационарное решение уравнения:

$$\frac{dx}{dt} = f(x)$$

Пусть система, первоначально находившаяся в стационарном состоянии, отклонилась от него и перешла в близкую точку с координатой $x = \bar{x} + \xi$, причем $\xi/\bar{x} \ll 1$.

Перейдем в уравнении (1) от переменной x к переменной ξ , т.е. новой переменной будет *отклонение системы от стационарного состояния*.

Получим:

$$\frac{d(\bar{x} + \xi)}{dt} = f(\bar{x} + \xi).$$

Правую часть разложим в ряд Тейлора в точке $\bar{x}$:

$$\frac{d\xi}{dt} = f(\bar{x}) + f'(\bar{x})\xi + \frac{1}{2} f''(\bar{x})\xi^2 + \dots$$

или

$$\frac{d\xi}{dt} = a_1\xi + \frac{1}{2}a_2\xi^2 + \dots$$

где $a_1 = f'(\bar{x}), a_2 = f''(\bar{x}), \dots$

Отбросим члены порядка 2 и выше. Останется линейное уравнение:

$$\frac{d\xi}{dt} = a_1\xi, \quad (4)$$

которое носит название *линеаризованного уравнения* или уравнения первого приближения. Интеграл этого уравнения для $\xi(t)$:

$$\xi(t) = c \cdot \exp(\lambda t), \quad (5)$$

где $\lambda = a_1 = f'(\bar{x})$, c – произвольная постоянная.

Если $\lambda < 0$, то $\xi \rightarrow 0$ и, следовательно, первоначальное отклонение ξ от состояния равновесия со временем затухает. Это означает, по определению, что состояние равновесия устойчиво.

Если же $\lambda > 0$, то $\xi \rightarrow \infty$ и 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.

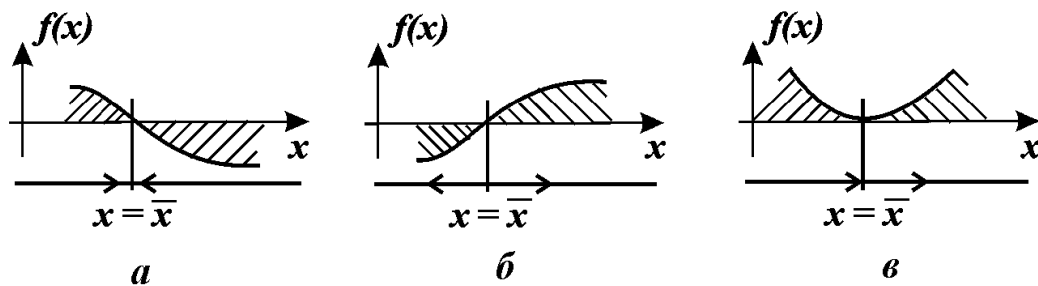
Аналогичные рассуждения проводятся при рассмотрении устойчивости стационарных состояний более сложных динамических систем.

Итак, устойчивость стационарного состояния $\bar{x}$ уравнения $dx/dt = f(x)$ определяется знаком производной правой части в стационарной точке.

В случае одного уравнения вопрос об устойчивости состояния равновесия нетрудно решить, рассматривая график функции $f(x)$.

По определению, в стационарной точке правая часть уравнения (1) - функция $f(x)$ - обращается в нуль.

Здесь возможны три случая (рис. 4 а, б, в).



а - стационарное состояние $\bar{x}$ устойчиво;
б, в - стационарное состояние $\bar{x}$ неустойчиво.

Рис. 4 - Определение устойчивости стационарного состояния по графику функции $f(x)$

1. Вблизи состояния равновесия функция $f(x)$ меняет знак с плюса на минус при возрастании x (рис. 4 а).

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 уменьшается, т.к. 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 (рис. 4 б).

Изображающая точка в области $x < \bar{x}$ и в области $x > \bar{x}$ удаляется от состояния равновесия. Стационарное состояние неустойчиво.

3. Вблизи состояния равновесия функции $f(x)$ не меняет знак (рис. 4 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. Состояние равновесия в случае 3 неустойчивое.

Пример. Рост колонии микроорганизмов.

За время Δt прирост численности равен:

$$\Delta x = R - S,$$

где R - число родившихся и S - число умерших за время Δt особей, пропорциональные этому промежутку времени.

$$R(\Delta t, x) = R(x)\Delta t, \quad S(\Delta t, x) = S(x)\Delta t.$$

В дискретной форме:

$$\Delta x = \{R(x) - S(x)\} \Delta t.$$

Разделив на Δt и переходя к пределу при $\Delta t \rightarrow 0$, получим дифференциальное уравнение:

$$\frac{dx}{dt} = R(x) - S(x). \quad (6)$$

В простейшем случае, когда рождаемость и смертность пропорциональны численности:

$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta x; \alpha - \beta = r, \\ \frac{dx}{dt} &= rx. \end{aligned} \quad (7)$$

Разделим переменные и проинтегрируем:

$$\frac{dx}{x} = r dt, \quad \ln x = rt + C$$

Переходя от логарифмов к значениям переменной x и определяя произвольную постоянную C из начальных условий, получим экспоненциальную форму динамики роста.

$$x = x_0 e^{rt}; \quad x_0 = x(t=0). \quad (8)$$

График функции (8) при положительных (размножение) и отрицательных (вымирание) значениях константы скорости роста представлен на рис. 5.

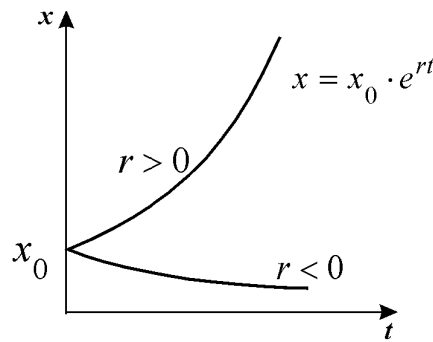


Рис. 5 - Экспоненциальная форма динамики роста численности колонии микроорганизмов в соответствии с системой уравнений (7).

Лишь для ограниченных классов дифференциальных уравнений разработаны аналитические методы решения. Подробно они изучаются в курсах дифференциальных уравнений. Отметим основные из них:

1. Уравнения с разделяющимися переменными решаются в интегралах. К ним относятся оба приведенные выше примера.

2. Линейные дифференциальные уравнения (не обязательно автономные).

3. Некоторые специальные виды уравнений.

Решение линейного уравнения

Линейным дифференциальным уравнением 1-го порядка называют уравнение, линейное относительно искомой функции и ее производной. Оно имеет вид:

$$A \frac{dx}{dt} + Bx + C = 0. \quad (9)$$

Здесь A, B, C - заданные непрерывные функции от t .

Пусть $A \neq 0$ в некотором интервале изменения t . Тогда на него можно разделить все члены уравнения. При этом получим:

$$\frac{dx}{dt} + Px = Q. \quad (10)$$

Если $Q = 0$, уравнение (10) называется однородным, если $Q \neq 0$ - неоднородным. Решим сначала однородное уравнение:

$$\frac{dx}{x} = -P dt.$$

Общее решение линейного однородного уравнения имеет вид:

$$x = C \exp(-\int P dt). \quad (11)$$

Чтобы найти решение неоднородного уравнения, применим метод вариации постоянной. Будем считать C неизвестной функцией t . Подставляя правую часть выражения (11) в уравнение (10), имеем:

$$\frac{dC}{dt} e^{-\int P dt} - C P e^{-\int P dt} + C P e^{-\int P dt} = Q, \quad \frac{dC}{dt} = Q e^{\int P dt}$$

Теперь C находим интегрированием: $C = \int Qe^{\int Pdt} dt + C_1$. Здесь C_1 - произвольная постоянная.

Итак, общее решение линейного неоднородного уравнения первого порядка:

$$x = \left(\int Qe^{\int Pdt} dt + C_1 \right) e^{-\int Pdt}. \quad (12)$$

Таким образом, решение уравнения (10) представляет собой сумму двух слагаемых:

- 1) общее решение однородного уравнения (11) и
- 2) частное решение неоднородного уравнения, которое получается из общего решения, если $C_1 = 0$.

Рассмотрим еще один пример, который относится к классическим моделям математической экологии. *Логистическое уравнение* было предложено Ферхюльстом в 1838 г. Оно имеет вид:

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K} \right). \quad (13)$$

Это уравнение обладает двумя важными свойствами. При малых t численность x возрастает, при больших - приближается к определенному пределу K .

Уравнение (13) можно решить аналитически. Ход решения следующий. Произведем разделение переменных:

$$\frac{Kdx}{x(K-x)} = rdt. \quad (14)$$

Представим левую часть в виде суммы и проинтегрируем:

$$\left(\frac{1}{x} + \frac{1}{K-x} \right) dx = rdt. \quad \ln x - \ln(K-x) = rt + \ln C.$$

Переходя от логарифмов к переменным, получим:

$$\frac{x}{(K-x)} = Ce^{rt}. \quad (15)$$

Здесь C - произвольная постоянная, которая определяется начальным значением численности x_0 :

$$x(t=0) = x_0, C = \frac{x_0}{(K-x_0)}.$$

Подставим это значение C в формулу (17):

$$\frac{x}{(K-x)} = \frac{x_0}{(K-x_0)} e^{rt}.$$

Отсюда получим решение - зависимость численности от времени:

$$x(t) = \frac{x_0 K e^{rt}}{(K - x_0 + x_0 e^{rt})}. \quad (16)$$

График функции (16) при разных начальных значениях численности популяции представлен на рис. 6.

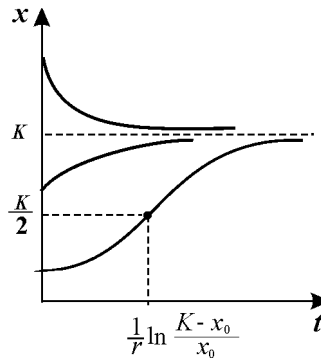


Рис. 6 - Динамика численности в логистической модели (18) при разных начальных значениях численности.

Если начальное значение $x_0 < K/2$, кривая роста имеет точку перегиба. Если $x_0 > K$, то численность со временем убывает.

В приведенных примерах в правой части уравнения стоят полиномы первой и второй степени. Если и правой части - более сложная нелинейная функция, алгебраическое уравнение для стационарных значений может иметь несколько корней. Какое из этих решений реализуется в этом случае, будет зависеть от начальных условий.

В дальнейшем мы, как правило, не будем искать аналитическое решение для наших моделей. Для более сложных нелинейных уравнений это и невозможно. Однако важные заключения относительно свойств моделей можно сделать и на основании качественного их исследования, в первую очередь, путем исследования устойчивости стационарных значений численности, парных состояний и типов поведения системы вблизи этих состояний. При этом 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. Для описания более сложного поведения необходимо либо переходить к системам большей размерности (2, 3 порядка и выше), либо вводить время в явном виде в правую часть уравнения.

Лекция 14. МОДЕЛИ БИОЛОГИЧЕСКИХ СИСТЕМ, ОПИСЫВАЕМЫХ СИСТЕМАМИ ДВУХ АВТОНОМНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

1. Системы двух автономных дифференциальных уравнений.
2. Исследование устойчивости стационарных состояний моделей биологических систем, описываемых системами двух автономных дифференциальных уравнений.

1. Системы двух автономных дифференциальных уравнений

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. Рассмотрим систему двух автономных обыкновенных дифференциальных уравнений общего вида

$$\frac{dx}{dt} = P(x, y), \quad \frac{dy}{dt} = Q(x, y) \quad (1)$$

$P(x, y)$, $Q(x, y)$ - 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) 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 (1), так что каждому состоянию системы соответствует пара значений переменных (x, y) .

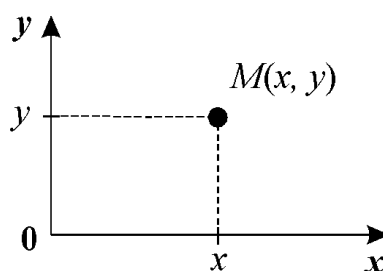


Рис.1 – Изображающая точка на фазовой плоскости

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 носит название фазовой плоскости и изображает совокупность всех состояний системы. Точка $M(x, y)$ называется изображающей или представляющей точкой.

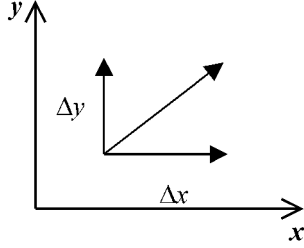
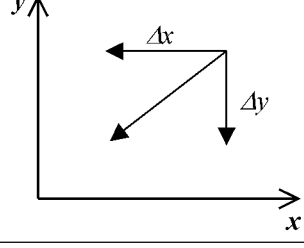
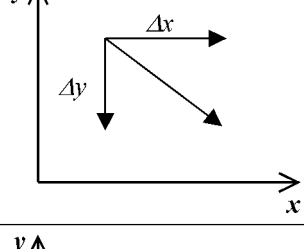
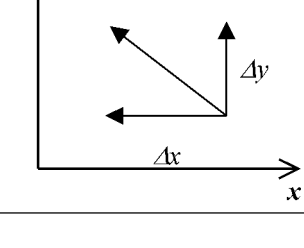
Пусть в начальный момент времени $t=t_0$ координаты изображающей точки $M_0(x(t_0), y(t_0))$. В каждый следующий момент времени t изображающая точка будет смещаться в соответствии с изменениями значений переменных $x(t), y(t)$. Совокупность точек $M(x(t), y(t))$ на фазовой плоскости, положение которых соответствует состояниям системы в процессе изменения во времени переменных $x(t), y(t)$ согласно уравнениям (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 (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 Δx и Δy из выражений:

$$\Delta x = P(x, y) \Delta t, \quad \Delta y = Q(x, y) \Delta t.$$

Направление вектора dy/dx в точке (x, y) зависит от знака функций $P(x, y), Q(x, y)$ и может быть задано таблицей:

$P(x, y) > 0, Q(x, y) > 0$	
$P(x, y) < 0, Q(x, y) < 0$	
$P(x, y) > 0, Q(x, y) < 0$	
$P(x, y) < 0, Q(x, y) > 0$	

Задача построения векторного поля упрощается, если получить выражение для фазовых траекторий в аналитическом виде. Для этого разделим второе из уравнений системы (1) на первое:

$$\frac{dy}{dx} = \frac{Q(x, y)}{P(x, y)} . \quad (2)$$

Решение этого уравнения $y = y(x, c)$, или в неявном виде $F(x, y) = c$, где c – постоянная интегрирования, дает семейство интегральных кривых уравнения (2) – фазовых траекторий системы (1) на плоскости x, y .

Метод изоклин

Для построения фазового портрета пользуются методом изоклин – на фазовой плоскости наносят линии, которые пересекают интегральные кривые под одним определенным углом. Уравнение изоклин легко получить из (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 (2) величину A получим уравнение изоклин:

$$A = \frac{Q(x, y)}{P(x, y)}. \quad (3)$$

Уравнение (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. В ней одновременно обращаются в нуль производные по времени переменных x и y .

$$\left. \frac{dx}{dt} \right|_{\bar{x}, \bar{y}} = P(\bar{x}, \bar{y}) = 0, \quad \left. \frac{dy}{dt} \right|_{\bar{x}, \bar{y}} = Q(\bar{x}, \bar{y}) = 0.$$

Таким образом, в особой точке скорости изменения переменных равны нулю. Следовательно, особая точка дифференциальных уравнений фазовых траекторий (2) соответствует стационарному состоянию системы (1), а ее координаты – суть стационарные значения переменных x , y .

Особый интерес представляют главные изоклины:

$dy/dx=0$, $P(x, y) = 0$ – изоклина горизонтальных касательных и

$dy/dx=\infty$, $Q(x, y) = 0$ – изоклина вертикальных касательных.

Построив главные изоклины и найдя точку их пересечения (x, y) , координаты которой удовлетворяют условиям:

$$P(\bar{x}, \bar{y}) = 0, \quad Q(\bar{x}, \bar{y}) = 0.$$

мы найдем тем самым точку пересечения всех изоклин фазовой плоскости, в которой направление касательных к фазовым траекториям неопределенно. Это – особая точка, которая соответствует стационарному состоянию системы (рис. 2).

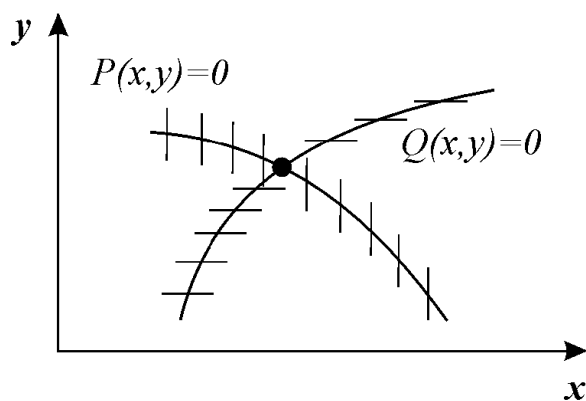


Рис. 2 – Пересечение главных изоклин на фазовой плоскости

Система (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.

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.

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 стационарного состояния. Применительно к системе из двух уравнений определение устойчивости выглядит следующим образом.

Состояние равновесия устойчиво, если для любой заданной области отклонений от состояния равновесия (ϵ) можно указать область $\delta(\epsilon)$, окружающую состояние равновесия и обладающую тем свойством, что ни одна траектория, которая начинается внутри области δ , никогда не достигнет границы ϵ (рис. 3).

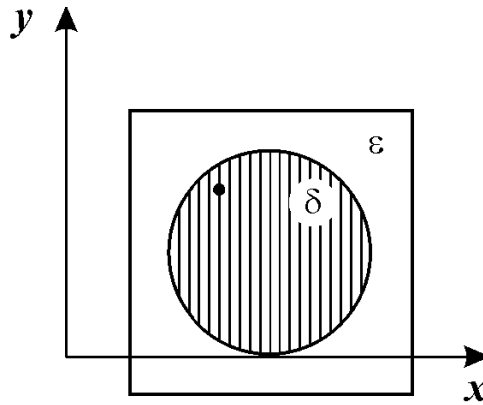


Рис.3 - Иллюстрация к определению устойчивости на плоскости (x, y)

Для большого класса систем – грубых систем – характер 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.

Линейные системы

Рассмотрим систему двух линейных уравнений:

$$\frac{dx}{dt} = ax + by, \quad \frac{dy}{dt} = cx + dy. \quad (4)$$

Здесь a, b, c, d - константы, x, y - декартовы координаты на фазовой плоскости.

Общее решение будем искать в виде:

$$x = Ae^{\lambda t}, \quad y = Be^{\lambda t}. \quad (5)$$

Подставим эти выражения в (4) и сократим на $e^{\lambda t}$:

$$\lambda A = aA + bB, \quad \lambda B = cA + dB \quad (6)$$

Алгебраическая система уравнений (6) с неизвестными A, B имеет ненулевое решение лишь в том случае, если ее определитель, 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:

$$\lambda^2 + (a + d)\lambda + (ad - bc) = 0. \quad (7)$$

Решение этого уравнения дает значения показателя $\lambda_{1,2}$, при которых возможны ненулевые для A и B решения уравнения (6). Эти значения суть

$$\lambda_{1,2} = \frac{a + d}{2} \pm \sqrt{\frac{(a + d)^2 - 4(ad - bc)}{4}}. \quad (8)$$

Если подкоренное выражение отрицательно, то $\lambda_{1,2}$ комплексно сопряженные числа. Предположим, что оба корня уравнения (7) имеют отличные от нуля действительные части и что нет кратных корней. Тогда общее решение системы (4) можно представить в виде линейной комбинации экспонент с показателями λ_1, λ_2 :

$$x = c_{11}e^{\lambda_1 t} + c_{12}e^{\lambda_2 t}, \quad y = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t} \quad (9)$$

Рассмотрим различные случаи, которые могут здесь представиться.

Корни λ_1, λ_2 – действительны и одного знака

Выясним направление движений изображающей точки вдоль фазовых траекторий.

Если λ_1, λ_2 – отрицательны, то, изображающая точка приближается к особой точке, никогда, однако, не достигая ее.

Такая особая точка, через которую проходят интегральные кривые, подобно тому, как семейство парабол проходит через начало координат, носит название узла.

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 особой точке (состоянию равновесия). Это устойчивый узел. Если же $\lambda_1, \lambda_2 > 0$, то с течением времени и изображающая точка удаляется от состояния равновесия. В этом случае особая точка – неустойчивый узел.

Корни λ_1, λ_2 – действительны и разных знаков.

Рассмотрим характер движения изображающей точки по фазовым траекториям вблизи состояния равновесия. Пусть, например, $\lambda_1 > 0, \lambda_2 < 0$. Тогда изображающая точка, помещенная на оси ξ , будет удаляться от начала координат, а помещенная на оси η – будет неограниченно приближаться к началу координат, не достигая его за конечное время. Где бы ни находилась изображающая 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 в начале 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. Если считать, что все начальные состояния системы на фазовой плоскости равновероятны, то вероятность такого начального состояния, которое соответствует движению по

направлению к особой точке, равна нулю. Поэтому всякое реальное движение будет удалять систему от состояния равновесия.

Корни λ_1, λ_2 – комплексные сопряженные

В этом случае при действительных x и y мы будем иметь комплексные сопряженные ξ, η .

На фазовой плоскости мы имеем дело с семейством спиралей, каждая из которых имеет асимптотическую точку в состоянии равновесия. Особая точка, которая является асимптотической точкой всех интегральных кривых, имеющих вид спиралей, вложенных друг в друга, называется фокусом.

Рассмотрим характер движения изображающей точки по фазовым траекториям.

Пусть $\operatorname{Re} \lambda < 0$. Изображающая точка тогда непрерывно приближается к особой точке, не достигая ее в конечное время. Это означает, что фазовые траектории представляют собой скручивающиеся спирали и соответствуют затухающим колебаниям переменных. Это – устойчивый фокус.

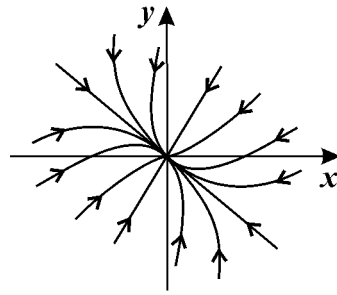
В случае устойчивого фокуса, как и в случае устойчивого узла, выполнено не только условие Ляпунова, но и более жесткое требование. Именно, при любых начальных отклонениях система по прошествии времени вернется как угодно близко к положению равновесия. Такая устойчивость, при которой начальные отклонения не только не нарастают, но затухают, стремясь к нулю, называют абсолютной устойчивостью.

Если $\operatorname{Re} \lambda > 0$, то изображающая точка удаляется от особой точки, и мы имеем дело с неустойчивым фокусом.

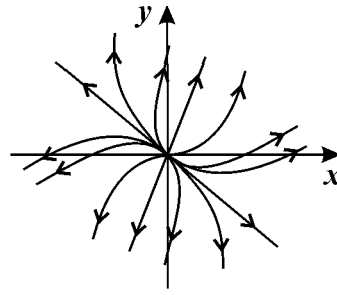
Рассмотрим теперь случай, когда $\operatorname{Re} \lambda = 0$. Фазовыми траекториями на плоскости x, y будут эллипсы:

Таким образом, при $\operatorname{Re} \lambda = 0$ через особую точку 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.

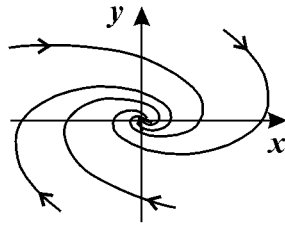
Таким образом, возможны шесть типов состояния равновесия в зависимости от характера корней характеристического уравнения (7). Вид фазовых траекторий на плоскости x, y для этих шести случаев изображен на рис. 4.



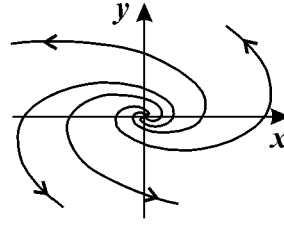
Устойчивый узел.
(λ_1, λ_2 действительны и отрицательны)



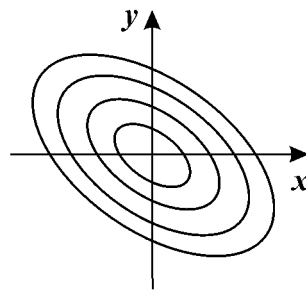
Неустойчивый узел.
(λ_1, λ_2 действительны и положительны)



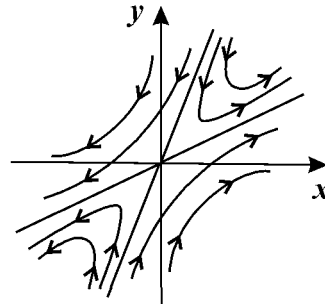
Устойчивый фокус
(λ_1, λ_2 - комплексны,
 $\text{Re } \lambda_{1,2} < 0$)



Неустойчивый фокус
(λ_1, λ_2 - комплексны,
 $\text{Re } \lambda_{1,2} > 0$)



Центр.
(λ_1, λ_2 - чисто мнимые)



Седло.
(λ_1, λ_2 - действительны и разных знаков)

Рис. 4. - Типы фазовых портретов в окрестности стационарного состояния для системы линейных уравнений (4)

Пять типов состояния равновесия грубые, их характер не изменяется при достаточно малых изменениях правых частей уравнений (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.

Бифуркационная диаграмма

Введем обозначения:

$$\sigma = -(a+d); \quad \Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix}. \quad (10)$$

Тогда характеристическое уравнение запишется в виде:

$$\lambda^2 + \sigma\lambda + \Delta = 0. \quad (11)$$

Рассмотрим плоскость с прямоугольными декартовыми координатами σ , Δ и отметим на ней области, соответствующие тому или иному типу состояния равновесия, который определяется характером корней характеристического уравнения

$$\lambda_{1,2} = \frac{-\sigma \pm \sqrt{\sigma^2 - 4\Delta}}{2}. \quad (12)$$

Условием устойчивости состояния равновесия будет наличие отрицательной действительной части у λ_1 и λ_2 . Необходимое и достаточное условие этого – выполнение неравенств $\sigma > 0$, $\Delta > 0$. На диаграмме (рис.5) этому условию соответствуют точки, расположенные в первой четверти плоскости параметров. Особая точка будет фокусом, если λ_1 и λ_2 комплексны. Этому условию соответствуют те точки плоскости, для которых $\sigma^2 < 4\Delta$, т.е. точки между двумя ветвями параболы $\sigma^2 = 4\Delta$. Точки полуоси $\sigma = 0$, $\Delta > 0$, соответствуют состояниям равновесия типа центр. Аналогично, λ_1 и λ_2 – действительны, но разных знаков, т.е. особая точка будет седлом, если $\Delta < 0$, и т.д. В итоге мы получим диаграмму разбиения плоскости параметров σ , Δ на области, соответствующие различным типам состояния равновесия.

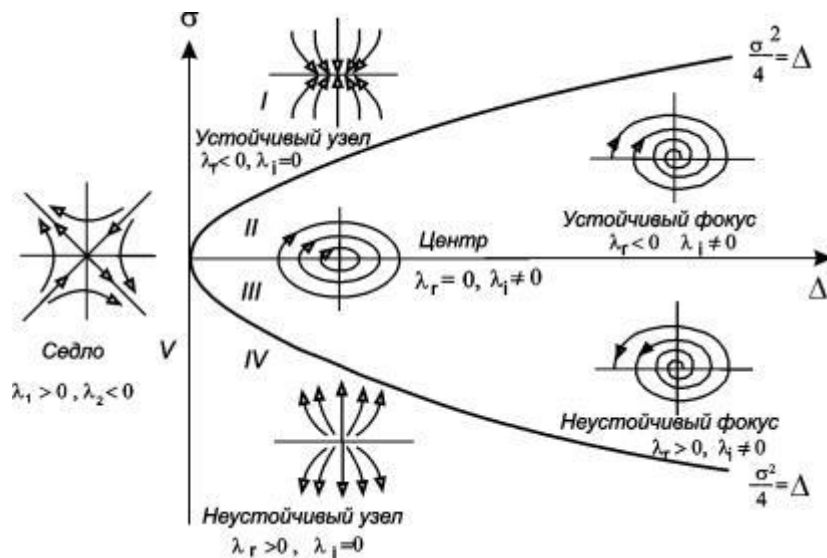


Рис. 5 - Бифуркационная диаграмма для системы линейных уравнений (4)

Если коэффициенты линейной системы a , b , c , d зависят от некоторого параметра, то при изменении этого параметра будут меняться и величины σ , Δ . При переходе через границы характер фазового портрета качественно

меняется. Поэтому такие границы называются бифуркационными – по разные стороны от границы система имеет два топологически 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 топология фазового пространства при этом не меняется.

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.

Пусть биологическая система описывается системой двух автономных дифференциальных уравнений второго порядка общего вида:

$$\frac{dx}{dt} = P(x, y), \quad \frac{dy}{dt} = Q(x, y) \quad (13)$$

Стационарные значения переменных системы определяются из алгебраических уравнений:

$$P(\bar{x}, \bar{y}) = 0, \quad Q(\bar{x}, \bar{y}) = 0 \quad (14)$$

Стационарные состояния соответствуют особым точкам дифференциального уравнения первого порядка, определяющего интегральные кривые:

$$\frac{dy}{dx} = \frac{Q(x, y)}{P(x, y)} \quad (15)$$

Математический анализ поведения траекторий этой системы на фазовой плоскости связан с именами французского математика Анри Пуанкаре 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 характер поведения переменных при некотором небольшом отклонении системы от состояния равновесия. Введем вместо переменных x, y новые независимые переменные ξ, η , определив их как смещения относительно равновесных значений переменных

$$x = \bar{x} + \xi, \quad y = \bar{y} + \eta \quad (16)$$

Подставив эти выражения в (13), получим:

$$\frac{d\bar{x}}{dt} + \frac{d\xi}{dt} = P(\bar{x} + \xi, \bar{y} + \eta), \quad \frac{d\bar{y}}{dt} + \frac{d\eta}{dt} = Q(\bar{x} + \xi, \bar{y} + \eta) \quad (17)$$

$$\frac{d\bar{x}}{dt} = \frac{d\bar{y}}{dt} = 0, \text{ так как } \bar{x}, \bar{y} - \text{координаты особой точки.}$$

Предположим, что функции P и Q непрерывны и имеют непрерывные производные не ниже первого порядка. Тогда мы можем разложить правые части уравнений (17) в ряд Тейлора по переменным ξ, η :

$$\frac{d\xi}{dt} = P(\bar{x}, \bar{y}) + a\xi + b\eta + \dots, \quad \frac{d\eta}{dt} = Q(\bar{x}, \bar{y}) + c\xi + d\eta + \dots \quad (18)$$

где

$$a = P'_x(\bar{x}, \bar{y}), \quad b = P'_y(\bar{x}, \bar{y}), \quad c = Q'_x(\bar{x}, \bar{y}), \quad d = Q'_y(\bar{x}, \bar{y}). \quad (19)$$

Учтем, что по определению особой точки

$$P(\bar{x}, \bar{y}) = 0, \quad Q(\bar{x}, \bar{y}) = 0.$$

и отбросим в уравнениях (18) нелинейные члены. Получим систему линейных уравнений с постоянными коэффициентами - систему первого приближения:

$$\frac{d\xi}{dt} = a\xi + b\eta, \quad \frac{d\eta}{dt} = c\xi + d\eta. \quad (20)$$

Решение этой системы было рассмотрено ранее. Оно определяется корнями характеристического уравнения системы:

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0 \quad (21)$$

Ляпунов показал, что в случае, если оба корня уравнения (21):

$$\lambda_{1,2} = \frac{(a + d) \pm \sqrt{(a + d)^2 - 4(ad - bc)}}{2}. \quad (22)$$

имеют отличные от нуля действительные части, исследование уравнений первого приближения (20) всегда дает правильный ответ на вопрос о типе устойчивости состояния равновесия в системе (13). А именно:

- если оба корня имеют отрицательную действительную часть и, следовательно, все решения уравнений первого приближения (20) затухают, то состояние равновесия устойчиво;

- если хотя бы один корень имеет положительную действительную часть, то есть система (20) имеет нарастающие решения, то состояние равновесия неустойчиво.

Если действительные части обоих корней характеристического уравнения равны нулю или если один корень равен нулю, а другой отрицателен, то уравнения (20) 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 (18).

В случае, когда оба корня характеристического уравнения имеют отличные от нуля действительные части (грубые системы), уравнение первого приближения определяют 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 равновесий удобно пользоваться диаграммой, изображенной на рис. 5. Для системы (13):

$$\sigma = P'_x(\bar{x}, \bar{y}) + Q'_y(\bar{x}, \bar{y}) \quad (23)$$

$$\Delta = \begin{vmatrix} P'_x(\bar{x}, \bar{y}) & Q'_x(\bar{x}, \bar{y}) \\ P'_y(\bar{x}, \bar{y}) & Q'_y(\bar{x}, \bar{y}) \end{vmatrix}. \quad (24)$$

Грубым состояниям равновесия соответствуют все точки плоскости параметров σ , Δ 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 (13) 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 (20). В отличие от линейных систем, уже при небольших изменений в правых частях содержащихся там нелинейных членов может произойти качественное изменение фазового портрета - бифуркация.

Лекция 15. МУЛЬТИСТАЦИОНАРНЫЕ СИСТЕМЫ

1. Силовое и параметрическое переключение систем.
2. Примеры триггерных систем.

1. Силовое и параметрическое переключение систем.

Важная особенность биологических систем – переключение из одного режима функционирования в другой. Приведем простые примеры переключения процессов в живых системах:

- Сон и бодрствование – это разные типы метаболизма. Переключение происходит периодически и синхронизируется геофизическим ритмом.

- У большинства насекомых один и тот же организм может существовать в виде гусеницы, куколки, бабочки. Переключение происходит последовательно в соответствии с генетической программой.

- Дифференцировка тканей – клетки получают путем деления из одного типа клеток, но впоследствии каждая выполняет свои функции.

На фазовой плоскости триггерной системе в простейшем случае соответствует два или несколько устойчивых стационарных решений, разделенных сепаратрисами. Напомним, что все особые точки (устойчивые и седло) лежат на пересечении главных изоклин – изоклин вертикальных и горизонтальных касательных.

На рис.1 представлен относительно простой фазовый портрет триггерной системы, описывающей явление конкуренции двух одинаковых видов.

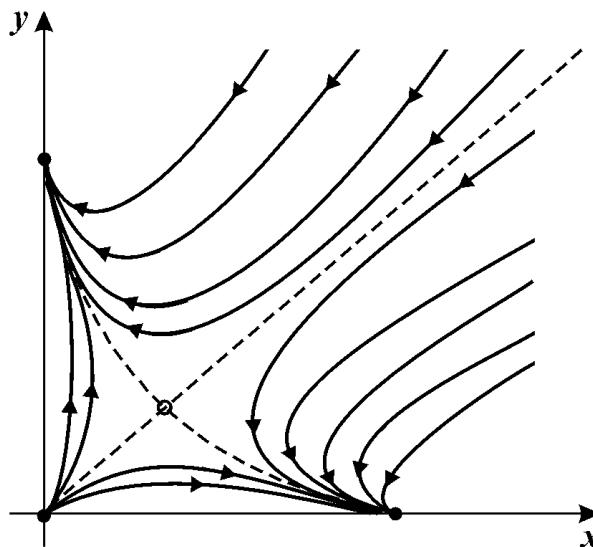


Рис.1 - Фазовый портрет триггерной системы, описывающей явление конкуренции между двумя одинаковыми видами

Соответствующая система уравнений имеет вид:

$$\frac{dx_1}{dt} = x_1 - x_1x_2 - ax_1^2, \quad \frac{dx_2}{dt} = x_2 - x_1x_2 - ax_2^2. \quad (1)$$

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. $x_1 = \frac{1}{a}, \quad x_2 = 0$ – устойчивый узел;
4. $x_1 = 0, \quad x_2 = \frac{1}{a}$ – устойчивый узел.

Бистабильная система может иметь гораздо более сложную структуру фазового портрета. Пример такой системы – движение шарика в ложбине с двумя лунками в присутствии трения (Д.С. Чернавский).

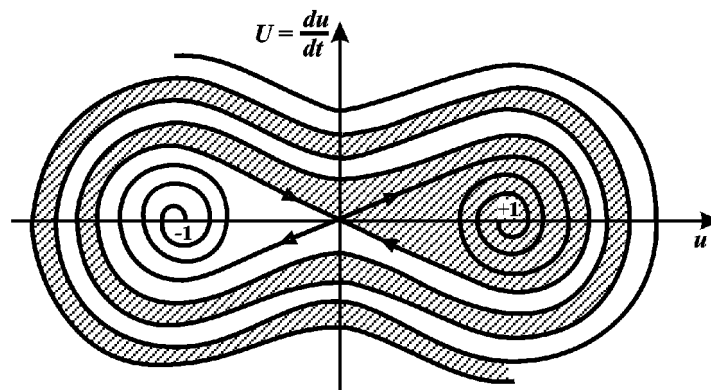


Рис. 2. Фазовый портрет системы (2) (шарик в ложбине с двумя лунками). Темным обозначена область притяжения стационарного состояния (+1)

Система описывается уравнениями:

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = -ay + a(x - x^3). \quad (2)$$

В такой системе три стационарных состояния. Состояние $x=y=0$ – седло. Два других стационарных состояния – 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 выше примеров, в триггерных системах и поведение во времени и стационарное решение зависит не только от параметров, но и от начальных условий.

Способы переключения триггера

Слово триггер означает переключатель. Встает вопрос, как можно переключить триггер из одного в другое стационарное состояние?

Рассмотрим фазовый портрет системы, обладающей двумя устойчивыми стационарными состояниями (рис. 3). Здесь a , c – устойчивые стационарные состояния, b – седло.

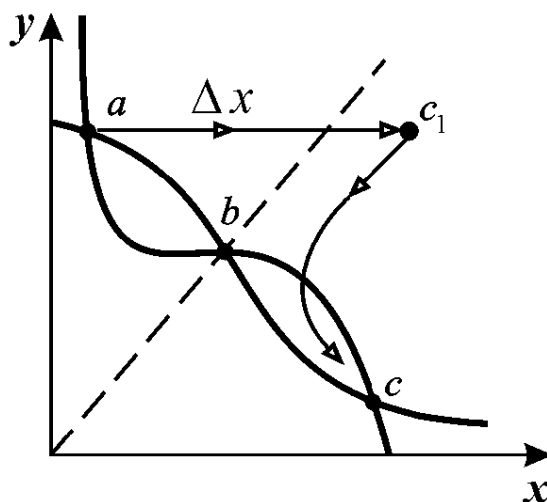


Рис. 3 - Триггерная система. Жирными линиями показаны главные изоклины. Пунктирной линией обозначена сепаратриса, отделяющая области влияния двух устойчивых стационарных состояний a и c . Двойная стрелка показывает процесс силового переключения триггера.

Если начальное положение изображающей точки расположено левее сепаратрисы седла (пунктирная линия), 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 .

1. *Силовое переключение.* Можно изменить значения концентраций (например, добавить определенное количество вещества x_1 , так что система «перепрыгнет» через сепаратрису, например 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 рис. 3 силовое (специфическое) переключение показано двойной стрелкой. Кинетика переменных при таком переключении показана на рис. 4.

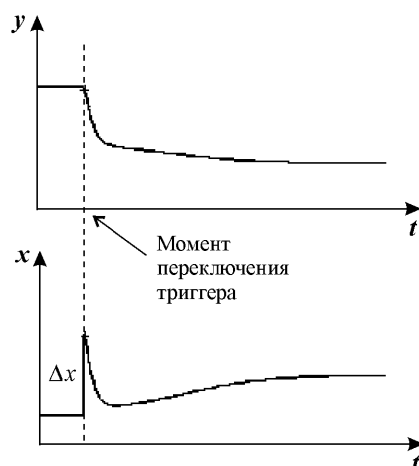


Рис. 4. Поведение переменных во времени при силовом переключении после добавления в систему вещества x в количестве, достаточном для переключения системы из режима a в режим c (смотри рис. 3).

2. Параметрическое переключение.

Другой – неспецифический способ переключения показан на рис. 5, 6.

При таком способе переключения 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 фазовый портрет претерпевает последовательность превращений, показанных на рис. 5 ($a - z$). На стадии ($в$) устойчивый узел (a) и седло (b) сливаются в одну полуустойчивую точку седло-узел. На стадии ($с$) в системе остается лишь одно устойчивое стационарное состояние, к которому и сходятся все фазовые траектории.

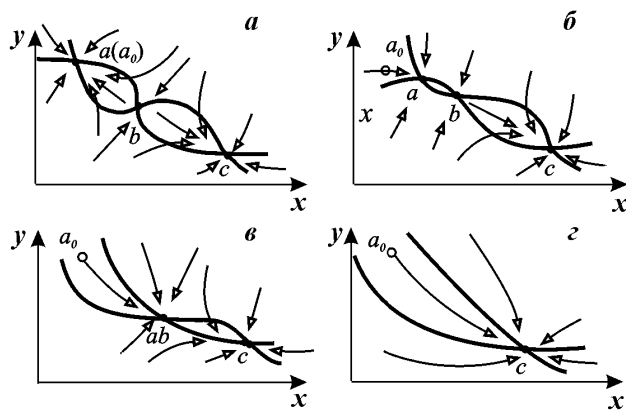


Рис. 5 - Параметрическое переключение триггера. Последовательные стадии трансформации фазового портрета. Стрелками обозначено направление фазовых траекторий.

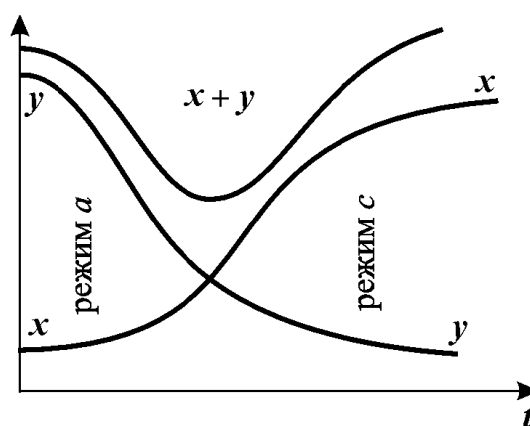


Рис. 6 – Кинетика изменения переменных в процессе параметрического переключения триггера

Тогда система, находившаяся в начале процесса переключения в стационарном режиме a , в результате параметрического переключения окажется в области притяжения единственного устойчивого стационарного режима c , куда с течением времени и перейдет (рис. 6).

Параметрический способ переключения реализуется при изменении любой генетической программы, он может также иметь место при изменении внешних условий, приводящих к изменению управляющего параметра системы.

2. Примеры триггерных систем.

Модели отбора

Как мы видели выше, 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.

Существующий генетический код не связан с физико-химическими свойствами аминокислот и кодонов. Число равноправных кодов - $20!$ а реализован только один. Вероятность случайного возникновения именно существующего кода крайне мала.

На вопрос «Как же произошел отбор»? возможно несколько ответов:

Кастлер: начальный код возник случайно, другие комбинации не успели возникнуть.

Эйген: возникло несколько разных кодов, но отобраны наилучшие.

Чернавский: произошел отбор одного из равноправных.

Модель образования единого кода

Можно выделить четыре стадии эволюции формирования единого генетического кода.

1. Образование первичного бульона.
2. Образование белково-нуклеотидных комплексов, способных к авторепродукции.
3. Образование единого кода в результате отбора.
4. Образование разных видов на основе единого кода.

Рассмотрим 3-й этап. Существует три возможных механизма:

а) Один объект возникает раньше других и развивается так быстро, что другие не успевают возникнуть.

б) В результате конкуренции между объектами с различными свойствами выжили и отобрались наилучшие, обеспечив наибольшую скорость епликации.

в) В результате антагонистического взаимодействия между равноправными объектами (с одинаковой скоростью репликации), но разными последовательностями нуклеотидов, выживает один вид объектов.

Действие каждого из этих механизмов может привести к возникновению совокупности полностью одинаковых объектов, в которой одной последовательности нуклеотидов соответствует одна последовательность аминокислот - однозначный код.

Отбор одного из равноправных

Общая модель такого отбора имеет вид:

$$\frac{dX_i}{dt} = aX_i - \gamma \sum_{j=1, j \neq i}^N X_i X_j; i = 1, 2, \dots, N \quad (3)$$

Здесь a - эффективный коэффициент репродукции, γ - вероятность гибели в результате встречи.

Пусть $N=2$, $X_1 = x$, $X_2 = y$. Система уравнений имеет вид:

$$\frac{dx}{dt} = ax - \gamma xy; \quad \frac{dy}{dt} = ay - \gamma xy. \quad (4)$$

Стационарные решения находятся из алгебраических уравнений, полученных приравнением правых частей нулю.

$$ax - \gamma xy = 0; \quad ay - \gamma xy = 0. \quad (5)$$

Система имеет два стационарных решения:

$$\begin{aligned} 1) \quad \bar{x} &= 0; \quad y = 0. \\ 2) \quad \bar{x} &= \frac{a}{\gamma}; \quad \bar{y} = \frac{a}{\gamma}. \end{aligned} \quad (6)$$

Характеристический определитель имеет вид

$$\begin{vmatrix} a - \gamma y - \lambda & -\gamma x \\ -\gamma y & a - \gamma x - \lambda \end{vmatrix} = 0. \quad (7)$$

Характеристическое уравнение:

$$\lambda^2 - \lambda(2a - \gamma x - \gamma y) + (a - \gamma y)(a - \gamma x) - \gamma^2 xy = 0$$

Для второго – нетривиального симметричного стационарного состояния характеристический определитель системы имеет вид:

$$\lambda^2 + a^2 - 2a^2 = 0 \quad (8)$$

Выражения для характеристических чисел находятся из уравнения:

$$\lambda^2 - a^2 = 0; \quad \lambda_{1,2} = \pm a. \quad (9)$$

Корни положительны и разных знаков. Это означает, что симметричное стационарное состояние представляет собой седло.

Аналогичный анализ показывает, что нулевая особая точка представляет собой неустойчивый узел.

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.

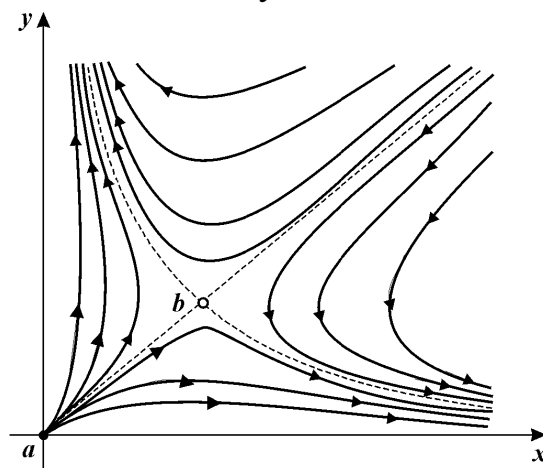


Рис. 7 - Фазовый портрет системы (4), описывающей отбор одного из двух равноправных видов в отсутствие ограничений роста. a (начало координат) – неустойчивый узел, b – седло

Биологический смысл модели.

Модель (4) 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 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. То же – при небиологическом синтезе. По-видимому, и первичный бульон был рацемической смесью.

Рассмотренная модель описывает выживание одних 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 (4) не может описывать 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 виде ферхюльстовских членов. Тогда мы придем к модели (1). 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 (9). График этой функции приведен на рис. 3.

$$a = \frac{a_0 S}{K_s + S}. \quad (10)$$

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:

$$\frac{dS}{dt} = -\alpha \frac{a_0 S}{K_s + S} (X + Y) + v. \quad (11)$$

Здесь $\alpha > 1$ - экономический коэффициент - указывает, сколько субстрата идет на образование единицы биомассы.

Уравнения для концентраций объектов типа x и y :

$$\begin{aligned} \frac{dX}{dt} &= \frac{a_0 S}{K_s + S} X - \beta X - \gamma XY, \\ \frac{dY}{dt} &= \frac{a_0 S}{K_s + S} Y - \beta Y - \gamma XY. \end{aligned} \quad (12)$$

Введем безразмерные переменные:

$$\begin{aligned} t' = \beta t, \quad x = \frac{\gamma X}{\beta}, \quad y = \frac{\gamma Y}{\beta}, \quad z = \frac{\gamma S}{\beta}, \\ v' = \frac{\gamma v}{\beta^2}, \quad f(z) = \frac{a_0 z}{K_z + z}, \quad K_z = \frac{\gamma K_s}{\beta}. \end{aligned} \quad (13)$$

Система в безразмерном виде:

$$\begin{aligned} \frac{dx}{dt} &= f(z) - x - xy, \\ \frac{dy}{dt} &= f(z)y - y - xy, \\ \frac{dz}{dt} &= -\alpha f(z)(x + y) + v. \end{aligned} \quad (14)$$

Пусть процессы поглощения субстрата существенно более быстрые, чем процессы репродукции. В этом случае может быть использован метод квазистационарных концентраций, и дифференциальное уравнение для быстрой переменной $z(S)$ — концентрации субстрата — заменено алгебраическим. Тогда субстрат на интересующих нас временах достигнет квазистационарной концентрации: $dz/dt = 0$.

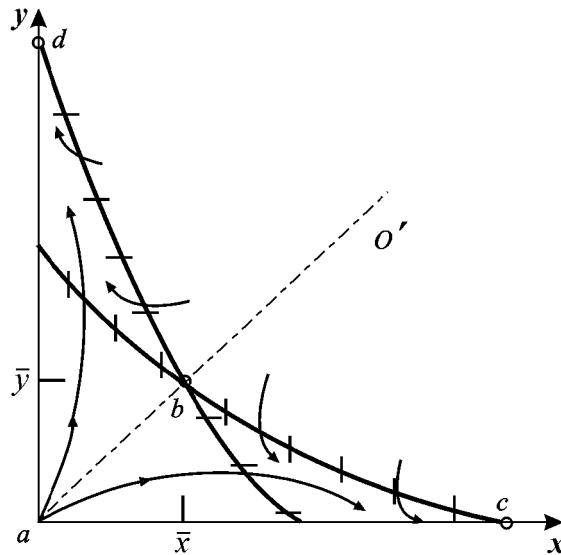


Рис. 8 - Фазовый портрет системы (15), описывающей отбор одного из двух равноправных видов когда субстрат поступает в систему с постоянной скоростью. a (начало координат) – неустойчивый узел, b – седло, c, d – устойчивые узлы.

Отсюда

$$f(z) = \frac{v}{\alpha(x+y)} = \frac{v_0}{x+y}. \quad (15)$$

В итоге получается система двух безразмерных уравнений

$$\frac{dx}{dt} = x \left[\frac{v_0}{x+y} - (1+y) \right], \quad \frac{dy}{dt} = y \left[\frac{v_0}{x+y} - (1+x) \right]. \quad (16)$$

Построим фазовый портрет системы (рис.8). Изоклины вертикальных касательных:

$x=0$ (ось ординат) и кривая $\frac{v_0}{x+y} - (1+y) = 0$.

Изоклины горизонтальных касательных:

$y=0$ (ось абсцисс) и $v_0 - (1+x)(x+y) = 0$ или

$y = \frac{v_0}{1+y} - x$ – гипербола.

Переменные x и y симметричны, поэтому изоклина вертикальных касательных симметрична изоклине горизонтальных касательных.

Система имеет четыре особые точки:

- 1) $x=0, y=0$ - неустойчивый узел;
- 2) $x=0, y=v_0$ - устойчивый узел;
- 3) $x=v_0, y=0$ - устойчивый узел;

4) симметричную точку - седло

$$x = y = \frac{\sqrt{1 + 2v_0} - 1}{2}. \quad (17)$$

В такой системе выживет один из видов: x или y . Его стационарная концентрация определяется скоростью притока субстрата и экономическим коэффициентом a . Как и в предыдущей системе (4) здесь причина отбора – неустойчивость симметричного состояния.

Лекция 16. КОЛЕБАНИЯ В БИОЛОГИЧЕСКИХ СИСТЕМАХ

1. Предельный цикл.
2. Теоремы, определяющие существование предельного цикла.
3. Анализ модели «брюсселятор».

1. Предельный цикл.

Для биологических систем характерно периодическое изменение различных характеристик. Период этих колебаний может быть связан с периодическими изменениями условий жизни на Земле - смена времен года, смена дня и ночи. Существуют и другие геофизические ритмы – 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.

С некоторыми из типов периодических движений мы уже имели дело при рассмотрении 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, система называется автоколебательной.

Незатухающие колебания в таких системах устойчивы, так как отклонения от стационарного колебательного режима затухают. 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 и сообществах.

Предельный цикл. В фазовом пространстве такому типу поведения соответствует притягивающее множество (аттрактор), называемое предельным циклом.

Предельный цикл есть изолированная замкнутая кривая на фазовой плоскости, к которой в пределе при $t \rightarrow \infty$ стремятся все интегральные кривые. Предельный цикл представляет стационарный режим с определенной амплитудой, не зависящий от начальных условий, а определяющийся только организацией системы. Существование предельного цикла на фазовой плоскости есть 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:

$$\frac{dx}{dt} = P(x, y), \quad \frac{dy}{dt} = Q(x, y). \quad (1)$$

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 спиралям, эта изолированная замкнутая траектория есть предельный цикл.

Простые примеры позволяют убедиться, что система общего вида (1) допускает в качестве траекторий предельные циклы.

Например, для системы

$$\frac{dx}{dt} = y + x[1 - (x^2 + y^2)], \quad \frac{dy}{dt} = -x + y[1 - (x^2 + y^2)] \quad (2)$$

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:

$$x = \frac{\cos(t - t_0)}{\sqrt{1 + Ce^{-2(t-t_0)}}}, \quad y = \frac{\sin(t - t_0)}{\sqrt{1 + Ce^{-2(t-t_0)}}}. \quad (3)$$

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.

Предельный цикл называется устойчивым, если существует такая область на фазовой плоскости, содержащая этот предельный цикл, - окрестность ε , что все фазовые траектории, начинающиеся в окрестности ε , асимптотически при $t \rightarrow \infty$ приближаются к предельному циклу.

Если же, наоборот, в любой сколь угодно малой окрестности ε 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.

На рис. 1 изображены устойчивый предельный цикл (а) и неустойчивые (б) и (в).

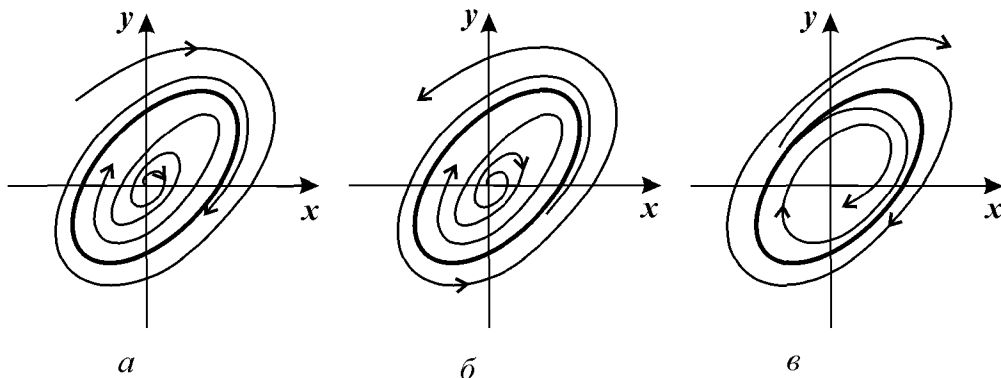


Рис. 1 – Устойчивый (а) и неустойчивые (б и в) предельные циклы

на фазовой плоскости

Неустойчивые предельные циклы, подобные изображенному на рис. 1б, такие, что все траектории с одной стороны (например, изнутри) приближаются к ним, а с другой стороны (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.

А.М.Ляпунов показал, что для исследования устойчивости периодического движения $x = \varphi(t)$, $y = \psi(t)$ можно идти по пути линеаризации уравнений, подобно тому, как мы это делали при исследовании устойчивости состояний равновесия. Если положить

$$x = \varphi(t) + \xi, \quad y = \psi(t) + \eta,$$

подставить эти выражения в уравнения (1), разложить правые части этих уравнений - функции

$$P(\varphi + \xi, \psi + \eta), \quad Q(\varphi + \xi, \psi + \eta)$$

в ряды по степеням ξ и η и отбросить нелинейные члены, то мы получим линейные уравнения (уравнения первого приближения) для координат возмущения ξ и η :

$$\frac{d\xi}{dt} = a\xi + b\eta, \quad \frac{d\eta}{dt} = c\xi + d\eta.$$

Коэффициенты в правой части:

$$\begin{aligned} a &= P'_x(\varphi(t), \psi(t)), & b &= P'_y(\varphi(t), \psi(t)), \\ c &= Q'_x(\varphi(t), \psi(t)), & d &= Q'_y(\varphi(t), \psi(t)). \end{aligned}$$

Это система линейных дифференциальных уравнений с периодическими коэффициентами периода T , поскольку a , b , c , d суть функции от φ , ψ - периодических функций времени с периодом T . Общий вид ее решения

$$\begin{aligned} \xi &= -c_1 f_{11}(t)e^{h_1 t} + c_2 f_{12}(t)e^{h_2 t}, \\ \eta &= -c_1 f_{21}(t)e^{h_1 t} + c_2 f_{22}(t)e^{h_2 t}. \end{aligned}$$

Здесь f_{ij} - некоторые периодические функции с периодом T . От показателей и которые носят название «характеристических показателей», зависят свойства решений для отклонений от стационарного периодического решения ξ и η . А именно, знаки их действительных частей определяют, являются ли эти решения нарастающими или затухающими. Можно показать,

что в силу автономности исходной системы (1) один из характеристических показателей равен нулю, а другой равен h .

$$h = \frac{1}{T} \int_0^T \{P'_x(\varphi(t), \psi(t)) + Q'_y(\varphi(t), \psi(t))\} dt, \quad (4)$$

где $x = \varphi(t)$, $y = \psi(t)$ - любое периодическое решение, соответствующее рассматриваемому предельному циклу, T - период решения.

Таким образом, устойчивость предельного цикла (и устойчивость в смысле Ляпунова соответствующих периодических движений) определяется знаком характеристического показателя. Предельный цикл устойчив, если $h < 0$ и неустойчив, если $h > 0$. Если же $h = 0$, уравнения первого приближения не решают вопроса об устойчивости периодического движения.

Для нахождения предельных циклов не существует таких простых аналитических методов, как для нахождения стационарных точек и исследования их устойчивости. Однако исследование фазовой плоскости системы позволяет ответить на вопрос, есть в данной системе предельный цикл, или нет.

2. Теоремы, определяющие существование предельного цикла.

Сформулируем несколько теорем, определяющих наличие предельного цикла по топологическому строению фазовой плоскости. Они могут быть полезны как при аналитическом, так и при компьютерном анализе системы.

Теорема 1. Пусть на фазовой плоскости существует область, из которой фазовые траектории не выходят, и в которой нет положений равновесия (особых точек). Тогда в этой области обязательно существует предельный цикл, причем все остальные траектории обязательно наматываются на него.

На рис. 2 изображена такая область G , из которой фазовые траектории не выходят. Это означает, что фазовые траектории либо входят, пересекая границу, внутрь области, либо сама граница является траекторией. Легко видеть, что такая область не может быть односвязной. Поскольку траектория наматывается на предельный цикл изнутри, это означает, что внутри этого предельного цикла на фазовой плоскости существует либо неустойчивая особая точка, либо неустойчивый предельный цикл, очевидно, не принадлежащие рассматриваемой области G .

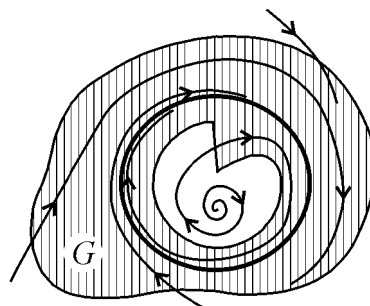


Рис. 2 – Иллюстрация к теореме 1. Жирная кривая – предельный цикл

Таким образом, если найти на фазовой плоскости такую двусвязную область, что направления фазовых траекторий на всей границе обращены внутрь этой области, то можно утверждать, что внутри этой области имеется предельный цикл.

Теорема 2. Если существует на фазовой плоскости некоторая замкнутая область, такая, что все фазовые траектории, пересекающие границу этой области, входят в нее, и внутри этой области находится неустойчивая особая точка, то в этой области обязательно имеется хотя бы один предельный цикл (рис. 3).

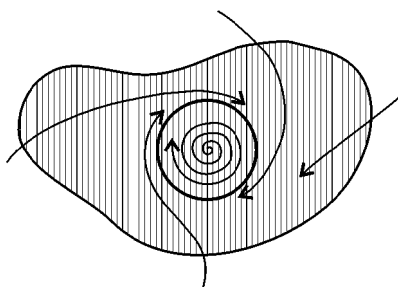


Рис. 3 – Иллюстрация к теореме 2

Приведем также некоторые критерии отсутствия замкнутых фазовых траекторий (в том числе предельных циклов).

1. Если в системе не существует особых точек, то в ней не может быть и замкнутых фазовых траекторий.
2. Если в системе существует только одна особая точка, отличная от узла, фокуса и центра (например, седло), то такая система не допускает замкнутых фазовых траекторий.
3. Если в системе имеются только простые особые точки, причем через все точки типа узел и фокус проходят интегральные кривые, уходящие на бесконечность, то в такой системе нет замкнутых фазовых траекторий.

В случае, если критерии 1–3 выполнены, можно с уверенностью утверждать, что в системе нет предельных циклов. Однако невыполнение

этих критериев еще не позволяет сделать вывод о наличии в системе предельных циклов и, следовательно, автоколебаний.

Неустойчивый предельный цикл также может содержаться в фазовом портрете грубых систем. Однако такой предельный цикл не соответствует реальному периодическому процессу, он играет лишь роль «водораздела», по обе стороны которого траектории имеют различное поведение. Например, на рис. 4 представляет собой сепаратрису, отделяющую область тяготения траекторий к устойчивой особой точке, с одной стороны, и к устойчивому предельному циклу, с другой.

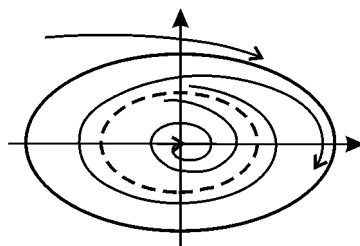


Рис. 4 – Фазовый портрет системы, имеющий устойчивый и неустойчивый (пунктир) предельные циклы

3. Анализ модели «брюсселятор».

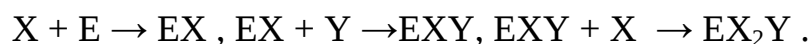
Простейшим классическим примером существования автоколебаний в системе химических реакций является тримолекулярная модель «Брюсселятор», предложенная в Брюсселе Пригожиным и Лефевром (1965). Основной целью при изучении этой модели было установление качественных типов поведения, совместимых с фундаментальными 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 содержит простейшую реализацию кубической нелинейности посредством химической реакции

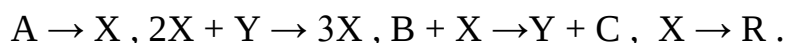


Хотя тримолекулярная стадия в химической кинетике не столь распространена, как бимолекулярные процессы, выражения для скорости ряда биохимических реакций в определенных случаях можно свести к кубическому виду. В качестве примера приведем следующую последовательность ферментативных реакций:



Здесь предполагается что фермент E имеет по крайней мере три каталитических центра, способных одновременно фиксировать две молекулы X и одну молекулу Y . Если образующиеся комплексы распадаются с достаточно большой скоростью, а ферменты присутствуют в небольших количествах, легко показать, что всю последовательность реакций можно свести к одной стадии, дающей нелинейный член типа X^2Y в выражении для скорости реакции.

Брюсселятор представляет собой следующую схему гипотетических химических реакций:



Здесь A, B - исходные вещества, C, R - продукты, X, Y - промежуточные вещества.

Пусть конечные продукты C и R немедленно удаляются из реакционного пространства. Это означает, что обратные константы $k_{-3} = k_{-4} = 0$. Если субстрат A находится в избытке, $k_{-1} = 0$. Предположим также, что $k_{-2} = 0$. Значения остальных констант положим равными единице. Тогда схема реакций (в случае точечной системы) описывается системой уравнений:

$$\frac{dx}{dt} = A + x^2y - (B+1)x, \quad \frac{dy}{dt} = Bx - x^2y. \quad (6)$$

Модель имеет одну особую точку с координатами:

$$\bar{x} = A, \quad \bar{y} = \frac{B}{A}. \quad (7)$$

Исследуем стационарное решение (6) на устойчивость по методу Ляпунова. Введем переменные, характеризующие отклонения от особой точки:

$$\xi = x - \bar{x}, \quad \eta = y - \bar{y}.$$

Линеаризованная система имеет вид:

$$\frac{d\xi}{dt} = (B-1)\xi + A^2\eta, \quad \frac{d\eta}{dt} = -B\xi - A^2\eta.$$

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:

$$\lambda_{1,2} = -\frac{1}{2}(A^2 + 1 - B) \pm \frac{1}{2}\sqrt{(A^2 + 1 - B)^2 - 4A^2}. \quad (8)$$

Напомним, что особая точка является устойчивой, если действительные части корней характеристического уравнения отрицательны. Из выражения (8) видно, что при $B < 1 + A^2$ особая точка (7) устойчива. Если же $B > 1 + A^2$ особая точка становится неустойчивой, и у системы (6) появляется устойчивый предельный цикл. Значение $B = 1 + A^2$ является бифуркационным. Если величина B лишь немного превосходит бифуркационный порог, автоколебания в системе носят квазигармонический характер. Таким образом, брюсселятор при выполнении условия

$$B > 1 + A^2 \quad (8)$$

является автоколебательной системой. Фазовый портрет брюсселятора при разных значениях параметров изображен на рис. 5.

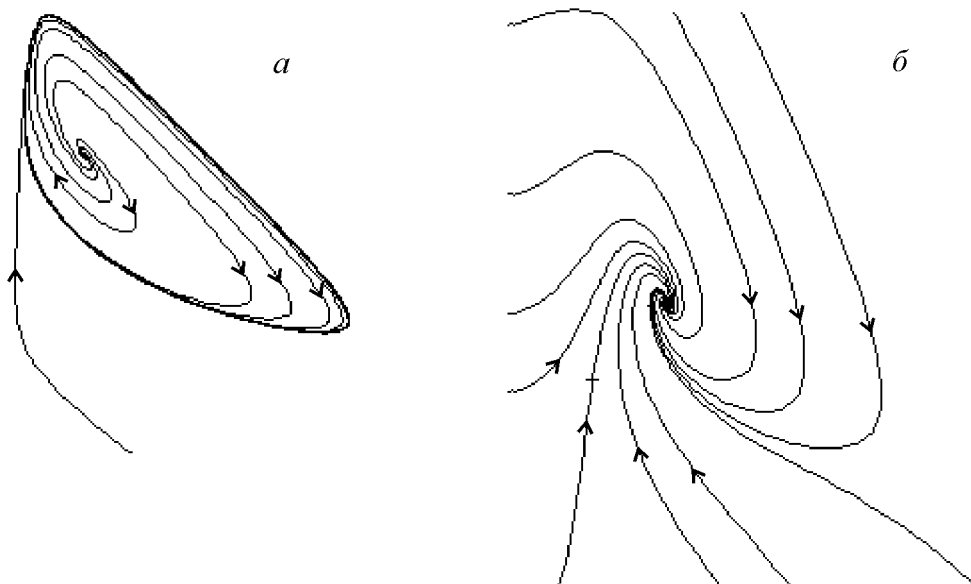


Рис.5 – Фазовый портрет системы брюсселятор при $B > 1 + A^2$ (а) и $B < 1 + A^2$ (б)

Лекция 17. МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ДВУХ ВИДОВ

1. Модели взаимодействия двух видов
2. Уравнения конкуренции и их анализ.
3. Уравнения системы «хищник-жертва» и их анализ.

1. Модели взаимодействия двух видов

Основателем современной математической теории популяций справедливо считается итальянский математик Вито Вольтерра, разработавший математическую теорию биологических сообществ, аппаратом которой служат дифференциальные и интегро-дифференциальные уравнения. (Vito Volterra. *Lecons sur la Theorie Mathematique de la Lutte pour la Vie*. Paris, 1931). В последующие десятилетия популяционная динамика развивалась, в основном, в русле высказанных в этой книге идей. Русский перевод книги Вольтерра вышел в 1976 г. под названием: «Математическая теория борьбы за существование» с послесловием Ю.М. Свирежева, в котором рассматривается история развития математической экологии в период 1931-1976 гг.

Книга Вольтерра написана так, как пишут книги по математике. В ней сначала сформулированы некоторые предположения о математических объектах, которые предполагается изучать, а затем проводится математическое исследование свойств этих объектов.

Системы, изученные Вольтерра, состоят из двух или нескольких видов. В отдельных случаях рассматривается запас используемой пищи. В основу уравнений, описывающих взаимодействие этих видов, положены следующие представления.

Гипотезы Вольтерра

1. Пища либо имеется в неограниченном количестве, либо ее поступление с течением времени жестко регламентировано.
2. Особи каждого вида отмирают так, что в единицу времени погибает постоянная доля существующих особей.
3. Хищные виды поедают жертв, причем в единицу времени количество съеденных жертв всегда пропорционально вероятности встречи особей этих двух видов, т.е. произведению количества хищников на количество жертв.
4. Если имеется пища в ограниченном количестве и несколько видов, которые способны ее потреблять, то доля пищи, потребляемой видом в

единицу времени, пропорциональна количеству особей этого вида, взятому с некоторым коэффициентом, зависящим от вида (модели межвидовой конкуренции).

5. Если вид питается пищей, имеющейся в неограниченном количестве, прирост численности вида в единицу времени пропорционален численности вида.

6. Если вид питается пищей, имеющейся в ограниченном количестве, то его размножение регулируется скоростью потребления пищи, т.е. за единицу времени прирост пропорционален количеству съеденной пищи.

Эти гипотезы имеют близкие параллели с химической кинетикой. В уравнениях популяционной динамики, как и в уравнениях химической кинетики, используется “принцип соударений”, когда скорость реакции пропорциональна произведению концентраций реагирующих компонентов.

Действительно, согласно гипотезам Вольтерра, скорость процесса отмирания каждого вида пропорциональна численности вида. В химической кинетике это соответствует мономолекулярной реакции распада некоторого вещества, а в математической модели – отрицательным линейным членам в правых частях уравнений.

Согласно представлениям химической кинетики, скорость бимолекулярной реакции взаимодействия двух веществ пропорциональна вероятности столкновения этих веществ, т.е. произведению их концентрации. Точно так же, в соответствии с гипотезами Вольтерра, скорость размножения хищников (гибели жертв) пропорциональна вероятности встреч особей хищника и жертвы, т.е. произведению их численностей. И в том и в другом случае в модельной системе появляются билинейные члены в правых частях соответствующих уравнений.

Наконец, линейные положительные члены в правых частях уравнений Вольтерра, отвечающие росту популяций в неограниченных условиях, соответствуют автокаталитическим членам химических реакций. Такое сходство уравнений в химических и экологических моделях позволяет применить для математического моделирования кинетики популяций те же методы исследований, что и для систем химических реакций.

Классификация типов взаимодействий

В соответствии с гипотезами Вольтерра взаимодействие двух видов, численности которых x_1 и x_2 , могут быть описаны уравнениями:

$$\begin{aligned}\frac{dx_1}{dt} &= a_1x_1 + b_{12}x_1x_2 - c_1x_1^2, \\ \frac{dx_2}{dt} &= a_2x_2 + b_{21}x_1x_2 - c_2x_2^2.\end{aligned}\tag{1}$$

Здесь параметры a_i - константы собственной скорости роста видов, c_i - константы самоограничения численности (внутривидовой конкуренции), b_{ij} - константы взаимодействия видов, ($i, j = 1, 2$). Знаки этих коэффициентов определяют тип взаимодействия.

В биологической литературе обычно классифицируют взаимодействия по участвующим в них механизмам. Разнообразие здесь огромно: различные трофические взаимодействия, химические взаимодействия, существующие между бактериями и планктонными водорослями, взаимодействия грибов с другими организмами, сукцессии растительных организмов, связанные в частности, с конкуренцией за солнечный свет и с эволюцией почв и т.д. Такая классификация кажется необозримой.

Е. Одум, учитывая предложенные В.Вольтерра модели, предложил классификацию не по механизмам, а по результатам. Согласно этой классификации, оценивать взаимоотношения следует как положительные, отрицательные или нейтральные в зависимости от того, возрастает, убывает или остается неизменной численность одного вида в присутствии другого вида. Тогда основные типы взаимодействий могут быть представлены в виде

ТИПЫ ВЗАИМОДЕЙСТВИЯ ВИДОВ:

1. СИМБИОЗ: $b_{12}, b_{21} > 0$
2. КОММЕНСАЛИЗМ: $b_{12} > 0, b_{21} = 0$
3. ХИЩНИК-ЖЕРТВА: $b_{12} > 0, b_{21} < 0$
4. АМЕНСАЛИЗМ: $b_{12} = 0, b_{21} < 0$
5. КОНКУРЕНЦИЯ: $b_{12}, b_{21} < 0$
6. НЕЙТРАЛИЗМ: $b_{12}, b_{21} = 0$

Рассмотрим основные типы взаимодействий.

2. Уравнения конкуренции и их анализ.

Уравнения конкуренции имеют вид:

$$\begin{aligned}\frac{dx_1}{dt} &= x_1(a_1 - b_{12}x_2 - c_1x_1), \\ \frac{dx_2}{dt} &= x_2(a_2 - b_{21}x_1 - c_2x_2).\end{aligned}\tag{2}$$

Стационарные решения системы:

$$(1) \quad \bar{x}_1^{(1)} = 0, \quad \bar{x}_2^{(1)} = 0.$$

Начало координат, при любых параметрах системы представляет собой неустойчивый узел.

$$(2) \quad \bar{x}_1^{(2)} = 0, \quad \bar{x}_2^{(2)} = \frac{a_2}{c_2}. \quad (3)$$

Стационарное состояние (3) представляет собой седло при $a_1 > b_{12} / c_2$ и устойчивый узел при $a_1 < b_{12} / c_2$. Это условие означает, что вид вымирает, если его собственная скорость роста меньше некоторой критической величины.

$$(3) \quad \bar{x}_1^{(3)} = \frac{a_1}{c_1}, \quad \bar{x}_2^{(3)} = 0. \quad (4)$$

Стационарное решение (4) - седло при $a_2 > b_{21} / c_1$ и устойчивый узел при $a_2 < b_{21} / c_1$.

$$(4) \quad \bar{x}_1^{(4)} = \frac{a_1 c_2 - a_2 b_{12}}{c_1 c_2 - b_{12} b_{21}}, \quad \bar{x}_2^{(4)} = \frac{c_1 b_{12} - a_1 b_{21}}{c_1 c_2 - b_{12} b_{21}}. \quad (5)$$

Стационарное состояние (5) характеризует сосуществование двух конкурирующих видов и представляет собой устойчивый узел в случае выполнения соотношения:

$$\frac{a_1 b_{12}}{c_2} < a_1 < \frac{a_2 c_1}{b_{21}}.$$

Отсюда следует неравенство:

$$b_{12} b_{21} < c_1 c_2, \quad (6)$$

позволяющее сформулировать условие сосуществования видов:

Произведение коэффициентов межпопуляционного взаимодействия меньше произведения коэффициентов внутри популяционного взаимодействия.

Действительно, пусть естественные скорости роста двух рассматриваемых видов a_1 , a_2 одинаковы. Тогда необходимым для устойчивости условием будет

$$c_2 > b_{12}, \quad c_1 > b_{21}.$$

Эти неравенства показывают, что увеличение численности одного из конкурентов сильнее подавляет его собственный рост, чем рост другого конкурента. Если численность обоих видов ограничивается, частично или полностью, различными ресурсами, приведенные выше неравенства справедливы. Если же оба вида имеют совершенно одинаковые потребности, то один из них окажется более жизнеспособным и вытеснит своего конкурента.

Поведение фазовых траекторий системы дает наглядное представление о возможных исходах конкуренции. Приравняем нулю правые части уравнений системы (2):

$$x_1 (a_1 - c_1 x_1 - b_{12} x_2) = 0 \quad (dx_1/dt = 0),$$

$$x_2 (a_2 - b_{21} x_1 - c_2 x_2) = 0 \quad (dx_2/dt = 0).$$

При этом получим уравнения для главных изоклин системы

$$x_2 = -b_{21} x_1 / c_2 + a_2 / c_2, \quad x_2 = 0$$

– уравнения изоклин вертикальных касательных,

$$x_2 = -c_1 x_1 / b_{12} + a_1 / b_{12}, \quad x_1 = 0$$

– уравнения изоклин горизонтальных касательных. Точки попарного пересечения изоклин вертикальных и горизонтальных касательных систем представляют собой стационарные решения системы уравнений (2), а их координаты суть стационарные численности конкурирующих видов.

Возможное расположение главных изоклин в системе (2) изображено на рис.1.

Рис. 1а соответствует выживанию вида x_1 , рис. 1 б – выживанию вида x_2 , рис. 1 в – сосуществованию видов при выполнении условия (6). Рис. 1 г демонстрирует триггерную систему. Здесь исход конкуренции зависит от начальных условий. Ненулевое для обоих видов стационарное состояние (5) – неустойчивое. Это – седло, через которое проходит сепаратриса, отделяющая области выживания каждого из видов.

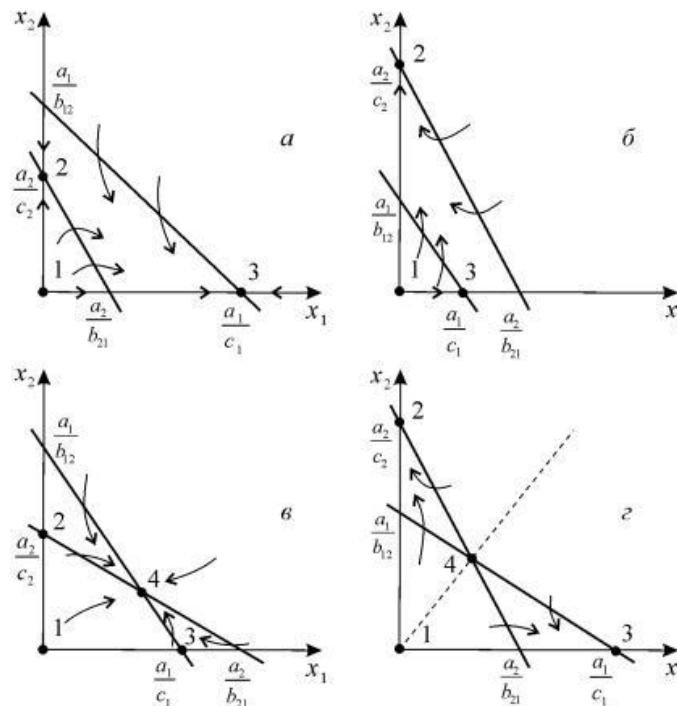


Рис. 1. Расположение главных изоклин на фазовом портрете вольтерровской системы конкуренции двух видов (2) при разном соотношении параметров.

Для изучения конкуренции видов ставились эксперименты на самых различных организмах. Обычно выбирают два близкородственных вида и выращивают их вместе и по отдельности в строго контролируемых условиях. Через определенные промежутки времени проводят полный или выборочный учет численности популяции. Регистрируют данные по нескольким повторным экспериментам и анализируют. Исследования проводили на простейших (в частности, инфузориях), многих видах жуков рода *Tribolium*, дрозофилах, пресноводных ракообразных (дафниях). Много экспериментов проводилось на микробных популяциях. В природе также проводили эксперименты, в том числе на планариях (Рейнольдс), двух видах муравьев (Понтин) и др. На рис. 2. изображены кривые роста диатомовых водорослей, использующих один и тот же ресурс (занимающих одну и ту же экологическую нишу).

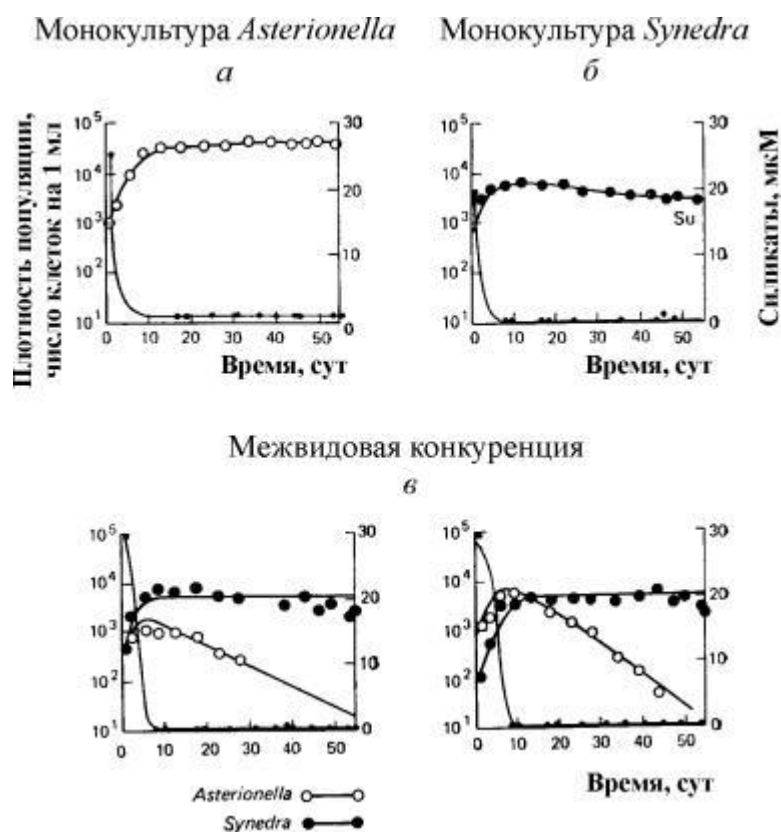


Рис. 2. Конкуренция у диатомовых водорослей. а - при выращивании в монокультуре *Asterionella Formosa* выходит на постоянный уровень плотности и поддерживает концентрацию ресурса (силиката) на постоянно низком уровне. б - при выращивании в монокультуре *Synedra* ведет себя сходным образом и поддерживает концентрацию силиката на еще более низком уровне. в - при совместном культивировании (в двух повторностях) *Synedra* вытесняет *Asterionella Formosa*. По-видимому, *Synedra*

выигрывает конкуренцию благодаря своей способности к более полному использованию субстрата.

При выращивании в монокультуре *Asterionella Formosa* выходит на постоянный уровень плотности и поддерживает концентрацию ресурса (силиката) на постоянно низком уровне. Б. При выращивании в монокультуре *Synedrauina* ведет себя сходным образом и поддерживает концентрацию силиката на еще более низком уровне. В. При совместном культивировании (в двух повторностях) *Synedrauina* вытесняет *Asterionella Formosa*. По-видимому, *Synedra* выигрывает конкуренцию благодаря своей способности к более полному использованию субстрата.

Широко известны эксперименты по изучению конкуренции Г. Гаузе, продемонстрировавшие выживание одного из конкурирующих видов и позволившие ему сформулировать «закон конкурентного исключения». Закон гласит, что в одной экологической нише может существовать только один вид. На рис. 3. приведены результаты экспериментов Гаузе для двух видов *Paramecium*, занимающих одну экологическую нишу (рис. 3 а, б) и видами, занимающими разные экологические ниши (рис. 3. в).

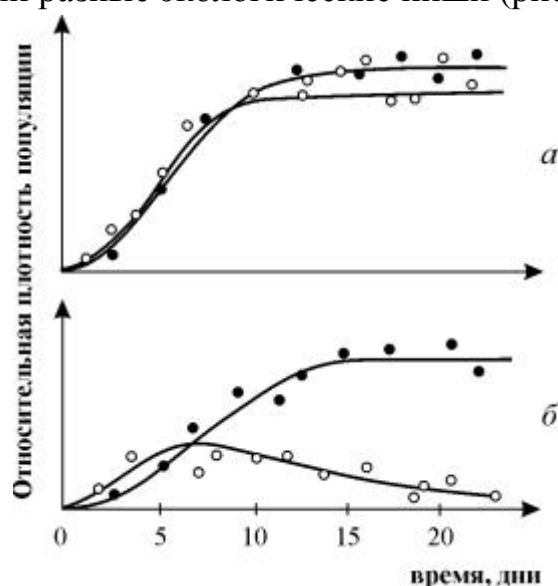


Рис. 3. а - Кривые роста популяций двух видов *Paramecium* в одновидовых культурах. Черные кружки – *P. Aurelia*, белые кружки – *P. Caudatum* ; б - Кривые роста *P. Aurelia* и *P. Caudatum* в смешанной культуре. По Gause, 1934

Модель конкуренции (2) имеет недостатки, в частности, из нее следует, что сосуществование двух видов возможно лишь в случае, если их численность ограничивается разными факторами, но модель не дает указаний, насколько велики должны быть различия для обеспечения длительного сосуществования. В то же время известно, что для длительного сосуществования в изменчивой среде необходимо различие, достигающее

определенной величины. Внесение в модель стохастических элементов (например, введение функции использования ресурса) позволяет количественно исследовать эти вопросы.

3. Уравнения системы «хищник-жертва» и их анализ.

Для взаимоотношений типа хищник-жертва или паразит-хозяин система уравнений (1) принимает вид:

$$\begin{aligned}\frac{dx_1}{dt} &= x_1(a_1 - b_{12}x_2 - c_1x_1), \\ \frac{dx_2}{dt} &= x_2(a_2 + b_{21}x_1 - c_2x_2).\end{aligned}\tag{7}$$

Здесь, в отличие от (2) знаки b_{12} и b_{21} - разные. Как и в случае конкуренции, начало координат

$$\bar{x}_1^{(1)} = 0, \quad \bar{x}_2^{(1)} = 0.\tag{8}$$

является особой точкой типа неустойчивый узел. Три других возможных стационарных состояния:

$$\bar{x}_1^{(2)} = 0, \quad \bar{x}_2^{(2)} = \frac{a_2}{c_2}.\tag{9}$$

$$\bar{x}_1^{(3)} = \frac{a_1}{c_1}, \quad \bar{x}_2^{(3)} = 0.\tag{10}$$

$$\bar{x}_1^{(4)} = \frac{a_1c_2 + a_2b_{12}}{c_1c_2 + b_{12}b_{21}}, \quad \bar{x}_2^{(4)} = \frac{c_1b_{12} - a_1b_{21}}{c_1c_2 + b_{12}b_{21}}\tag{11}$$

Таким образом, возможно выживание только жертвы (10), только хищника (9) (если у него имеются и другие источники питания) и сосуществование обоих видов (11). Возможные типы фазовых портретов для системы хищник-жертва представлены на рис. 4.

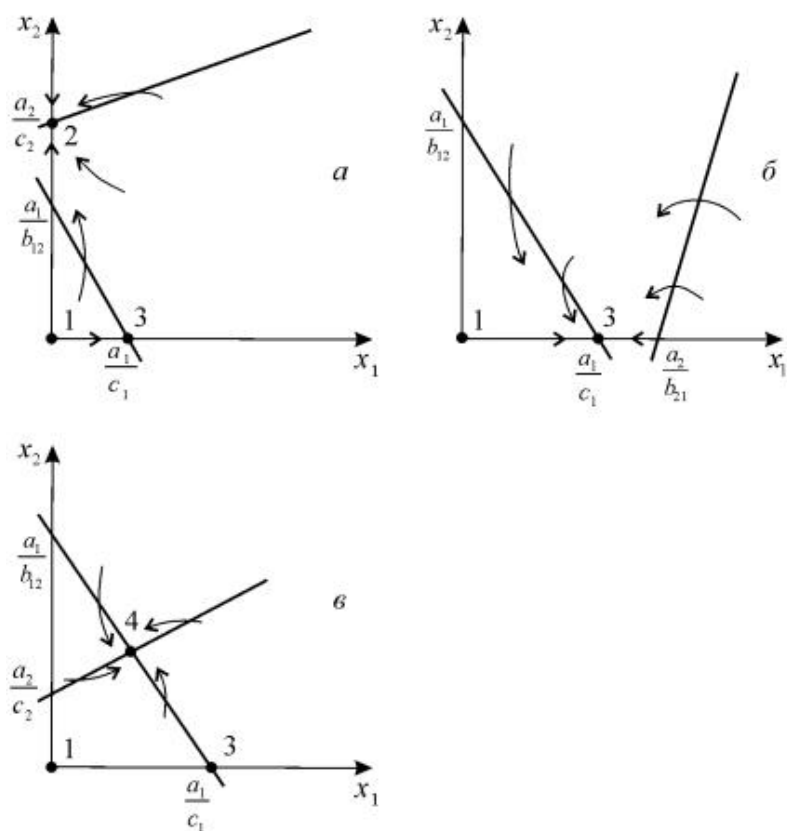


Рис. 4. Расположение главных изоклин на фазовом портрете вольтерровской системы хищник-жертва (7) при различных соотношениях параметров. Стрелками указано направление фазовых траекторий.

Изоклины горизонтальных касательных представляют собой прямые

$$x_2 = -b_{21}x_1/c_2 + a_1/c_2, \quad x_2 = 0,$$

а изоклины вертикальных касательных – прямые

$$x_2 = -c_1x_1/b_{12} + a_2/b_{12}, \quad x_1 = 0.$$

Стационарные точки лежат на пересечении изоклин вертикальных и горизонтальных касательных.

Из рис. 4 видно следующее. Система хищник – жертва (7) может иметь устойчивое положение равновесия, в котором популяция жертв полностью вымерла ($x_1=0$) и остались только хищники (точка 2 на рис. 4 а). Очевидно, такая ситуация может реализоваться лишь в случае, если кроме рассматриваемого вида жертв x_1 хищник x_2 – имеет дополнительные источники питания. Этот факт в модели отражается положительным членом в правой части уравнения для x_2 . Особые точки (1) и (3) (рис. 4 а) являются неустойчивыми. Вторая возможность – устойчивое стационарное состояние, в котором популяция хищников полностью вымерла и остались одни жертвы – устойчивая точка (3) (рис. 4 б). Здесь особая точка (1) – также неустойчивый узел.

Наконец, третья возможность – устойчивое сосуществование популяций хищника и жертвы (рис. 4 в), стационарные численности которых выражаются формулами (11).

Как и в случае одной популяции, для модели (7) можно разработать стохастическую модель, но для нее нельзя получить решение в явном виде. Поэтому мы ограничимся общими рассуждениями. Допустим, например, что точка равновесия находится на некотором расстоянии от каждой из осей. Тогда для фазовых траекторий, на которых значения x_1 , x_2 остаются достаточно большими, вполне удовлетворительной будет детерминистическая модель. Но если в некоторой точке фазовой траектории какая-либо переменная не очень велика, то существенное значение могут приобрести случайные флуктуации. Они приводят к тому, что изображающая точка переместится на одну из осей, что означает вымирание соответствующего вида.

Таким образом, стохастическая модель оказывается неустойчивой, так как стохастический “дрейф” рано или поздно приводит к вымиранию одного из видов. В такого рода модели хищник в конечном счете вымирает, это может произойти либо случайно, либо вследствие того, что сначала элиминируется популяция его жертвы. Стохастическая модель системы хищник – жертва хорошо объясняет эксперименты Гаузе (1934 г.), в которых инфузория *Paramecium caudatum* служила жертвой для другой инфузории *Didinium nasutum* – хищника. Ожидавшиеся согласно детерминистическим уравнениям (7) равновесные численности в этих экспериментах составляли примерно всего по пять особей каждого вида, так что нет ничего удивительного в том, что в каждом повторном эксперименте довольно быстро вымирали либо хищники, либо жертвы (а за ними и хищники). Результаты экспериментов представлены на рис. 5.

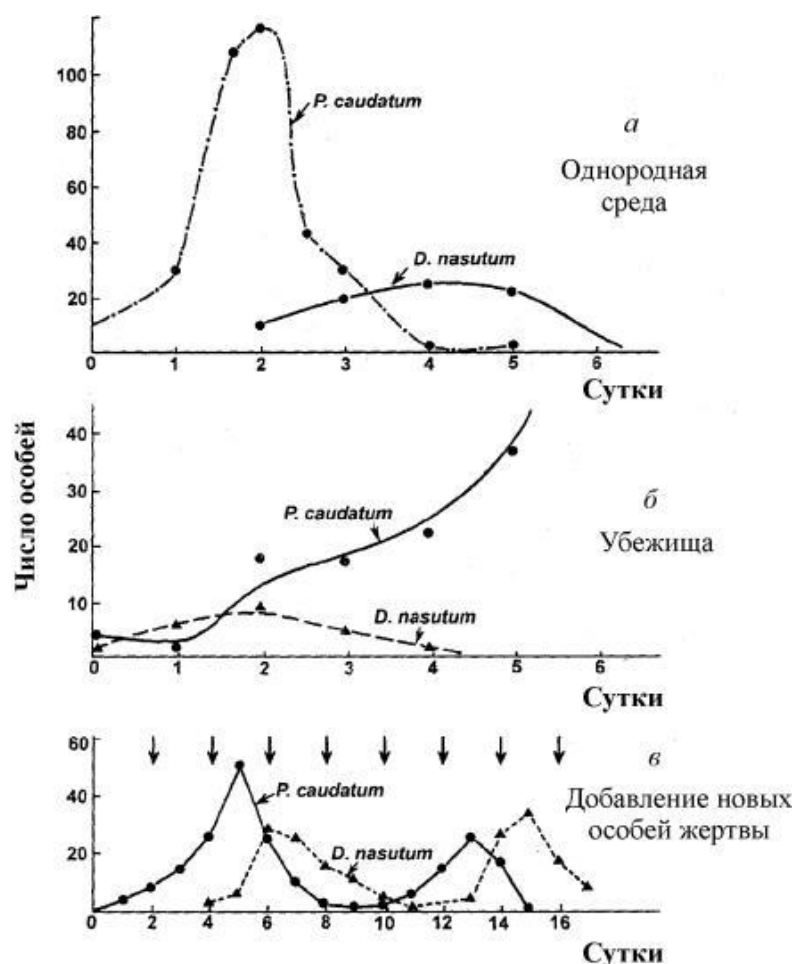


Рис. 5. Рост *Parametium caudatum* и хищной инфузории *Dadinium nasutum*. Из: Gause G.F. The struggle for existence. Baltimore, 1934.

Итак, анализ вольтерровских моделей взаимодействия видов показывает, что, несмотря на большое разнообразие типов поведения таких систем, незатухающих колебаний численности в модели конкурирующих видов не может быть вовсе. Однако в природе и в эксперименте такие колебания наблюдаются. Необходимость их теоретического объяснения послужила одной из причин для формулировки модельных описаний в более общем виде.

Было предложено большое число моделей, описывающих взаимодействие видов, правые части уравнений которых представляли собой функции численностей взаимодействующих популяций. Решался вопрос о выработке общих критериев, позволяющих установить, какого вида функции могут описать особенности поведения временного численности популяции, в том числе устойчивые колебания. Наиболее известные из этих моделей принадлежат Колмогорову (1935, переработанная статья - 1972) и Розенцвейгу (1963).

Итак, мы рассмотрели автономные непрерывные математические модели, описывающие взаимодействие двух видов. Сделаем некоторые выводы. При моделировании биоценоза из двух видов система Вольтерра (1) дает возможность для описания устойчивого сосуществования видов в условиях конкуренции, симбиоза и хищничества (паразитизма). При попытке описать устойчивые колебания численности видов мы сталкиваемся с трудностями. Система уравнений (7), описывающая взаимодействия хищник-жертва без учета самоограничения численности популяций и имеющая особую точку типа центр, - негрубая и, следовательно, неустойчива к случайным флуктуациям численности. Предельных же циклов, являющихся фазовыми траекториями устойчивых автоколебаний, система типа Вольтерра (1) иметь не может. Для получения предельных циклов в модельных системах приходится выходить за рамки гипотез Вольтерра и учитывать более тонкие эффекты взаимодействия между видами. Правые части уравнений при этом становятся существенно нелинейными.

Дальнейшее углубление математической теории взаимодействия видов идет по линии детализации структуры самих популяций и учета временных и пространственных факторов.

Лекция 18. МОДЕЛИРОВАНИЕ МИКРОБНЫХ ПОПУЛЯЦИЙ

1. Теория хемостата.

2. Анализ модели Моно.

1. Теория хемостата.

Микробные популяции как объект моделирования и управления. Непрерывная культура микроорганизмов. Модель Моно.

Микробиология является одной из немногих областей современной биологии, где математическое моделирование стало действенным средством научного исследования. Более того, математические модели прочно вошли в практику биотехнологического производства микроорганизмов как инструмент управления биотехнологическими процессами.

Мы остановимся на моделях, которые не только лежат в основе моделей микробиологических систем, но являются базовыми моделями всей математической биологии, в том числе используются в популяционной динамике, при моделировании иммунных процессов и проч.

В большинстве своем микроорганизмы - одноклеточные организмы, они имеют высокое отношение поверхности к объему и поэтому высокие интенсивности обмена с окружающей средой. С этим связаны:

- высокие скорости размножения микроорганизмов,
- большой прирост биомассы,
- высокая скорость роста микробных популяций
- высокая скорость микроэволюционных процессов в микробных сообществах.

Все это делает микробные популяции чрезвычайно привлекательными как в практическом отношении для биотехнологии, так и в качестве научного объекта для изучения популяционных и эволюционных процессов.

Для математического описания микробных популяций обычно используют аппарат обыкновенных дифференциальных уравнений. В отношении микробиологических систем такое описание гораздо более обосновано, чем применительно к наземным и водным высшим организмам. Из-за многочисленности микробных популяций к ним применимо понятие концентрации.

Действительно, даже в лабораторных исследованиях, *in vitro* приходится иметь дело с количеством особей порядка 10^{10} и выше. В большом промышленном ферментере могут одновременно жить 10^{16} - 10^{17} дрожжевых клеток.

Напомним, что отклонение численности от средних значений, вызванное случайными обстоятельствами, пропорционально $1/\sqrt{N}$, где N - численность популяции. Таким образом, для многочисленных популяций можно строить модель в терминах средних численностей, или концентраций.

Второй фактор - относительная однородность культуры микроорганизмов в объеме культиватора. Это позволяет пренебречь пространственными эффектами.

Для управления биотехнологическим процессом необходимо:

- сформулировать модель, описывающую рост управляемой культуры микроорганизмов,
- указать параметры, по которым производится управление,
- определить цель, которая при этом преследуется.

Например, целью может быть максимальная скорость роста культуры, или получение максимальной биомассы в течение всего срока выращивания, или минимизация времени выхода культиватора на стационарный режим работы. В зависимости от этого должна быть математически сформулирована соответствующая целевая функция. Нахождение значений управляющих параметров, которые позволяют достичь экстремума этой целевой функции, и составляют задачу управления.

Непрерывные культуры микроорганизмов.

Характерная кривая роста микроорганизмов приведена на рис.1

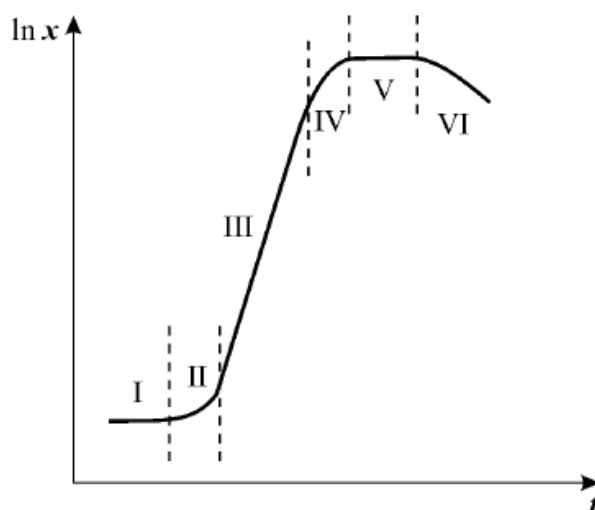


Рис. 1. Кривая роста микроорганизмов при периодическом культивировании. I – лаг-фаза; II – фаза ускорения роста; III – фаза экспоненциального роста; IV – фаза замедления роста; V – фаза стационарная; VI – фаза отмирания культуры

Процессы культивирования разделяют на периодические и непрерывные. При периодическом режиме в культиватор одновременно закладывают все необходимое для роста микроорганизмов (субстраты) и некоторую “затравку” биомассы, после чего популяция микроорганизмов растет и развивается по своим законам. В некоторый момент времени производится изъятие биомассы. Затем процесс повторяется. Таким образом, снятие урожая производится периодически, и каждый раз популяция проходит через все стадии роста.

Непрерывные культуры микроорганизмов - это культуры, в которые все время добавляется питательная среда, а часть содержимого, в том числе живые организмы - биомасса - постоянно удаляется. Эти условия имитируют естественные проточные системы. Однако в отличие от естественных систем, условия среды и развития микроорганизмов в установках непрерывного культивирования в лабораториях и на промышленных предприятиях находятся под контролем и могут быть стабилизированы. Это позволяет проводить эксперименты с культурами микроорганизмов по изучению популяционных законов развития видов и их сообществ, наблюдать процессы микроэволюции.

Для микроорганизмов, особенно автотрофных бактерий и дрожжей, условия выращивания довольно просты. Их выращивают в жидкой среде, представляющей собой раствор солей и простых органических соединений. Культуру содержат при постоянной температуре и перемешивают, причем из резервуара в нее постоянно поступает стерильная среда. (рис.2)

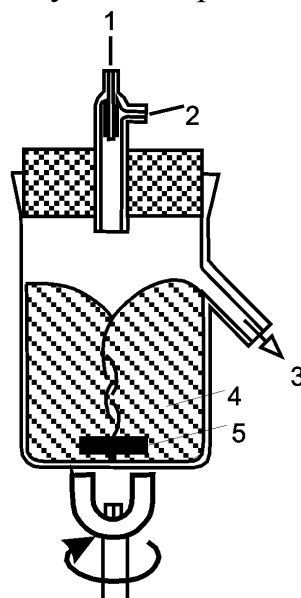


Рис.2. 1 – регулятор, 2 – поступление субстрата, 3 – отток (вымывание) смеси субстрата и биомассы, 4 – культура внутри культиватора, 5 – мешалка

При построении моделей в микробиологии в качестве равноправных переменных используют как концентрации микроорганизмов, так и концентрации различных растворимых органических и неорганических веществ: субстратов, ферментов, продуктов. В микробиологии общепринят эмпирический подход к построению моделей. Из всех факторов, влияющих на рост клетки, выбирают лимитирующий, и опытным путем находят зависимость скорости роста от его концентрации. Особый класс составляют задачи, где в процессе роста происходит смена лимитирования.

В общем виде кинетика концентрации клеток в непрерывной культуре описывается уравнением:

$$\frac{dx}{dt} = x(\mu - \nu) . \quad (1)$$

Здесь x – концентрация клеток в культиваторе; μ – функция, описывающая размножение популяции. Она может зависеть от концентрации клеток x , концентрации субстрата (обычно обозначается S), температуры, pH среды и прочих факторов; ν – скорость вымывания.

В хорошо перемешиваемой культуре скорость вымывания зависит только от скорости потока. Если объем культиватора равен V , а скорость притока f , то величина, называемая разбавлением, определяется как $D=f / V$ и тогда скорость вымывания микроорганизмов из культиватора

$$\nu = - D . \quad (2)$$

Без учета вымывания клеток рост биомассы описывается уравнением:

$$\frac{dx}{dt} = \mu x . \quad (3)$$

При неограниченных ресурсах питательных веществ величина μ постоянна, и уравнение (2) описывает экспоненциальный рост популяции клеток. Если же какие-либо причины начинают лимитировать рост, величина μ будет уменьшаться. Для микробиологических систем обычно величина, лимитирующая рост, это концентрация субстрата. Наиболее распространенная форма записи, учитывающая насыщение скорости роста культуры по питательному субстрату, предложена Моно:

$$\frac{dx}{dt} = \frac{\mu_m S}{K_S + S} x . \quad (4)$$

Здесь μ_m – максимальная скорость роста микроорганизмов при данных условиях; K_S – константа, численно равная концентрации субстрата, при которой скорость роста культуры равна половине максимальной. График

функции величины скорости роста от концентрации субстрата приведен на рис. 3.

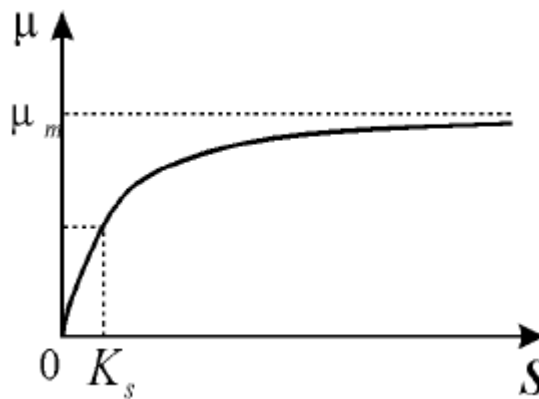


Рис. 3. График зависимости скорости роста от концентрации субстрата в соответствии с формулой Моно (4)

Вид уравнения Моно аналогичен формуле Михаэлиса-Ментен из ферментативной кинетики. И это не только формальное сходство. В основе жизнедеятельности любой клетки лежат ферментативные процессы. Скорость роста биомассы в конечном счете определяется скоростью переработки лимитирующего субстрата ферментом узкого места в метаболической сети. Пусть концентрация фермента на единицу биомассы равна E_0 . Тогда по закону Михаэлиса, скорость переработки субстрата единицей биомассы определяется формулой:

$$\frac{1}{x} \frac{dS}{dt} = - \frac{kE_0 S}{K_m + S}. \quad (5)$$

Здесь K_m - константа Михаэлиса, k - константа скорости реакции. Вся биомасса концентрации x обладает количеством фермента $E_0 x$, Следовательно, суммарная скорость убыли субстрата равна

$$\frac{dS}{dt} = - \frac{kE_0 S}{K_m + S} x \quad (6)$$

Предположим, что прирост биомассы пропорционален убыли субстрата:

$$\frac{dx}{dt} = - \frac{1}{\alpha} \frac{dS}{dt} \quad (7)$$

Обозначив $K_0 = K_m$ и $\mu_m = kE_0/\alpha$, получим формулу (4).

В формулах (4) и (6) имеются важные различия. Формула Михаэлиса-Ментен (6) относится к отдельной ферментативной реакции, все входящие в нее константы выражаются через скорости соответствующих биохимических реакций. В формуле Моно (4) константы скоростей K_S и μ_m являются эффективными величинами и определяются по эмпирической зависимости скорости роста культуры от концентрации питательного субстрата.

При моделировании конкретной культуры микроорганизмов часто нелегко выделить лимитирующий фактор. Здесь может играть роль соотношение коэффициентов растворимости различных веществ или проницаемости мембран клеток по отношению к этим веществам. Только специально поставленные эксперименты могут выделить управляющее звено - лимитирующий субстрат, который входит в формулу (4)

В стационарном состоянии процессы размножения популяции и вымывания должны быть уравновешены. При непрерывном культивировании подбором скорости потока можно стабилизировать скорость роста популяции в любой точке на восходящей ветви кривой роста популяции. Для этого применяются различные способы управления скоростью потока. Основное их свойство - обратная связь между приростом концентрации биомассы и удалением части популяции из ферментера. В различных культурах применяются разные физико-химические методы поддержания плотности культуры на разном уровне: турбидостатный, основанный на регулировании оптической плотности культуры, рН-статный для процессов, в которых имеется связь между приростом биомассы и изменениями рН, оксигенатный - для аэробных микроорганизмов. Эти способы управления дают возможность поддерживать культуру в условиях нелимитированного роста, когда скорость прироста биомассы определяется лишь собственной генетически обусловленной способностью популяции к размножению. При этом достигаются очень высокие скорости, которые особенно важны при изучении микроэволюционных процессов. Например, бактерии могут размножаться в турбидостате со скоростью, соответствующей средней продолжительности поколения - около 5 мин.

Для поддержания культуры в области нелимитированного роста требуются внешние регуляторы. В случае лимитирования роста внешним фактором, например, недостатком субстрата, стационарный режим работы культиватора устанавливается путем саморегуляции. Это имеет место в природных проточных системах и в наиболее распространенном типе непрерывных культиваторов - хемостате, где задается скорость разбавления культуры, или скорость потока.

Наиболее устойчиво работает хеостат в пределах скорости потока, малых по сравнению с максимальной удельной скоростью роста культуры. В области сравнимых значений этих величин система становится неустойчивой: малые колебания скорости потока могут приводить к заметным изменениям концентрации биомассы и даже к вымыванию культуры из культиватора. Теория хеостата впервые была разработана Моно (1950) и Гербертом (1956) и с той поры постоянно совершенствуется. Однако основы ее остались незыблемыми. На них мы и сосредоточим свое внимание.

2. Анализ модели Моно.

При непрерывном перемешивании можно считать весь объем культиватора однородно заполненным, концентрации субстрата и клеток в каждой точке культиватора одинаковыми, и описывать поведение этих концентраций во времени с помощью системы обыкновенных дифференциальных уравнений:

$$\begin{aligned}\frac{dx}{dt} &= \mu(S)x - Dx, \\ \frac{dS}{dt} &= DS_0 - \alpha\mu(S)x - DS, \\ \mu(S) &= \frac{\mu_m S}{K_m + S}.\end{aligned}\tag{8}$$

Здесь S - концентрация субстрата, x - концентрация клеток в культиваторе, S_0 - концентрация субстрата, поступившего в культиватор, D - скорость потока (разбавления) культуры, α - “экономический” коэффициент, показывающий, какая часть поглощенного субстрата идет на приращение биомассы.

Поясним смысл членов, входящих в правые части уравнений. В первом уравнении: $\mu(S)x$ - прирост биомассы за счет поглощения субстрата, $-Dx$ - отток биомассы из культиватора.

Во втором уравнении: $-\alpha\mu(S)x$ - количество субстрата, поглощенного клетками культуры, DS_0 - приток субстрата в культиватор, $-DS$ - отток неиспользованного субстрата из культиватора.

Скорость роста биомассы предполагается зависящей только от концентрации субстрата в соответствии с формулой Моно (третье уравнение).

Исследуем тип стационарных режимов и переходных процессов в культиваторе, используя методы, изученные ранее.

Введем безразмерные концентрации, время и скорость потока

$$x' = \alpha x / K_S, y = S / K_S, y_0 = S_0 / K_S, t' = t \mu_m, D' = D / \mu_m.$$

Штрихи у новых переменных опустим. В новых переменных система имеет вид:

$$\begin{aligned} \frac{dx}{dt} &= \mu(y)x - Dx, \\ \frac{dy}{dt} &= -\mu(y)x + D(y_0 - y), \\ \mu(y) &= \frac{y}{1+y}. \end{aligned} \quad (9)$$

Найдем стационарные концентрации биомассы и субстрата. Приравняем правые части уравнений нулю

$$\left(\frac{\bar{y}}{1+\bar{y}} - D \right) \bar{x} = 0, \quad \frac{\bar{y}}{1+\bar{y}} \bar{x} + D(y_0 - \bar{y}) = 0. \quad (10)$$

Система алгебраических уравнений (10) имеет два решения, следовательно, система дифференциальных уравнений (9) имеет два стационарных состояния

$$\bar{x}_1 = 0, \quad \bar{y}_1 = y_0; \quad (11)$$

$$\bar{x}_2 = y_0 - \frac{D}{1-D}, \quad \bar{y}_2 = \frac{D}{1-D}. \quad (12)$$

Примем во внимание, что безразмерная концентрация клеток x имеет смысл только при значениях $x > 0$, а безразмерная концентрация субстрата y ограничена сверху значением $y_0 = S_0 / K_S$ - концентрацией притекающего субстрата. Легко видеть, что ненулевое стационарное значение биомассы (12) имеет смысл только в случае, когда безразмерная скорость потока D меньше определенной величины

$$D \leq \frac{y_0}{1 + y_0} = D_B. \quad (13)$$

Граничное значение скорости протока называется скоростью вымывания. В размерном виде его величина равна:

$$D_B = \frac{\mu_m y_0}{K_S + y_0}. \quad (14)$$

При скоростях протока, больших D_B , прирост биомассы не может компенсировать ее отток, и культура полностью вымывается из культиватора.

Определим характер устойчивости стационарных состояний системы, используя метод линеаризации системы в окрестности стационарного состояния.

Характеристический определитель системы (9) имеет вид.

$$\begin{vmatrix} \mu(\bar{y})x - D - \lambda & \frac{\bar{x}}{(1 + \bar{y})^2} \\ -\mu(\bar{y}) & \frac{\bar{x}}{(1 + \bar{y})^2} - D - \lambda \end{vmatrix} \quad (15)$$

Исследуем характер устойчивости режима вымывания - особой точки с координатами (11). В этом случае

$$\mu(y_0) = \frac{y_0}{1 + y_0} = D_B. \quad (16)$$

и характеристический определитель принимает вид

$$\begin{vmatrix} D_B - D - \lambda & 0 \\ -D_B & -D - \lambda \end{vmatrix} = 0. \quad (17)$$

Корни характеристического уравнения (17)

$$\lambda_1 = -D, \quad \lambda_2 = D_B - D, \quad (18)$$

действительны и имеют различные знаки при $D < D_B$, то есть при скоростях разбавления, меньших скоростей вымывания. При этом точка $(0, y_0)$ неустойчива - седло.

Если же $D > D_B$ - оба корня отрицательны, и особая точка (11) является устойчивым узлом. Этот режим называется режимом вымывания.

Концентрация субстрата в культиваторе равна при этом концентрации поступающего субстрата S_0 , а концентрация биомассы равна нулю. Если в такой культиватор заложить “затравку”, микроорганизмы будут вымыты из культиватора, не успев размножиться.

Для второй особой точки с координатами (12) корни характеристического уравнения равны

$$\lambda_1 = -D, \quad \lambda_2 = -(D_B - D)(1 + y_0)(1 - D), \quad (19)$$

Напомним, что это ненулевое по биомассе состояние равновесия существует в положительном квадранте фазовой плоскости лишь при значениях скорости разбавления $D < D_B$. Так как

$$D_B = \frac{y_0}{1 + y_0} < 1.$$

то все три сомножителя, входящие в выражение для λ_2 в (19) положительны. Следовательно, $\lambda_2 < 0$, и точка (12) - устойчивый узел. Это и есть рабочее состояние проточного культиватора.

Фазовые портреты системы для двух значений скоростей протока $D < D_B$ и $D > D_B$ приведены на рис.4 (а,б)

Уравнение изоклины горизонтальных касательных получим, приравняв правую часть второго уравнения (9) нулю

$$x = \frac{D(y_0 - y)(1 + y)}{y}. \quad (20)$$

Изоклины вертикальных касательных на рисунках 4: ось $x=0$ и прямая

$$y = \frac{D}{1 - D}. \quad (21)$$

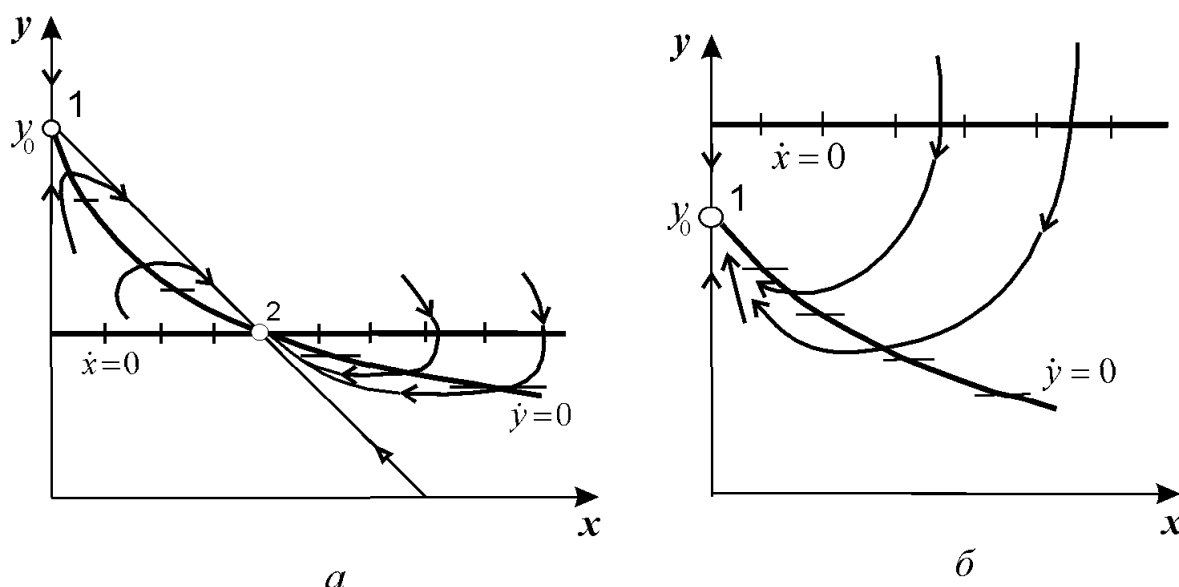


Рис. 4. Фазовые портреты системы (9). а – стационарный режим работы, б – режим вымывания.

В случае, когда $D < D_B$ главные изоклины (20) и (21) пересекаются в положительном квадранте, и точка их пересечения является устойчивым узлом, а точка пересечения кривой (20) с осью $x = 0$ – седлом (рис.4 а)

В случае $D > D_B$ главные изоклины (20) и (21) пересекаются вне положительного квадранта, и устойчивым узлом будет особая точка (11), соответствующая режиму вымывания (рис.4б).

Рассмотренная модель является упрощенной и для описания реальных процессов требует дополнений. Например, при больших концентрациях субстрат может оказывать ингибирующее действие, и тогда формулу для скорости роста следует записывать в виде:

$$\mu(S) = \frac{\mu_m S}{K_S + S + AS^2}. \quad (22)$$

В системе, где существует такая зависимость скорости роста от субстрата, возможны триггерные режимы - наличие двух устойчивых стационарных состояний и зависимость стационарных значений концентраций субстрата и биомассы от начальных условий (от величины затравки и начальной концентрации биомассы).

На скорость роста биомассы может оказывать влияние концентрация продуктов метаболизма в среде, окружающей клетку. Тогда к двум уравнениям, описывающим динамику концентрации биомассы и субстрата в непрерывном процессе культивирования, следует добавить третье уравнение, выражающее динамику концентрации продуктов метаболизма. При этом

скорость роста биомассы будет зависеть как от концентрации субстрата. Так и от концентрации продукта. Наиболее известную формулу такой зависимости предложил Иерусалимский:

$$\mu(S) = \frac{\mu_m S}{(K_S + S)(K_P + P)} . \quad (23)$$

Формула (23) известна как формула Моно-Иерусалимского.

Исследование модели, учитывающей ингибирующее действие продукта показывает, что значение скорости вымывания в такой системе совпадает с величиной DB , полученной выше для модели Моно. В то же время ингибирующее влияние продукта ведет к значительному уменьшению стационарных концентраций биомассы.

Учебно-методическое обеспечение дисциплины

Учебная литература (основная)

- 1) Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика : учеб. пособие / В. Е. Гмурман. - 12-е изд., перераб.. - М. : Высш. шк., 2008. - 479 с. : ил.. - (Основы наук)
- 2) Гмурман, Владимир Ефимович. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для студентов вузов / В. Е. Гмурман. - 11-е изд., перераб.. - М. : Высшее образование, 2009 [т.е. 2008]. - 403, [1] с. : ил. ; 22. -
- 3) Крючков, Александр Васильевич. Биометрия : учебное пособие / А. В. Крючков, И. В. Маракулин ; ВятГУ, БФ, каф. МБ. - Киров : [б. и.], 2011. - 115 с.. - Библиогр.: с. 115
- 4) Даровских, Д. А. Математическая статистика [Электронный ресурс] : задания к типовому расчету по дисциплине "Математика": для студентов химико-биологических специальностей / Д. А. Даровских ; ВятГУ, ФПМТ, каф. ВМ. - Киров : [б. и.], 2010. - 29 с.
- 5) Теория вероятностей и математическая статистика. Математические модели : учеб. пособие для студентов высших учебных заведений, обучающихся по направлению "Биология" / В.Д. Мятлев, Л.А. Панченко, Г.Ю. Ризниченко, А.Т. Терехин. - М. : Академия, 2009. - 314, [1] с. : ил.. - (Университетский учебник. Высшая математика и ее приложения к биологии). - Библиогр.: с. 307-311

Учебная литература (для углубленного изучения)

- 1) Гмурман, Владимир Ефимович. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие / В. Е. Гмурман. - 10-е изд., стер.. - М. : Высш. шк., 2005. - 405 с. : ил.
- 2) Дюк, Вячеслав Анатольевич. Информационные технологии в медико-биологических исследованиях / В. А. Дюк, В. Л. Эмануэль. - СПб. : Питер, 2003. - 528 с. : ил.. - Библиогр.: с. 528
- 3) Ризниченко, Галина Юрьевна. Лекции по математическим моделям в биологии [Текст] / Г. Ю. Ризниченко. - М. : НИЦ "Регулярная и хаотическая динамика". Ч. 1 : Описание процессов в живых системах во времени. - 2002. - 232 с.
- 4) Рубин, Андрей Борисович. Биофизика [Текст] : учеб. / А. Б. Рубин. - 3-е изд., испр. и доп.. - М. : Изд-во МГУ : Наука. - (Классический университетский учебник). Т. 1 : Теоретическая биофизика. - 2004. - 448 с. : ил.
- 5) Вуколов, Эдуард Александрович. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учебное пособие / Э. А. Вуколов. - 2-е изд., испр. и доп.. - М. : ФОРУМ, 2011. - 463 с

б) Туганбаев, Аскар Аканович. Теория вероятностей и математическая статистика : учеб. пособие / А. А. Туганбаев, В. Г. Крупин. - СПб. ; М. ; Краснодар : Лань, 2011. - 223 с.. - Библиогр.: с. 221

Учебно-методические издания

1) Методические указания к практическим занятиям по дисциплине "Биометрия" [Электронный ресурс] : специальность 012400, курс 4. Специальность 070100, курс 3 / ВятГУ, БФ, каф. МБ ; сост. А. В. Крючков. - Киров : [б. и.], 2006

2) Крючков, Александр Васильевич. Сборник задач по дисциплине "Биометрия" [Электронный ресурс] : для студентов, обучающихся по специальностям 240901, 020209 и направлению 020400 / А. В. Крючков ; ВятГУ, БФ, каф. МБ. - Киров : [б. и.], 2011. - 35 с.

3) Крючков, Александр Васильевич. Сборник задач по дисциплине "Биометрия" : для студентов, обучающихся по специальностям 240901, 020209 и направлению 020400 / А. В. Крючков ; ВятГУ, БФ, каф. МБ. - Киров : [б. и.], 2011. - 32 с.. - Библиогр.: с. 32

Периодические издания

1) Вопросы статистики [Электронный ресурс]. - Электрон. журн.. - М. : АНО Информационно-издательский центр Статистика России. - . - Загл. с титул. экрана. - Электрон. версия печ. публикации Полный текст в электронном виде доступен на платформе eLIBRARY.RU. Для доступа к журналу необходима персональная регистрация. (2014г., N1-5; 2013г., N1-12; 2012г., N1-12)