**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования**

**«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ЛИНГВИСТИЧЕСКИЙ УНИВЕРСИТЕТ им. Н.А. ДОБРОЛЮБОВА»**

Д.Ю. Акатьев, Н.Д. Чикова

# СБОРНИК ЗАДАЧ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Нижний Новгород

2015

Печатается по решению редакционно-издательского совета ФГБОУ ВПО «НГЛУ».

Направление подготовки: 09.06.01– *Информатика и вычислительная техника* (профиль 05.13.17 – *Теоретические основы информатики).*

УДК 519.2
ББК 22
А 383

Акатьев Д.Ю., Чикова Н.Д. Сборник задач по теории вероятностей и математической статистике. – Н. Новгород: ФГБОУ ВПО «НГЛУ», 2015. – 96 с.

Предлагаемое учебное пособие выстроено согласно действующему государственному стандарту по дисциплине «Основы теории вероятностей» и в совокупности с конспектом лекций профессора В.В. Савченко по данному курсу [1] обеспечивает самостоятельную работу студентов над учебным материалом. Приведены необходимые формулы и определения для решения задач основных разделов теории вероятностей и математической статистики. Рассмотрено решение типовых задач, а также даны вопросы и задачи для самостоятельной проработки.

Пособие предназначено для самостоятельной работы студентов экономических и лингвистических направлений подготовки в НГЛУ, а также для аспирантов по направлению подготовки 09.06.01– *Информатика и вычислительная техника*, профиль 05.13.17 – *Теоретические основы информатики*.

УДК 519.2
ББК 22

Авторы: Д.Ю. Акатьев, канд. техн. наук, доцент, профессор кафедры математики и информатики ФГБОУ ВПО «НГЛУ»

Н.Д. Чикова, старший преподаватель кафедры математики и информатики ФГБОУ ВПО «НГЛУ»

Рецензент: В.В. Савченко, д-р техн. наук, профессор, зав. кафедрой математики и информатики ФГБОУ ВПО «НГЛУ»

## Введение

Задача любой науки состоит в выявлении и исследовании закономерностей, которым подчиняются реальные процессы. Найденные закономерности, относящиеся к управлению и экономике, имеют не только теоретическую ценность, они широко применяются на практике – в планировании, управлении и прогнозировании.

**Теория вероятностей** – математическая наука, изучающая закономерности случайных явлений.

**Математическая статистика –** раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений с целью выявления статистических закономерностей.

Математическая статистика опирается на теорию вероятностей. Теория вероятностей позволяет находить вероятности «сложных» событий или явлений через вероятности «простых» событий (связанных с ними каким-либо образом). Математическая статистика по наблюдаемым значениям (выборочным данным или выборке) оценивает вероятности этих событий либо осуществляет проверку предположений (гипотез) относительно этих вероятностей.

При большом числе наблюдений случайные воздействия в значительной мере удаётся определить и устранить, и получаемый результат оказывается практически неслучайным, предсказуемым. Этот принцип и является базой для практического использования вероятностных и математико-статистических методов исследования. **Цель указанных методов** состоит в том, чтобы, минуя сложное исследование отдельного случайного явления, изучить закономерности массовых случайных явлений, прогнозировать их характеристики, влиять на ход этих явлений, контролировать их, ограничивать область действия случайности. Развитие информационных технологий и применение статистических программных пакетов сделали эти методы более доступными и наглядными, так как трудоёмкую работу по расчету различных статистик, параметров, характеристик, построению таблиц и графиков в основном стал выполнять компьютер, а исследователю осталась главным образом творческая работа: постановка задачи, выбор методов её решения и интерпретация результатов. Наиболее популярные в России универсальные и специализированные статистические пакеты: отечественные *STADIA*, Эвриста, американские *STATGRAPHICS*, *STATISTICA* и др.

## 1. События и определение вероятности события

### 1.1. Классификация событий

Наблюдаемые нами **события** (явления) можно **подразделить на** следующие **три вида**: достоверные, невозможные и случайные.

**Достоверным** называют событие, которое обязательно произойдёт, если будет осуществлена определённая совокупность условий *S*.

***Пример 1***. Если в сосуде содержится вода при нормальном атмосферном давлении и температуре $20^0$, то событие $A$ = {вода в сосуде находится в жидком состоянии} есть достоверное. В этом примере заданны атмосферное давление и температура воды, которые составляют совокупность условий *S*.

Невозможным **называют событие, которое заведомо не произойдёт, если будет осуществлена совокупность условий S.**

***Пример 2.*** Событие $B$ = {вода в сосуде находится в твёрдом состоянии} заведомо не произойдет, если будет осуществлена совокупность условий предыдущего примера.

**Случайным** называют событие, которое при осуществлении совокупности условий *S* может либо произойти, либо не произойти. В дальнейшем, вместо того чтобы говорить «совокупность условий *S* осуществлена», будем говорить кратко: «произведено испытание». Таким образом,

**событие** будет рассматриваться как **результат испытания (опыта)**. События обозначаются прописными буквами латинского алфавита: $A, B, C\ldots$.

***Пример 3.*** Стрелок стреляет по мишени, разделённой на области. Выстрел – это испытание. Попадание в определённую область мишени – событие.

***Пример 4.*** Проведем испытание, состоящее в угадывании буквы при следующем комплексе условий: угадываемой букве предшествует цепочка *которо*, текст русский без ошибок и опечаток. Это испытание может дать события:

$A$ = {появление буквы *г* (*которого*)},

$B$ = {появление буквы *й* (*которой*)},

$C$ = {появление буквы *м* (*котором*)},

$D$ = {появление буквы *е* (*которое*)}.

События называются ***несовместными***, если появление одного из них исключает появление других событий в одном и том же опыте.

***Пример 5.*** Из ящика с деталями наудачу извлечена деталь. Появление стандартной детали исключает появление нестандартной детали. События $A$ = {появилась стандартная деталь} и $B$ = {появилась нестандартная деталь} – несовместные.

***Пример 6.*** Брошена монета. Появление «реверса» исключает появление «аверса». События $A$ = {появился реверс} и $B$ = {появился аверс} – несовместные.

События называются **совместными**, если появление одного из них не исключает появления другого события в одном и том же опыте.

***Например.*** Подбрасываются две монеты. События $A$ = {появился реверс на первой монете} и $B$ = {появился аверс на второй монете} – совместные.

***Пример 7.*** Получение студентом на экзамене по одной дисциплине оценок «отлично», «хорошо», «удовлетворительно» – события несовместные, а получение тех же оценок на экзаменах по трём дисциплинам – события совместные.

Несколько событий образуют **полную группу (полную систему),** если они являются единственно возможными и несовместными исходами испытания. Это означает, что в результате испытания обязательно должно произойти одно и только одно из этих испытаний.

***Пример 8.*** Приобретены два билета денежно-вещевой лотереи. Обязательно произойдёт одно и только одно из следующих событий:

$A$ = {выигрыш выпал на первый билет и не выпал на второй},

$B$ = {выигрыш не выпал на первый билет и выпал на второй},

$C$ = {выигрыш выпал на оба билета},

$D$ = {на оба билета выигрыш не выпал}.

Эти события образуют полную группу несовместных событий.

Частным случаем событий, образующих полную группу, являются противоположные события. Два несовместных события, из которых одно должно обязательно произойти, называются **противоположными**. Событие противоположное событию $A$, будем обозначать $\overline{A}$.

***Например.*** События $A$ = {отсутствие бракованных изделий} и $\overline{A}$ = {наличие хотя бы одного бракованного изделия} – события противоположные.

События называют **равновозможными**, если есть основания считать, что ни одно из них не является более возможным, чем другое.

***Пример 9.*** Появление того или иного числа очков на брошенной игральной кости – равновозможные события. Предполагается, что игральная кость изготовлена из однородного материала, имеет форму правильного многогранника, и наличие очков не оказывает влияния на выпадение любой грани.

### 1.2. Классическое определение вероятности

Для практической деятельности важно уметь сравнивать события по степени возможности их наступления. Поэтому для сравнения событий нужна определённая мера.

Численная мера степени объективной возможности наступления события называется **вероятностью события**.

Это определение, **качественно** отражающее понятие вероятности события, не является математическим. Чтобы оно стало таким, определим его **количественно**.

Пусть исходы некоторого испытания образуют полную группу событий и равновозможные. Такие исходы называются **элементарными исходами, случаями** или **шансами**. Случай называется **благоприятствующим (благоприятным)** событию $A$, если появление этого случая влечет за собой появление события $A$.

**Классическое определение вероятности.** Вероятность события равна отношению числа элементарных исходов, благоприятствующих данному событию, к общему числу элементарных исходов опыта:

$$P(A)=\frac{m}{n},$$

где $P(A)$ – вероятность события $A$; $m$ – число элементарных исходов, благоприятствующих событию А; $n$ – общее число элементарных исходов.

***Пример 10.*** Текст «Капитанской дочки» А.С. Пушкина состоит из 29343 словоупотреблений. Формы слова <u>*быть*</u> встречаются здесь 430 раз. Определить вероятность появления в тексте «Капитанской дочки» форм слова <u>*быть*</u>?

***Решение.*** Пусть событие $A$ = {появление в тексте форм слова *быть*}, тогда

$$P(A)=\frac{m}{n}=\frac{430}{29343}\approx 0{,}0147\ .$$

Из классического определения следуют **свойства** вероятности:

1. *Вероятность достоверного события всегда равна* 1.
2. *Вероятность невозможного события равна* 0.
3. *Вероятность случайного события заключена между нулём и единицей, т. е.*

$$0\le P(A)\le 1.$$

4. *Вероятность противоположного события* $\overline{A}$ *равна дополнению вероятности данного события A до* 1*, т. е.*

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

При расчёте вероятности по классическому методу часто применяют комбинаторный анализ или комбинаторику.

**Основные формулы комбинаторики.** Комбинаторика изучает количество комбинаций, которое можно составить из элементов любой природы, заданного конечного множества. Приведём наиболее употребительные соотношения.

**Число комбинаций**. Пусть имеется $m$ множеств по $n_1\cdot n_2\cdot\ldots\cdot n_m$ элементов в каждом. Выбрать по одному объекту из каждого множества можно $n_1\cdot n_2\cdot\ldots\cdot n_m$ способами.

***Пример 11.*** В группе 16 человек. Необходимо выбрать старосту, его заместителя и профорга. Сколько существует способов это сделать?

***Решение.*** Старостой может быть выбран любой из 16 студентов, его заместителем – любой из оставшихся 15, а профоргом – любой из оставшихся 14 студентов, т. е. $n_1$ = 16, $n_2$ = 15, $n_3$ = 14. Общее число способов выбора старосты, его заместителя и профорга равно $n_1 \cdot n_2 \cdot n_3 = 16 \cdot 15 \cdot 14 = 3360$.

**Перестановками** называют комбинации, состоящие из одних и тех же *n* различных элементов и отличающиеся только порядком их расположения. Число всех возможных перестановок:

$$P_n = n!,$$

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

***Пример 12.*** Сколько трёхсловных предложений можно построить из трёх слов: <u>*сегодня*</u>, <u>*идет*</u>, <u>*дождь*</u>?

***Решение.*** Число предложений равно числу перестановок из трёх элементов:

$$P_n = 3! = 1 \cdot 2 \cdot 3 = 6.$$

**Перестановки с повторениями**. Если в перестановках из общего числа *n* элементов есть *m* различных элементов, при этом 1-й элемент повторяется $n_1$ раз, 2-й элемент – $n_2$ раз, *m*-й элемент – $n_k$ раз, причем $n_1 + n_2 + \ldots + n_k = n$, то такие перестановки называют **перестановками с повторениями** из *n* элементов. Число перестановок с повторениями из *n* элементов равно:

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! n_2! \ldots n_k!}$$

**Размещениями** называют комбинации, составленные из *n* различных элементов по *m* элементов, которые отличаются либо составом элементов, либо их порядком. Число всех возможных размещений равно:

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1) = \frac{n!}{(n-m)!}.$$

При $n = m$, $A_n^m = P_n$.

***Пример 13.*** Сколько имеется вариантов занятия трех призовых мест 8 спортсменами одного уровня?

***Решение.*** Искомое число вариантов:

$$A_8^3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = 8 \cdot 7 \cdot 6 = 336 .$$

**Сочетаниями** называют комбинации, составленные из $n$ различных элементов по $m$ элементов, которые отличаются хотя бы одним элементом. Число сочетаний рассчитывается по соотношению:

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

***Пример 14.*** В урне 4 белых и 2 чёрных шара. Из урны вынимают наудачу два шара. Найти вероятность того, что а) оба шара будут белыми; b) хотя бы один будет белым.

***Решение.*** а. Обозначим событие $A$ = {два белых шара}. Общее число возможных случаев равно:

$$C_6^2 = \frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 4} = 15 .$$

Число исходов благоприятствующих появлению события $A$ равно:

$$C_4^2 = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{2 \cdot 3 \cdot 4}{2 \cdot 2} = 6 .$$

Тогда по классическому определению вероятности

$$P(A) = \frac{m}{n} = \frac{6}{15} = \frac{2}{5} = 0{,}4 .$$

б. Обозначим событие $B$ = {два черных шара}. Число исходов благоприятствующих появлению события $B$ равно $C_2^2 = 1$. Следовательно, $P(B) = \frac{m}{n} = \frac{1}{15}$. Определим вероятность события С = $\overline{B}$ {хотя бы один шар белый}. По свойству (4) противоположных событий получим:

$$P(C)=1-\frac{1}{15}=\frac{14}{15}.$$

## 1.3. Статистическое определение вероятности

**Статистической вероятностью** события $A$ называется относительная частота появления события $A$ в $n$ произведенных испытаниях

$$\widetilde{P}(A)=\lim_{n\to\infty} w(A)=\frac{m}{n},$$

где $m$ – число испытаний, в которых события $A$ наступило;
$n$ – общее число произведенных испытаний.

При статистическом определении в качестве вероятности события принимают его относительную частоту при количестве испытаний стремящихся к бесконечности.

## 1.4. Геометрическое определение вероятности

При классическом определении вероятности не всегда можно определить числа $m$ и $n$ для вычисления вероятности событий, и поэтому непосредственно пользоваться формулой $P(A)=m/n$ не удается. В таких случаях вводят понятие геометрической вероятности, т. е. вероятности попадания точки в область (отрезок, часть плоскости, часть тела и др.)

Пусть, например, имеется некоторая область $G$ и в ней содержится другая область $g$. Требуется найти вероятность того, что точка, взятая наудачу в области $G$, попадет в область $g$. Тогда вероятность попадания точки в какую-либо часть области $g$ пропорциональна мере ($mes$) этой части (длине, площади, объему и т. п.) и не зависит от ее расположения и формы. В результате получим геометрическое определение вероятности:

$$P=\frac{mes g}{mes G}.$$

***Пример 15.*** Внутри эллипса $\frac{x^2}{25}+\frac{y^2}{16}=1$ расположен круг $x^2+y^2=9$. Найти вероятность попадания точки в кольцо, ограниченное эллипсом и кругом.

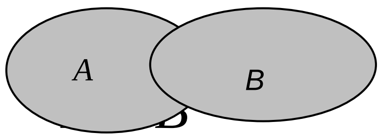
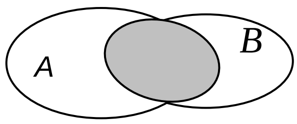
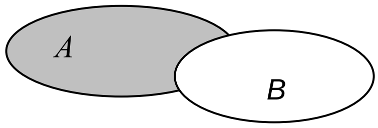
***Решение.*** Пусть событие А = {попадание точки в кольцо, ограниченное эллипсом и кругом}, тогда:

$$P(A)=S_{кол}/S_{эл},\text{ где } S_{кол}=S_{эл}-S_{кр}=\pi ab-\pi r^2.$$

Так как *a = 5, b = 4, r = 3*, то $P(A)=(20\pi-9\pi)/(20\pi)=11/20=0{,}55$.

## 1.5. Действия над событиями

Ведём понятия суммы, произведения, разности событий и дадим их геометрическую интерпретацию с помощью диаграмм Эйлера – Венна.

| ОПРЕДЕЛЕНИЕ | ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ |
|---|---|
| **Суммой** двух событий (*A*+*B*) называется событие, состоящее из всех исходов, входящих либо в *A*, либо в *B*. |  |
| **Произведением** двух событий (*AB*) называется событие, состоящее из тех исходов, которые входят как в *A*, так и в *B*. Т. е. события *A* и *B* наступают одновременно. |  $A\cap B$ |
| **Разностью** двух событий (*A*-*B*) называется событие, состоящее из исходов, входящих в *A*, но не входящих в *B*. Т. е. событие *A* наступает, а событие *B* не наступает. |  $A\setminus B$ |

***Пример 16***. Двое играют в шахматы. Событие $A$ означает, что выиграл первый игрок, событие $B$ – выиграл второй игрок. Что означают события:

1) $\overline{B}$ ;
2) $\overline{A}+\overline{B}$ ;
3) $\overline{B}$ − А?

***Решение.*** Учтём, что $AB=\varnothing$, так как одновременное наступление событий $A$ и $B$ невозможно. Тогда:

1. $\overline{B}$ = {либо А, либо НИЧЬЯ};
2. $\overline{A}+\overline{B}$ = {либо А, либо В, либо НИЧЬЯ};
3. $\overline{B}$ − А = {НИЧЬЯ}.

## Вопросы и задачи

1. Что называется событием?
2. Монета подбрасывается три раза подряд. Исходом каждого бросания служит выпадение «реверса» – Р или «аверса» – А. Описать пространство элементарных событий данного эксперимента и описать событие $B$, состоящее в том, что выпало не менее двух «реверсов».
3. Приведите пример достоверного события.
4. Чему равна вероятность невозможного события?
5. На полке стоят 8 учебников. Из них 3 – по математике. Наугад берём два учебника. Найдите вероятность того, что они оба по математике. (р = 0,107).
6. Цифры 0, 1, 2, 3 записаны на 4 карточках. Сколько различных 4-значных чисел можно составить из этих карточек? (18)
7. Что называется частотой события?
8. Приведите пример произведения двух событий.

9. Из таблицы чисел взято одно число. Событие $A$ = {выбранное число делиться на 5}; событие $B$ = {выбранное число оканчивается нулём}. Что означают события $A - B$ и $\overline{B}A$? (Оканчивается на 5).

10. Пусть при бросании игральной кости событие $A$ = {выпало четное число}, событие $B$ = {выпало число кратное трем}. Найти **a)** $\boldsymbol{A + B}$**, b)** $\boldsymbol{AB}$**, c)** $\boldsymbol{A - B}$**, d)** $\overline{B}$ **.**

11. Ещё в XIII веке француз Шевалье де Мере задался вопросом: какая сумма очков имеет больше шансов выпасть при бросании двух игральных кубиков – 11 или 12? На первый взгляд шансы равны. Так ли это?

12. В коробке 5 одинаковых изделий, причем 3 из них окрашены. Наудачу извлечены 2 изделия. Найти вероятность того, что среди отобранных изделий:
    1) одно окрашено;
    2) два окрашены;
    3) хотя бы одно окрашено.

    (1. $p$ = 0,6; 2. $p$ = 0,3; 3. $p$ = 0,9).

## 2. Основные теоремы теории вероятностей

### 2.1. Условная вероятность событий

Вероятность $P(B)$ как мера степени объективной возможности наступления события $B$ имеет смысл при выполнении определённого комплекса условий. При изменении условий вероятность события $B$ может измениться. Так, если к комплексу условий, при котором изучалась вероятность $P(B)$, добавить новое условие $A$, то полученная вероятность события $B$, найденная при условии, что событие $A$ произошло, называется **условной вероятностью события** $B$ и обозначается $P(B/A)$.

***Пример 17.*** В ящике 10 деталей, среди которых 6 стандартные и 4 бракованные. Поочерёдно из него извлекается по одной детали (с возвратом и без возврата). Найти вероятность извлечения во второй раз стандартной детали.

***Решение.*** Пусть событие $A$ = {извлечение стандартной детали в 1-й раз}, событие $B$ = {извлечение стандартной детали во 2-й раз}. Очевидно, что $P(A)=\frac{6}{10}=0{,}6$. Если вынутая деталь вновь возвращается в ящик, то вероятность извлечения стандартной детали во второй раз $P(B)=\frac{6}{10}=0{,}6$. Если вынутая деталь в ящик не возвращается, то вероятность извлечения стандартной детали во второй раз $P(B)$ зависит от того, какая деталь была извлечена в первый раз – стандартная (событие $A$) или бракованная (событие $\overline{A}$). В первом случае $P(B/A)=\frac{5}{9}$, во втором случае $P\left(B/\overline{A}\right)=\frac{6}{9}=\frac{2}{3}$. Таким образом, вероятность события $B$ существенным образом изменяется в зависимости от того, произошло или нет событие $A$.

### 2.2. Теорема (правило) умножения вероятностей

Вероятность совместного появления двух событий равна произведению вероятности одного из них на условную вероятность другого, найденную в предположении, что первое событие произошло:

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

Теорема умножения вероятностей принимает наиболее простой вид, когда события независимы. Событие $B$ называется **независимым от события** $A$, если его вероятность не меняется от того, произошло событие $A$ или нет, т. е.

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

$$(P(B/\overline{A}) = P(B)).$$

В противном случае, если $P(B/A) \neq P(B)$, $(P(B/\overline{A}) \neq P(B))$, событие $B$ называется **зависимым** от $A$.

Теорема умножения вероятностей для двух или нескольких независимых событий примет вид:

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

$$P(ABC...KL) = P(A)P(B)P(C)...P(K)P(L).$$

При решении ряда задач требуется найти вероятность суммы двух или нескольких **совместных** событий, т. е. вероятность появления **хотя бы одного** из этих событий.

## 2.3. Теорема сложения вероятностей

**Теорема сложения вероятностей 1.** Вероятность появления одного из двух, трёх, четырёх, и т. д. несовместных событий $A$, $B$, ..., $C$ равна сумме вероятностей каждого события в отдельности, т. е.

$$P(A+B+...+K) = P(A) + P(B) +...+ P(K).$$

**Следствие 1.** Если несовместные события $A_1$, $A_2$, ..., $A_n$ образуют полную группу событий, то сумма их вероятностей равна 1:

$$\sum_{i=1}^{n} P(A_i) = 1.$$

**Следствие 2.** Сумма вероятностей противоположных событий равна 1:

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

**Теорема сложения вероятностей 2.** Вероятность суммы двух совместных событий равна сумме вероятностей этих событий без вероятности их произведения, т. е.

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

В случае трёх и более совместных событий соответствующая формула для вероятности суммы $P(A+B+\ldots+K)$ громоздка и проще перейти к противоположным событиям. Рассмотрим типичные случаи:

| | |
|---|---|
| Вероятность суммы **нескольких совместных событий** равна разности между единицей и вероятностью произведения противоположных событий | $P(A+B+\ldots+K)=1-P(\overline{A}\overline{B}\ldots\overline{K})$ |
| Для **совместных и независимых** события $A,B,\ldots,K$ вероятность равна | $P(A+B+\ldots+K)=1-P(\overline{A})P(\overline{B})\ldots P(\overline{K})$ |
| Для независимых событий **с равными** вероятностями. $P(A)=P(B)=\ldots=P(K)=p$ сумма вероятностей равна | $P(A+B+\ldots+K)=1-(1-p)^n$ |

***Пример 18.*** Вероятность поражения цели первым стрелком (событие А) равна 0,8, а вероятность поражения цели вторым стрелком (событие В) равна 0,7. Какова вероятность того, что цель будет поражена хотя бы одним стрелком?

***Решение.*** Пусть событие С = {цель поражена хотя бы одним стрелком}, противоположное событие $\overline{C}$ = {оба стрелка промахнулись}. Таким образом, $\overline{C}=\overline{A}\overline{B}$. По теореме умножения вероятностей, с учётом того, что событий $\overline{A}$ и $\overline{B}$ независимы, найдем :

$$P(\overline{C}) = P(\overline{A})P(\overline{B}) = [1 - P(A)] \cdot [1 - P(B)] = (1 - 0{,}8) \cdot (1 - 0{,}7) = 0{,}2 \cdot 0{,}3 = 0{,}06.$$

Вероятность того, что цель будет поражена хотя бы одним стрелком:

$$P(C) = 1 - P(\overline{C}) = 1 - 0{,}06 = 0{,}94.$$

Следствием двух основных теорем теории вероятностей – теоремы сложения и теоремы умножения – является формула полной вероятности.

### 2.4. Формула полной вероятности

**Теорема.** Если событие $A$ может произойти только при условии появления одного из событий (гипотез) $H_1, H_2, \ldots, H_n$, образующих полную группу, то вероятность события $A$ равна сумме произведений вероятностей каждого из этих событий (гипотез) на соответствующие условные вероятности события $A$:

$$P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+\ldots+P(H_n)P(A/H_n).$$

Это и есть **формула полной вероятности**.

***Пример 19.*** Имеется английский научно-технический текст общей длиной в 400 тыс. словоупотреблений. По тематике этот текст распадается на следующие выборки разной длины:

1) радиоэлектроника – 200 тыс. словоупотреблений;
2) автомобилестроение – 100 тыс. словоупотреблений;
3) судовые механизмы – 50 тыс. словоупотреблений;
4) строительные материалы – 50 тыс. словоупотреблений.

Словоформа <u>*are*</u> – множественное число настоящего времени глагола *to be* употреблена в 1-й выборке 1610, во 2-й – 1273, в 3-й – 469 и в 4-й – 346 раз. Определить вероятность того, что извлечённое наугад из нашего текста словоупотребление будет словоформой <u>*are*</u>.

***Решение.*** Пусть событие $A$ = {появление словоформы <u>*are*</u>}. Рассмотрим следующие четыре гипотезы: $H_1$ = {словоформа принадлежит к текстам по радиоэлектронике}, $H_2$ = {словоформа принадлежит к текстам по автомобилестроению}, $H_3$ = {словоформа принадлежит к текстам по судовым механизмам}, $H_4$ = {словоформа принадлежит к текстам по строительным материалам}. Считая доли указанных текстов в общей выборке вероятностями наших гипотез, находим:

$$P(H_1)=\frac{200000}{400000}=0{,}5;\ P(H_2)=\frac{100000}{400000}=0{,}25;\ P(H_3)=P(H_4)=\frac{50000}{400000}=0{,}125.$$

Условные вероятности события *A* при этих гипотезах соответственно равны:

$$P(A/H_1)=\frac{1610}{200000}=0{,}008;\ P(A/H_2)=\frac{1273}{100000}=0{,}012;$$

$$P(A/H_3)=\frac{469}{50000}=0{,}009;\ P(A/H_4)=\frac{346}{50000}=0{,}007.$$

Применяя формулу полной вероятности, вычислим вероятность события A = {появление словоформы <u>*are*</u>}:

$$P(A)=P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3) + P(H_4)P(A/H_4) =$$

$$= 0{,}5\cdot0{,}008 + 0{,}25\cdot0{,}012 + 0{,}125\cdot0{,}009 + 0{,}125\cdot0{,}007 = 0{,}009.$$

Следствием теоремы умножения и формулы полной вероятности является формула Байеса.

## 2.5. Теорема гипотез (Формула Байеса)

Пусть событие *A* (которое может произойти только при условии появления одной из (гипотез) $H_1$, $H_2$, …, $H_n$, образующих полную группу), ***произошло*** и необходимо произвести количественную переоценку ***априорных*** вероятностей гипотез $P(H_1)$, $P(H_2)$, ..., $P(H_n)$, известных до испытания, т. е. надо найти ***апостериорные*** (после опытные) условные вероятности гипотез $P(H_i/A)$ ($i$=1,2,…,$n$) по формуле:

$$P(H_i/A)=\frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n}P(H_i)P(A/H_i)}.$$

Значение формулы Байеса состоит в том, что при наступлении события *A*, т. е. по мере получения новой информации, мы можем проверять выдвинутые до испытания гипотезы. Такой подход дает возможность корректировать управленческие решения в экономике.

***Пример 20.*** В магазин для продажи поступает продукция трех поставщиков, относительные доли 1-го – 50 %, 2-го – 30 %, 3-го – 20 %. Брак составляет у 1-го поставщика – 2 %, 2-го – 3 %, 3-го – 5 %. Какова вероятность того, что покупатель купит качественную продукцию от первого поставщика?

***Решение.*** Здесь возможны три гипотезы: $H_1$ = {куплена продукция, поступившая от 1-го поставщика}, $H_2$ = {куплена продукция, поступившая от 2-го поставщика}, $H_3$ = {куплена продукция, поступившая от 3-го поставщика}. Очевидно, система этих гипотез полная, их вероятности равны $P(H_1) = 0{,}5$, $P(H_2) = 0{,}3$, $P(H_3) = 0{,}2$.

Пусть событие $A$ = {покупка качественной продукции}. Тогда соответствующие условные вероятности события $A$ равны:

$$P(A/H_1) = 1 - 0{,}02 = 0{,}98,$$

$$P(A/H_2) = 1 - 0{,}03 = 0{,}97,$$

$$P(A/H_3) = 1 - 0{,}05 = 0{,}95.$$

По формуле полной вероятности рассчитаем:

$$P(A) = 0{,}5 \cdot 0{,}98 + 0{,}3 \cdot 0{,}97 + 0{,}2 \cdot 0{,}95 = 0{,}971.$$

По формуле Байеса получим:

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)},$$

$$P(H_1/A) = \frac{0{,}5 \cdot 0{,}98}{0{,}971} = 0{,}505.$$

## 2.6. Схема Бернулли

На практике часто приходится сталкиваться с задачами, которые можно представить в виде многократно повторяющихся независимых испытаний, в которых представляет интерес вероятность числа $m$ наступлений некоторого события $A$ в $n$ испытаниях. Если независимые повторные испытания проводятся при одном и том же комплексе условий, то вероятность наступления события $A$ в каждом испытании одна и та же. Такая последовательность испытаний называется **схемой Бернулли.**

**Теорема.** Если вероятность $p$ наступления события $A$ в каждом испытании постоянна, то вероятность $P_{m,n}$ того, что событие $A$ наступит $m$ раз в $n$ независимых испытаниях, равна:

$$P_{m,n} = C_n^m p^m q^{n-m}, \quad p = 1 - q\,.$$

Вероятность того, что в $n$ испытаниях событие $A$ наступит: **менее** $m$ **раз** равна:

$$P_{0,n} + P_{1,n} + \ldots + P_{m-1,n}\ .$$

Вероятность того, что в $n$ испытаниях событие $A$ наступит **более** $m$ **раз** равна:

$$P_{m+1,n} + P_{m+2,n} + \ldots + P_{n,n}\ .$$

Вероятность того, что в n испытаниях событие $A$ наступит **не более** $m$ **раз** равна:

$$P_{0,n} + P_{1,n} + \ldots + P_{m-1,n} + P_{m,n}\ .$$

***Пример 21.*** Два равносильных противника играют в шахматы. Что вероятнее: выиграть не менее двух партий из четырех или не менее трех партий из пяти? Ничьи во внимание не принимаются.

***Решение.*** Играют равносильные шахматисты, поэтому вероятность выигрыша $p$ = 1/2, следовательно, вероятность проигрыша $q$ = 1/2. Так как во всех партиях вероятность выигрыша постоянна, то применима формула Бернулли.  Найдем вероятность того, что не менее двух партий из четырёх будут выиграны:

$$P(\geq 2) = C_4^2 p^2 q^2 + C_4^3 p^3 q^1 + C_4^4 p^4 q^0 = \frac{4 \cdot 3}{1 \cdot 2}\left(\frac{1}{2}\right)^4 + \frac{4}{1}\left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4 = \frac{11}{16} = 0{,}69.$$

Найдем вероятность того, что не менее трёх партий из пяти будут выиграны:

$$P(\geq 3) = C_5^3 p^3 q^2 + C_5^4 p^4 q^1 + C_5^5 p^5 q^0 = \frac{4 \cdot 5}{1 \cdot 2}\left(\frac{1}{2}\right)^5 + \frac{5}{1}\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{16}{32} = 0{,}5.$$

Так как $P(\geq 2) > P(\geq 3)$, то вероятнее выиграть не менее двух партий из четырёх, чем не менее трёх из пяти.

В случаях, когда $n$ велико, а вероятность мала, использование формулы Бернулли вызывает затруднения. Тогда применяется **формула Пуассона**:

$$P_n(m) \approx \frac{\lambda^m}{m!} \ell^{-\lambda}, \quad \lambda = np .$$

## Вопросы и задачи

1. Какие события называются независимыми?
2. Приведите пример события $A$, зависящего от события $B$.
3. Что называется условной вероятностью?
4. Сформулируйте теорему умножения вероятностей.
5. Производительности трёх станков, обрабатывающих одинаковые детали, относятся как 1:3:6. Из нерассортированной партии обработанных деталей взяты наудачу две. Какова вероятность того, что: a) одна из них обработана на 3-м станке; b) обе обработаны на одном станке?
   (a) $p = 0{,}48$; b) $p = 0{,}46$)
6. Вероятность того, что студент сдаст первый экзамен, равна 0,9; второй – 0,9; третий – 0,8. Найти вероятность того, что студентом будут сданы: a) только один экзамен; b) хотя бы один экзамен; c) по крайней мере, два экзамена; d) три экзамена.

(a) $p = 0{,}044$;   b) $p = 0{,}998$);   c) $p = 0{,}954$;   d) $p = 0{,}648$)

7. Запишите формулу полной вероятности.
8. Вероятность поражения самолета при одиночном выстреле для первого ракетного расчета равно 0,2, а для второго – 0,1. Каждый ракетный комплекс производит по одному выстрелу, причем зарегистрировано одно попадание в самолет. Какова вероятность, что удачный выстрел принадлежит первому расчету?
9. Вероятность того, что расход электроэнергии в продолжении одних суток не превысит установленной нормы, равна $p = 0{,}75$. Найти вероятность того, что в ближайшие 5 суток расход электроэнергии в течение 3 суток не превысит нормы.

   ($p_5(3) = 0{,}26$)

# 3. Случайные величины

## 3.1. Основные понятия

В исследуемых социально-экономических системах нас часто интересуют такие величины, которые от наблюдениях к наблюдению принимают разные числовые значения. Например, у каждого человека имеется много числовых характеристик: рост, возраст, вес и т. д. Если мы выбираем человека случайно (например, из группы или из толпы), то случайными будут и значения указанных характеристик.

**Случайной** называется величина, которая в результате испытания принимает одно возможное значение, заранее не известное и зависящее от случайных причин. Случайные величины обозначаются заглавными буквами $X, Y, Z, \ldots$ латинского алфавита, а их возможные значения – соответствующими строчными буквами $x, y, z, \ldots$ В практических задачах обычно используются два вида случайных величин – дискретные и непрерывные.

**Дискретной** называется случайная величина (СВ), которая может принимать конечное или счетное множество значений (счетным называют множество, элементы которого можно пронумеровать).

***Примеры:***

1. Число родившихся детей в течение суток в Нижнем Новгороде.

2. Число очков, выпадающее при бросании игральной кости.

3. Выбирая наугад слова из текста, мы встречаем слова разной длины, с различным количеством гласных или согласных фонем.

**Непрерывной** называют случайную величину, бесконечное множество значений которой есть некоторый интервал (конечный или бесконечный) числовой оси.

***Примеры:***

1. Расход электроэнергии на предприятии за месяц.

2. Интенсивность звука, которая колеблется обычно в определённых пределах.

**Законом распределения** случайной величины $X$ называется определённое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

***Примеры:***

1. Вероятность появления букв, цифр и символов в принимаемом сообщении.

2. Ежедневная прибыль фирмы, рассчитанная по многолетним статистическим данным

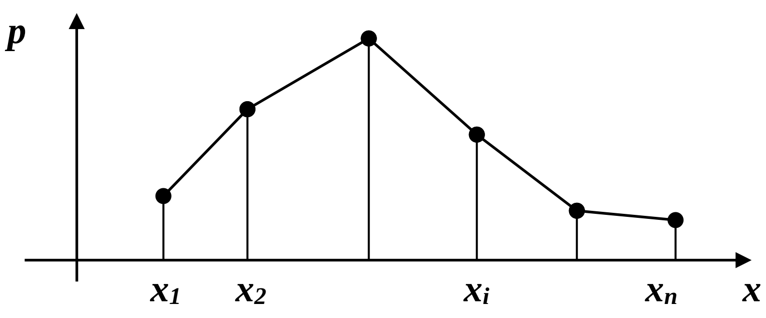
## 3.2. Способы задания закона распределения дискретной СВ

***1.*** **Табличный.** Простейшей формой задания закона распределения дискретной СВ является таблица (матрица),

| $X$ | $x_1$ | $x_2$ | … | $x_i$ | … | $x_n$ | |
|---|---|---|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | … | $p_i$ | … | $p_n$ | $\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1$ |

в которой перечислены в порядке возрастания все возможные значения случайной величины и соответствующие им вероятности.

***2.*** **Графический.** Ряд распределения может быть изображен графически, если по оси абсцисс откладывать значения случайной величины, по оси ординат – соответствующие им вероятности. Соединение полученных точек образует ломаную линию, называемую **многоугольником распределения вероятностей**.




***3.*** **Аналитический *(в виде формулы).*** Закон распределения может быть задан или **функцией распределения** или **плотностью распределе-**

**ления.** Например, формула Бернулли описывает Биномиальный закон через плотность распределения. Для описания закона распределения случайной величины можно рассматривать вероятности события $X < x$, где $x$ – текущая переменная, т. е. некоторую функцию от $x$.

**Функцией распределения** случайной величины $X$ называется функция $F(x)$, выражающая для каждого $x$ вероятность того, что случайная величина $X$ примет значения, меньшее $x$.

$$F(x)=P(X<x), \text{ или } F(x)=\sum_{x_i<x} P(X<x).$$

Функцию $F(x)$ называют также **интегральной функцией распределения** или **интегральным законом распределения**.

**Свойства** функции распределения:

1. *Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей:*
$$0 \le F(x) \le 1.$$
2. *Функция распределения случайной величины есть неубывающая функция на всей числовой оси.*
3. *На минус бесконечности функция распределения равна нулю, на плюс бесконечности равна единице, т. е.*
$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,\ F(+\infty)=\lim_{x\to+\infty}F(x)=1.$$

***Пример 22.*** Пусть $X$ – число реверсов, выпавших при четырех бросаниях правильной монеты. Требуется: а) составить закон распределения дискретной случайной величины $X$; б) найти функцию распределения и начертить её график.

***Решение.*** А. Случайная величина $X$ может принимать значения 0, 1, 2, 3, 4. Испытания независимы, и вероятность события $A$ = {выпадение реверса} в каждом испытании одна и та же. Следовательно, вероятности значений случайной величины $X$ находим по формуле Бернулли:

$$P_{0,4}=C_4^0p^0q^4=q^4=\frac{1}{16};\ P_{1,4}=C_4^1pq^3=\frac{4}{16};\ P_{2,4}=C_4^2p^2q^2=\frac{6}{16}$$

$$P_{3,4}=C_4^3p^3q=\frac{4}{16};\ P_{4,4}=C_4^4p^4q^0=p^4=\frac{1}{16}.$$

Запишем искомый закон распределения $X$:

| ***X*** | **0** | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|
| ***p*** | $\frac{1}{16}$ | $\frac{4}{16}$ | $\frac{6}{16}$ | $\frac{4}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}=1$ |

Б. Будем задавать различные значения $x$ и находить для них $F(x)=P(X<x)$.

1. Если $x\le 0$, то $F(x)=0$ (в том числе и при $x=0$ $F(0)=P(X<0)=0$).
2. Пусть $0<x\le 1$ (например, $x=0{,}5$), $F(x)=P(X=0)=\frac{1}{16}$.

   Очевидно, что и $F(1)=P(X<1)=\frac{1}{16}$.
3. Пусть $1<x\le 2$ (например, $x=1{,}5$),

   $$F(x)=P(X<x)=P(X=0)+P(X=1)=\frac{1}{16}+\frac{4}{16}=\frac{5}{16}.$$

   Очевидно, что и $F(2)=P(X<2)=\frac{5}{16}$.
4. Пусть $2<x\le 3$,

   $$F(x)=[P(X=0)+P(X=1)]+P(X=2)=\frac{5}{16}+\frac{6}{16}=\frac{11}{16}.$$

   Очевидно, что и $F(3)=P(X<3)=\frac{11}{16}$.
5. Пусть $3<x\le 4$ (например, $x=3{,}5$),

   $$F(x)=[P(X=0)+P(X=1)+P(X=2)]+P(X=3)=\frac{11}{16}+\frac{4}{16}=\frac{15}{16}.$$

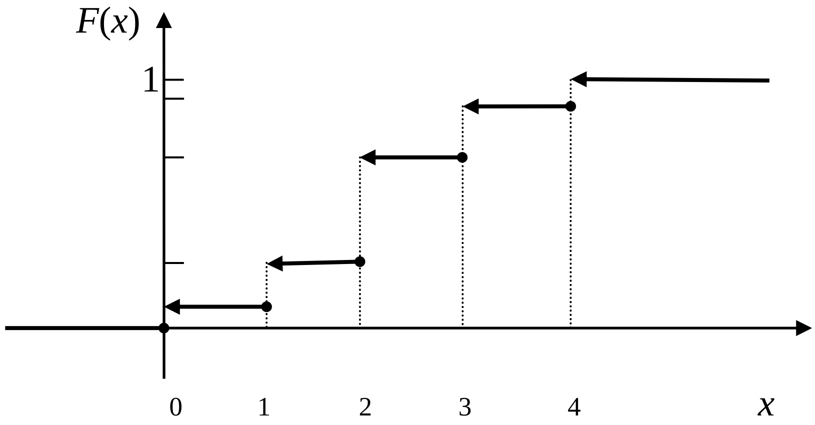
   Очевидно, что и $F(4)=\frac{15}{16}$.
6. Пусть $x>4$,

$$F(x)=[P(X=0)+P(X=1)+P(X=2)+P(X=3)]+P(X=4)=\frac{15}{16}+\frac{1}{16}=\frac{16}{16}=1.$$

Итак, функция распределения имеет вид:

$$F(x)=\begin{cases} 0; & x\le 0\\ \frac{1}{16}; & 0<x\le 1\\ \frac{5}{16}; & 1<x\le 2\\ \frac{11}{16}; & 2<x\le 3\\ \frac{15}{16}; & 3<x\le 4\\ 1; & x>4 \end{cases}$$

Построим график полученной функции:




### 3.3. Способы задания закона распределения непрерывной *СВ*

1. С помощью **функции распределения.**

Дополним свойства функции распределения $F(x)$ для непрерывной величины следующим требованием – это дифференцируемая функция в заданном интервале изменения аргумента *x*.

**Теорема.** Вероятность попадания случайной величины в интервал $[x_1, x_2]$ равна приращению ее функции распределения на этом интервале, т. е. $P(x_1 \le X \le x_2)=F(x_2)-F(x_1)$.

**Замечание.** Справедливы равенства

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 < X < x_2) = P(x_1 \le X < x_2).$$

**Следствие.** Вероятность любого отдельно взятого значения непрерывной случайной величины равна нулю.

2. С помощью **плотности распределения.**

Введем понятие плотности вероятности непрерывной *СВ*. Рассмотрим вероятность попадания непрерывной случайной величины *X* в интервал $[x, x+\Delta x]$. По предыдущей теореме имеем $P(x \le X \le x+\Delta x) = F(x+\Delta x) - F(x)$, т. е. вероятность равна приращению функции распределения *F(x)* на этом интервале. Тогда вероятность, приходящая на единицу длины равна $\frac{P(x \le X \le x+\Delta x)}{\Delta x} = \frac{F(x+\Delta x) - F(x)}{\Delta x}$. Переходя к пределу при $\Delta x \to 0$, получим плотность вероятности в точке *x*, представляющую производную функции распределения *F(x)*:

$$\lim_{\Delta x \to 0} \frac{P(x \le X \le x+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x).$$

На основании определения производной получим выражение плотности вероятности через функцию распределения:

$$f(x) = F'(x).$$

Отсюда следует, что функция распределения является первообразной для плотности распределения.

Функцию $f(x)$ называют **плотностью вероятности** или **дифференциальным законом распределения.** График плотности распределения называют **кривой распределения**.

**Свойства** плотности распределения:

1. Плотность распределения – неотрицательная функция:

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

2. Несобственный интеграл от плотности распределения в пределах от $-\infty$ до $+\infty$ равен единице:

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

Геометрически это означает, что вся площадь криволинейной трапеции, ограниченной осью ***Ox*** и кривой распределения, равна единице.

**Теорема.** Вероятность попадания случайной величины в интервал $[x_1, x_2]$ равна определённому интегралу от плотности распределения, взятому в пределах от $x_1$ до $x_2$:

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

### 3.4. Числовые характеристики случайных величин

Часто, полное описание распределения вероятностей при помощи плотности или функции распределения заменяют заданием небольшого числа числовых характеристик. Они указывают, как правило, наиболее типичные (в том или ином отношении) значения случайной величины и степень рассеяния значений случайной величины около некоторого типичного значения. Из этих характеристик наиболее употребительны математическое ожидание (среднее значение) и дисперсия (рассеяние), а также начальные и центральные моменты.

**Определение математического ожидания и дисперсии**

| | Дискретная величина | Непрерывная величина |
|---|---|---|
| Математическое ожидание $M(X) = a$ | $\sum_{i=1}^{n} x_i p_i$ | $\int_{-\infty}^{+\infty} xf(x)dx$ |
| Дисперсия $D(X) = \sigma^2$ | $\sum_{i=1}^{n} (x_i - a)^2 p_i =$ <br> $= \sum_{i=1}^{n} x_i^2 p_i - a^2$ | $\int_{-\infty}^{+\infty} (x-a)^2 f(x)dx =$ <br> $= \int_{-\infty}^{+\infty} x^2 f(x)dx - a^2$ |

Где $x_i$ – значение дискретной случайной величины $X$, $p_i$ – соответствующая значению $x_i$ вероятность, $f(x)$ – плотность распределения непрерывной случайной величины $X$.

**Свойства математического ожидания и дисперсии**

| | Математическое ожидание | Дисперсия | Среднеквадратическое отклонение $\sigma = \sqrt{D(x)}$ |
|---|---|---|---|
| Случайная величина – константа | $M(C) = C$ | $D(C) = 0$ | $\sigma(C) = 0$ |
| Случайная величина умножается на константу | $M(C \cdot X) = CM(X)$ | $D(C \cdot X) = C^2 D(X)$ | $\sigma(CX) = \|C\|\sigma(X)$ |
| Случайные величины складываются | $M\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} M(X_i)$ | $D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i)$ (для независимых величин) | $\sigma(X_1 + X_2) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2)}$ (для независимых величин) |

Кроме математического ожидания и дисперсии, в теории вероятностей применяется ещё ряд числовых характеристик.

**Модой** $Mo(X)$ случайной величины $X$ называется её наиболее вероятное значение (для которого вероятность $p_i$ или плотность вероятности $f(x)$ достигает максимума).

**Медианой** $Me(X)$ непрерывной случайной величины $X$ называется такое её значение, для которого:

$$P(X < Me(X)) = P(X > Me(X)) = 0{,}5.$$

Геометрически это означает, что вертикальная прямая $x = Me(X)$ делит площадь фигуры под кривой распределения на две равные части. Очевидно, что в точке $x = Me(X)$ функция распределения равна 0,5.

## Определение начальных и центральных моментов

| | | |
|---|---|---|
| **Начальным моментом порядка *k*** случайной величины $X$ называют математическое ожидание величины $X^k$ | Дискретная СВ $\nu_k = M\left(X^k\right) = \sum_{i=1}^{n} x_i^k p_i$ Непрерывная СВ $\nu_k = \int_{-\infty}^{+\infty} x^k f(x) dx$ | |
| Начальный момент первого порядка | $\nu_1 = M(X)$ | Равен математическому ожиданию СВ $X$ |
| Начальный момент второго порядка | $\nu_2 = M\left(X^2\right)$ | |
| Начальный момент третьего порядка | $\nu_3 = M\left(X^3\right)$ | |
| Начальный момент четвёртого порядка | $\nu_4 = M\left(X^4\right)$ | |

| | | |
|---|---|---|
| **Центральным моментом порядка** $k$ случайной величины $X$ называют математическое ожидание величины $[X - \mathrm{M}(X)]^k$ | Дискретная СВ $\mu_к = M\left[X - M(X)\right]^k$ Непрерывная СВ $\mu_к = \int_{-\infty}^{+\infty} (x-a)^k f(x) dx$ | |
| Центральный момент первого порядка | $\mu_1 = M\left[X - M(X)\right] = 0$ | $\mu_1 = 0$ |
| Центральный момент второго порядка | $\mu_2 = M\left[X - M(X)\right]^2$ | Равен дисперсии СВ $X$. Характеризует степень рассеяния значений $X$ относительно $\boldsymbol{M}(X)$ $\mu_2 = \nu_2 - \nu_1^2$ |
| Центральный момент третьего порядка | $\mu_3 = M\left[X - M(X)\right]^3$ | $\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3$ $A = \frac{\mu_3}{\sigma^3}$ – **коэффициент асимметрии распределения** |
| Центральный момент | | $\mu_4 = \nu_3 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4$ |

| | | |
|---|---|---|
| четвёртого порядка | $\mu_4 = M[X - M(X)]^4$ | $E = \frac{\mu_4}{\sigma^4} - 3$ – **эксцесс.** Характеризует островершинность (E > 0) или плосковершинность (E < 0) распределения |

***Пример 23.*** Случайная величина *X* задана законом распределения:

| $X$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $p$ | $\frac{1}{16}$ | $\frac{4}{16}$ | $\frac{6}{16}$ | $\frac{4}{16}$ | $\frac{1}{16}$ |

Найти математическое ожидание, центральные моменты первого, второго, третьего порядков.

***Решение.*** Вычислим математическое ожидание СВ *X*:

$$M(X) = 0\cdot\frac{1}{16} + 1\cdot\frac{4}{16} + 2\cdot\frac{6}{16} + 3\cdot\frac{4}{16} + 4\cdot\frac{1}{16} = 2.$$

Для вычисления центральных моментов, предварительно найдем начальные моменты:

$$\nu_1 = M(X) = 2;\ \nu_2 = M(X^2) = 0\cdot\frac{1}{16} + 1\cdot\frac{4}{16} + 4\cdot\frac{6}{16} + 9\cdot\frac{4}{16} + 16\cdot\frac{1}{16} = 5;$$

$$\nu_3 = M(X^3) = 0\cdot\frac{1}{16} + 1\cdot\frac{4}{16} + 8\cdot\frac{6}{16} + 27\cdot\frac{4}{16} + 64\cdot\frac{1}{16} = 14.$$

Найдем центральные моменты:

$$\mu_1 = 0;\ \mu_2 = \nu_2 - \nu_1^2 = 5 - 2^2 = 1 \text{ – дисперсия;}$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3 = 14 - 3\cdot 2\cdot 5 + 2\cdot 2^3 = 0.$$

***Пример 24.*** Дана функция распределения случайной величины *X*:

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 0 \\ x^2/4 & \text{при} \quad 0 < x \le 2 \\ 1 & \text{при} \quad x > 2 \end{cases}.$$

a. Найти плотность вероятности $f(x)$.

b. Построить графики функций $f(x)$ и $F(x)$.

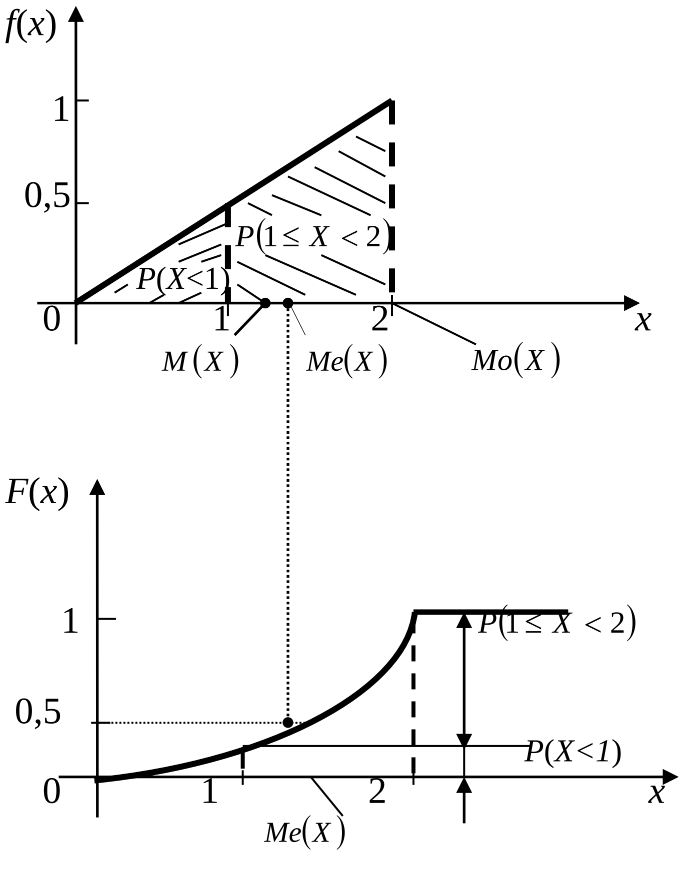
c. Найти вероятности $P(X=1)$, $P(X<1)$, $P(1\le X<2)$.

d. Вычислить математическое ожидание $M(X)$, дисперсию $D(X)$, моду $Mo(X)$, медиану $Me(X)$.

***Решение.*** a. Плотность вероятности найдём по определению:

$$f(x)=F'(x)=\begin{cases}0 & при \quad x\le 0\\ x/2 & при \quad 0<x\le 2\\ 0 & при \quad x>2\end{cases}.$$

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




c. $P(X=1)=0$ как вероятность отдельно взятого значения непрерывной случайной величины. $P(X<1)$ можно найти либо по определению функции распределения:

$$P(X<1)=F(1)=\frac{1^2}{4}=\frac{1}{4}$$ (см. график функции *F(x)*),

либо через плотность вероятности:

$$P(X<1)=\int_{-\infty}^{1} f(x)dx=\int_{-\infty}^{0} 0\cdot dx+\int_{0}^{1}\frac{x}{2}dx=0+\left.\frac{x^2}{4}\right|_0^1=\frac{1}{4}$$

(площадь под кривой распределения $f(x)$ на отрезке [0; 1]).

$P(1\le X<2)$ можно найти либо как приращение функции распределения:

$$P(1\le X<2)=F(2)-F(1)=\frac{2^2}{4}-\frac{1^2}{4}=\frac{3}{4},$$

либо через плотность вероятности:

$$P(1\le X<2)=\int_{1}^{2}\frac{x}{2}dx=\left.\frac{x^2}{4}\right|_1^2=\frac{2^2}{4}-\frac{1^2}{4}=\frac{3}{4}.$$

d. $a=M(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_{-\infty}^{0} 0\cdot dx+\int_{0}^{2} x\frac{x}{2}dx+\int_{2}^{+\infty} 0\cdot dx=\frac{4}{3}.$

$$D(X)=M\left(X^2\right)-a^2=2-\left(\frac{4}{3}\right)^2=\frac{2}{9}.$$

По графику видно, что плотность вероятности $f(x)$ достигает максимального значения при $x=2$, следовательно $Mo(X)=2$.

Медиану $Me(X)=m$ найдем либо из условия $F(m)=0{,}5$, т. е. $\frac{m^2}{4}=\frac{1}{2}$, откуда $m=\sqrt{2}$, либо через плотность вероятности $\int_{-\infty}^{m} f(x)dx=\frac{1}{2}$,

т. е. $0+\int_{0}^{m}\frac{x}{2}dx=\left.\frac{x^2}{4}\right|_0^m=\frac{m^2}{4}=\frac{1}{2}$, получаем $m=Me(X)=\sqrt{2}$.

## 3.5. Примеры распределений случайных величин

### Примеры распределений дискретных случайных величин

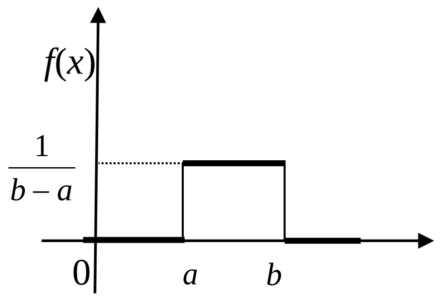
| **Биномиальное** распределение (БР) | |
|---|---|
| Закон распределения определяется выражением $P = C_n^k p^k q^{n-k}$, где $C_n^k = \frac{n!}{k!(n-k)!}$, $p$ – вероятность наступления события $A$ в каждом из $n$ независимых испытаний, $q = 1 - p$ | Числовые характеристики $M(X) = np$, $D(X) = npq$, $A = \frac{q-p}{\sqrt{npq}}$, $E = \frac{1-6pq}{npq}$ |
| Локальная теорема Муавра-Лапласа $\lim\limits_{n\to\infty} P_{kn} = \frac{1}{\sqrt{2\pi D(x)}} \exp\left[-\frac{1}{2D(x)}(k - M(x))^2\right]$ | пределом БР является нормальный закон распределения с параметрами $a = np$, $\sigma^2 = npq$ |

| Распределение **Пуассона** | |
|---|---|
| Число испытаний $n$ – велико, вероятность $p$ – постоянна и мала, число $\lambda = np$ незначительно, $P_{kn} = \frac{\lambda^k e^{-\lambda}}{k!}$ | $M(X) = D(X) = \lambda$ |

## Примеры распределений непрерывных случайных величин

<table>
<tr><td colspan="2">Равномерное распределение</td></tr>
<tr><td>Плотность вероятности

$$f(x)=\begin{cases} 0 & \text{при} \quad x<a \\ \dfrac{1}{b-a} & \text{при} \quad a\le x\le b \\ 0 & \text{при} \quad x>b \end{cases}$$


</td><td rowspan="2">Числовые характеристики

$M(X)=\dfrac{a+b}{2}$ – средина отрезка $[a,b]$

$Me(X)=M(X)$

$Mo(X)$ нет

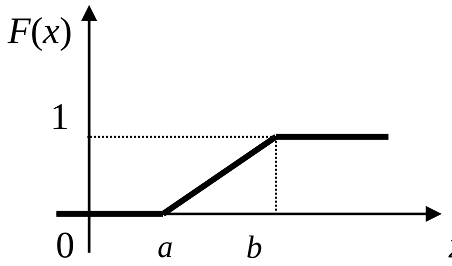
$D(X)=\dfrac{(b-a)^2}{12}$

$\sigma=\dfrac{(b-a)}{2\sqrt{3}}$

$A=0$

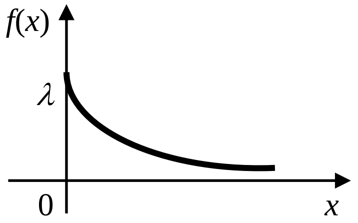
$E=-1{,}2<0$</td></tr>
<tr><td>Функция распределения

$$F(x)=\int\limits_{-\infty}^{x} f(x)dx=\begin{cases} 0 & \text{при} \quad x<a \\ \dfrac{x-a}{b-a} & \text{при} \quad a\le x\le b \\ 1 & \text{при} \quad x>b \end{cases}$$


</td></tr>
<tr><td>Вероятность попадания в интервал $(\alpha,\beta)\subset[a,b]$</td><td>$P(\alpha\le X\le\beta)\overset{\Delta}{=}\int\limits_{\alpha}^{\beta}\dfrac{dx}{b-a}=\dfrac{\beta-\alpha}{b-a}$</td></tr>
</table>

<table>
<tr><td colspan="2">Показательный (экспоненциальный) закон распределения</td></tr>
<tr><td>Плотность вероятности

$$f(x)=\begin{cases}0 & x<0\\ \lambda e^{-\lambda x} & x\ge 0\end{cases}$$



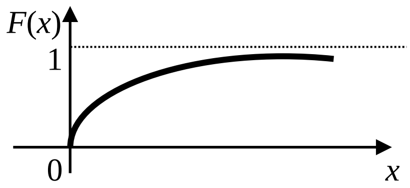
</td><td rowspan="2">Числовые характеристики

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

$$D(X)=\frac{1}{\lambda^2}$$

$$\sigma=M(X)=\frac{1}{\lambda}$$
</td></tr>
<tr><td>Функция распределения

$$F(x)=\begin{cases}0 & x<0\\ 1-e^{-\lambda x} & x\ge 0\end{cases}$$



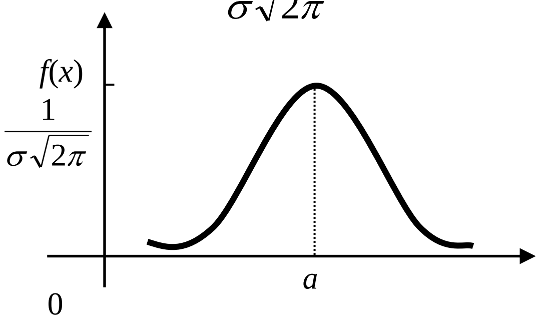
</td></tr>
<tr><td>Вероятность попадания в интервал $[a,b]$ СВ $X$</td><td>

$$P(a\le X\le b)=e^{-\lambda a}-e^{-\lambda b}$$
</td></tr>
</table>

<table>
<tr><td colspan="2">Нормальный (Гаусса) закон распределения</td></tr>
<tr><td>Плотность вероятности для СВ

$$X\sim N(a,\sigma^2)$$

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



</td><td>Числовые характеристики

$$M(X)=a$$

$$D(X)=\sigma^2$$

$$Mo(X)=Me(X)=M(X)=a$$

$$\mu_k=(k-1)\sigma^2\mu_{k-2},\ k=2,3,\ldots$$

$$A=E=0$$
</td></tr>
</table>

Если $\sigma = const$ и параметр $a$ меняется, то кривая распределения будет смещаться вдоль оси абсцисс, не меняя формы.

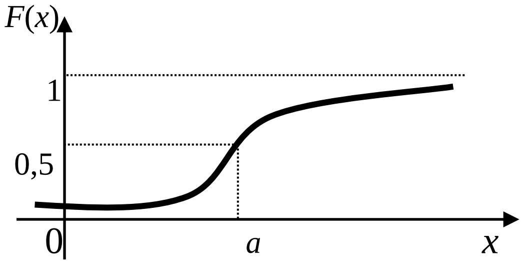
Если $a = const$ и параметр $\sigma^2$ меняется, то меняется ордината максимума кривой $f_{\max}(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$. При увеличении $\sigma^2$ кривая становится более плоской, растягиваясь вдоль оси абсцисс; при уменьшении $\sigma^2$ кривая вытягивается вверх, одновременно сжимаясь с боков (площадь под любой кривой распределения должна оставаться равной единице).

<table>
<tr>
<td>

Функция распределения

для *СВ* $X \sim N\left(a, \sigma^2\right)$

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




</td>
<td>

Плотность вероятности

для *СВ* $T = \dfrac{X-a}{\sigma} \sim N(0{,}1)$

$$f(t) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right]$$

Функция распределения

для *СВ* $T = \dfrac{X-a}{\sigma} \sim N(0{,}1)$

$F(t) = \dfrac{1}{2} + \dfrac{1}{2}\Phi(t)$, где $\Phi(t)$ – **функция Лапласа**

</td>
</tr>
</table>

| | |
|---|---|
| Вероятность того, что отклонение *СВ X*, распределенной по нормальному закону, от математического ожидания $a$ не превысит величину $\Delta$ (по абсолютной величине)<br><br>$P(\lvert X-a\rvert\le\Delta)=\Phi(t)$, где $t=\dfrac{\Delta}{\sigma}$<br>$\Delta=\sigma \quad P(\lvert X-a\rvert\le\sigma)=\Phi(1)=0{,}68$<br>$\Delta=2\sigma \quad P(\lvert X-a\rvert\le 2\sigma)=\Phi(2)=0{,}95$<br>$\Delta=3\sigma \quad P(\lvert X-a\rvert\le 3\sigma)=\Phi(3)=0{,}997$<br>**Правило трёх сигм:**<br>Если *СВ X* ~ $N(a,\sigma^2)$, то практически достоверно, что её значения заключены в интервале<br>$(a-3\sigma, a+3\sigma)$ | Функция Лапласа<br><br>$\Phi(t)\overset{\Delta}{=}\dfrac{2}{\sqrt{2\pi}}\int\limits_0^t \exp\left[-\dfrac{t^2}{2}\right]d\tilde{t}$<br><br>$\Phi(-t)=-\Phi(t)$<br><br>$\Phi(t)$ – монотонно возрастающая функция<br>$\lim\limits_{t\to+\infty}\Phi(t)=1$<br><br>функция $\Phi(t)$ табулирована |
| Вероятность попадания в интервал $[\alpha,\beta]$ *СВ X* | $P(\alpha\le X\le\beta)=\dfrac{1}{2}\left[\Phi\left(\dfrac{\beta-a}{\sigma}\right)-\Phi\left(\dfrac{\alpha-a}{\sigma}\right)\right]$ |

***Пример 25.*** Случайная величина *X* распределена по нормальному закону, математическое ожидание равно 10, среднее квадратичное отклонение равно 2. Найти вероятность того, что в результате испытания *СВ X* примет значение, заключенное в интервале (12; 14).

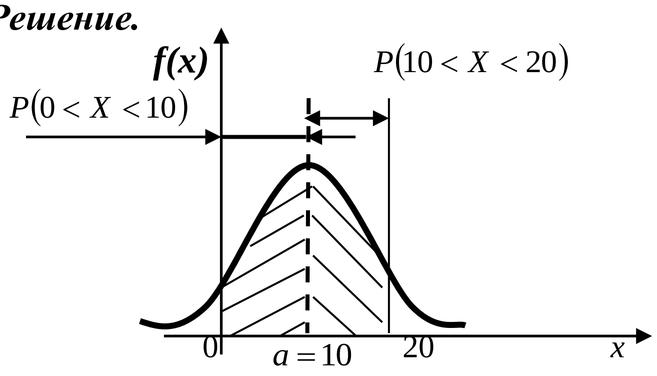
***Решение***. Найдём вероятность попадания *СВ X* в интервал $[12;14]$:

$$P(12<X<14)=\frac{1}{2}\left[\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right)\right]=$$

$$=\frac{1}{2}\left[\Phi\left(\frac{14-10}{2}\right)-\Phi\left(\frac{12-10}{2}\right)\right]=\frac{1}{2}[\Phi(2)-\Phi(1)]=0{,}1359$$

***Пример 26.*** Случайная величина $X$ распределена нормально с $M(X)=10$. Вероятность попадания $X$ в интервал (10; 20) равна 0,3. Чему равна вероятность попадания $X$ в интервал (0; 10).

***Решение.***




Так как нормальная кривая симметрична относительно прямой $x=a=10$, то площади ограниченные сверху нормальной кривой и снизу – интервалами (0; 10) и (10; 20) равны между собой. Поскольку эти площади числено равны вероятности попадания $X$ в соответствующий интервал, то

$$P(0<X<10)=P(10<X<20)=0{,}3.$$

***Пример 27.*** Дана плотность вероятности

$$f(x)=\begin{cases} 0 & x<0 \\ \lambda e^{-\lambda x} & x\geq 0 \end{cases}.$$

Доказать, что $a=\dfrac{1}{\lambda}$, $\sigma^2=\dfrac{1}{\lambda^2}$.

***Решение.*** По определению математического ожидания непрерывной случайной величины:

$$a=M(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{0}x\cdot 0dx+\int_{0}^{\infty}x\lambda e^{-\lambda x}dx$$

Интегрируя по частям, положив $\begin{cases} u=x \\ dv=e^{-\lambda x}dx \end{cases}$, найдём:

$$M(X)=\lambda\int_{0}^{\infty}xe^{-\lambda x}dx=-xe^{-\lambda x}\Big|_0^{\infty}+\int_{0}^{\infty}e^{-\lambda x}dx=-xe^{-\lambda x}\Big|_0^{\infty}-\frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty}=\frac{1}{\lambda}$$

Найдём дисперсию случайной величины *X* по формуле:

$$D(X)=M(X^2)-M^2(X).$$

Интегрируя по частям дважды, получим:

$$M(X^2)=\int_0^\infty x^2\lambda e^{-\lambda x}dx=\lambda\int_0^\infty x^2 e^{-\lambda x}dx=\left\{\begin{matrix} u=x^2 & du=2xdx \\ dv=e^{-\lambda x}dx & v=-\frac{1}{\lambda}e^{-\lambda x}\end{matrix}\right\}=$$

$$=-x^2e^{-\lambda x}\Big|_0^\infty+2\int_0^\infty xe^{-\lambda x}dx=0+2\int_0^\infty xe^{-\lambda x}dx=\left\{\begin{matrix} u=x & du=dx \\ dv=e^{-\lambda x}dx & v=-\frac{1}{\lambda}e^{-\lambda x}\end{matrix}\right\}=-\frac{2x}{\lambda}e^{-\lambda x}\Big|_0^\infty+$$

$$+\frac{2}{\lambda}\int_0^\infty e^{-\lambda x}dx=-\frac{2}{\lambda^2}e^{-\lambda x}\Big|_0^\infty=\frac{2}{\lambda^2}$$

Учитывая, что $M(X)=\frac{1}{\lambda}$, вычислим:

$$\sigma^2=D(X)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$ , что и требовалось доказать.

## 3.6. Распределения, наиболее близко соответствующие нормальному закону распределения

Распределения, близкие к нормальному закону, играют большую роль в проверке статистических гипотез. Рассмотрим более подробно ряд наиболее важных распределений.

### 1. Гамма-распределение

Тесно связано с нормальным так называемое $\Gamma_{\alpha,\lambda}$-распределение с плотностью:

$$f(x)=\frac{\alpha^\lambda}{\Gamma(\lambda)}x^{\lambda-1}e^{-x}, \qquad x\ge 0.$$

Здесь $\alpha,\lambda$ – положительные параметры, а $\Gamma(\lambda)=\int_0^\infty x^{\lambda-1}e^{-x}dx$ – гамма-функция **(интеграл Эйлера)**, в частности, $\Gamma(\lambda+1)=\lambda\Gamma(\lambda)$.

Пусть $x_1, x_2, ..., x_n$ независимы и нормально распределены с математическим ожиданием, равным нулю, и дисперсией, равной единице. Существует теорема, согласно которой распределение случайной величины $\chi_n^2 = x_1^2 + x_2^2 + ... + x_n^2$ есть частный случай гамма-распределения: $\chi_n^2 = \Gamma_{\frac{1}{2},\frac{n}{2}}$.

## 2. Распределение $\chi^2$ (хи-квадрат)

Рассмотрим случайную величину $V$, распределенную по нормальному закону с математическим ожиданием $a$ и средним квадратическим отклонением $\sigma$: $Y \in N(a,\sigma)$. Тогда случайная величина $W = \dfrac{Y-a}{\sigma}$ распределена по нормальному закону с параметрами $M(W) = 0$ и $\sigma(W) = 1$, т. е. $W \in N(0,1)$.

Квадрат такой стандартизованной случайной величины

$$W^2 = \left(\frac{Y-a}{\sigma}\right)^2 = \chi^2$$

называется *случайной величиной* $\chi^2$ *с одной степенью свободы.*

Рассмотрим $n$ независимых случайных величин $Y_1, Y_2, \ldots, Y_n$, распределенных по нормальному закону с $M(Y_i) = a_i$ и средними квадратическими отклонениями $\sigma_i$ $i = 1, ..., n$. Образуем для каждой из этих случайных величин стандартизованную случайную величину:

$$W_i = \frac{Y_i - a_i}{\sigma_i}, \; i = 1, ..., n.$$

Сумма квадратов стандартизованных переменных

$$\chi^2 = W_1^2 + W_2^2 + \ldots + W_n^2 =$$
$$= \left(\frac{Y_1 - a_1}{\sigma_1}\right)^2 + \left(\frac{Y_2 - a_2}{\sigma_2}\right)^2 + \ldots + \left(\frac{Y_n - a_n}{\sigma_n}\right)^2$$

называется *случайной величиной* $\chi^2$ *с* $\nu = n$ *степенями свободы.*

## 3. Распределение Стьюдента (*t*-распределение)

**Распределение Стьюдента** – имеет важное значение при статистических вычислениях, связанных с нормальным законом, а именно тогда, когда среднее квадратическое отклонение $\sigma$ неизвестно и подлежит определению по опытным данным.

Пусть $Y_1, Y_2, \ldots, Y_n$ – независимые случайные величины, имеющие нормальное распределение с параметрами $M(Y) = M(Y_i) = 0$ и $\sigma_Y = \sigma_{Y_i} = 1$, $i = 1, \ldots, n$.

Случайная величина

$$t_n = \frac{Y}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}} = \frac{Y}{\sqrt{\frac{1}{n}\chi_n^2}}, \tag{1}$$

являющаяся функцией нормально распределенных случайных величин, называется **распределением Стьюдента** с $n$ степенями свободы.

Распределение случайной величины $t$ зависит только от одного параметра – числа степеней свободы $\nu$, равного числу слагаемых в подкоренном выражении распределения Стьюдента (1).

Математическое ожидание и дисперсия случайной величины $t$ соответственно равны:

$$M(t) = 0; \quad D(t) = \frac{\nu}{\nu - 2} \quad (\nu > 2).$$

Функция плотности распределения Стьюдента симметрична относительно замены $t \to -t$, а также слабо сходится к нормальному распределению при $\nu \to \infty$.

Распределение Стьюдента – одно из наиболее известных распределений среди используемых при анализе реальных данных. Его применяют при оценивании математического ожидания, прогнозируемого значения и других характеристик с помощью доверительных интервалов, по проверке гипотез однородности выборок и т.д.

## 4. Распределение Фишера (*F*-распределение)

**Распределение Фишера** – это распределение случайной величины

$$F = \frac{\frac{1}{n_1} X_1}{\frac{1}{n_2} X_2},$$

где случайные величины $X_1$ и $X_2$ независимы и имеют распределения хи-квадрат с числом степеней свободы $n_1$ и $n_2$ соответственно. При этом пара $(n_1, n_2)$ – пара «чисел степеней свободы» распределения Фишера.

Распределение Фишера используют при проверке гипотез об адекватности модели в регрессионном анализе, о равенстве дисперсий и в других задачах прикладной статистики.

Выражения для функций распределения хи-квадрат, Стьюдента и Фишера, их плотностей и характеристик, а также таблицы, необходимые для их практического использования, можно найти в специальной литературе (см., например, [8]).

**Вопросы и задачи**

1. Что называется случайной величиной?
2. Привести пример дискретной случайной величиной.
3. В магазин поступила обувь с двух фабрик в соотношении 2:3. Куплено 4 пары обуви. Найти закон распределения числа купленных пар обуви, изготовленной первой фабрикой. Найти математическое ожидание и среднеквадратическое отклонение этой случайной величины.
4. Дискретная случайная величина X задана законом распределения:

| | | | |
|---|---|---|---|
| **X** | 1 | 4 | 7 |
| **p** | 0,5 | 0,4 | 0,1 |

Построить график функции распределения этой величины.

5. Что называется плотностью вероятности случайной величиной?
6. Дана функция распределения *СВ X*

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

Найти: a) плотность вероятности *f(x)*;

b) математическое ожидание *M(X);*

c) дисперсию *D(X);*

d) медиану *Me(X),*моду *Mo(X)*;

e) вероятности *P(X=0,5), P(X<0,5),* $P(0,5\le X\le 1)$;

g) построить графики функций *f(x)* и *F(x).*

7. Установлено, что время ремонта телевизоров есть случайная величина *X*, распределенная по показательному закону. Определить вероятность того, что на ремонт телевизора потребуется не менее 20 дней, если среднее время ремонта телевизоров составляет 15 дней. Найти плотность вероятности, функцию распределения и среднеквадратическое отклонение случайной величины *X*.

8. Дана функция

$$f(x) = \begin{cases} 0, & x < 0 \\ Cxe^{-x} & x \geq 0 \end{cases}$$

При каком значении параметра *C* эта функция является плотностью распределения некоторой непрерывной случайной величины *X*? Найти математическое ожидание и дисперсию *СВ X*.

9. Исследовать функцию

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

10. Показать, что для функции из примера 9, параметр $a$ есть математическое ожидание, $\sigma$ – среднеквадратическое отклонение нормального распределения.

11. Случайная величина *X*, распределена по нормальному закону. Математическое ожидание и среднеквадратическое отклонение этой величины соответственно равны 30 и 10. Найти вероятность того, что X примет значение, принадлежащее интервалу (10; 50).

## 4. Многомерные случайные величины

### 4.1. Основные понятия

Часто результат испытания характеризуется не одной случайной величиной, а системой случайных величин $X_1$, $X_2$, …, $X_n$, которую называют многомерной случайной величиной.

***Пример 28.*** Успеваемость студента характеризуется системой случайных величин – оценками по различным дисциплинам.

***Пример 29.*** Деталь в виде стержня характеризуется 2 случайными величинами – длиной и диаметром.

***Пример 30.*** Погода в данном месте в данное время характеризуется системой случайных величин $X_1$ – температура, $X_2$ – влажность воздуха, $X_3$ – давление.

Случайные величины $X_1$, $X_2$, …, $X_n$, входящие в систему, могут быть дискретными (пример 28) или непрерывными (примеры 29, 30).

В случае ***многомерной случайной величины*** каждому элементарному событию ставится в соответствие несколько действительных чисел $x_1$, $x_2$,…, $x_n$, которые приняли случайные величины $X_1$, $X_2$, …, $X_n$ в результате испытания.

Будем рассматривать наиболее простой случай – двумерные случайные величины.

### 4.2. Дискретные случайные величины

Распределение двумерной случайной величины $(X, Y)$ можно представить в виде таблицы, в каждой клетке $(i, j)$ которой располагаются вероятности событий $p_{ij} = P(X = x_i,\ Y = y_j)$ и выполняется условие нормировки $\sum\limits_{i,j} p_{ij} = 1$.

Таблица 1

| $X$ / $Y$ | $x_1$ | $x_2$ | … | $x_n$ | $\sum_{j=1}^{n} p_{ij} = p_i$ |
|---|---|---|---|---|---|
| $y_1$ | $p_{11}=P(x_1,y_1)$ | $p_{12}=P(x_2,y_1)$ | … | $p_{1n}=P(x_n,y_1)$ | $\sum_{j=1}^{n} p_{1j}$ |
| $y_2$ | $p_{21}=P(x_1,y_2)$ | $p_{22}=P(x_2,y_2)$ | … | $p_{2n}=P(x_n,y_2)$ | $\sum_{j=1}^{n} p_{2j}$ |
| … | … | … | … | … | … |
| $y_m$ | $p_{m1}=P(x_1,y_m)$ | $p_{m2}=P(x_2,y_m)$ | … | $p_{mn}=P(x_n,y_m)$ | $\sum_{j=1}^{n} p_{mj}$ |
| $\sum_{i=1}^{m} p_{ij} = p_j$ | $\sum_{i=1}^{m} p_{m1}$ | $\sum_{i=1}^{m} p_{m2}$ | … | $\sum_{i=1}^{m} p_{mn}$ | $\sum_{i,j} p_{ij} = 1$ |

Полученные вероятности $p_i$ и $p_j$ дают возможность рассчитать значения математического ожидания и дисперсии для каждой из случайных величин:

| | |
|---|---|
| $a_x = M(X) = \sum_j x_j p_j$ | $a_y = M(Y) = \sum_i x_i p_i$ |
| $D(X) = \sum_j (x_j - a_x)^2 p_j$ | $D(Y) = \sum_i (y_i - a_y)^2 p_i$ |
| $\sigma_x = \sqrt{D(X)}$ | $\sigma_y = \sqrt{D(Y)}$ |

***Пример 31.*** Группа студентов в 30 человек сдает два экзамена. В таблице дано число студентов, получивших определенные оценки на двух экзаменах.

| 1 экз. / 2 экз. | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 2 | 4 | 7 | 0 | 0 |
| 3 | 9 | 14 | 3 | 0 |
| 4 | 0 | 6 | 10 | 5 |
| 5 | 0 | 0 | 0 | 2 |

Построить таблицу распределения и найти математическое ожидание (средний балл) и дисперсию для каждого экзамена.

***Решение.*** Построим таблицу распределения, разделив каждое из чисел таблицы на 60 – общее число оценок на двух экзаменах.

| Y \ X | 2 | 3 | 4 | 5 | $p_i$ |
|---|---|---|---|---|---|
| 2 | 0,07 | 0,12 | 0 | 0 | 0,19 |
| 3 | 0,15 | 0,23 | 0,05 | 0 | 0,43 |
| 4 | 0 | 0,10 | 0,17 | 0,08 | 0,35 |
| 5 | 0 | 0 | 0 | 0,03 | 0,03 |
| $p_j$ | 0,22 | 0,45 | 0,22 | 0,11 | 1 |

Запишем закон распределения случайной величины $X$ – оценок, полученных на первом экзамене, сложив вероятности «по столбцам»:

| ***X*** | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ***p*** | 0,22 | 0,45 | 0,22 | 0,11 |

Закон распределения для случайной величины $Y$ – оценок, полученных на втором экзамене, имеет вид

| ***Y*** | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ***p*** | 0,19 | 0,43 | 0,35 | 0,03 |

Вычислим математическое ожидание, дисперсию, среднеквадратическое отклонение для каждой из величин:

$$M(X) = 2\cdot 0{,}22 + 3\cdot 0{,}45 + 4\cdot 0{,}22 + 5\cdot 0{,}11 = 3{,}22\,,$$

$$M(Y) = 2\cdot 0{,}19 + 3\cdot 0{,}43 + 4\cdot 0{,}35 + 5\cdot 0{,}03 = 3{,}22\,,$$

$$D(X) = (2-3{,}22)^2\cdot 0{,}22 + (3-3{,}22)^2\cdot 0{,}45 + (4-3{,}22)^2\cdot 0{,}22 + (5-3{,}22)^2\cdot 0{,}11 = 0{,}83\,,$$

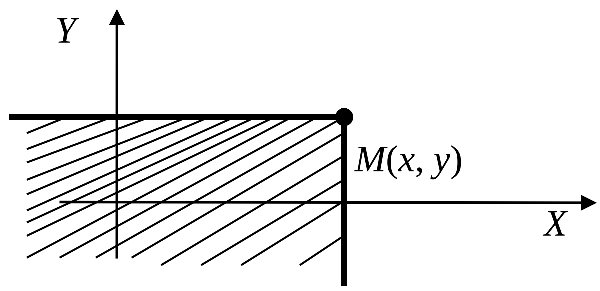
$$D(Y) = (2-3{,}22)^2\cdot 0{,}19 + (3-3{,}22)^2\cdot 0{,}43 + (4-3{,}22)^2\cdot 0{,}35 + (5-3{,}22)^2\cdot 0{,}03 = 0{,}61\,,$$

$$\sigma_x = \sqrt{D(X)} = \sqrt{0{,}83} = 0{,}91,\ \sigma_y = \sqrt{D(Y)} = \sqrt{0{,}61} = 0{,}78.$$

**Функцией распределения** случайной величины ($X$, $Y$) называют функцию $F(x,y)$, определяющую для каждой пары чисел ($x$, $y$) вероятность того, что $X$ примет значения, меньшее $x$, и при этом $Y$ примет значение, меньшее $y$:

$$F(x, y) = P(X < x, Y < y).$$

Геометрически это равенство можно пояснить так: $F(x, y)$ есть вероятность того, что случайная точка $(X, Y)$ попадет в бесконечный квадрант с вершиной $M(x, y)$, расположенный левее и ниже этой вершины.




Для дискретной случайной величины функция распределения определяется по формуле:

$$F(x, y) = \sum_i \sum_j p_{ij} \ ,$$

где суммирование вероятностей распространяется на все $i$, для которых $x_i < x$, и все $j$, для которых $y_j < y$.

## Свойства функции распределения

**1.** *Функция распределения F(x, y) есть неотрицательная функция, заключенная между нулем и единицей:*

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

**2.** *Функция распределения F(x, y) есть неубывающая функция по каждому из аргументов*, т. е.

$$\begin{aligned} &при \quad x_2 > x_1 \quad F(x_2, y) \ge F(x_1, y), \\ &при \quad y_2 > y \quad F(x, y_2) \ge F(x, y_1). \end{aligned}$$

**3.** *Если хотя бы один из аргументов обращается в $-\infty$, то функция распределения F(x, y) равна нулю*, т. е.

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

**4.** *Если один из аргументов обращается в* $+\infty$, *то функция распределения F(x, y) становится равной функции распределения случайной величины, соответствующей другому аргументу:*

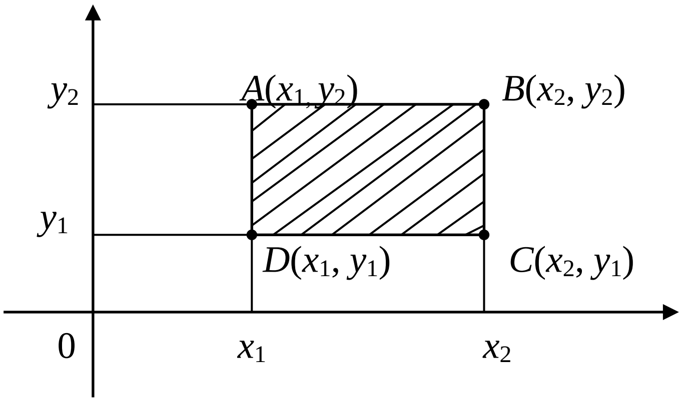
$$F(+\infty, y) = F_2(y) \overset{\Delta}{=} P(Y < y),$$

$$F(x, +\infty) = F_1(x) \overset{\Delta}{=} P(X < x).$$

**5.** *Если оба аргумента равны* $+\infty$, *то функция распределения равна единице:*

$$F(+\infty, +\infty) = 1$$

**6.** *Зная функцию распределения F(x, y), можно найти вероятность попадания случайной точки (X, Y) в пределы прямоугольника ABCD:*




$$P[(x_1 \le X < x_2)(y_1 \le Y < y_2)] =$$
$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

***Пример 32***. Задана функция распределения двумерной случайной величины:

$$F(x, y) = \begin{cases} \sin x \cdot \sin y, & 0 \le x \le \pi/2, \quad 0 \le y \le \pi/2 \\ 0 & x < 0, \qquad y < 0 \end{cases}.$$

Найти вероятность попадания случайной точки (X, Y) в прямоугольник, ограниченный прямыми $x = 0, \quad x = \pi/4, \quad y = \pi/6, \quad y = \pi/3$.

***Решение.*** Используя **свойство 6**, положим $x_1 = 0$, $x_2 = \pi/4$, $y_1 = \pi/6$, $y_2 = \pi/3$, найдем:

$$
\begin{aligned}
&P(0 < X < \pi/4, \pi/6 < Y < \pi/3) = \sin(\pi/4)\sin(\pi/3) - \\
&- \sin(0)\sin(\pi/3) - \sin(\pi/4)\sin(\pi/6) + \sin 0 \cdot \sin(\pi/6) = \\
&= \left(\sqrt{6} - \sqrt{2}\right)/4 = 0{,}26.
\end{aligned}
$$

### 4.3. Двумерные непрерывные случайные величины

Двумерная случайная величина ($X$, $Y$) называется **непрерывной**, если её функция распределения $F(x, y)$ – непрерывная функция, дифференцируемая по каждому из аргументов, и существует вторая смешанная производная $F''_{xy}(x, y)$.

**Плотностью вероятности** (плотностью распределения или совместной плотностью) непрерывной двумерной случайной величины ($X$, $Y$) называется вторая смешанная частная производная её функции распределения, т. е.

$$
f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = F''_{xy}(x, y).
$$

Геометрически плотность вероятности двумерной непрерывной случайной величины ($X$, $Y$) представляет собой поверхность распределения $f(x, y)$, в пространстве $Oxyz$.

**Свойства плотности вероятности *f*(*x*, *y*)**

1. *Плотность вероятности* $f(x, y)$ *есть неотрицательная функция, т. е.*

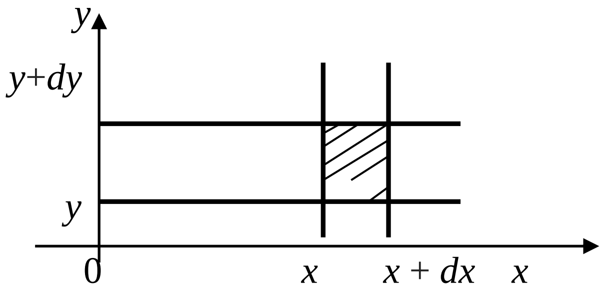
$$
f(x, y) \geq 0 .
$$

2. *Вероятность попадания непрерывной двумерной величины* ($X$, $Y$) *в область D равна*

$$P\left[\left(X,Y\right)\in D\right]=\iint\limits_{D} f\left(x,y\right)dxdy$$

Поясним геометрически эту формулу.

Введем понятие «элемент вероятности» для двумерной непрерывной случайной величины ($X$, $Y$), равный $f(x, y)dxdy$. Он представляет вероятность попадания случайной точки ($X$, $Y$) в элементарный прямоугольник со сторонами $dx$ и $dy$ .




Эта вероятность приближенно равна объему элементарного параллелепипеда с высотой $f(x, y)$, опирающегося на элементарный прямоугольник со сторонами $dx$ и $dy$. Итак, вероятность попадания непрерывной двумерной величины ($X$, $Y$) в область $D$ на плоскости $Oxy$ геометрически изображается объёмом цилиндрического тела, ограниченного сверху поверхностью распределения $f(x, y)$ и опирающегося на область $D$, а аналитически – двойным интегралом.

**3.** *Функция распределения непрерывной двумерной случайной величины может быть выражена через её плотность вероятности $f(x, y)$ по формуле:*

$$F\left(x,y\right)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f\left(x,y\right)dxdy$$

**4.** *Двойной несобственный интеграл в бесконечных пределах от плотности вероятности двумерной случайной величины равен единице:*

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy=1.$$

В этой формуле несобственный интеграл есть вероятность попадания во всю плоскость *Oxy*, т. е. вероятность достоверного события, равная единице. Это означает, что *полный объём тела, ограниченного поверхностью распределения и плоскостью Oxy, равен 1.*

***Пример 33.*** Задана функция распределения двумерной случайной величины:

$$F(x,y)=\begin{cases}1-3^{-x}-3^{-y}+3^{-x-y}, & x\ge 0, y\ge 0\\ 0 & x<0, y<0\end{cases}.$$

Найти двумерную плотность вероятности системы.

***Решение***. Используем формулу:

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}.$$

Найдем частные производные:

$$\frac{\partial F}{\partial x}=\left(1-3^{-x}-3^{-y}+3^{-x-y}\right)'_x=3^{-x}\ln 3-3^{-x-y}\ln 3=\left(3^{-x}-3^{-x-y}\right)\ln 3,$$

$$\frac{\partial^2 F}{\partial x dy}=\ln 3\left(3^{-x}-3^{-x-y}\right)'_y=\ln 3\cdot\left(3^{-x-y}\right)\ln 3=3^{-x-y}\ln^2 3.$$

Итак, искомая двумерная плотность вероятности:

$$f(x,y)=\begin{cases}3^{-x-y}\ln^2 3, & x\ge 0, y\ge 0\\ 0 & x<0, y<0\end{cases}.$$

### 4.4. Выявление связи между величинами

Во многих задачах требуется установить и оценить зависимость случайной величины $Y$ от случайной или неслучайной величины $X$ (или от нескольких величин).

Две случайные величины могут быть зависимыми, если они связаны либо функциональной зависимостью (значение $Y$ однозначно вычисляется по значению $X$), либо статистической зависимостью, когда кроме физической зависимости обе величины или одна из них подвержены влиянию случайных факторов.

**Статистической** называют зависимость, при которой изменение одной величины влечет изменение распределения другой. Самый простой случай – при изменении одной величины изменяется среднее значение (математическое ожидание) другой. В этом случае статистическую зависимость называют **корреляционной.**

***Пример***. Пусть $Y$ – урожай зерна, $X$ – количество удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т. е. $Y$ не является функцией $X$. Это объясняется влиянием различных случайных факторов (осадки, уход, тип почвы и др.). Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т. е. $Y$ связан с $X$ корреляционной зависимостью.

В терминах теории вероятностей при статистической зависимости $X$ и $Y$ как случайные величины задаются своим совместным распределением вероятностей (см. таблицу 1).

**Условным законом распределения** одной из одномерных составляющих двумерной случайной величины $(X, Y)$ называется её закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение.

То есть, зафиксировав значение одного из аргументов $X = x_j$ ($j$ сохраняет одно и тоже значение при всех возможных значениях $Y$) можно получить условное распределение $P(Y / X = x_j)$ по соотношениям:

$$P(Y = y_1 / X = x_j) = \frac{P(x_j, y_1)}{P(x_j)}, \ldots, P(Y = y_i / X = x_j) = \frac{P(x_j, y_i)}{P(x_j)}, \ldots,$$

$$P(Y = y_m / X = x_j) = \frac{P(x_j, y_m)}{P(x_j)}.$$

Аналогично находим условное распределение $P(X / Y = y_i)$, $P(X = x_j / Y = y_i) = \dfrac{P(x_j, y_i)}{P(y_i)}$, где $j$ принимает значения от 1 до $n$. Для контроля вычислений целесообразно убедиться, что сумма вероятностей условного распределения равна единице.

По условным вероятностям можно сосчитать условные математические ожидания $M(X/Y = y)$ и $M(Y/X = x)$ и условные дисперсии $D(X/Y = y)$ и $D(Y/X = x)$. Они вычисляются по формулам:

$$M(Y / X = x_j) = \sum_{i=1}^{m} y_i P(Y = y_i / X = x_j),$$

$$D(Y / X = x_j) = \sum_{i=1}^{m} (y_i - M(Y / X = x_j))^2 P(Y = y_i / X = x_j),$$

$$M(X / Y = y_i) = \sum_{j=1}^{n} x_j P(X = x_j / Y = y_i),$$

$$D(X / Y = y_i) = \sum_{j=1}^{n} (x_j - M(X / Y = y_i))^2 P(X = x_j / Y = y_i).$$

***Пример 34.*** Условие то же, что и у примера 31. Найти распределение оценок у студентов, сдавших первый экзамен на «отлично».

***Решение.*** Распределение оценок у студентов, сдавших первый экзамен на отлично – это условная вероятность распределения величины $Y$ при условии, что $X = 5$.

По формуле $P\left(Y = y_i \,/\, X = x_j\right) = \dfrac{P\left(x_j, y_i\right)}{P\left(x_j\right)}$, $x_j = 5$, $P\left(x_j\right) = 0{,}11$.

Таким образом

$$P\left(Y = 4 \,/\, X = 5\right) = \frac{0{,}08}{0{,}11} = 0{,}73,\ P\left(Y = 5 \,/\, X = 5\right) = \frac{0{,}03}{0{,}11} = 0{,}27\,.$$

Из студентов, сдавших первый экзамен на «отлично», 73 % сдали второй на 4, а 27 % – на 5.

**Функция регрессии.** Условное математическое ожидания $M(Y/X = x)$ есть функция от $x$:

$$M(Y/X = x) = f(x).$$

Эту функцию называют **функцией регрессии *Y* на *X***, а её график называется **линией регрессии**. Аналогично определяется регрессия $X$ на $Y$. Если случайные величины $X$ и $Y$ *независимы,* то линии регрессии $Y$ по $X$ и $X$ по $Y$ *соответственно параллельны координатным осям* $0x$ и $0y$.

***Пример 35.*** Условие то же, что и у примера 31. Найти средний балл, который получили на втором экзамене студенты, сдавшие первый экзамен на a) 2, b) 3, c) 4, d) 5.

***Решение.*** Надо найти регрессию $Y$ на $X$ (средний балл на втором экзамене) $M(Y/X = x)$ для значений $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, $x_4 = 5$. Воспользуемся ранее рассчитанной таблицей распределения вероятностей и формулой

$$M\left(Y \,/\, X = x_j\right) = \sum_{i=1}^{m} y_i P\left(Y = y_i \,/\, X = x_j\right):$$

$$M\left(Y \,/\, X = 2\right) = \frac{2 \cdot 0{,}07 + 3 \cdot 0{,}15}{0{,}22} = 2{,}68\,,$$

$$M(Y/X=3)=\frac{2\cdot 0{,}12+3\cdot 0{,}23+4\cdot 0{,}10}{0{,}45}=2{,}96,$$

$$M(Y/X=4)=\frac{3\cdot 0{,}05+4\cdot 0{,}17}{0{,}22}=3{,}77,$$

$$M(Y/X=5)=\frac{4\cdot 0{,}08+5\cdot 0{,}03}{0{,}11}=4{,}27.$$

Мы получили, что линия регрессии $M(2) = 2{,}68$, $M(3) = 2{,}96$, $M(4) = 3{,}77$, $M(5) = 4{,}27$ не параллельна оси $X$, следовательно, случайные величины $X$ и $Y$ не будут независимыми. Значит средний балл, который студенты получили на втором экзамене, не является независимой величиной, а зависит от того, какой средний балл они получили на первом экзамене.

Случайные величины $X$ и $Y$ называются **независимыми**, если их совместная функция распределения $F(x, y)$ представляется в виде произведения функций распределения $F(x)$ и $F(y)$:

$$F(x, y) = F(x)F(y).$$

В противном случае случайные величины $X$ и $Y$ называются зависимыми. Из этой формулы следует, что для дискретных величин:

$$P(X=x_j, Y=y_i)=P(X=x_j)P(Y=y_i),$$

а для непрерывных величин:

$$f(x, y)= f(x)f(y).$$

Для удобства анализа зависимости между двумя случайными величинами вместо исследования линий регрессии рассматривается значение ковариации случайных величин или их коэффициент корреляции.

**Ковариация (корреляционный момент)** двух случайных величин определяется как математическое ожидания их произведения минус произведение их математических ожиданий:

$$K_{xy} = M[(X - M(X))(Y - M(Y))] = M(XY) - M(X)M(Y).$$

Для дискретных величин используют формулу:

$$K_{xy} = \sum_{j=1}^{n}\sum_{i=1}^{m} [x_j - M(X)]\cdot[y_i - M(Y)]P(x_j, y_i)$$

Для непрерывных величин используют формулу:

$$K_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x - M(X)]\cdot[y - M(Y)]f(x, y)dxdy.$$

Ковариация – величина размерная (равная произведению размерностей величин $X$ и $Y$), поэтому вместо неё обычно рассматривают безразмерную величину $\rho$ – коэффициент корреляции.

**Коэффициентом корреляции** двух случайных величин называется отношение их ковариации к произведению среднеквадратических отклонений этих величин:

$$\rho = \frac{K_{xy}}{\sigma_x \sigma_y}.$$

**Свойства коэффициента корреляции**

1. $-1 \le \rho \le 1$.
2. *Если коэффициент корреляции случайных величин X и Y равен нулю, то такие величины называются* **некоррелированными**. *Если случайные величины независимы, то они некоррелированы*. Обратное утверждение, вообще говоря, неверно: из того, что $\rho = 0$, еще не

следует независимость случайных величин. Тем не менее в важном частном случае совместного нормального распределения величин *X* и *Y* некоррелированность и независимость являются равносильными понятиями (далее этот случай рассматривается более подробно).

3. *Если у двух случайных величин* $|\rho| = 1$, *то между этими случайными величинами существует линейная функциональная зависимость*; причем если связь по возрастающей (большему значению X соответствует большее значение *Y*), то $\rho = 1$, а если связь по убывающей (большему значению *X* соответствует меньшее значение *Y*), то $\rho = -1$.

***Пример 36.*** Условие то же, что и у примера 31. Определить коэффициент корреляции случайных величин *X* и *Y*.

***Решение.*** Ранее мы получили *M*(*X*)=3,22, *M*(*Y*)=3,22, $\sigma_x = 0{,}91$, $\sigma_y = 0{,}78$.

Теперь рассчитаем $K_{xy}$ и $\rho$.

| ***X – M(X)*** / Y – M(Y) | 2 – 3,22 | 3 – 3,22 | 4 – 3,22 | 5 – 3,22 |
|---|---|---|---|---|
| 2 – 3,32 | 0,07 | 0,12 | 0 | 0 |
| 3 – 3,22 | 0,15 | 0,23 | 0,05 | 0 |
| 4 – 3,22 | 0 | 0,10 | 0,17 | 0,08 |
| 5 – 3,22 | 0 | 0 | 0 | 0,03 |

$$K_{xy} = (2-3{,}22)(2-3{,}22)0{,}07 + (2-3{,}22)(3-3{,}22)0{,}12 + (3-3{,}22)(2-3{,}22)0{,}15 +$$
$$+ (3-3{,}22)(3-3{,}22)0{,}23 + (3-3{,}22)(4-3{,}22)0{,}05 + (4-3{,}22)(3-3{,}22)0{,}1 +$$
$$+ (4-3{,}22)(4-3{,}22)0{,}17 + (4-3{,}22)(5-3{,}22)0{,}08 + (5-3{,}22)(5-3{,}22)0{,}03 = 0{,}52$$

$$\rho = \frac{0{,}52}{0{,}91 \cdot 0{,}78} = 0{,}73$$

Таким образом, величины *X* и *Y* – оценки, полученные на двух экзаменах нельзя считать независимыми. Об этом говорит и коэффициент корреляции и линия регрессии, полученная ранее.

### 4.5. Двумерное нормальное распределение

Случайная величина (*X*, *Y*) распределена по двумерному нормальному закону, если совместная плотность величин *x* и *y* имеет вид:

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-a_x}{\sigma_x}\right)^2-2\rho\frac{x-a_x}{\sigma_x}\cdot\frac{y-a_y}{\sigma_y}+\left(\frac{y-a_y}{\sigma_y}\right)^2\right]}.$$

Таким образом, двумерный нормальный закон распределения определяется пятью параметрами: $a_x,\ a_y,\ \sigma_x,\ \sigma_y,\ \rho.$ Параметры $a_x,\ a_y$ выражают математические ожидания величин *X* и *Y*, параметры $\sigma_x,\ \sigma_y$ – их среднеквадратические отклонения, $\rho$ – коэффициент корреляции случайных величин *X* и *Y*. Можно показать, что условные математические ожидания и дисперсии, сосчитанные по этому закону распределения, выражаются с помощью следующих формул:

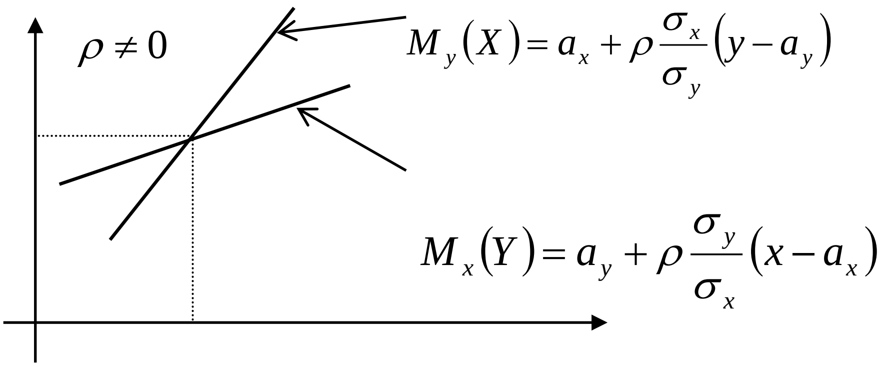
$$M(X/Y=y)=M_y(X)=a_x+\rho\frac{\sigma_x}{\sigma_y}\left(y-a_y\right),$$

$$D(X/Y=y)=\sigma_x^2\left(1-\rho^2\right),$$

$$M(Y/X=x)=M_x(Y)=a_y+\rho\frac{\sigma_y}{\sigma_x}\left(x-a_x\right),$$

$$D(Y/X=x)=\sigma_y^2\left(1-\rho^2\right).$$

Из этих формул вытекает, что линии регрессии нормально распределенных случайных величин представляют собой прямые линии.

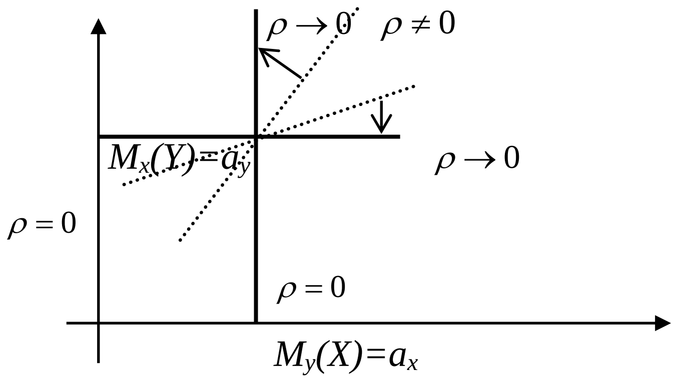



Условные дисперсии постоянны и не зависят от значения второй случайной величины. Заметим, что из вида плотности двумерного нормального распределения следует, что если $\rho = 0$, то компоненты *X* и *Y* независимы. Действительно, в этом случае:

$$f(x,y) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{\frac{(x-a_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} e^{\frac{(y-a_y)^2}{2\sigma_y^2}} = f(x) \cdot f(y),$$

где *f(x)* и *f(y)* – плотности вероятностей одномерных случайных величин *X* и *Y*. Уравнения регрессии обратятся в $M_y(X)=a_x$ и $M_x(Y)=a_y$.

Линии регрессии при $\rho = 0$ приобретут вид двух прямых, перпендикулярных координатным осям.




С уменьшением $\rho$ ножницы становятся все шире, пока не обратятся в две линии, перпендикулярные осям координат.

## Вопросы и задачи

1. Что называется многомерной случайной величиной?
2. Как можно представить распределение двумерной дискретной случайной величины?
3. Что называется условным законом распределения одной из составляющих двумерной случайной величины?
4. Задана дискретная двумерная случайная величина $(X, Y)$:

| $\boldsymbol{Y / X}$ | $x_1 = 3$ | $x_2 = 6$ |
|---|---|---|
| $y_1 = 10$ | 0,25 | 0,10 |
| $y_2 = 14$ | 0,15 | 0,05 |
| $y_3 = 18$ | 0,32 | 0,13 |

Найти: a) безусловные законы распределения составляющих;

b) условный закон распределения составляющей $X$ при условии, что $Y = 10$;

c) условный закон распределения составляющей $Y$ при условии, что $X = 6$.

*Ответ*: a)

| $X$ | 3 | 6 |
|---|---|---|
| $p$ | 0,72 | 0,28 |

| $Y$ | 10 | 14 | 18 |
|---|---|---|---|
| $p$ | 0,35 | 0,20 | 0,45 |

b)

| $X$ | 3 | 6 |
|---|---|---|
| $P(X/10)$ | 5/7 | 2/7 |

c)

| $Y$ | 10 | 14 | 18 |
|---|---|---|---|
| $P(Y/6)$ | 5/14 | 5/28 | 13/28 |

5. Что называется функцией регрессии случайной величины?
6. Задано распределение двумерной случайной величины ($X$, $Y$):

| $Y / X$ | $x_1 = 1$ | $x_2 = 3$ | $x_3 = 4$ | $x_4 = 8$ |
|---|---|---|---|---|
| $y_1 = 3$ | 0,15 | 0,06 | 0,25 | 0,04 |
| $y_2 = 6$ | 0,30 | 0,10 | 0,03 | 0,07 |

Найти: a) безусловные законы распределения составляющих;

b) условный закон распределения величины $X$ при условии, что $Y = 6$;

c) условный закон распределения составляющей $Y$ при условии, что $X = 1$;

d) условное математическое ожидание $Y$ при условии, что $X = 4$;

e)условное математическое ожидание $X$ при условии, что $Y = 3$.

(d) $M(Y/X = 4)$=3,32 …; e) $M(X/Y = 3) = 3{,}3$)

7. Какой вид имеют линии регрессии для независимых случайных величин?
8. Какие случайные величины называются независимыми?
9. Как определяется коэффициент корреляции?
10. Что можно сказать о случайных величинах $X$ и $Y$, если для них коэффициент корреляции равен 0,5?

11. По закону распределения двумерной случайной величины ($X$, $Y$):

| **Y/X** | **2** | **3** |
|---|---|---|
| **3** | 0,3 | 0,4 |
| **5** | 0,1 | 0,2 |

Найти: а) математические ожидания и среднеквадратические отклонения $X$ и $Y$;

b) корреляционный момент и коэффициент корреляции $X$ и $Y$;

c) значения регрессии $Y$ на $X$.

(a) $M(X) = 2{,}6$, $M(Y) = 3{,}6$, $\sigma_x = 0{,}4898$, $\sigma_y = 0{,}9165$;

b) $K_{xy} = 0{,}04$, $\rho = 0{,}0891$;

c) $M(Y/X = 2) = 3{,}5$, $M(Y/X = 3) = 3{,}667$)

## 5. Статистический анализ данных

### 5.1. Понятие вариационного ряда

Анализируемые данные могут быть заданы последовательностью полученных значений $x_1, x_2, \ldots, x_n$, либо таблицей частот:

| Значения $x_i$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| Частоты $m_i$ | $m_1$ | $m_2$ | … | $m_n$ |

либо таблицей относительных частот, т. е. таблицей эмпирического распределения выборки:

| Значения $x_i$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| Относительные частоты $m_i/n$ | $m_1/n$ | $m_2/n$ | | $m_n/n$ |

Формулы, по которым в зависимости от описания данных выборки вычисляются среднее значение и разброс выборки (дисперсия наблюдений), приведены в таблице

| | Вариационный ряд задан последовательностью | Задана таблица частот вариационного ряда | Задана таблица относительных частот вариационного ряда |
|---|---|---|---|
| Среднее значение выборки $\overline{x}$ | $\frac{1}{n}\sum_{i=1}^{n} x_i$ | $\frac{1}{n}\sum_{i=1}^{k} x_i m_i$ | $\sum_{i=1}^{k} x_i \frac{m_i}{n}$ |
| Дисперсия (разброс) выборки $S^2$ | $\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2 = \overline{x^2}-\overline{x}^2$ | $\frac{1}{n}\sum_{i=1}^{k} x_i m_i - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^2 m_i$ | $\sum_{i=1}^{k} x_i^2 \frac{m_i}{n} - \overline{x}^2 = \sum_{i=1}^{k}\left(x_i-\overline{x}\right)^2 \frac{m_i}{n}$ |

***Пример 37.*** Найти $\bar{x}$ и $S^2$ по таблице выборки:

| | | | | |
|---|---|---|---|---|
| ***Варианты*** | 2 | 6 | 12 | |
| $m_i$ | 3 | 10 | 7 | $\boldsymbol{n} = 20$ |

***Решение.***

$$\bar{x} = (2 \cdot 3 + 6 \cdot 10 + 12 \cdot 7)/20 = 7{,}5\,;$$

$$S^2 = \frac{1}{20}(4 \cdot 3 + 36 \cdot 10 + 144 \cdot 7) - (7{,}5)^2 =$$

$$= \frac{1}{20} \cdot 1380 - 56{,}25 = 69 - 56{,}25 = 12{,}75.$$

## 5.2. Статистические оценки параметров распределения

В практике статистических исследований различают два вида наблюдений: *сплошное*, когда изучаются все объекты совокупности (перепись населения), и *выборочное*, когда изучается часть объектов (социологические исследования). Вся подлежащая изучению совокупность объектов (наблюдений) называется генеральной совокупностью. Та часть объектов, которая отобрана для непосредственного изучения из генеральной совокупности, называется выборкой.

Важнейшей задачей выборочного метода является *оценка параметров (характеристик)* генеральной совокупности по данным выборки.

Пусть $x_1, x_2, \ldots, x_n$ – выборка, отвечающая случайной величине $X$ с рядом распределения $p(x_i, \Theta) = P(X = x_i)$ (для дискретной случайной величины $X$) или плотностью вероятности $f(x, \Theta)$ (для непрерывной случайной величины $X$), которая зависит от некоторого неизвестного параметра $\Theta$. Таким параметром может быть генеральная средняя, т. е. математическое ожидание $M(X)$, или генеральная дисперсия, т. е. дисперсия $D(X)$. Требуется оценить параметр $\Theta$.

**Статистической оценкой** $\Theta^*$ неизвестного параметра $\Theta$ теоретического распределения называют функцию $\Theta(X_1,X_2,\ldots,X_n)$ от наблюдаемых случайных величин $X_1, X_2, \ldots, X_n$.

**Точечной** называют статистическую оценку, которая определяется одним числом $\Theta^* = f(x_1, x_2, \ldots, x_n)$, где $x_1, x_2, \ldots, x_n$ – результаты $n$ наблюдений над количественным признаком $X$ (выборка).

Для того чтобы хорошо аппроксимировать $\Theta$, оценка $\Theta^*$ в соответствии с требованиями математической статистики должна удовлетворять ряду критериев, основными из которых являются требования *состоятельности и несмещенности*.

**Оценка** $\Theta^*$ **называется** состоятельной **оценкой параметра** $\Theta$**, если**:

$$\Theta^* \underset{n\to\infty}{\overset{P}{\to}} \Theta,$$

т. е. $P(|\Theta^* - \Theta| > \varepsilon) \to 0$ при любом $\varepsilon > 0$.

**Несмещенной** называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки:

$$M(\Theta^*) = \Theta$$

Это означает, что отклонение $\Theta^*$ от $\Theta$ не содержит систематической ошибки.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя:

$$\overline{x} = \frac{\sum_{i=1}^{k} m_i x_i}{n},$$

где $x_i$ – варианта выборки, $m_i$ – частота варианты $x_i$, $n=\sum_{i=1}^{k} m_i$ – объем выборки.

Несмещенной оценкой генеральной дисперсии является исправленная выборочная дисперсия:

$$\hat{S}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 .$$

Можно показать, что выборочная средняя $\bar{x}$ и исправленная выборочная дисперсия $\hat{S}^2$ являются ***несмещенными состоятельными оценками***.

**Смещенной** называют точечную оценку, математическое ожидание которой не равно оцениваемому параметру.

**Оценкой генеральной дисперсии** может служить и ***выборочная дисперсия***:

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

В отличие от исправленной выборочной дисперсии, эта оценка является смещенной состоятельной оценкой и связана с $\hat{S}^2$ соотношением:

$$\hat{S}^2=\frac{n}{n-1}S^2 .$$

***Пример 38.*** По статистическому распределению выборки из примера 35 найти несмещенные оценки генеральной средней и генеральной дисперсии.

***Решение.*** Несмещенной оценкой генеральной средней служит выборочная средняя

$$\bar{x}=7{,}5$$

Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия:

$$\hat{S}^2 = \frac{20}{19} \cdot 12{,}75 = 13{,}42 .$$

## 5.3. Метод максимального правдоподобия

Одним из методов, получения оценок параметров генеральной совокупности по данным выборки, является **метод максимального правдоподобия, предложенный Р. Фишером.**

Основу метода составляет **функция правдоподобия**, выражающая вероятность совместного появления результатов выборки $x_1, x_2, \ldots, x_n$ (для дискретной случайной величины $X$)

$$L(x_1, x_2, \ldots, x_n, \Theta) = p(x_1, \Theta) \cdot p(x_2, \Theta) \cdot \ldots \cdot p(x_n, \Theta),$$

или *плотность вероятности* совместного появления результатов выборки $x_1, x_2, \ldots, x_n$ (для непрерывной случайной величины $X$)

$$L(x_1, x_2, \ldots, x_n, \Theta) = f(x_1, \Theta) \cdot f(x_2, \Theta) \cdot \ldots \cdot f(x_n, \Theta).$$

Согласно методу максимального правдоподобия в качестве оценки неизвестного параметра $\Theta$ принимается такое значение $\Theta^*$, которое максимизирует функцию $L$. Нахождение оценки $\Theta^*$ упрощается, если максимизировать не саму функцию $L$, а $\ln L$, поскольку максимум обеих функций достигается при одном и том же значении $\Theta$. Поэтому для отыскания оценки параметра $\Theta$ надо решить уравнение правдоподобия $\frac{d \ln L}{d\Theta} = 0$. Затем найти вторую производную $\frac{d^2 \ln L}{d\Theta^2}$; если вторая производная при $\Theta = \Theta^*$ отрицательна, то $\Theta^*$ – точка максимума. Найденную точку максимума $\Theta^*$ принимают в качестве оценки наибольшего правдоподобия параметра $\Theta$.

Если плотность вероятности *f(x)* непрерывной случайной величины определяется двумя неизвестными параметрами $\Theta_1$ и $\Theta_2$, то функция правдоподобия есть функция двух независимых аргументов $\Theta_1$ и $\Theta_2$:

$$L = f(x_1, \Theta_1, \Theta_2) \cdot f(x_2, \Theta_1, \Theta_2) \cdot \ldots \cdot f(x_n, \Theta_1, \Theta_2).$$

Далее находят логарифмическую функцию правдоподобия и для отыскания её максимума составляют и решают систему:

$$\begin{cases} \dfrac{\partial \ln L}{\partial \Theta_1} = 0 \\ \dfrac{\partial \ln L}{\partial \Theta_2} = 0. \end{cases}$$

***Пример 39.*** Найти методом максимального правдоподобия по выборке $x_1, x_2, \ldots, x_n$ точечную оценку неизвестного параметра $\lambda$ показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$.

***Решение.*** Запишем функцию правдоподобия:

$$L = f(x_1, \Theta) \cdot f(x_2, \Theta) \cdot \ldots \cdot f(x_n, \Theta),$$

и учитывая, что $\Theta = \lambda$, получим:

$$L = \left(\lambda e^{-\lambda x_1}\right) \cdot \left(\lambda e^{-\lambda x_2}\right) \cdot \ldots \cdot \left(\lambda e^{-\lambda x_n}\right) = \lambda^n e^{-\lambda \sum x_i},$$

или после логарифмирования:

$$\ln L = n \ln \lambda - \lambda \sum x_i .$$

Найдем первую производную по $\lambda$:

$$\frac{d \ln L}{d\lambda} = \frac{n}{\lambda} - \sum x_i .$$

***и решив уравнение правдоподобия определим оценку неизвестного параметра*** $\lambda$***:***

$$\frac{n}{\lambda} - \sum x_i = 0\,,$$

$$\lambda = \frac{1}{\left(\sum x_i\right)/n} = \frac{1}{\overline{x}}\,.$$

Вторая производная по $\lambda$:

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{n}{\lambda^2}\,.$$

при $\lambda = 1/\overline{x}$ отрицательна, следовательно, эта точка максимума. Следовательно, в качестве оценки максимального правдоподобия надо принять величину $\lambda^* = 1/\overline{x}$ .

### 5.4. Метод доверительных интервалов

Рассмотренные выше численные оценки МО и дисперсии по конечной выборке наблюдений представляют собой примеры точечных оценок неизвестных параметров. Точечных в том смысле, что в качестве оценки параметра получают то или иное конкретное число, т. е. точки на числовой оси. Наряду с точечным оцениванием статистическая теория оптимальных оценок занимается и вопросами интервального оценивания неизвестных параметров. *Доверительным интервалом* некоторого параметра $\Theta$ называют такой интервал $\left[\Theta_n^{(1)}, \Theta_n^{(2)}\right]$, для которого выполняется требование:

$$P\left(\Theta_n^{(1)} \le \Theta \le \Theta_n^{(2)}\right) = p = \mathrm{const} > 0\,.$$

При этом константа $p > 0$ определяет *доверительную вероятность* попадания в доверительный интервал истинного значения неизвестного параметра. Чем больше вероятность $p$, тем надежнее интервальная оценка. На практике, как правило, выбирают $p \approx 1$, поэтому константу $p$ часто задают следующим образом: $p = 1 - \alpha$, где $\alpha << 1$ – *уровень значимости интервальной оценки.*

## 5.5. Построение доверительного интервала для оценки математического ожидания по выборке из нормальной генеральной совокупности

**Случай А.** При известной дисперсия $\sigma^2 = \text{const}$ доверительный интервал для оценки неизвестного МО $m$ вычисленной по выборке объёма $n$ определяется следующим образом:

$$\overline{X}_n - \frac{z_p \sigma}{\sqrt{n}} < m < \overline{X}_n + \frac{z_p \sigma}{\sqrt{n}}.$$

где $\overline{X}_n$ – несмещённая оценка МО, $z_p$ такое значение аргумента интеграла вероятности (функция Лапласа) $\Phi(z_p)$, при котором $\Phi(z_p) = p = \text{const}$. По таблицам нормального распределения (приложение 1) находим соответствующее значение порогового уровня $z_p$, где индекс $p$ указывает на очевидную взаимосвязь порогового уровня с константой $p$. Его длина $\Delta_p = 2\frac{z_p \sigma}{\sqrt{n}}$ асимптотически стремится к нулю при $n \to \infty$.

**Случай Б.** Если дисперсия $\sigma^2 = ?$– неизвестна, то доверительный интервал для оценки неизвестного МО $m$

$$\overline{X}_n - \frac{t_{n,p} \hat{S}}{\sqrt{n}} < m < \overline{X}_n + \frac{t_{n,p} \hat{S}}{\sqrt{n}}$$

определяется законом распределения Стьюдента и исправленной выборочной дисперсией $\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$. Задаваясь некоторым значением константы $p$ как доверительной вероятностью и пользуясь таблицей $t$-распределения (приложение 2), находим соответствующее значение статистики $t_{n,p}$, где $n$ – объём выборки наблюдений (параметр распределения).

Таким образом, с доверительной вероятностью $p$ можно утверждать, что неизвестное значение МО находится в пределах доверительного интервала:

$$\left[\overline{X}_n - \frac{t_{n,p}\hat{S}}{\sqrt{n}}; \quad \overline{X}_n + \frac{t_{n,p}\hat{S}}{\sqrt{n}}\right].$$

***Пример 40.*** Результаты 20 опытов над случайной гауссовой величиной $X$ (нормальный закон распределения) приведены в таблице:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| $X_i$ | 10,5 | 10,8 | 11,2 | 10,9 | 10,4 | 10,6 | 10,9 | 11,0 | 10,3 | 10,8 |
| $i$ | *11* | *12* | *13* | *14* | *15* | *16* | *17* | *18* | *19* | *20* |
| $X_i$ | 10,6 | 11,3 | 10,5 | 10,7 | 10,8 | 10,9 | 10,8 | 10,7 | 10,9 | 11,0 |

Найти оценку МО и построить доверительный интервал, соответствующий доверительной вероятности $p$ = 0,8.

**Решение.** Оценка математического ожидания, рассчитанная по выборке объёмом $n$ = 20, равна $\overline{X}_n = \frac{1}{20}\sum_{i=1}^{20} X_i = 10{,}78$. Исправленная выборочная дисперсия $\hat{S}^2 = \frac{1}{19}\sum_{i=1}^{20}\left(X_i - \overline{X}_n\right)^2 = 0{,}064$, а $\hat{S} = 0{,}253$. По таблице $t$-распределения Стьюдента (приложение 2) при $n$ = 20 и $p$ = 0,8 находим $t_{n,p} = 1{,}325$ и получаем $\frac{t_{n,p}\hat{S}}{\sqrt{n}} = 0{,}075$, откуда

$$\left[\overline{X}_n - \frac{t_{n,p}\hat{S}}{\sqrt{n}};\quad \overline{X}_n + \frac{t_{n,p}\hat{S}}{\sqrt{n}}\right] = [10{,}705;\ 10{,}855].$$

## 5.6. Построение доверительного интервала для оценки дисперсии по выборке из нормальной генеральной совокупности

Доверительный интервал для оценки неизвестной дисперсии $\sigma^2$ вычисленной по выборке объёма *n*

$$\frac{n\hat{S}^2}{\chi_2^2} < \sigma^2 < \frac{n\hat{S}^2}{\chi_1^2}$$

определяется законом распределения Пирсона ( $\chi^2$ – распределение) с *n* или *n-1* степенями свободы в зависимости от рассматриваемого случая: известно или неизвестно априори МО *M(X),* а также исправленной выборочной дисперсией $\hat{S}^2$. С учётом несимметричности распределения Пирсона, для расчета значений $\chi_1^2$ и $\chi_2^2$ зададим вероятности $P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2}$, и $P\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2}$, где $\alpha = 1 - p$. По таблицам $\chi^2$ – распределения (приложение 3) определим соответствующие значения двух пороговых уровней $\chi_1^2$, $\chi_2^2$, и подставим эти найденные значения в выражение для доверительного интервала.

Таким образом, с доверительной вероятностью $p = 1 - \alpha$ значение неизвестной дисперсии не выходит за пределы доверительного интервала:

$$\left[\frac{nS^2}{\chi_2^2};\quad \frac{nS^2}{\chi_1^2}\right].$$

Его длина $\Delta_n = nS^2\left(\chi_1^{-2} - \chi_2^{-2}\right)$ стремится к нулю при $n \to \infty$.

***Пример 41.*** Найти доверительный интервал для оценки неизвестной дисперсии, соответствующий доверительной вероятности $p$ = 0,8 в условиях примера 40.

**Решение.** Для доверительной вероятности $p$ = 0,8 значение $\alpha = 1 - p = 0{,}2$ и $\alpha/2 = 0{,}1$. По таблице распределения Пирсона (приложение 3) для числа степеней свободы $n - 1 = 19$ и $\alpha/2 = 0{,}1$ находим $\chi_2^2$ = 27,2, а для $1 - \alpha/2 = 0{,}9$ получаем $\chi_1^2$ = 11,65. С учётом ранее вычисленного значения $\hat{S}^2 = 0{,}064$ находим доверительный интервал для дисперсии:

$$\left[\frac{nS^2}{\chi_2^2};\quad \frac{nS^2}{\chi_1^2}\right] = [0{,}045;\ 0{,}104].$$

## Вопросы и задачи

1. Что характеризуют среднее значение $\overline{x}$ и разброс выборки (дисперсия наблюдений) $S^2$?
2. Найти $\overline{x}$ и $S^2$ по выборке

   0.3, 0.33, 0.37, 0.36, 0.31, 0.29, 0.34, 0.39, 0.37, 0.38, 0.35, 0.32, 0.39, 0.3, 0.32, 0.32, 0.38, 0.37, 0.38, 0.33, 0.37, 0.33, 0.34, 0.33, 0.3, 0.34, 0.36, 0.33, 0.34, 0.36, 0.29, 0.3, 0.33, 0.32, 0.32, 0.38, 0.37, 0.34, 0.35, 0.36.
3. Каким критериям должна удовлетворять статистическая оценка?
4. Найти по методу максимального правдоподобия точечную оценку неизвестных параметра $m$ и σ нормального распределения.

5. В условиях примера 40 найти доверительный интервал, если дисперсия известна и равна 0,064. Сравнить результат с ранее полученным.

6. Независимая выборка наблюдений объемом $n = 40$:

0.3, 0.33, 0.37, 0.36, 0.31, 0.29, 0.34, 0.39, 0.37, 0.38, 0.35, 0.32, 0.39, 0.3, 0.32, 0.32, 0.38, 0.37, 0.38, 0.33, 0.37, 0.33, 0.34, 0.33, 0.3, 0.34, 0.36, 0.33, 0.34, 0.36, 0.29, 0.3, 0.33, 0.32, 0.32, 0.38, 0.37, 0.34, 0.35, 0.36.

взята из нормального распределения с неизвестными параметрами. Найти доверительный интервал для оценки математического ожидания при доверительной вероятности $p = 0{,}8$. (0.332 0.351)

7. Найти доверительный интервал в условиях примера 41 для доверительной вероятности $p = 0{,}9$. Сравнить с ранее полученным результатом.

## 6. Проверка статистических гипотез

Статистическая гипотеза – любое высказывание или суждение о статистических свойствах или законе распределения анализируемой СВ. Основной является задача проверки нулевой гипотезы $H_0$ против альтернативы $H_1$ на основании имеющихся выборочных данных $X_1, X_2, \ldots, X_n$ конечного объёма $n < \infty$.

### 6.1. Проверка гипотез о математическом ожидании нормальной генеральной совокупности

Пусть требуется проверить исходную гипотезу

$$H_0 \colon X \sim N\left(m_0, \sigma^2\right)$$

против альтернативы

$$H_1 \colon X \sim N\left(m_1, \sigma^2\right),$$

где $\sigma^2 = D(X) = \text{const}$.

Задача относится к случаю проверки простых гипотез. Используя критерий максимального правдоподобия, при дополнительном предположении $m_1 > m_0$ имеем решающее (оптимальное) правило:

$$W: \quad \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n} \geq \frac{m_1 + m_0}{2}.$$

Здесь $\overline{X_n}$ – минимальная решающая статистика, которая называется достаточной статистикой задачи; $\dfrac{m_1 + m_0}{2}$ – пороговый уровень.

Полученное правило принятия решения в задаче проверки простых гипотез о МО по критерию максимального правдоподобия (МП) формулируется следующим образом. *Если средняя выборочная величина* $\overline{X_n}$ *пре-*

*вышает порог* $\frac{m_1 + m_0}{2}$, *то гипотеза* $H_0$ *отвергается (принимается* $H_1$*). В противном случае принимается гипотеза* $H_0$.

Анализ эффективности. Эффективность полученного алгоритма оптимального решения определяется вероятностью ошибки первого рода

$$\alpha = P\left\{ \overline{X_n} > \frac{m_1 + m_0}{2} \middle| H_0 \right\}.$$

Поскольку $\overline{X_n} \middle| H_0 \sim N\left(m_0, \sigma^2/n\right)$, где $n$ – объём выборки, то вероятность случайного события

$$P\left\{ \overline{X_n} > \frac{m_1 + m_0}{2} \middle| H_0 \right\} == \frac{1}{2}\left[1 - \Phi\left(z_p\right)\right],$$

где $\Phi\left(z_p\right)$ – значение интеграла вероятности или функции Лапласа (приложение 1) в точке

$$z_p = \frac{m_1 - m_0}{2\sigma}\sqrt{n}.$$

Аналогично получаем выражение для вероятности ошибки второго рода

$$P\left\{ \overline{X_n} < \frac{m_1 + m_0}{2} \middle| H_1 \right\} = \frac{1}{2}\left[1 - \Phi\left(\frac{m_1 - m_0}{2\sigma}\sqrt{n}\right)\right] = \alpha.$$

***Пример 42.*** Из нормальной генеральной совокупности извлечена выборка объёма $n = 8$:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $X_i$ | 2 | 0 | -2 | -4 | 0 | 4 | 6 | 2 |

.

1. При заданном значении $\sigma^2 = 16$ принять решение по критерию максимального правдоподобия о значении математического ожидания $m_0 = 0{,}2$ или $m_1 = 2$.

2. Дать количественную характеристику эффективности принятого решения.

***Решение.*** 1. Подставляя в решающее правило для критерия МП $\overline{X_n} \geq \frac{m_1 + m_0}{2} = l_0$ рассчитанные значения $\frac{1}{8}\sum_{i=1}^{8} X_i = \overline{X_8} = 1$ и $l_0 = 1{,}1$, получаем $\overline{X_n} < l_0$. Таким образом, принимаем решение, что справедлива гипотеза $H_0$, т. е. выборочным данным $X_1, X_2, \ldots, X_n$ соответствует МО $m_0 = 0{,}2$.

2. Т. к. приняли гипотезу $H_0$, то отвергли гипотезу $H_1$, т. е. можно рассчитать ошибку второго рода, определяющую вероятность «необнаружения» гипотезы $H_1 \sim m_1$:

$$\beta = \frac{1}{2}\left[1 - \Phi\left(\frac{m_1 - m_0}{2\sigma}\sqrt{n}\right)\right] = \frac{1}{2}\left[1 - \Phi(0{,}64)\right] = \frac{1}{2}\left[1 - 0{,}48\right] = 0{,}26.$$

Вывод: Вероятность ошибки достаточно большая, т. е. выборочным данным может соответствовать и гипотеза $H_1$. Для уточнения необходимо увеличить объём выборки *n*.

## 6.2. Проверка гипотезы о равенстве математических ожиданий двух независимых нормальных генеральных совокупностей

Рассмотрим две СВ $X \sim N\left(m_x, \sigma_x^2\right)$ и $Y \sim N\left(m_y, \sigma_y^2\right)$, причём будем полагать, что их дисперсии одинаковы $\sigma_x^2 = \sigma_y^2 = \sigma^2$, но $\sigma^2$ – в общем случае неизвестна. Пусть имеются две независимые выборки из данных генеральных совокупностей $X_1, X_2, \ldots, X_n$ объёма $n_1 > 1$ и $Y_1, Y_2, \ldots, Y_{n_2}$ объёма $n_2 \neq n_1$. Поставим задачу проверки гипотезы $H_0 : m_x = m_y$ против альтернативы $H_1 : m_x \neq m_y$, используя средние выборочные величины $\overline{X}_{n_1}$ и $\overline{Y}_{n_2}$.

Для решения задачи вводят дополнительно СВ $T = X - Y$. При этом $T \sim N\left(m_T, \sigma_T^2\right)$, где $m_T = m_x - m_y$, а $\sigma_T^2 = \sigma_x^2 + \sigma_y^2 = 2\sigma^2$.

Тогда задача проверки рассматриваемых гипотез может быть сформулирована так:

$$H_0 : m_T = 0;$$

$$H_1 : |m_T| > 0.$$

Её решение по аналогии с решением задачи проверки простых гипотез принимает следующий вид:

$$W: \left|\overline{T}_{n_1+n_2}\right| > t_{n_1+n_2-2,p},$$

где $\overline{T}_{n_1+n_2}$ – среднее выборочное значение СВ $T$ по выборке суммарного объёма $n_1 + n_2$:

$$\overline{T}_{n_1+n_2} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{(n_1-1)\hat{S}_x^2 + (n_2-1)\hat{S}_y^2}{n_1+n_2-2}}},$$

нормированная СВ, имеющая распределение Стьюдента с $k, n_1, n_2$ степенями свободы. При этом $t_{n_1+n_2-2,p}$ определяет $p$ % точку распределения Стьюдента, заданную таблицами распределения (приложение 2) для уровня значимости решения $\alpha = 1 - p$.

***Пример 43.*** Имеются две независимые выборки из нормальных генеральных совокупностей объёма $n_1 = 8$ и $n_2 = 16$:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $X_i$ | 2 | 0 | -4 | -6 | 0 | 4 | 6 | 2 |

,

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $Y_i$ | 4 | -2 | -8 | 0 | 2 | -4 | 2 | -2 |
| $i$ | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| $Y_i$ | 2 | 0 | -4 | -6 | 0 | 4 | 6 | 2 |

.

Проверить гипотезу $H_0 : m_x = m_y$ против альтернативы $H_1 : m_x \neq m_y$ при заданной вероятности ошибки первого рода $\alpha = 0{,}02$.

***Решение.*** Перейдём к новой СВ $T = X - Y$ и проверим гипотезы $H_0 : m_T = 0$ и $H_1 : |m_T| > 0$. Решающее правило, с учётом $p = 1 - \alpha = 0{,}98$ и $n_1 + n_2 = 24$ имеет следующий вид:

$$W : \left|\overline{T}_{n_1+n_2}\right| > t_{22,\ 0{,}98},$$

где $$\overline{T}_{n_1+n_2} = \frac{\overline{X}_8 - \overline{Y}_{16}}{\sqrt{\left(\frac{1}{8} + \frac{1}{16}\right)\frac{7\hat{S}_x^2 + 15\hat{S}_y^2}{22}}} \approx -0{,}5,$$

рассчитана по средним выборочным величинам $\overline{X_8}$ и $\overline{Y_{16}}$ и исправленным выборочным дисперсиям $\hat{S}_x^2 = \frac{1}{7}\sum_{i=1}^{8}\left(X_i - \overline{X}_8\right)^2 = 15{,}71$ и $\hat{S}_y^2 = \frac{1}{15}\sum_{i=1}^{16}\left(Y_i - \overline{Y}_{16}\right)^2 = 14{,}87$. По таблицам распределения Стьюдента для числа степеней свободы 22 определяем критическую точку $t_{22,\ 0{,}98} = 2{,}51$, и на основе решающего правила принимаем гипотезу о равенстве математических ожиданий. Другими словами, средние выборочные величины $\overline{X}_{n_1}$ и $\overline{Y}_{n_2}$ различаются незначительно.

## 6.3. Проверка гипотезы о равенстве дисперсий двух нормальных генеральных совокупностей

Пусть

$$X \sim N\left(m_x, \sigma_x^2\right);$$

$$Y \sim N\left(m_y, \sigma_y^2\right)$$

две нормальные генеральные совокупности, МО и дисперсии которых заранее неизвестны. Проверим гипотезу о равенстве дисперсий этих генеральных совокупностей в предположении, что $\sigma_x^2 \geq \sigma_y^2$.

Также имеются две выборки наблюдений:

$$X_1, X_2, ..., X_{n_1} \text{ объёма } n_1 < \infty;$$

$$Y_1, Y_2, ..., Y_{n_2} \text{ объёма } n_2 \neq n_1.$$

С учётом этого сформулируем задачу в терминах проверки двух статистических гипотез:

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad ;$$

$$H_1 : \sigma_x^2 > \sigma_y^2 \quad .$$

Решение такой задачи основывается на двух независимых оценках неизвестных дисперсий. Обозначим их как

$$\hat{S}_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n} \left( X_i - \overline{X}_n \right)^2;$$

$$\hat{S}_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n} \left( Y_i - \overline{Y}_n \right)^2$$

и после этого воспользуемся решающим правилом вида

$$W : \frac{\hat{S}_x^2}{\hat{S}_y^2} > F_{\alpha, k_1, k_2} = \text{const},$$

где $W : \dfrac{\hat{S}_x^2}{\hat{S}_y^2} \equiv F$ случайная решающая статистика, подчинённая закону $F$-распределения Фишера с $k_1, k_2$ степенями свободы. При этом $k_1 = n_1 - 1$, $k_2 = n_2 - 1$, $F_{\alpha,k_1,k_2}$ – $\alpha\%$ или критическая точка $F$-распределения Фишера (приложение 4), заданная вероятностью ошибки первого рода: $P\left\{F > F_{\alpha,k_1,k_2}\right\} = \alpha$.

***Пример 44***. Для двух независимых выборок $X$ и $Y$ из нормальных генеральных совокупностей, каждая объёма $n = 8$:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $X_i$ | 4 | -2 | -8 | 0 | 2 | -4 | 2 | -2 |

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $Y_i$ | 2 | 0 | -4 | -6 | 0 | 4 | 6 | 2 |

,

1) оценить дисперсии $\hat{S}_x^2$ и $\hat{S}_y^2$ при неизвестных МО;

2) проверить гипотезу о равенстве дисперсий при заданной вероятности ошибки первого рода (уровня значимости) $\alpha = 0{,}05$.

***Решение.*** 1. Оценки неизвестных дисперсий

$\hat{S}_x^2 = \dfrac{1}{7}\sum_{i=1}^{8}\left(X_i - \overline{X}_8\right)^2 = 14{,}86$ и $\hat{S}_y^2 = \dfrac{1}{7}\sum_{i=1}^{8}\left(Y_i - \overline{Y}_8\right)^2 = 15{,}71$ получены с учётом значений выборочных средних величины $\overline{X}_8$ и $\overline{Y}_8$.

2. Найдём отношение большей исправленной оценки дисперсии к меньшей

$F = \dfrac{\hat{S}_y^2}{\hat{S}_x^2} = \dfrac{15{,}71}{14{,}86} = 1{,}057$. По таблице (приложение 4), по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_1 - 1 = 7$, $k_2 = n_2 - 1 = 7$ находим критическую точку $F$-распределения Фишера (приложение 4) $F_{\alpha,k_1,k_2} = 3{,}79$. Так как $F < F_{\alpha,k_1,k_2}$, то нет оснований отвергнуть гипотезу о равенстве дисперсий.

## Вопросы и задачи

1. Что называется оптимальным решающим правилом?
2. Привести примеры задач проверки простых и сложных гипотез.
3. Из нормальной генеральной совокупности извлечена выборка объёма $n$ = 16:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $X_i$ | 4 | -2 | -8 | 0 | 2 | -4 | 2 | -2 |
| $i$ | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| $X_i$ | 2 | 0 | -4 | -6 | 0 | 4 | 6 | 2 |

1) при заданном значении $\sigma^2 = 9$ принять решение по критерию максимального правдоподобия о значении математического ожидания $m_0 = 0{,}2$ или $m_1 = 2$. 2) дать количественную характеристику эффективности принятого решения.

4. Какое решение будет принято в примере 43, если вероятность ошибки первого рода $\alpha = 0{,}2$?
5. Что определяют ошибки первого и второго рода?
6. Имеются две независимые выборки из нормальных генеральных совокупностей равных объёма $n_1 = n_2 = 8$:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $X_i$ | 2 | 0 | -4 | -6 | 0 | 4 | 6 | 2 |

,

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $Y_i$ | 4 | -2 | -8 | 0 | 2 | -4 | 2 | -2 |

.

Проверить гипотезу $H_0: m_x = m_y$ против альтернативы $H_1: m_x \neq m_y$ при заданной вероятности ошибки первого рода $\alpha = 0{,}1$.

7. Пусть два множества некоторых объектов, обладающих количественным признаком, подвергнуты выборочному контролю. Значе-

ния количественного признака есть распределенные по нормальному закону случайные величины. Из первого множества сделана выборка объема $n_1 = 21$ и подсчитана исправленная выборочная дисперсия равная 0,75. Из второго множества сделана выборка объема $n_2 = 11$. Эта выборка дала значение исправленной выборочной дисперсии, равное 0,25. Для выбранного уровня значимости $\alpha$ = 0,05 проверить гипотезу о равенстве дисперсий (гипотеза о равенстве дисперсий должна быть отвергнута).

## Библиографический список


1. **Савченко В.В.** Теория вероятностей: Конспект лекций. Н. Новгород: НГТУ, 1997.

2. **Савченко В.В.** Теория вероятностей и математическая статистика. Конспект лекций. Н. Новгород: НГЛУ, 2005.

3. **Вуколов Э.А.** Основы статистического анализа: Практикум по статистическим методам и исследованию операций с использованием пакетов *STATISTICA* и *EXCEL*. 2-е изд., испр. и доп. М.: Форум; ИНФРА-М, 2014.

4. Высшая математика для экономического бакалавриата: Учебник и практикум: углубл. курс / **Н.Ш. Кремер** [и др.]; под ред. Н.Ш. Кремера. 4-е изд., перераб. и доп. М.: Юрайт, 2015.

5. **Агапов Г.И.** Задачник по теории вероятностей. М.: Высшая школа, 2012.

6. **Палий И.А.** Введение в теорию вероятностей: Учебное пособие. М., 2005.

7. **Вентцель Е.С.** Теория вероятностей. М.: Высшая школа, 2004.

8. **Гмурман В.Е.** Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 2008.

9. **Гусева Е.Н.** Теория вероятностей и математическая статистика: Учебное пособие. М.: Флинта, 2011 // Электронный ресурс Интернет: http://biblioclub.ru/index.php?page=book&id=83543.

# ПРИЛОЖЕНИЯ

## Приложение I

**Значение функции Лапласа** $\Phi(t) \stackrel{\Delta}{=} \frac{2}{\sqrt{2\pi}} \int_0^t \exp\left[-\frac{t^2}{2}\right] dt.$

| ***t*** | ***Ф(t)*** | ***t*** | ***Ф(t)*** |
|---|---|---|---|
| 0,0 | 0,000 | 2,1 | 0,9643 |
| 0,1 | 0,0797 | 2,2 | 0,9722 |
| 0,2 | 0,1585 | 2,3 | 0,9786 |
| 0,3 | 0,2358 | 2,4 | 0,9836 |
| 0,4 | 0,3108 | 2,5 | 0,9876 |
| 0,5 | 0,3839 | 2,6 | 0,9907 |
| 0,6 | 0,4515 | 2,7 | 0,9931 |
| 0,7 | 0,5161 | 2,8 | 0,9949 |
| 0,8 | 0,5763 | 2,9 | 0,9963 |
| 0,9 | 0,6319 | 3,0 | 0,9973 |
| 1,0 | 0,6827 | 3,1 | 0,9981 |
| 1,1 | 0,7287 | 3,2 | 0,9986 |
| 1,2 | 0,7699 | 3,3 | 0,9990 |
| 1,3 | 0,8064 | 3,4 | 0,9993 |
| 1,4 | 0,8385 | 3,5 | 0,9995 |
| 1,5 | 0,8664 | 3,6 | 0,9997 |
| 1,6 | 0,8904 | 3,7 | 0,9998 |
| 1,7 | 0,9109 | 3,8 | 0,9999 |
| 1,8 | 0,9281 | 3,9 | 0,9999 |
| 1,9 | 0,9426 | 4,0 | 0,999936 |
| 2,0 | 0,9545 | 4,5 | 0,999994 |

## Приложение II

Распределение Стьюдента $T_k$

Значение функции $t_{k,p}$ определяется равенством $P\{\ |T_k| < t_{k,p}\} = p$.

| $k \backslash p$ | 0.2 | 0.6 | 0.8 | 0.9 | 0.95 | 0.98 | 0.99 | 0.999 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.325 | 1.376 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 636.619 |
| 2 | 0.289 | 1.061 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 31.598 |
| 3 | 0.277 | 0.978 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 12.941 |
| 4 | 0.271 | 0.941 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |
| 5 | 0.267 | 0.920 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 6.859 |
| 6 | 0.265 | 0.906 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | 0.263 | 0.896 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 5.405 |
| 8 | 0.262 | 0.889 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | 0.261 | 0.883 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | 0.260 | 0.879 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | 0.260 | 0.876 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | 0.259 | 0.873 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | 0.259 | 0.870 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | 0.258 | 0.868 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | 0.258 | 0.866 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | 0.258 | 0.865 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | 0.257 | 0.863 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | 0.257 | 0.862 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | 0.257 | 0.861 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | 0.257 | 0.860 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.850 |
| 21 | 0.257 | 0.859 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.819 |
| 22 | 0.256 | 0.858 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.792 |
| 23 | 0.256 | 0.858 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.767 |
| 24 | 0.256 | 0.857 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.745 |
| 25 | 0.256 | 0.856 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.725 |
| 26 | 0.256 | 0.856 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.707 |
| 27 | 0.256 | 0.855 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.690 |
| 28 | 0.256 | 0.855 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.674 |
| 29 | 0.256 | 0.854 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.659 |
| 30 | 0.256 | 0.854 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.646 |
| 40 | 0.255 | 0.851 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.551 |
| 60 | 0.254 | 0.848 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.460 |
| 120 | 0.254 | 0.845 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.373 |
| $\infty$ | 0.253 | 0.842 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

## Приложение III

Распределение $\chi^2$

Значение функции $\chi^2_{k,\nu}$ определяется равенством $P\left(\chi^2_k > \chi^2_{k,\nu}\right) = \nu$ .

| $k$ \ $\nu$ | 0.95 | 0.90 | 0.80 | 0.70 | 0.50 | 0.30 | 0.20 | 0.10 | 0.05 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.02 | 0.06 | 0.15 | 0.46 | 1.07 | 1.64 | 2.71 | 3.84 |
| 2 | 0.10 | 0.21 | 0.45 | 0.71 | 1.39 | 2.41 | 3.22 | 4.60 | 5.99 |
| 3 | 0.35 | 0.58 | 1.01 | 1.42 | 2.37 | 3.66 | 4.64 | 6.25 | 7.82 |
| 4 | 0.71 | 1.06 | 1.65 | 2.20 | 3.36 | 4.88 | 5.99 | 7.78 | 9.49 |
| 5 | 1.15 | 1.61 | 2.34 | 3.00 | 4.35 | 6.06 | 7.29 | 9.24 | 11.07 |
| 6 | 1.65 | 2.20 | 3.07 | 3.83 | 5.35 | 7.23 | 8.56 | 10.64 | 12.59 |
| 7 | 2.17 | 2.83 | 3.82 | 4.67 | 6.35 | 8.38 | 9.80 | 12.02 | 14.07 |
| 8 | 2.73 | 3.49 | 4.59 | 5.53 | 7.34 | 9.52 | 11.03 | 13.36 | 15.51 |
| 9 | 3.32 | 4.17 | 5.38 | 6.39 | 8.34 | 10.66 | 12.24 | 14.68 | 16.92 |
| 10 | 3.94 | 4.86 | 6.18 | 7.27 | 9.34 | 11.78 | 13.44 | 15.99 | 18.31 |
| 11 | 4.58 | 5.58 | 6.99 | 8.15 | 10.34 | 12.90 | 14.63 | 17.28 | 19.68 |
| 12 | 5.23 | 6.30 | 7.81 | 9.03 | 11.34 | 14.01 | 15.81 | 18.55 | 21.03 |
| 13 | 5.89 | 7.04 | 8.63 | 9.93 | 12.34 | 15.12 | 16.98 | 19.81 | 22.36 |
| 14 | 6.57 | 7.79 | 9.47 | 10.82 | 13.34 | 16.22 | 18.15 | 21.06 | 23.69 |
| 15 | 7.26 | 8.55 | 10.31 | 11.72 | 14.34 | 17.32 | 19.31 | 22.31 | 25.00 |
| 16 | 7.96 | 9.31 | 11.15 | 12.62 | 15.34 | 18.42 | 20.47 | 23.54 | 26.29 |
| 17 | 8.67 | 10.08 | 12.00 | 13.53 | 16.34 | 19.51 | 21.62 | 24.78 | 27.60 |
| 18 | 9.39 | 10.86 | 12.86 | 14.44 | 17.34 | 20.60 | 22.76 | 25.59 | 28.87 |
| 19 | 10.11 | 11.65 | 13.72 | 15.35 | 18.34 | 21.70 | 23.90 | 27.20 | 30.14 |
| 20 | 10.85 | 12.44 | 14.58 | 16.27 | 19.34 | 22.80 | 25.04 | 28.41 | 31.41 |
| 21 | 11.59 | 13.24 | 15.44 | 17.18 | 20.30 | 23.90 | 26.17 | 29.61 | 32.67 |
| 22 | 12.34 | 14.04 | 16.31 | 18.10 | 21.30 | 24.90 | 27.30 | 30.81 | 33.92 |
| 23 | 13.09 | 14.85 | 17.19 | 19.02 | 22.30 | 26.00 | 28.43 | 32.01 | 35.17 |
| 24 | 13.85 | 15.66 | 18.06 | 19.94 | 23.30 | 27.10 | 29.55 | 33.20 | 36.42 |
| 25 | 14.61 | 16.47 | 18.94 | 20.90 | 24.30 | 28.20 | 30.78 | 34.38 | 37.65 |
| 26 | 15.38 | 17.29 | 19.82 | 21.80 | 25.30 | 29.20 | 31.80 | 35.56 | 38.89 |
| 27 | 16.15 | 18.11 | 20.70 | 22.70 | 26.30 | 30.30 | 32.91 | 36.74 | 40.11 |
| 28 | 16.93 | 18.94 | 21.60 | 23.60 | 27.30 | 31.40 | 34.03 | 37.92 | 41.34 |
| 29 | 17.71 | 19.77 | 22.50 | 24.60 | 28.30 | 32.50 | 35.14 | 39.09 | 42.56 |
| 30 | 18.49 | 20.60 | 23.40 | 25.50 | 29.30 | 33.50 | 36.25 | 40.26 | 43.77 |

## Приложение IV

Распределение Фишера

Значение критической точки $F_{\alpha,k_1,k_2}$, определяется вероятностью ошибки первого рода: $P\left\{F > F_{\alpha,k_1,k_2}\right\} = \alpha$.

$k_1$ – число степеней свободы большей дисперсии,

$k_2$ – число степеней свободы меньшей дисперсии,

$\alpha = 0{,}05$.

| $k_2$ \ $k_1$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 161,446 | 199,499 | 215,707 | 224,583 | 230,160 |
| 2 | 18,513 | 19,000 | 19,164 | 19,247 | 19,296 |
| 3 | 10,128 | 9,552 | 9,277 | 9,117 | 9,013 |
| 4 | 7,709 | 6,944 | 6,591 | 6,388 | 6,256 |
| 5 | 6,608 | 5,786 | 5,409 | 5,192 | 5,050 |
| 6 | 5,987 | 5,143 | 4,757 | 4,534 | 4,387 |
| 7 | 5,591 | 4,737 | 4,347 | 4,120 | 3,972 |
| 8 | 5,318 | 4,459 | 4,066 | 3,838 | 3,688 |
| 9 | 5,117 | 4,256 | 3,863 | 3,633 | 3,482 |
| 10 | 4,965 | 4,103 | 3,708 | 3,478 | 3,326 |
| 11 | 4,844 | 3,982 | 3,587 | 3,357 | 3,204 |
| 12 | 4,747 | 3,885 | 3,490 | 3,259 | 3,106 |
| 13 | 4,667 | 3,806 | 3,411 | 3,179 | 3,025 |
| 14 | 4,600 | 3,739 | 3,344 | 3,112 | 2,958 |
| 15 | 4,543 | 3,682 | 3,287 | 3,056 | 2,901 |
| 16 | 4,494 | 3,634 | 3,239 | 3,007 | 2,852 |
| 17 | 4,451 | 3,592 | 3,197 | 2,965 | 2,810 |
| 18 | 4,414 | 3,555 | 3,160 | 2,928 | 2,773 |
| 19 | 4,381 | 3,522 | 3,127 | 2,895 | 2,740 |
| 20 | 4,351 | 3,493 | 3,098 | 2,866 | 2,711 |
| 21 | 4,325 | 3,467 | 3,072 | 2,840 | 2,685 |
| 22 | 4,301 | 3,443 | 3,049 | 2,817 | 2,661 |
| 23 | 4,279 | 3,422 | 3,028 | 2,796 | 2,640 |
| 24 | 4,260 | 3,403 | 3,009 | 2,776 | 2,621 |
| 25 | 4,242 | 3,385 | 2,991 | 2,759 | 2,603 |
| 26 | 4,225 | 3,369 | 2,975 | 2,743 | 2,587 |
| 27 | 4,210 | 3,354 | 2,960 | 2,728 | 2,572 |
| 28 | 4,196 | 3,340 | 2,947 | 2,714 | 2,558 |
| 29 | 4,183 | 3,328 | 2,934 | 2,701 | 2,545 |
| 30 | 4,171 | 3,316 | 2,922 | 2,690 | 2,534 |
| 34 | 4,130 | 3,276 | 2,883 | 2,650 | 2,494 |
| 40 | 4,085 | 3,232 | 2,839 | 2,606 | 2,449 |
| 50 | 4,034 | 3,183 | 2,790 | 2,557 | 2,400 |
| 100 | 3,936 | 3,087 | 2,696 | 2,463 | 2,305 |

продолжение таблицы

| | 6 | 7 | 12 | 24 | 50 |
|---|---|---|---|---|---|
| 1 | 233,988 | 236,767 | 243,905 | 249,052 | 251,774 |
| 2 | 19,329 | 19,353 | 19,412 | 19,454 | 19,476 |
| 3 | 8,941 | 8,887 | 8,745 | 8,638 | 8,581 |
| 4 | 6,163 | 6,094 | 5,912 | 5,774 | 5,699 |
| 5 | 4,950 | 4,876 | 4,678 | 4,527 | 4,444 |
| 6 | 4,284 | 4,207 | 4,000 | 3,841 | 3,754 |
| 7 | 3,866 | 3,787 | 3,575 | 3,410 | 3,319 |
| 8 | 3,581 | 3,500 | 3,284 | 3,115 | 3,020 |
| 9 | 3,374 | 3,293 | 3,073 | 2,900 | 2,803 |
| 10 | 3,217 | 3,135 | 2,913 | 2,737 | 2,637 |
| 11 | 3,095 | 3,012 | 2,788 | 2,609 | 2,507 |
| 12 | 2,996 | 2,913 | 2,687 | 2,505 | 2,401 |
| 13 | 2,915 | 2,832 | 2,604 | 2,420 | 2,314 |
| 14 | 2,848 | 2,764 | 2,534 | 2,349 | 2,241 |
| 15 | 2,790 | 2,707 | 2,475 | 2,288 | 2,178 |
| 16 | 2,741 | 2,657 | 2,425 | 2,235 | 2,124 |
| 17 | 2,699 | 2,614 | 2,381 | 2,190 | 2,077 |
| 18 | 2,661 | 2,577 | 2,342 | 2,150 | 2,035 |
| 19 | 2,628 | 2,544 | 2,308 | 2,114 | 1,999 |
| 20 | 2,599 | 2,514 | 2,278 | 2,082 | 1,966 |
| 21 | 2,573 | 2,488 | 2,250 | 2,054 | 1,936 |
| 22 | 2,549 | 2,464 | 2,226 | 2,028 | 1,909 |
| 23 | 2,528 | 2,442 | 2,204 | 2,005 | 1,885 |
| 24 | 2,508 | 2,423 | 2,183 | 1,984 | 1,863 |
| 25 | 2,490 | 2,405 | 2,165 | 1,964 | 1,842 |
| 26 | 2,474 | 2,388 | 2,148 | 1,946 | 1,823 |
| 27 | 2,459 | 2,373 | 2,132 | 1,930 | 1,806 |
| 28 | 2,445 | 2,359 | 2,118 | 1,915 | 1,790 |
| 29 | 2,432 | 2,346 | 2,104 | 1,901 | 1,775 |
| 30 | 2,421 | 2,334 | 2,092 | 1,887 | 1,761 |
| 34 | 2,380 | 2,294 | 2,050 | 1,843 | 1,713 |
| 40 | 2,336 | 2,249 | 2,003 | 1,793 | 1,660 |
| 50 | 2,286 | 2,199 | 1,952 | 1,737 | 1,599 |
| 100 | 2,191 | 2,103 | 1,850 | 1,627 | 1,477 |

## Содержание

**Дмитрий Юрьевич Акатьев**

**Нана Джимшеровна Чикова**

# СБОРНИК ЗАДАЧ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Редакторы: А.О. Кузнецова
Д.В. Носикова
А.С. Паршаков

Лицензия ПД № 18-0062 от 20.12.2000

---

| | | | |
|---|---|---|---|
| Подписано к печати | | | Формат 60 x 90 1/16 |
| Печ. л. | Тираж | экз. | Заказ |
| Цена договорная | | | |

---

Типография ФГБОУ ВПО «НГЛУ»
603155, Н. Новгород, ул. Минина, 31а