**ФГБ ОУ ВПО**
**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**
**ПУТЕЙ СООБЩЕНИЯ»**

**Кафедра «Математика»**

**Л.Ф. Кочнева, З.С. Липкина, В. И. Новосельцева**

# Теория вероятностей
# и математическая статистика

**МОСКВА - 2014**

# Оглавление

## 1. Основные понятия теории вероятностей

**1.** Опыт (*испытание, эксперимент*) – осуществление на практике определённого комплекса условий, который обозначим ***S***.

При этом рассматривают такие опыты, которые можно повторять теоретически многократно. Например, при бросании монеты условия ***S***: монета правильной формы, без дефектов и не может падать на ребро.

**2**. Событие, случайное событие – любой исход опыта, который может произойти или не произойти. События обозначаются заглавными буквами латинского алфавита: ***A***, ***B***, ***C***, …..

Например, при бросании игрального кубика условия ***S*** – кубик правильной формы, без дефектов из однородного материала. Можно рассматривать следующие события: ***A*** - выпадение 3 очков; ***B*** – выпадение числа очков не меньше 4; ***C*** – выпадение чётного числа очков; ***D*** – выпадение целого числа очков; ***E*** – выпадение 10 очков, и т.д.

**3.** Исходы, элементарные события рассматриваются как взаимоисключающие исходы опыта.

При бросании монеты – два исхода, при бросании игрального кубика 6 исходов.

Событие называется достоверным, если в результате опыта происходит всегда.

Событие называется невозможным если в результате рассматриваемого опыта никогда не произойдет, будем обозначать ***Ø***. Так при бросании кубика событие ***D*** – выпадение целого числа очков – достоверное; событие ***E*** – выпадение 10 очков – невозможное.

**5.** Два события называются несовместными, если появление одного из них исключает появление другого в одном и том же опыте. ***В*** противном случае события называются совместными. При бросании кубика события ***A*** – «выпадение 3 очков» и ***C*** - «выпадение чётного числа очков» события несовместные; события ***A*** и ***D*** - совместные.

**6.** События $A_1$ , $A_2$ , ....., $A_n$ называются попарно-несовместными, если любые два из них в результате опыта несовместны.

Попарно несовместные события образуют полную группу если в результате опыта происходит только одно из них.

В опыте «бросание кубика» исходы «1», «2»…»6»,образуют полную группу событий, так как попарно несовместны и в результате каждого опыта происходит только одно из них.

**7.** События называются равновозможными, если есть основания считать, что ни одно из них не является более возможным, чем другие в данном опыте.

При бросании монеты появление герба или решки – равновозможные события. При бросании кубика элементарные события «1»,…»6» равновозможны.

## 2. Алгебра событий

Пусть имеются события ***A, B.***

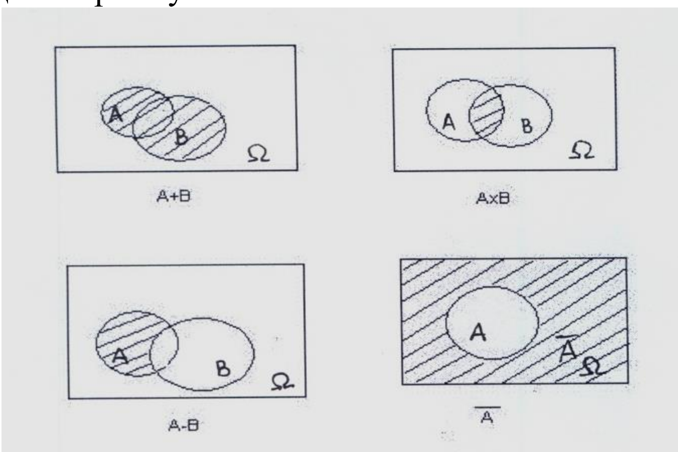
**1.** Суммой событий ***A*** и ***B*** называется такое событие ***C***, которое состоит в том, что происходит или ***A*** или ***B***, или оба вместе (хотя бы одно из них происходит). Обозначается $A + B = C$.

**2.** Произведением событий ***A*** и ***B*** называется такое событие ***C***, которое состоит в том, что происходит и ***A*** и ***B*** одновременно. Обозначается $C = A \cdot B$.

**3.** Разностью событий ***A*** и ***B*** называется такое событие ***C*,** которое состоит в том, что происходит событие ***A*** и не происходит событие ***B*.**
Обозначается $C = A - B$.

**4.** Отрицанием события ***A*** называется событие $\overline{A}$ - «не ***A***» (то есть $\overline{A}$ - событие противоположное событию ***A***).

Действия над событиями можно наглядно изобразить с помощью диаграмм Эйлера-Венна. Достоверное событие событие - множество исходов изображается прямоугольником; элементарные события – точки прямоугольника; случайные события ***A*, *B*** – области, входящие в прямоугольник:



## 3. Классическое определение вероятности

Классическое определение вероятности события основано на равновозможности конечного числа исходов опыта. Если проводится опыт с $n$ исходами, то их можно представить в виде полной группы несовместных равновозможных событий. Такие исходы называют элементарными событиями (случаями).

Случай, который приводит к наступлению события $A$, называется благоприятным (или благоприятствующим) событию $A$.

**Определение.** Вероятностью события $A$ называется отношение числа $m$ случаев, благоприятствующих событию $A$, к общему числу $n$ случаев:

$$P(A) = \frac{m}{n}, \quad 0 \le m \le n.$$

***Из определения следуют свойства:***

**1.** Вероятность любого события заключена между нулём и единицей:

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

**2.** Вероятность невозможного события равна нулю:

$$P(\varnothing) = 0.$$

**3.** Вероятность достоверного события равна единице:

$$P(\Omega) = 1.$$

**4.** Вероятность суммы несовместных событий равна сумме вероятностей:

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

**Пример 1.** В урне (в ящике) 5 шаров; 2 белых и 3 черных. Наугад вынимаем 1 шар. Какова вероятность того, что он черный – событие $\boldsymbol{A}$?
Решение: Элементарное событие – вынимание шара. Таких событий 5. Все они равновозможны. Из $\boldsymbol{n}$ = 5 случаев событию $\boldsymbol{A}$ благоприятствуют 3.
Следовательно, $\boldsymbol{P(A)}$ = 3/10.

**Пример 2.** Бросаем две монеты. Найти вероятность события $\boldsymbol{A}$ появление двух орлов} Здесь четыре элементарных исхода: (**О,О**)={два орла}, (**О,Р**)={на первой монете орел, на второй решка}, (**Р,Р**)={обе решки}, (**Р,О**)={на первой решка, на второй орел}. Эти исходы равновозможны в силу симметрии монеты; $\boldsymbol{n}$ = 4.
Появлению двух орлов благоприятствует только один исход, $\boldsymbol{m_A}$=1, Следовательно $\boldsymbol{P(A)}$ = 1/4.

**Пример 3.** В урне 5 шаров: 2 белых и 3 черных. Наугад вынимаем два шара (одновременно либо последовательно). Найти вероятность того, что они оба белые: $\boldsymbol{B}$ = {оба шара белые} и что оба черные: $\boldsymbol{C}$ = {оба черные}.
Решение: С примером 1 здесь общим является урна и шары, но опыт иной – вынимаем пару шаров, а не один. Занумеруем шары 1,2 –белые, 3,4,5 – черные. Элементарные события – разные пары (1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5) – всего их $\boldsymbol{n}$ = 10. Все пары равновозможны, а благоприятствует событию В лишь одна: (1,2) т.е. $\boldsymbol{P(B)}$ = 1/10. (Один шанс из десяти).
$\boldsymbol{m_c}$ = 3 (событие **С** происходит в трех случаях: (3,4), (3,5), (4,5)). Следовательно, $\boldsymbol{P(C)}$ = 3/10.

## 4. Статистическое определение вероятности

Статистический подход состоит в том, что вероятность приближенно заменяется относительной частотой события, т.е. отношением числа опытов, когда событие произошло к числу всех опытов. Если опыты производятся в одинаковых условиях и их число достаточно велико, то относительная частота обнаруживает свойство устойчивости, меняясь мало и колеблясь около некоторого постоянного числа. Из теоремы, доказанной Я.Бернулли (1713г.) следует, что относительная частота стремится к вероятности события при бесконечном увеличении числа опытов.

Относительную частоту события для большого числа опытов называют *статистической вероятностью* события.

Допустим, что в проведенных $\boldsymbol{n}$ опытах событие $\boldsymbol{A}$ зафиксировано $\boldsymbol{m}$ раз. Число $\boldsymbol{m}$ называется частотой события $\boldsymbol{A}$, а отношение

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

называется относительной частотой события $A$ в проведенных опытах.

***Относительная частота обладает следующими свойствами:***

**1.** $O \leq P(A) \leq 1$ , $O \leq m \leq n$ .

**2.** При $\boldsymbol{m} = 0$, событие $\boldsymbol{A} = \varnothing$ – невозможное, в данных опытах не зафиксировано ни разу, то есть $\boldsymbol{P(\varnothing)} = 0$.

**3.** Если в данных ***n*** опытах событие ***A*** наступило ***n*** раз, то ***A*** = $\Omega$ **-** достоверное событие и $\boldsymbol{P(\Omega)} = 1$.

**4.** Относительная частота суммы двух несовместных событий в данных опытах равна сумме относительных частот, то есть если $\boldsymbol{A \cdot B = \emptyset}$, то

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

**5.** Относительная частота обладает свойством статистической устойчивости: с увеличением числа опытов ***n*** она принимает значения, близкие к некоторому постоянному числу, которое и называют статистической вероятностью.

**<u>Пример:</u>** Некоторые ученые провели испытания с реальными монетами для нахождения частоты появления орла при ***n*** подбрасываниях монеты.

Вот лишь некоторые результаты:

| исследователь | n | m/n | из таблицы видно, что относительная частота колеблясь приближается к числу $\frac{1}{2}$. |
|---|---|---|---|
| Бюффон | 4040 | 0,507 | |
| Романовский | 80640 | 0,4923 | |
| Пирсон К | 24000 | 0,5005 | |
| Феллер | 10000 | 0,4979 | |

## 5. Основные формулы комбинаторики

***Комбинаторика*** – это раздел математики, который решает задачи о вычислении количества комбинаций различного вида из конечного числа элементов. Формулы комбинаторики используются при непосредственном вычислении вероятностей событий.

Рассмотрим наиболее часто встречающиеся виды комбинаций из **n** элементов: перестановки, размещения и сочетания.

**1.** Перестановки – это комбинации из **n** элементов, которые отличаются друг от друга только порядком расположения. Число всех возможных перестановок находится по формуле:
$P_n = n!$ – читается эн факториал, где $n! = 1\cdot2\cdot3\cdot\ldots\cdot(n-1)\cdot n$.

Очевидно, что $1! = 1$, $2!=2$, $3! = 6$, $4! = 24$, …….
Кроме того, по определению принимается $0! = 1$.

**Пример:** Сколькими способами можно поставить в очередь 5 человек?

Решение: Каждая реализация очереди из 5 человек характеризуется только порядком.
Поэтому количество комбинаций находится по формуле перестановок, т.е.

$$P_5 = 5! = 1\cdot2\cdot3\cdot4\cdot5 = 120.$$

**2.** Размещениями называются комбинации из ***n*** элементов по ***m*** элементов
$(0 < m \le n)$, которые отличаются друг от друга составом элементов или их порядком. Число размещений из ***n*** элементов по ***m*** элементов обозначается $A_n^m$ и вычисляется по формуле: $A_n^m = n(n-1)(n-2)\ldots\ldots[n-(m-1)]$.

Заметим, что, используя факториалы, число размещений из $n$ элементов по $m$ элементов можно вычислить по формуле:

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

**Пример:** Сколько двузначных чисел можно составить из трех цифр $\{1,2,3\}$?

Решение: Очевидно, что числа: 12; 13; 21; 31; 23; 32. Комбинации цифр отличаются элементами или порядком, то есть это размещения из трех элементов по 2 элемента и можно вычислить по формуле размещений:

$$A_3^2 = 3 \cdot 2 = 6.$$

**3. Сочетания** – это комбинации из $n$ элементов по $m$, которые отличаются друг от друга хотя бы одним элементом. Число сочетаний из $n$ элементов по $m$ элементов вычисляется по формуле:

$$C_n^m = \frac{A_n^m}{P_m},$$

то есть из всех размещений $A_n^m$, удаляются комбинации, которые отличаются одно от другого только порядком.

На практике число сочетаний из $n$ элементов по $m$ можно вычислять через факториалы:

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

**Пример:** Сколькими способами можно назначить трех дежурных из группы студентов из 10 человек?

*Решение:* Группы дежурных из 3-х человек должны отличаться друг от друга хотя бы одним человеком. Поэтому количество вариантов дежурных вычисляется как число сочетаний из 10 по 3: $C_{10}^{3} = \dfrac{10!}{3!(10-3)!} = 120$ .

**4.** Размещения с повторениями – это комбинации из ***n*** элементов по ***m*** элементов, которые отличаются друг от друга элементами, их порядком и количеством повторений элементов. Число всех таких размещений обозначается $\overline{A}_n^m$ и вычисляется по формуле:

$$\overline{A}_n^m = n^m .$$

**Пример:** Составить все размещения по 2 из трех элементов с повторениями элементов ***a, b, c*.** Число таких размещений по формуле равно $\overline{A}_3^2 = 3^2 = 9$.

И перечислим их для наглядности:

*(a,a), (b,b), (c,c),*
*(a,b), (a,c), (b,c),*
*(b,a), (c,a), (c,b).*

**Задача 1.** В урне жетоны с номерами 1, 2, …., 100. Опыт – выбор наугад жетона из урны. Какова вероятность того, что номер выбранного жетона:

а) не делится на 5 (событие *A*);
б) не содержит цифру 5 в своей записи (событие *B*)?

Решение: Элементарные события – выбор каждого номера и они – равновозможны. Число их ***n*** = 100. Из них 20 случаев (номера 5, 10, 15, …., 100) не благоприятствуют событию *A*, → 80 случаев благоприятствуют событию ***A***.

$$P(A) = m_A/n = 80/100 = 0{,}80$$

Цифра 5 содержится в 19 числах из данных 100 (5, 15, ….., 50, 51….., 59, 65 …., ….., 95) и не содержится в $\boldsymbol{m_в}$ = 81 числе.

$$P(B) = m_в / n = 81/100 = 0{,}81$$

**Задача 2.** Из колоды в 36 карт наугад берем 4 карты. Какова вероятность того, что вынутые карты будут:

а) разной масти (событие ***A***);

б) одной масти (событие ***B***).

Решение: Любая четверка карт – это элементарное событие. Все они равновозможные и всего таких случае $\boldsymbol{n}$ = $C_{36}^{4}$ число сочетаний из 36 по 4 – столькими способами можно взять 4 карты из 36.

а) первая карта – 9 способов;
вторая карта – 9 способов;
третья карта – 9 способов;
четвертая карта – 9 способов.

Итого: $P = (9 \cdot 9 \cdot 9 \cdot 9) / C_{36}^{4} = \dfrac{6561}{58905} = 0{,}11$

б) Выбор четырех карт одной масти $C_9^4$

Выбор одной масти из четырех: $C_4' = 4$.

$P = (C_4' \cdot C_9^4) / C_{36}^4 = \dfrac{504}{58905} = 0{,}0085$

## 6. Теоремы сложения и умножения

Если события ***A*, *B*** несовместны, то вероятность их суммы равна сумме вероятностей:

$$P\,(A+B) = P(A) + P(B).$$

**Пример:** В урне имеется 10 шаров, из них 3 красных шара 5 синих и 2 белых. Какова вероятность вынуть цветной шар (или красный, или синий)?

*Решение.* Пусть событие $A$ – «вынуть красный шар», событие ***B*** – «вынуть синий шар», событие ***C*** – «вынуть цветной шар». По определению ***C*=*A*+*B***.

$$P\,(C) = \frac{3}{10} + \frac{5}{10} = \frac{8}{10} = 0{,}8.$$

Введем понятие условных вероятностей. Пусть рассматриваются события ***A*** и ***B*** в данном опыте, и событие $\boldsymbol{C = A \cdot B}$.

Вероятность события *A* при условии, что событие *B* произошло, будем называть условной вероятностью и обозначать ***P(A/B)***.

Аналогично вероятность события В при условии, что ***A*** произошло будем называть условной вероятность и обозначать ***P(B/A).***

**Пример:** В примере с шарами из §6 пусть случайным образом вынимается подряд 2 шара. Обозначим событие ***A*** – «вынимается красный шар», событие ***B*** – «вынимается синий шар».

Вероятность вынуть 1-й шар красный ***P(A)*** = $\frac{3}{10}$, вероятность вынуть 1-й шар синий ***P(B)*** = $\frac{5}{10}$. Пусть теперь извлекается второй шар. Обозначим событие ***C*** – «вынуть 2-ой шар красный», событие ***D*** – «вынуть 2-ой шар синий». Очевидно, что вероятности событий ***C, D*** зависят от того, какое событие произошло первым ***A*** или ***B***.

Действительно, если 1-й шар вынули красный – то $P(C/A) = \frac{2}{9}$ . Если 1-ый шар вынули синий, то $P(C/B) = \frac{3}{9}$. Аналогично $P(D/A) = \frac{5}{9}$, $P(D/B) = \frac{4}{9}$. Из приведенных рассуждений следует, что вероятности событий ***C, D*** зависят от того, какое событие произошло ***A*** или ***B***.

Рассмотрим ещё один пример.

**Пример**. По мишени стреляют 2 стрелка по одному разу. Обозначим событие ***A*** – «1-ый стрелок попадает в цель», событие ***B*** – «2-ой стрелок попадает в цель». Событие ***C*** – «и ***A,*** и ***B***», то есть оба попадают в цель. ***C = A · B.***
Вероятность попадания в цель 1-го стрелка ***P(A)*** не зависит от того, попал ли в цель 2-ой стрелок и наоборот. То есть эти события независимые.

Вероятность совместного появления двух независимых событий равна произведению их вероятностей:

$$P(A \cdot B) = P(A) \cdot P(B).$$

Если ***P(A)*** = 0,7; ***P(B)*** = 0,8, то вероятность того, что оба попадут в цель ***P(A · B)*** = 0,7 · 0,8 = 0,56.

В примере с шарами вероятность того, что 1-ый вынутый шар красный, а 2-ой синий находится по формуле:

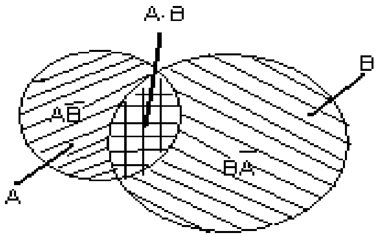
$$P(A \cdot D) = P\ (A) \cdot P(D/A) = \frac{3}{10} \cdot \frac{5}{9}$$

Таким образом, для зависимых событий ***A***, ***B*** вероятность произведения находится по формуле:

$$P(A \cdot B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B).$$

Если события ***A*** и ***B*** совместные, то вероятность суммы событий находится по формуле:

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




Поскольку исходы, входящие в пересечение ***AB*** учитываются 2 раза.

**Пример.** Вероятность попадания в цель 1-ым стрелком равна 0,6, а 2-ым стрелком – 0,8. Какова вероятность того, что цель будет поражена при одновременном выстреле двух стрелков?

Решение: Обозначим событие ***A*** – «попадание в цель 1-го стрелка»; событие ***B*** – «попадание в цель 2-го стрелка». Событие ***C*** – «цель поражена», что означает или ***A***, или ***B***, или «и ***A*** и ***B***», то есть ***C = A+B*** сумма совместных событий. Значит

$$P(C) = P(A) + P(B) - P(A \cdot B) = 0,6 + 0,8 - 0,6 \cdot 0,8 = 0,92.$$

Пусть имеем $A_1$, $A_2$, ……., $A_n$ - независимые события, вероятности которых: $P(A_1) = P_1$, $P(A_2) = P_2$, …..$P(A_n) = P_n$.

Требуется определить вероятность «появления хотя бы одного из них» - событие C.

Рассмотрим событие $\overline{C}$ - «ни одно событие не произойдет», то есть:

$$\overline{C} = \overline{A}_1 \cdot \overline{A}_2 \cdot \dots \cdot \overline{A}_n .$$

Вероятность $P(\overline{C}) = P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot \dots \cdot P(\overline{A}_n).$

Используя формулу для вероятности противоположного события, получим:

$$P(C) = 1 - P(\overline{C}) = 1 - P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot \dots \cdot P(\overline{A}_n).$$

Для краткости: $P(\overline{A}_1) = 1 - p_1 = q_1$, $P(\overline{A}_2) = 1 - p_2 = q_2, \dots P(\overline{A}_n) = 1 - p_n = q_n$.

Тогда формула для вычисления вероятности появления хотя бы одного из n независимых событий примет вид:

$$P(C) = 1 - q_1 \cdot q_2 \cdot \dots \cdot q_n . \qquad (9.1)$$

**Пример 1.** Три стрелка по разу стреляют в мишень. Для первого стрелка вероятность попадания в мишень при выстреле равна $\boldsymbol{P_1} = 0{,}4$; для второго $\boldsymbol{P_2} = 0{,}5$; для третьего

$\boldsymbol{P_3}$ = 0,7. Найти вероятность а) трех попаданий (событие ***A***); б) хотя бы одного попадания.

Решение. Пусть $\boldsymbol{S_1}$ – попадание первым, $\boldsymbol{S_2}$ – попадание вторым, $\boldsymbol{S_3}$ – попадание третьим стрелком, $\boldsymbol{S_1 S_2}$ и $\boldsymbol{S_3}$ независимы.

Тогда

а) $P(A) = P\{S_1 S_2 S_3\} = P(S_1)\,P(S_2)\,P(S_3) = P_1 P_2 P_3 = 0{,}14$

б) по формуле (9.1)обозначив C – “хотя бы одно попадание” получим:

$$P(C) = 1 - P(\overline{S_1})\,P(\overline{S_2})\,P(\overline{S_3}) = 1 - (1-P_1)(1-P_2)(1-P_3)$$

$$= 1 - 0{,}6 \cdot 0{,}5 \cdot 0{,}3 = 0{,}91$$

**Пример 2.** Известны вероятности событий ***P(A)*** = 0,7; ***P(B)*** = 0,5. Можно ли сказать, что:

а) *A* и *B* независимы?

б) *A* и *B* несовместны?

Решение.

а) У нас нет информации о ***P(AB)***, а значит, и о независимости.

б) ***A*** и ***B*** совместны. Будь ***A*** и ***B*** несовместны, мы бы имели

$P(A+B) = P(A) + P(B) = 0{,}7 + 0{,}5 > 1$ - противоречие.

## 7. Формула полной вероятности

Пусть событие ***B*** в опыте может произойти с одним из ***n*** несовместных событий

$A_1, A_2, .., A_n$ , составляющих полную группу событий, то есть

Сумма ( $A_1 + A_2 + ... + A_n$)– достоверное событие, причем сумма вероятностей:

$$P(A_1) + P(A_2) + \ ... + P(A_n) = \sum_{i=1}^{n} P(A_i) = 1.$$

Тогда, очевидно, согласно свойствам операций над событиями,

*По теореме сложения вероятностей несовместных событий получаем:*

$$P(B) = P(BA_1) + P(BA_2) + ...... + P(BA_n) = \sum_{i=1}^{n} P(BA_i).$$

*По теореме умножения:*

$$P(B \cdot A_i) = P(A_i)\, P(B/A_i).$$

Тогда, $\boldsymbol{P(B) =} \sum_{i=1}^{n} P(A_i) P(B/A_i)$

– это и есть формула полной вероятности для события ***B***, которое может произойти с одним из событий $A_i$ , называемых гипотезами.

**Пример.** Электрические лампочки производятся на двух заводах, причем первый из них поставляет 70%, а второй 30% всей потребляемой продукции. Из каждых 100 ламп первого завода в среднем 83 стандартных, а из 100 ламп второго - лишь 63 стандартных. Найти вероятность того, что взятая наугад из всей продукции двух заводов лампочка будет стандартной,

Решение. Обозначим буквой $\boldsymbol{B}$ событие – «наугад взятая лампа стандартна». Мы не знаем, на каком заводе она изготовлена, поэтому выдвинем две гипотезы:

$\boldsymbol{A_1}$ = {лампа изготовлена первым заводом},

$\boldsymbol{A_2}$ = {лампа изготовлена вторым заводом}.

Заметим, что по условиям задачи других вариантов нет, то есть $\boldsymbol{A_1+A_2}$ - достоверное событие, $\boldsymbol{P(A_1)}=0{,}70$; $\boldsymbol{P(A_2)}=0{,}30$. $\boldsymbol{P(B|A_1)}=0{,}83$ - условная вероятность того, что лампа стандартна при условии, что она изготовлена первым заводом, т.е. доля стандартных ламп в продукции первого завода.

$P(B|A_2) = 0{,}63$ - доля стандартных ламп второго завода.

$P(A_1B) = P(A_1)P(B|A_1) = 0{,}7 \cdot 0{,}83 = 0{,}58$ - доля ламп на складе, изготовленных заводом 1 и притом стандартных.

$P(A_2B) = P(A_2)P(B|A_2) = 0{,}3 \cdot 0{,}63 = 0{,}19$ - вероятность того, что лампа на складе изготовлена 2 заводом и притом стандартна.

Тогда получим $P(B) = 0{,}58+0{,}19 = 0{,}77$.

## 8. Переоценка гипотез (Формула Байеса)

Пусть выполняются все условия, при которых применяется формула полной вероятности:

$\sum_{i=1}^{n} P(A_i) = 1$, событие ***B*** появляется с одной из гипотез ***$A_i$*** ***(i = 1, 2, …, n)*** и полная вероятность ***P(B)*** находиться по формуле $P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$.

Пусть теперь событие ***B*** произошло и требуется переоценить вероятности гипотез ***$A_i$*** по результатам проведенного опыта, то есть нужно найти условные вероятности ***$P(A_i/B)$.***

Для этого используется формула вероятности произведения событий:

$$P(B \cdot A_i) = P(B) \cdot P(A_i/B) = P(A_i) \cdot P(B/A_i).$$

Из равенства следует:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}, \qquad (i = 1, 2, \dots n).$$

*Это и есть* ***формула Байеса****, позволяющая определить вероятность гипотезы при условии, что событие* ***B*** *произошло.*

**Пример.** Предположим, что 96% изготовленных цехом деталей удовлетворяют стандарту. Изготовленные детали проходят контроль, допускающий ошибки. Он признает пригодной:

а) стандартную деталь с вероятностью 0,98;

б) нестандартную деталь с вероятностью 0,05.

Найти вероятность того, что:

а) деталь успешно прошла контроль и является стандартной (событие ***A***);

б) деталь успешно прошла контроль, значит, признана
стандартной (событие ***B***);

в) деталь является стандартной при условии, что она
успешно прошла контроль;

г) деталь является на деле стандартной при условии, что она забракована.

Решение. Все изготовленные детали перемешаны и «на глаз» не различимы.

Обозначим:

$A_1$ = { На контроль попадает стандартная деталь}; $P(A_1) = 0{,}96$

$A_2 = \overline{A_1}$ = {контролю подвергается нестандартная деталь}; $P(A_2) = 0{,}04$;

$B$ = { попадание детали в ящик №1}; $P(B|A_1)$ = О,98 - доля стандартных деталей, преодолевающих контроль, т.е. попадающих в ящик №1; $P(B|A_2) = 0{,}05$.

а) по формуле умножения вероятностей

$$P(A) = P(B \cdot A_1) = P(A_1) \cdot P(B|A_1) = 0{,}96 \cdot 0{,}98 = 0{,}9408$$

- на тысячу изготовленных деталей в ящик №1 в среднем попадает 940,8 стандартных деталей;

б) по формуле полной вероятности

$$P(B)=P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) = 0{,}9408 + 0{,}04 \cdot 0{,}05 = 0{,}9428$$

- на тысячу изготовленных деталей в ящик №1 попадает в среднем 942,8 деталей (помимо стандартных, немного нестандартных);

в) по формуле Байеса $P(A_1|B)=$

$$\frac{P(A_1B)}{P(B)}=\frac{0{,}9408}{0{,}9428}=0{,}998$$

- доля стандартных деталей в ящике №1, т.е. в среднем из 1000 деталей в ящике №1 998 на деле стандартные;

г) $P(\overline{B}) = 1 - P(B) = 0{,}057$ - из 1000 изготовленных деталей в ящик №2 в среднем попадает 57.

Находим далее

$$P(A_1/\overline{B}) = \frac{P(A_1)P(\overline{B}/A_1)}{P(\overline{B})} = \frac{0{,}96\cdot(1-0{,}98)}{0{,}057} = 0{,}337$$

– треть деталей в ящике №2 стандартные.

***Обратим внимание:***

$\boldsymbol{P(A_1)}$, $\boldsymbol{P(B)}$ - доля от всех произведенных деталей;

$\boldsymbol{P(A_1|B)}$ - доля стандартных деталей не от всех изготовленных деталей, а от попавших в ящик №1;

$\boldsymbol{P(A_1|\overline{B})}$ - доля стандартных деталей от всех деталей, попавших в ящик №2.

## 9. Схема Бернулли

Пусть проводится серия $n$ независимых испытаний, причем вероятность появления события $\boldsymbol{A}$ в каждом испытании постоянна и равна $\boldsymbol{p}$. Тогда вероятность того, что событие появится $\boldsymbol{k}$ раз в $\boldsymbol{n}$ испытаниях можно найти по формуле Бернулли:

$$P_n(k) = C_n^k\, p^k\, (1-p)^{n-k},$$

**Задача.** Монеты бросают 5 раз. Найти вероятность того, что герб появится:

а) два раза;

б) не больше четырех раз.

*Решение*. Вероятность выпадения герба при каждом броске монеты постоянна и равна 0,5.

а) $n = 5$, $p = 0{,}5$, $k = 2$..

*по формуле Бернулли* $P(2) = (4 \cdot 5/1 \cdot 2) \cdot (0{,}5)^2 \cdot (1 - 0{,}5)^3 =$ 5/16.

б) «герб выпал не больше четырех раз» - событие, противоположное событию«герб выпал 5 раз».

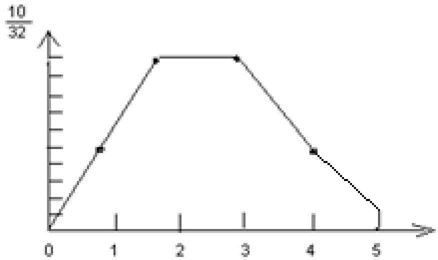
Следовательно, $P(k \leq 4) = 1 - P(5) = 1 - (0{,}5)^5 = 31/32$.

Правильная монета бросается 5 раз найти вероятности возможного числа появления гербов. Находим вероятности по формуле Бернулли. Поскольку $\boldsymbol{p} = \boldsymbol{q} = \frac{1}{2}$, то

$P_5(0) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$; $\qquad P_5(0) = \frac{1}{32}$ ;

$P_5(1) = C_5^1 \cdot \frac{1}{32} = \frac{5}{32} = P_5(4)$ ; $\qquad P_5(2) = C_5^2\, \frac{1}{32} = \frac{10}{32} = P_5(3)$ ·

Построим многоугольник или полигон распределения вероятностей, откладывая по оси $x$ значения $k$, а по оси $y$ $P_5(k)$ и соединяя эти точки.




Мы видим, что вероятности сначала возрастают, достигая максимума для $k_0 = 2$ и $k_0 = 3$, а затем убывают. Число $k_0$ наступления события $A$ в $n$ независимых испытаниях называется наивероятнейшим, если вероятность осуществления этого события $P_n(k_0)$ по крайней мере не меньше вероятностей других событий $P_n(k)$ при любом $k$.

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

## 9. Формула Пуассона

Решим формулами Бернулли такую задачу.

**Задача.** Завод выпускает в среднем 0,3% бракованных электролампочек. Найти вероятность того, что в партии из 1000 лампочек число бракованных будет:

а) равно три

в) не более трех.

Решение.

Здесь $n = 1000$, $p = 0{,}003$, $q = 0{,}997$.

а) $P_{1000}(3) = C^3_{1000}\ 0{,}003^3 \cdot 0{,}997^{997}$

в) $P_{1000}(0) + P_{1000}(1) + P_{1000}(2) + P_{1000}(3) = C^{0}_{1000}\ 0{,}003^0 \cdot 0{,}997^{1000} + C^{1}_{1000}\ 0{,}003 \cdot 0{,}997^{999} + C^{2}_{1000} \cdot 0{,}003^2 \cdot 0{,}997^{998} + C^{3}_{1000} \cdot 0{,}003^0 \cdot 0{,}997^{997}$.

Ясно, что непосредственное вычисление вероятностей технически сложно, поэтому желательно иметь более простые приближенные формулы. Такие формулы называются <u>*асимптотическими.*</u> Рассмотрим некоторые из них.

**<u>Теорема Пуассона.</u>** Если вероятность ***p*** наступления события ***A*** в каждом испытании стремится к нулю ***(p→0)***, причем произведение ***np*** стремится к постоянному числу ***λ*** при неограниченном увеличении числа ***n*** испытаний, то вероятность ***$P_n$ (k)*** того, что события ***A*** появится ровно ***k*** раз в ***n*** независимых испытаниях удовлетворяет предельному равенству

$$\lim_{n\to\infty} P_n(k) = P_k(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

(Строго говоря, условия теоремы Пуассона противоречат предпосылке схемы Бернулли, т.к. в каждом испытании ***p* = *const***. Однако, если вероятность ***p*** мала (< 0,01), ***n*** велико (***n*** > 100), а ***λ* = *np*** лежит в интервале (0,1; 10), то для ***k* << *n*/4** данная формула дает хорошие пр ибл иженные результаты)**.**.

Для приведенной выше задачи $\lambda = 0{,}003 \cdot 1000 = 3$

$P_3(3) = 0{,}2240$; $P_3(0) = 0{,}0498$; $P_3(1) = 0{,}1494$; $P_3(2) = 0{,}2240$.

Таким образом,

$P_{1000}(3) \approx 0{,}2240$; $P_{1000}(0 \le k \le 3) = 0{,}0498 + 0{,}1494 + 0{,}2240 \cdot 2 = 0{,}6472$

## 10. Локальная и интегральная теоремы Лапласа

Если вероятность $p$ наступления события $A$ в каждом испытании постоянна и отлична от 0 и 1, то вероятность $P_n(k)$ того, что события $A$ произойдет $k$ раз в $n$ независимых испытаниях при достаточно большом $n$ приближенно равна

$$P_n(k) = \frac{f(x)}{\sqrt{npq}}, \text{ где } f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{и} \quad x = \frac{k-np}{\sqrt{npq}}.$$

Погрешность тем меньше, чем больше $n$, причем должно быть выполнено условие $npq \geq 20$.

В приложениях даются таблицы значений функции $f(x)$.

Если требуется найти вероятность события $k_1 \leq k \leq k_2$, то удобнее пользоваться интегральной теоремой Муавра-Лапласа.

При тех же условиях, что и в локальной теореме Муавра-Лапласа

$$P_n(k_1 \leq k \leq k_2) = \Phi(x_2) - \Phi(x_1),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dx$, $x_1 = \frac{k_1 - np}{\sqrt{npq}}$, $x_2 = \frac{k_2 - np}{\sqrt{npq}}$

и $npq \geq 20$

Функция $\Phi(x)$ называется функцией Лапласа. Таблицы значений этой функции имеются в Приложении. Следует помнить, что

1. $\Phi(-x) = -\Phi(x)$; $\Phi(x)$ – нечетная функция и $\Phi(0) = 0$

2. $\Phi(x)$ монотонно возрастает, $\lim\limits_{x\to+\infty}\Phi(x)=0{,}5$; при этом при $x \geq 4$; $\Phi(x) \approx 0{,}5$

**Пример.** В среднем 20% пакетов акций на аукционе продаются по первоначально заявленной цене. Найти вероятность того, что из 200 пакетов акций в результате торгов по первоначально заявленной цене:

А) будет продано наивероятнейшее число пакетов ;

В) будет продано менее половины пакетов;

С) доля проданных пакетов будет меньше 25%;

Решение.

Для данной задачи $\boldsymbol{n} = 200$, $\boldsymbol{p} = 0{,}2$ $\boldsymbol{q} = 0{,}8$ $\boldsymbol{npq} = 0{,}2 \cdot 0{,}8 \cdot 200 = 32 > 20$

Значит, применимы формулы Муавра-Лапласа.

Найдем наивероятнейшее число проданных пакетов:

$$39{,}2 = 200 \cdot 0{,}2 - 0{,}8 \leq k_0 \leq 200 \cdot 0{,}2 + 0{,}2 = 40{,}2$$

$$k_0 = 40$$

$$P(A) = P_{200}(40) = \frac{f\left(\frac{40-40}{\sqrt{32}}\right)}{\sqrt{32}} = \frac{0{,}3989}{5{,}66} = 0{,}07 \quad \boldsymbol{x} = \frac{40-40}{\sqrt{32}} = 0$$

$$P(B) = P\,(0 \leq k \leq 100) = \Phi\left(\frac{100-40}{\sqrt{32}}\right) - \Phi\left(\frac{0-40}{\sqrt{32}}\right) = \Phi(10{,}6) - \Phi(-7{,}07) = 0{,}5 - (-0{,}5) = 1$$

т.е. практически менее половины пакетов будет продано по первоначально заявленной цене

С) $\alpha = 0 \leq \dfrac{k}{200} \leq 0{,}25 = \beta$,

$Z_1 = \dfrac{(0-0{,}2)\sqrt{200}}{\sqrt{0{,}16}} = -7{,}07;$

$Z_2 = \dfrac{(0{,}25-0{,}2)\sqrt{200}}{\sqrt{0{,}16}} = 1{,}77.$

$P(C) = \Phi(1{,}77) - \Phi(-7{,}07) = 0{,}4617 - (-0{,}5) = 0{,}9617$

Это означает, что в 96,17 % случаев доля проданных по первоначальной цене пакетов будет меньше 25 %

## Задачи для домашних заданий.

### *Вариант 1.*

**Задача 1.** Абонент забыл последнюю цифру номера телефона и набирает ее наудачу. Найти вероятность того, что ему придется сделать не более, чем две неудачные попытки (то есть, он наберет правильный номер или с первой попытки, или со второй, или с третьей).

**Задача 2.** Студент знает 45 из 60 экзаменационных вопросов. Билет содержит три вопроса. Найти вероятность того, что студент знает:

а) все три вопроса;

б) только один вопрос в экзаменационном билете.

**Задача 3.** Имеется 5 партий ламп; три партии по 8 штук в каждой из которых 6 стандартных и 2 нестандартных**,** и две партии по 10 штук, в каждой из которых 7 стандартных и 3 нестандартных. Из наугад взятой партии случайным образом выбрали одну деталь. Какова вероятность того, что она стандартная?

**Задача 4.** Имеем 2 команды первокурсников и 1 команду второкурсников. В каждой команде первокурсников 5 юношей и 3 девушки, в команде второкурсников - 6 юношей и 4 девушки. Выбрали одну из команд, а из нее одного участника:

а) какова вероятность того, что выбран юноша?

б) выбрали юношу, какова вероятность того, что он с первого курса?

**Задача 5.** Стрелок произвел 3 выстрела. Вероятность попадания в цель при одном выстреле 0,7. Найти вероятность поражения цели хотя бы одним выстрелом.

### *Вариант 2.*

**Задача 1.** Из корзины, в которой лежат 10 белых и 8 черных шаров, вынимают все шары. Найти вероятность того, что второй шар будет белый.

**Задача 2.** Семь из десяти посетителей кафе заказывают к кофе фирменное пирожное **.** Два пришедших посетителя заказывают кофе. Какова вероятность того, что они закажут:

а) два пирожных;

б) одно пирожное;

в) ни одного пирожного?

**Задача 3.** Если игрок выигрывает две партии подряд, то игра окончена. Вероятность выигрыша партии равна 0,6. Найти вероятность того, что игра окончена ранее четвертой?

**Задача 4.** Среди продукции завода брак составляет 2%. Найти вероятность того, что среди, 15 отобранных деталей не более 5 бракованных.

**Задача 5.** Брокер может приобрести акции одной из трех компаний А, В, С. Риск прогореть при покупке акций компании А составляет 50%, при покупке акций В – 40%, акций С – 20%. Брокер решает вложить все свои деньги в акции одной случайно выбранной компании. Какова вероятность того, что брокер прогорит?

## *Вариант 3.*

**Задача 1.** Из корзины в которой лежит **а** белых и **в** черных шаров, вынимают все шары. Найти вероятность того, что второй шар будет белый

**Задача 2.** Сигнализатор аварии содержит три независимо работающих устройства.

Вероятность того, что сработает первое устройство -0,9; второе -0,95; третье -0,85. Найти вероятность того, что при аварии сработает:

а) только одно устройство;

б) только два;

в) все три.

**Задача 3.** Если игрок выигрывает две партии подряд, то игра окончена. Вероятность выигрыша партии равна 0,5. Найти вероятность того, что игра окончена ранее шестой партии.

**Задача 4.** Среди продукции завода брак составляет 2%. Найти вероятность того, что среди 15 отобранных деталей не более 5 бракованных.

**Задача 5.** В школе 60% учащихся – девочки. 80% девочек и 70% мальчиков взяли билеты в театр. Принесли кем-то потерянный билет? Какова вероятность того, что билет потерял мальчик?

## 11. Ряд распределения вероятностей дискретной случайной величины.

Переменная величина, значение которой есть то или иное число, определяемое исходом случайного эксперимента, называется случайной величиной. Такие величины обозначаем X, Y, Z и т.д. и рассмотрим два основные их вида: 1) дискретные и 2) непрерывные случайные величины.

Если все возможные значения случайной величины X можно занумеровать, то X называют дискретной случайной величиной.

Дискретность выражается в отчетности множества значений, в изолированности возможных значений.

Возможные значения случайной величины X обозначают $x_1, x_2, ..., x_n$ и пишут в верхней строке таблицы, называемой рядом распределения вероятностей. В нижней строке числу $x_i$

сопоставляют вероятность p того, что X примет это значение $x_i$: $p_i=P\{X=x_i\}$ - вероятность события $\{X=x_i\}$.

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | ... | ... | $x_n$ |
| $p_i$ | $p_1$ | $p_2$ | ... | ... | $p_n$ |

Поскольку учтены все значения $x_1$ случайной величины X, то в нижней строке сумма $\sum_{i=1}^{n} p_i = 1$.

**Пример 1.** Опыт - бросание трех монет, X - число выпавших орлов. Ясно, что X может принять значения 0,1,2 и 3, но с какими вероятностями? Составим вспомогательную таблицу всех случаев: PPP (все решки), OPP (на первом монете орел, на других решки) и т.д. Всего случаев 8 и вероятность каждого 1/8.

| возможные случаи | PPP | OPP | POP | PPO | OOP | OPO | POO | OOO |
|---|---|---|---|---|---|---|---|---|
| число орлов | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |

Подсчитаем вероятности для построения ряда распределения:

$$p_1 = P\{X=0\} = \frac{1}{8}, p_2 = P\{X=1\} = P\{OOP + POP + PPO\} = \frac{3}{8}, \ldots$$

Ряд распределения случайной величины X

| $x_1$ (значения X) | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p_1=P\{X=x_1\}$ | 1/8 | 3/8 | 3/8 | 1/8 |

Ряд распределения, кроме таблицы, задают и графически в виде так называемого многоугольника распределения:

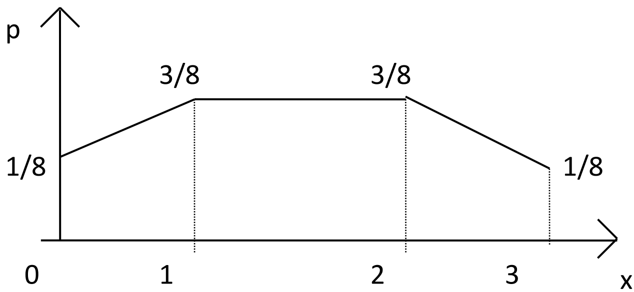



Рис.1

## 12. Непрерывная случайная величина. Функция распределения и плотность вероятности.

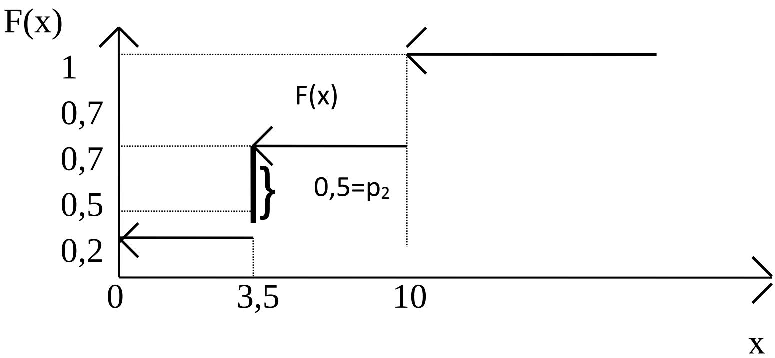
Всякую случайную величину X полностью характеризует ее функция распределения вероятности, обозначаемая F(x) либо $F_x(x)$. Взяв точку x на оси Ox, зададим в ней значение функции F(x) как вероятность $P\{X<x\}= P(\infty<X<x\}$ - вероятность того, что случайная величина X примет значение левее заданной точки x.

Таким образом, дискретную случайную величину X можно задать как рядом распределения, так и функцией распределения F(x), причем переход от одной формы закона распределения к другому совершается однозначно: по ряду легко построить функцию F(x) и наоборот.

**Пример 2.** Построить функцию распределения F(x) для рассмотренной в примере 3 случайной величины X, заданной рядом распределения

| | | | |
|---|---|---|---|
| $x_i$ | 0 | 3,5 | 10 |
| $p_i$ | 0,2 | 0,5 | 0,3 |

Функция F(x) показана на рисунке:



Например, $F(2)=0,2$; $F\{X<3,5\}=0,2$. А вот $F(3,501)=0,7$, т.к. $F(3,501)=P(X<3,501)=P(X=0)+P(X=3,5)=0,2+0,5=0,7$.

**Замечания.**

1. Для всякой случайной величины $F(-\infty)=0$, $F(\infty)=1$
2. Функция F(x) определена в любой точке x оси Ox, а ее значения лежат между 0 и 1.
3. Функция F(x) возрастает.

Обратимся теперь к непрерывной случайной величине X, которая в отличие от дискретной может принять любое значение из некоторого промежутка, т.е. ее значения сплошь заполняют некоторый интервал и потому их множество несчетно. Например:

1) размер (либо прочность, вес) детали массового производства;
2) урожай с одной сотки;
3) ошибка измерения;
4) продолжительность работы устройства (холодильника и т.д.) до момента отказа.

У непрерывной случайной величины X функция распределения F(x) всюду непрерывна и кусочно-дифференцируется.

## 13. Плотность вероятности.

Непрерывную случайную величину X можно задать либо функцией распределения F, либо ее производной

$f(x)=\dfrac{dF(x)}{dx}$ , называемой плотностью распределения вероятности или плотностью вероятности. В силу монотонности функции F(x) плотность $f$(x)≥0 всюду.

Зная F(x), можем найти плотность вероятности по формуле $f$(x)=F'(x), а зная $f$(x), найдем функцию распределения как $F(x)=\int\limits_{-\infty}^{x} f(t)dt$ .

Ясно, что $\int\limits_{-\infty}^{\infty} f(x)dx$ = F(∞)=1

Для непрерывной случайной величины X вероятность попадания ее в промежуток с концами a и b (неважно, открытый или замкнутый) равна

$$P\{a\le X\le b\}=F(b)-F(a)=\int\limits_{-\infty}^{в} f(x)dx-\int\limits_{-\infty}^{a} f(x)dx$$

Полезно помнить, что:

1) плотность вероятности $f$(x) это есть $\dfrac{\Delta P}{\Delta x}$ - вероятность попадания X в интервал (x, x+Δx), деленная на его длину Δx, когда длина Δx исчезающе мала;
2) вся площадь между графиком $f$(x) и осью Ox равна1.

## 14. Математическое ожидание.

Удобно представлять себе вероятность как массу (массу вероятности), распределенную вдоль оси абсцисс.

Вся масса равна 1. Если распределение дискретное, масса сосредоточена в изолированных точках $x_i$ оси, если распределение непрерывно - масса размазана вдоль оси Ox с плотностью $f$(x). Тогда абсциссу центра тяжести этой массы называют математическим ожиданием либо средним значением случайной величины.

Уточняя сказанное дадим определение математическим ожиданием (или средним значением) случайной величины X называют число

$$M(X)=\begin{cases}\sum_{i=1}^{n} p_i x_i & (*)\\ \int_{-\infty}^{\infty} xf(x)dx & (**)\end{cases}$$

(*) в дискретном случае, (**) - в непрерывном.

Вместо M(X) пишут также MX, $\mu$, E(X).

Число M(X) является неслучайной величиной той же размерности, что X, и характеризует распределение «в среднем», являясь абсциссой центра тяжести массы вероятности.

У симметричного закона среднее значение лежит на оси симметрии. Хотя среднее значение - неполная характеристика распределения, это важная характеристика.

Две другие характеристики положения распределения случайной величины: мода и медиана.

Мода - точка на оси абсцисс, в которой максимальна вероятность $p_i$, если X дискретна, и максимальна плотность $f$(x) в непрерывном случае.

Медиана - точка d, слева и справа от которой вероятность одинакова, т.е. $P\{X<d\}=P\{X>d\}$.

Основные свойства математического ожидания:

1. Если С = const, то М(С)=С - среднее значение постоянной С равно С;
2. М(СХ)=СМ(Х) - постоянный множитель можно выносить за знак математического ожидания;
3. М(Х+Y)=М(Х)+М(Y) - среднее значение суммы любых случайных величин, равно сумме их средних значений;
4. М(ХY)=М(Х)М(Y) - для независимых случайных величин.

Х, Y независимы, в п. 4.

Доказательство свойств можно найти в учебниках.

**Пример 3.** Вероятность поражения цели при выстреле ракетой равна р=0,7. Ракетный дивизион располагает 4 ракетами и стреляет ими в цель до поражения (поразив цель, не тратит лишних ракет). Для числа Х расходуемых ракет построить ряд распределения и найти среднее М(Х), а также вероятность поражения цели.

Начнем с последнего.

Р{поражения цели}=$p_1+p_2+p_3+p_4$, где вероятность поражения цели первым выстрелом, $p_1$==0,7;

$p_2$ - вероятность того, что первая ракета промахнется, но вторая поразит цель: $p_2=pq=0{,}3*0{,}7=0{,}21$.

Аналогично $p_3$=Р{два промаха, третий выстрел - поражение цели}$=q*q*p=0{,}3^2*0{,}7=0{,}063$, $p_4=q^3p=0{,}0189$.

$$P\{\textit{поражение}\cdot\textit{цели}\} = \sum^{4} p_i = 0{,}992\,.$$

Вероятность $P\{X=1\}=p_1$, $P\{X=2\}=p_2$, $P\{X=3\}=p_3$ - вероятность поразить цель только с третьего раза; но $P\{X=4\}>p_4$, т.к. при первых трех промахах последняя четвертая ракета будет, очевидно истрачена с поражением цели или без него. $P\{X=4\}=1-(p_1+p_2+p_3)$ - вероятность истратить все 4 ракеты.

Ряд распределения X

| | | | | |
|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | 4 |
| $p_i$ | 0,7 | 0,21 | 0,063 | 0,027 |

Среднее $MX=1*0{,}7+2*0{,}2+3*0{,}063+4*0{,}027=1{,}4$.□

## 15. Дисперсия.

Случайные величины при общем среднем могут быть совершенно разными, например, одна может в опытах меняться в узких пределах, а вторая - в широких при одинаковом среднем.

Чтобы охарактеризовать разброс, рассеяние, изменчивость в опытах случайной величины X есть несколько показателей, но чаще всего применяют дисперсию D(X) или среднеквадратическое (стандартное) отклонение $\sigma = \sqrt{D(X)}$

Пусть $\mu=M(X)$ - среднее значение X. Рассмотрим случайную величину $(X-\mu)^2$, т.е. величину, принимающую значение (х- $\mu$ )$^2$. Ее среднее значение и называется дисперсией случайной величины X: DX=M((X-μ)$^2$)

По формуле (16) получим:

$$D(X) = DX = \begin{cases} \sum_i (x_i - \mu)^2 p_i & (*) \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & (**) \end{cases}$$

(*) - в дискретном случае, (**) - в непрерывном.

Поскольку $(X-\mu)^2 \geq 0$, то $DX \geq 0$.

Важна формула

$DX=M[(X-\mu)^2]=M(X^2-2\mu X+\mu^2)=M(X^2)-2\mu*MX+\mu^2=M(X^2)-\mu^2$.

Словами: дисперсия равна среднему квадратов минус квадрат среднего.

**Пример 4.** О стрельбе ракетой и попадании точки в круг радиуса r = 2.

а)

$$DX = \sum_{t=1}^{4} (x_i - \mu)^2 p_i = (1-1{,}4)^2 * 0{,}7 + (2-1{,}4)^2 * 0{,}21 + (3-1{,}4)^2 *$$

$$\boxed{* 0{,}063 + (4-1{,}4)^2 * 0{,}027 = 0{,}531;}$$

б)

$$DX = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_0^2 \left(r - \frac{4}{3}\right)^2 \frac{r}{2} dr = \int_0^2 \left(\frac{r^3}{2} - \frac{4}{3}r^2 + \frac{8}{9}r\right) dr =$$

$$= \left[\frac{r^4}{8} - \frac{4r^3}{9} + \frac{4r^2}{9}\right]_0^2 = 2 - \frac{32}{9} + \frac{16}{9} = 0{,}22.$$

**Пример 5.** Для равномерного распределения:

X=R[0,1] найти среднее и дисперсию.

Решение. Плотность вероятности f(x) равна 1 в интервале (0,1) и 0 вне его. Поэтому:

$$\mu = MX = \int_1^0 x*1*dx = \frac{x^2}{2}\bigg|_0^1 = \frac{1}{2}.$$

$$DX = MX^2 - \mu^2 = \int_0^1 x^2*1*dx - \mu^2 = \frac{x^3}{3}\bigg|_0^1 - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

<u>Свойства дисперсии</u>

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

2. $D(CX)=C^2D(X)$ - постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат.

Действительно,

$D(CX)=M[(CX)^2]-[M(CX)]^2=C^2(MX^2-\mu^2)=C^2D(X)$

3. Для независимых X и Y D(X±Y)=D(X)+D(Y) - дисперсия суммы (разности) независимых случайных величин равна сумме их дисперсий.

Докажем свойство 3 для X+Y

$D(X+Y)=M[(X+Y)^2]-M^2[(X+Y)] \quad =M(X^2+2XY+Y^2)-(\mu_x+\mu_y)^2=M(X^2)+2M(XY)+M(Y^2)-\mu_x^2-2\mu_x\mu_y-\mu^2_y$ $==DX+DY+2[M(XY)-\mu_x\mu_y]$

Выражение $M(XY) - \mu_x\mu_y = cov(X,Y)=0$, а D(X+Y)=DX+DY.

**Замечание.** Из этих свойств следует D(-X)=D(X), D(X+C)=D(X).

**Пример 6.** Найти дисперсию D(Y) иD(Z) для Y=X+2 и Z=2X, если X =R[0,1]- равномерная стандартная величина

Решение.

$$DY = D(X+2) = D(X) + D(2) = \frac{1}{12} + 0 = \frac{1}{12}$$

DZ=D(2X)=4D(X)=1/3 - удвоение случайной величины дает вчетверо большую дисперсию. □

## 16. Биномиальное распределение.

Приступим теперь к рассмотрению одной из главных схем теории вероятностей. Рассматривается последовательность взаимно независимых испытаний (в дальнейшем будем называть их единичными), т.е. таких испытаний, что вероятность того или иного результата в каждом из них не зависит от того, какие результаты имеют место в остальных. В каждом из этих единичных испытаний интересующее нас событие А может наступить (или не наступить) с вероятностью р, не зависящей от номера испытания. Эта схема изучалась впервые известным ученым Яковом Бернулли в конце XVII века и потому получила название схемы Бернулли.

**Пример 7.** Среди волокон хлопка определенного сорта в среднем 75% имеют длину меньше 45мм и 25% - больше (или равную) 45мм. Найти вероятность того, что среди трех наудачу взятых волокон два будут короче, а одно длиннее 45мм (событие С).

Решение. Обозначим событие А={выбор волокна с длиной меньше 45мм }, $C = AA\overline{A} + A\overline{A}A + \overline{A}AA, \cdot где \cdot AA\overline{A} =$

{два первых выбранных волокна оказались короче, а третье длиннее 45 мм}.

Так как все три слагаемых - несовместные события, то

$$P(C) = P(AA\overline{A}) + P(A\overline{A}A) + P(\overline{A}AA)$$

По правилу умножения вероятностей для независимых событий

$$P(AA\overline{A}) = P(A)P(A)P(\overline{A}) = \left(\frac{3}{4}\right)^2 * \frac{1}{4} = \frac{9}{64}; P(A\overline{A}A) = \frac{9}{64};$$

$$P(\overline{A}AA) = \frac{9}{64}; P(C) = 3 * \frac{9}{64} = \frac{27}{64} \square$$

Обычно событие А в единичном испытании для краткости называют успехом, а $\overline{\text{A}}$ - неудачей.

**Пример 8.** В приборе 20 однотипных элементов. Вероятность отказа элемента за год работы равна 0,003. Считая отказы независимыми, найти вероятность того, что за год:

1) все 20 элементов проработают безотказно;
2) откажет ровно один элемент;
3) откажет более одного элемента.

Назовем кратко успехом безотказную работу элемента в течение года.

A={успех в единичном испытании}, P(A)=p=0,997, q=0,003, n=20.

X - число успехов в двадцати единичных испытаниях.

1) P{X=20}=P{все 20 элементов работают безотказно}=

$= P\left\{\underbrace{AAA\ldots A}_{n\cdot штук}\right\} = \left[P(A)\right]^{20}$ $=p^{20} = 0{,}997^{20} \approx 0{,}9417$ - в среднем 94% приборов не имеют отказов элементов;

2) P{X=19}=P{ровно 19 успехов}=

$= P\{\overline{A}A\ldots A + A\overline{A}A\ldots A + \ldots + AA\ldots A\overline{A}\} = P\{\overline{A}A\ldots A\} + P\{A\overline{A}A\ldots A\} + \ldots +$

$+ m\{AA\ldots A\overline{A}\} = p^{19}q + p^{19}q + \ldots + p^{19}q = 20\,p^{19}q = 0{,}0567$;

3) P{X<19}=P{успехов меньше 19}=1-P{X≥19}=1-0,9417-0,0567=0,0016.

Общий случай. Дано: p - вероятность успеха в единичном испытании, q=1-p, n - число единичных испытаний в опыте. Найти $P_{m,n}$ - вероятность получить ровно m успехов в опыте.

Вероятность получить любую конкретную комбинацию $A\overline{A}AA\overline{A}\ldots A$ , содержащую на конкретных местах m букв из A, означающих успех, и (n-m) букв $\overline{A}$ (неудач) равна $p^m q^{n-m}$.

Всего же различных таких комбинаций (отличающихся лишь порядком расположения A и $\overline{A}$) имеется $C_n^m$ - число сочетаний из n элементов по m. Отсюда

$$P_{m,n} = P\{X = m\} = C_n^m p^m q^{n-m}, m = 0,1,2,\ldots,n$$

Эта формула задает биномиальное распределение, т.е. вероятность получения m успехов и n-m неудач в схеме Бернулли. Название «биномиальное» связано с тем, что $P_{m,n}$ есть слагаемое в разложении бинома Ньютона:

$$1 = 1^n = (p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{2}p^{n-2}q + \ldots + q^n = \sum_{m=0}^{n} C_n^m p^m q^{n-m} =$$

$$= \sum_{ь=0}^{m} P_{m,n}.$$

**Пример 9.** Имеется урна с 10 шарами: 9 белых, один черный. 5 раз наудачу вынимаем из урны шар и возвращаем его в урну. Пусть X - число появлений белого шара. Составить ряд распределения для случайной величины X.

Решение. Параметры биномиальной величины X: n=5; p=9/10=0,9; q=0,1.

Заметим, что если бы мы в опыте не возвращали шар в урну, т.е. если бы вынимали 5 шаров без возвращения (неважно, сразу все 5 или последовательно), то имели бы дело с гипергеометрическим распределением.

В нашем же случае «работает» схема Бернулли P{X=0}=P{ни разу не вынут белый шар}= = $C_5^0\, p^0\, q^5 = 1*1*0{,}1^5 = 0{,}00001$; $P\{x=1\} = C^1{}_5\, 0{,}9*0{,}1^4 = 0.00045$; P(X=2)= $C_5^2\, 0{,}9^2 0{,}1^3 = 0{,}0081$ .

P{X=3}=P{белый шар появляется ровно 3 раза}= $C_5^3 p^3 q^{5-3} = \frac{5\cdot 4\cdot 3}{1\cdot 2\cdot 3} 0{,}9^3 0{,}1^2 = 0{,}0729$; P{X=4}= $C_5^4 p^4 q^1 =$ $5\cdot 0{,}9^4 0{,}1$=0,32805; P{X=5}=$p^5$=0,59049.

Ряд распределения вероятностей

| m | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P_{m,5}$ | 0,00001 | 0,00045 | 0,0081 | 0,0729 | 0,32805 | 0,59049 |

Проверка: $\sum_i p_i = \sum_{m=0}^{5} P_{m,5} = 1$ •

Дискретная случайная величина принимает с вероятностями $p_1, p_2, \ldots, p_n$ какие-то значения $x_1, x_2, \ldots, x_n$. То из этих значений, которому соответствует наибольшая из вероятностей $p_i$, называется модой случайной величины. В последнем примере наибольшей является вероятность 0,59049. Поэтому мода рассмотренной биномиальной величины равна 5. Иначе говоря, мода - наивероятнейшее значение случайной величины. Можно показать, что у биноминальной величины всегда одна или две моды, в зависимости от числа: k=np-q. Если же число k целое, то модой будут два целых числа: k и k+1. Если же число k нецелое, то мода одна и получается из k округлением в большую сторону до ближайшего целого. Так в предыдущем примере k=np-q=5*0,9-0,1= =4,4 - нецелое, мода=5.

У биномиальной величины X с параметрами n=5; p=0,5 получим две моды: k=np-q=5*0,5-0,5=2, мода равна 2 и мода равна 3. Примером такой величины может служить число выпавших орлов при бросании пяти монет. До точки x=2 вероятности $p_i$ будут возрастать, в точках x=2 и x=3 достигать максимума, а затем монотонно убывать.

Для вывода формул среднего и дисперсии биномиальной величины обычно выбирают один из двух путей:

1) непосредственное применение формул, дающих M(X), M($X^2$) и D(X)

Так, $M(X)=\sum_i x_i p_i=\sum_{m=0}^{n} mP_{m,n}$. Подставляя $P_{m,n}$ из формулы (31) и произведя преобразования, можем получить $\mu$=MX=np. Аналогично, $M(X^2)=\sum_{m=0}^{n} m^2 P_{m,n}$ оказывается равным n(np+q) и $D(X)=M(X^2)-\mu^2=npq$;

2) второй путь состоит в представлении биномиальной величины X как суммы независимых случайных величин и проще в вычислительном плане: $X=\sum_{j=1}^{n}\xi_j$, где $\xi_j$ принимает значение 0, если в j-ом единичном испытании наступит неудача, и 1, если успех.

Так, в опыте с результатом $AA\bar{A}AA\bar{A}\bar{A}$ с n=7 величины $\xi_1,\xi_2,\xi_4,\xi_5$ приняли значение 1, а $\xi_3,\xi_6,\xi_7$ - значение 0, т.к. в 3, 6 и 7 испытаниях A не наступило.

Поскольку каждая случайная величина $\xi_j$ имеет ряд распределения

| $\xi_j$ | 0 | 1 |
|---|---|---|
| вероятность | q | p |

то сразу получаем

$M\xi_j=p; M(X)=M\left(\sum_{j=1}^{n}\xi_j\right)=\sum_{j=1}^{n}M\xi_j=np$ - среднее значение биномиальной величины.

$D\xi_j=M\xi_j^2-p^2=0^2q+1^2p-p^2=p-p^2=pq$

$D(X)=D\sum_{j=1}^{n}\xi_j=\sum_{j=1}^{n}D\xi_j=npq$ - с учетом взаимной независимости $\xi_1,\xi_2,...,\xi_n$.

**Пример 10.** Найти среднее и дисперсию в примерах 14 и 15.

Решение. В примере14

M(X)=np=20*0,997=19,94;D(X)=npq= =19,94*0,003=0,06.

В примере 15 MX=5*0,9=4,5; DX=4,5*0,5=2,25.□

## 17. Распределение Пуассона.

Это широко применяемое в технических приложениях распределение можно рассматривать как предельный случай биномиального.

В биномиальном распределении величины Xn при заданных m, n и p вероятность $P\{X_n=m\}=P_{m,n}=C^m_n\, p^m q^{n-m}$, а среднее $M(X_n)=np$. Что, если не меняя среднего числа $np=\lambda=const$, мы будем увеличивать число n ( а, значит, $p=\frac{\lambda}{n}$ уменьшать). Как будет себя вести и к чему стремиться вероятность $P\{X_n=m\}$?

Формула дает распределение вероятностей для m=0,1,2,... так называемой пуассоновской величины

$$\sum_{m=0}^{\infty} p_m = e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}=1$$

Это распределение зависит только от одного параметра λ, который, как можно показать, равен как среднему значению, так и дисперсии: λ=M(X)=D(X).

Действительно, для биномиальной величины $MX_n=np=\lambda$ и $D(X_n)=npq=\lambda q$, где $q=1-p=1-\lambda/n$. Но M(X)= $\lim\limits_{n\to\infty} M(X_n)=\lambda$ и D(X)= $\lim\limits_{n\to\infty} D(X_n)=\lim\limits_{n\to\infty} \lambda\left(1-\frac{\lambda}{n}\right)=\lambda$ .

## 18. Функции дискретных случайных аргументов.

Пусть X - дискретная случайная величина, а $\varphi(x)$ заданная числовая функция одного аргумента. Тогда, приняв за аргумент случайную величину X, получим функцию случайного аргумента. Это будет новая случайная величина $Z=\varphi(X)$, которая в опытах, где X имеет значения $x_1,x_2,x_3,...$ принимает соответственно значениями $\varphi(x_1)$, $\varphi(x_2)$, $\varphi(x_3)$,..., некоторые из которых могут совпадать.

**Пример 11.** Случайная величина X задана рядом распределения

| $x_i$ | -1 | 0 | 1 | 2 | 10 |
|---|---|---|---|---|---|
| $p_i$ | 0,2 | 0,2 | 0,2 | 0,2 | 0,2 |

Ее легко моделировать, имея 5 шаров с числами $x_i$ на них. Требуется построить ряд распределения случайной величины $Y=2X_i+1$. В частности $X=-1 \Leftrightarrow Y=-1*2+1=-1$ и $P\{X=-1\}=P\{Y=-1\}=0,2$ $X=0 \Leftrightarrow Y=0+1=1$ и $P\{X=0\}=P\{Y=1\}=0,2$

Ряд распределения величины Y=2X+1 получается из ряда для X путем замены верхней строки: вместо $x_i$ надо написать $y_i=2\,x_i+1$:

| $y_i$ | -1 | 1 | 3 | 5 | 21 |
|---|---|---|---|---|---|
| $p_i$ | 0,2 | 0,2 | 0,2 | 0,2 | 0,2 |

Для $Z=X^2$ ряд распределения имеет меньше столбцов, т.к. P{Z=1}=P{$X^2$=1}=P{X=-1}+P{X=1}=0,4 - два разных значения (±1) величины X приводят к одному значению величины Z.

| $z_k$ | 0 | 1 | 4 | 100 |
|---|---|---|---|---|
| $p_k$ | 0,2 | 0,4 | 0,2 | 0,2 |

Находим среднее:

M(Y)=0,2(-1+1+3+5+21)=5,8

M(Z)=1*0,4+4*0,2+100*0,2=21,2 □

## 19. Равномерное распределение.

Переходя к непрерывным распределениям, рассмотрим простейшее - равномерное, называемое за вид его плотности вероятности f(x), также законом прямоугольника.

Точки a и b связаны лишь условием a<b. Плотность вероятности f(x) равна нулю вне интервала (a,b) и равна h внутри его. Высота h прямоугольника находится из условия единичной площади: $h = \dfrac{1}{b-a}$.

Таким образом,

$$f(x) \approx \begin{cases} h = \dfrac{1}{b-a} & (*) \\ h = 0 & (**) \end{cases}$$

(*) - при a<x<b; (* *) - при всех остальных значениях x.

Величину X с таким законом обозначим кратко так:

X=R[a,b] - величина X подчиняется равномерному закону на промежутке (a,b).

Этот закон в силу его простоты хорош для разных прикидок и демонстрации разных вероятностных свойств, случайная величина R[0,1], реализуемая рулеткой, служит основой моделирования других распределений в методе Монте-Карло. На практике встречаются величины, подчиняющиеся этому закону. Сюда относится, в частности, время ожидания в метро электрички пассажиром, прибывающим на платформу, а так же ошибки округления чисел при расчетах.

Функция распределения равномерной величины F(x)=P{X<x} равна нулю слева от точки a и 1 справа от b. Для точек x между a и b.

$$F(x) = \int_0^x \frac{1}{b-a} dt = \frac{t}{b-a}\Big|_a^x = \frac{x-a}{b-a}$$

Итак, F(x) внутри (a,b) линейно возрастает от 0 до 1. Найдем среднее MX и дисперсию DX для X=R[a,b].

$$\mu = M(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\frac{x^2}{2}\Bigg|_a^b = \frac{a+b}{2}$$

$$M(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = h\int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

$$D(X) = M(X^2) - \mu^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(a+b)^2}{3} - \frac{ab}{3} - \left(\frac{a+b}{2}\right)^2 =$$

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

Среднее квадратическое отклонение $\sigma = \sqrt{D(X)} = \dfrac{b-a}{\sqrt[2]{3}}$.

**Пример 12.** Найти вероятность попадания случайной величины X=R[0,8] в интервал (μ−σ,μ+σ).

Решение.

$$\mu = \frac{8}{2} = 4, \sigma = 4/\sqrt{3} \approx 2{,}31;$$

$$P\{\mu - \sigma < X < \mu + \sigma\} = \frac{2\sigma}{8} = 0{,}58.$$ □

## 20. Показательное распределение.

Как и распределение Пуассона, показательное распределение (называемое также экспоненциальным) широко применяется в теории надежности и теории массового обслуживания. Им описывается длительность безотказной работы многих технических устройств. Функция задается формулой

$$F(t)=\begin{cases}0\cdot при & t\le 0\\ 1-e\cdot^{-\lambda t} & t>0\end{cases}$$

Плотность вероятности

$$f(t)=\frac{dF(t)}{dt}=\begin{cases}0\cdot при & t\le 0\\ \lambda\cdot e^{-\lambda t} \quad при & t>0\end{cases}$$

Число λ>0 называется параметром распределения.

Среднее значение $\mu=M(T)=\int_0^\infty te^{-\lambda t}*\lambda dt=\frac{1}{\lambda}$.

Дисперсия $D(T)=\int_0^\infty (t-\lambda)^2 e^{-\lambda t}*\lambda dt=\frac{1}{\lambda^2},\sigma=\frac{1}{\lambda}$.

**Пример 13.** Положим, случайное время T работы прибора до момента отказа имеет плотность $f(t)=0{,}1e^{-0{,}1t}$ при t>0, параметр λ=0,1 характеризует интенсивность отказов. Найти вероятность того, что прибор откажет в первый год службы.

Решение. Средний срок службы M(T)=1/λ=1/0,1=10 лет.

P{T<1}=F(1)=1–$e^{-0{,}1}$=0,095, т.е. примерно 10% всех приборов данного типа выходят из строя в первый год работы.

## 21. Нормальное распределение.

Непрерывная случайная величина X называется распределенной по нормальному закону (или закону

Гаусса, а так же Гаусса-Лапласа), если ее плотность вероятности такова:

$$f(x) = c * \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$$

где константы c и $\sigma>0$, константа a - любое число, exp (t) обозначает $e^t$.

Нормирующий множитель С однозначно определяется из условия $\int_{-\infty}^{\infty} f(x)dx = 1$ и равен $c = \frac{1}{\sigma\sqrt{2\pi}}$.

Вычисления показывают, что

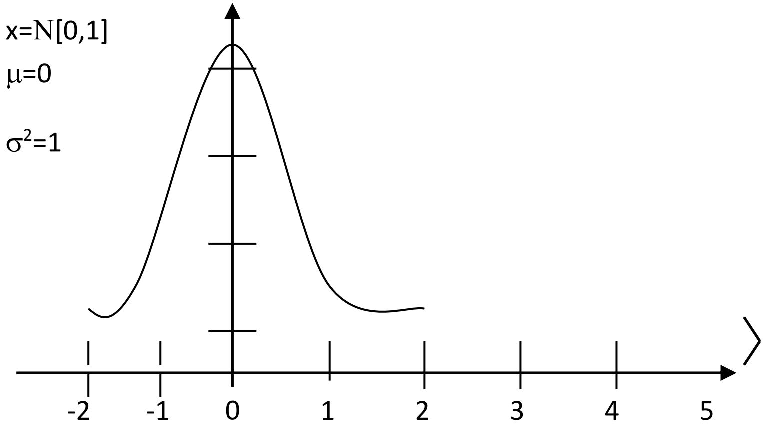
$\mu = \grave{I}\ \ (\tilde{O}) = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] dx = a; D(X) = \sigma^2$ , так что нормальное распределение зависит от двух параметров: a=M(X) и $\sigma^2$=D(X). Это кратко запишем так:X=N[a,$\sigma^2$] - нормальная величина X с M(X)=a, D(X)=$\sigma^2$. Величину $\xi$=N(0,1) с нулевым средним и единичной дисперсией называют стандартной нормальной. Ее плотность вероятности

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

Плотность вероятности

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

больше нуля в любой точке x оси абсцисс; она достигает максимума, равного $\frac{1}{\sigma\sqrt{2\pi}}$, в точке, лежащей на оси симметрии графика функции $f$. По обе стороны от точки x=μ кривая плоскости симметрично опускается, напоминая холм или колокол и имея перегиб в точках μ−σ и μ+σ




Уменьшение параметра μ означает сдвиг кривой f(x) влево на величину μ (без изменения формы). Уменьшение стандартного (или среднеквадратичного) отклонения σ означает растяжение графика функции $f$ по вертикали и сжатие его вдоль оси x (при сохранении единичной площади под кривой). Увеличение σ означает «сплющивание» распределения вертикали.

Нормальное распределение (или почти нормальное) часто встречается в окружающем мире, оно как бы норма (отсюда название).

Размеры механически обрабатываемых деталей, ошибки всевозможных приборов, производительность труда, урожайность сельхозкультуры с 1 га, величина снежного покрова на крыше здания 1 января каждого года

– случайные нормальные величины. Для стандартной нормальной величины ξ=N[0,1].

$$\Phi(x) = P\{\xi < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{i^2}{2}} \, \mathrm{dt}$$

Это - площадь под кривой $f$ слева от точки x.

Ее таблицы есть в большинстве книг по теории вероятностей. Вот некоторые ее значения.

| x | -3 | -2 | -1 | 0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,7 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ф (x) | 0,0 014 | 0,0 23 | 0,1 59 | 0,5 00 | 0,5 4 | 0,5 79 | 0,6 18 | 0,6 55 | 0,6 91 | 0,7 58 | 0,8 41 |

| x | 1,2 | 1,4 | 1,5 | 1,7 | 2 | 2,5 | 3 | 3,5 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| Ф(x) | 0,88 5 | 0,91 9 | 0,93 3 | 0,95 5 | 0,97 7 | 0,99 4 | 0,99 86 | 0,99 98 | 0,99 997 |

Часто в целях экономии места дают значения функции Ф лишь при x≥0, тогда при x<0 следует пользоваться формулой Ф(x)=1-Ф(-x). Например, Ф(-2)=1-Ф(2)=1-0,977=0,023.

Пользуясь таблицами, надо быть внимательным. Нередко через Ф(x) обозначают функцию Лапласа, отличающуюся от формулы функции распределения стандартной нормальной величины лишь тем, что вместо -∞ нижний предел равен 0. В старых книгах предпочитали пользоваться этой функцией. Важное свойство нормального закона - воспроизводимость: сумма (разность) нормальных величин является нормальной величиной.

### 11.1. *Вычисление вероятностей для нормальной величины.*

Пусть X=N(μ,σ) и заданы числа μ,σ,a,b. Надо найти вероятность попадания случайной величины X в промежуток (a,b): P{a<X<b}=P{a<X<b}=$F_x$(b)-$F_x$(a).

$$P\{a < X < b\} = P\{a \le X \le b\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

**Пример 14.** Допустим, что рост X мужчины подчиняется закону N[172,6], т.е. является нормальной величиной со средним значением 172 см. и стандартным отклонением σ=6см. Найти долю мужчин, чей рост лежит между 190 и 200 см., и долю мужчин с ростом меньше 180 см.

Решение.

1) $P\{190 < X < 200\} = \Phi\left(\frac{200-172}{6}\right) - \Phi\left(\frac{190-172}{}6\right) = \Phi(4{,}67) - \Phi(3) =$

$= 1 - 0{,}9986 = 0{,}0014$ - в среднем на 10000 мужчин приходится 14 человек с ростом X>190см.

Замечание. При x>4 можно считать Ф(x)=1.

2) $P\{X < 180\} = \Phi\left(\frac{180-172}{6}\right) = \Phi(1{,}333) = 0{,}909$ ,т.е. 90% мужчин имеют рост меньше 180 см.

Замечание. В жизни все обстоит сложнее, чем в схемах. Так, взятые здесь данные с M(X)=172 см., возможно устарели, поскольку последние десятилетия наблюдается акселерация и средний рост населения увеличивается. Безусловно, такие данные для мужчин РФ есть в военкоматах.

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

Рассмотрим n нормально распределенных независимых случайных величин, причем математическое ожидание каждой равно нулю, а дисперсия - единице. Сумму квадратов таких случайных величин обозначим:

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2, \quad где \quad \xi_i = N\,[0,1]$$

Закон распределения такой суммы называется законом «хи квадрат» с n степенями свободы. Если случайные величины связаны одним линейным соотношением, то число степеней свободы k=n-1. С увеличением числа степеней свободы распределение приближается к нормальному.

### *12.1. Распределение Стьюдента.*

Из двух независимых случайных величин $\xi$=N(0,1) и$\chi^2_n$, деля одну на корень из другой, создадим случайную величину $t_n = \dfrac{\xi}{\sqrt{\chi_n^2 / n}}$ . Распределение величины $t_n$ называют t-распределением (или распределением Стьюдента) с n степенями свободы. Поскольку в числителе стоит нормальная стандартная величина, то распределение $t_n$ симметрично относительно оси ординат. Это распределение при увеличении n неограниченно приближается к N[0,1]. При n>50 t-распределение практически совпадает с N[0,1].

### 12.2. F-распределение (Фишера - Снедекора)

Пусть $\chi_m^2 \grave{e} \chi_n^2$ независимые и случайные величины. Отношение случайных величин $\chi^2_m/m$ и $\chi^2_n/n$ называется дисперсионным отношением и обозначается

$$F_{m,n} = \frac{\chi_m^2}{m} \Big/ \left(\frac{\chi_n^2}{n}\right)$$

. F - случайная величина, подчиняющаяся F-распределению с m и n степенями свободы. Поскольку среднее величин $\chi_m^2 \chi_n^2$ равно 1, а дисперсия стремится к нулю с ростом m и n, то не удивительно, что F-распределение концентрируется вблизи 1 и становится все более иглообразным с ростом m и n.

Распределения $\chi_n^2$ , $t_n$ и F имеют чрезвычайно важное значение в статистике и для них составлены многочисленные подробные таблицы.

### 12.3. Закон больших чисел.

Этот закон выражает общий принцип, в силу которого совместное действие случайных факторов приводит при некоторых весьма общих условиях к результату, почти не зависящему от случая. Сближение частоты наступления случайного события с его вероятностью при возрастании числа испытаний (подмеченное сначала, по- видимому, на азартных играх) может служить первым примером действия этого закона.

На рубеже XVII и XVIII вв. Я.Бернулли доказал теорему, утверждающую, что в последовательности независимых испытаний по схеме Бернулли (т.е. когда вероятность p наступления интересующего нас события A не меняется от испытания к испытанию), верно соотношение

$P\left\{\left|\frac{m_n}{n}-p\right|>\varepsilon\right\}\to 0$ при любом $\varepsilon>0$ и $n\to\infty$;

Здесь $m_n$ число появлений события в первых n испытаниях, $\frac{m_n}{n}$ - частота появлений.

Грубо говоря, эта теорема - простейшая форма закона больших чисел - утверждает, что с ростом числа испытаний n частота приближается к вероятности, т.е. доказывается та устойчивость частоты, о которой упоминалось в самом начале и которая прежде была подмечена экспериментально. Образно говоря, «семь раз отмерь и один раз отрежь».

П.Л.Чебышев (1867г.) доказал закон больших чисел в более общей, чем у Бернулли, форме, а именно:

Пусть мы имеем n взаимно независимых случайных величин $X_1,X_2,...,X_n$ с одним и тем же математическим ожиданием $\mu$ и одним и тем же стандартным отклонением $\sigma$. Тогда среднее арифметическое значение этой случайных величин, то есть величина $\overline{X}=\frac{X_1+X_2+...+X_n}{n}$ при достаточно большом n будет с вероятностью, как угодно близкой к единице (практически достоверно), как угодно мало отличаться от $\mu$.

Это частный случай закона больших чисел.

В наше время доказаны куда более общие и сложные варианты этого закона. Но содержательный смысл этого замечательного закона доступен и неспециалисту и состоит в том, что, в то время как отдельная случайная величина может часто принимать значения, далекие от ее математического ожидания (иметь большое рассеяние), среднее арифметическое большого числа случайных величин ведет себя в этом отношении иначе: такая величина очень мало рассеяна и с очень высокой вероятностью принимает значения, близкие к ее среднему значению. На практике закон больших чисел используется, например, в том, что по сравнительно небольшой пробе (выборке) судят о качестве однородного материала.

Так, о качестве хлопка, находящегося в кипе, судят по нескольким маленьким его пучочкам (штапелям), выхваченным случайно из разных мест кипы. Или о качестве большой партии зерна судят по нескольким небольшим меркам (пуркам), наполненным зернами из разных случайных мест оцениваемой партии. Или, наконец, по опросу сравнительно небольшой, сделанной с учетом случайности выборки граждан, социологи делают достаточно верные предсказания к тем или иным проблемам.

## 13. Центральная предельная теорема.

Под этим названием в теории вероятностей существует ряд предельных теорем.

Смысл их в том, что при суммировании большего числа независимых или слабо зависимых случайных величин закон распределения их суммы становится близким к нормальному при выполнении некоторых условий. Суть этих условий состоит в том, что если слагаемое не является нормальной величиной, его дисперсия не должна быть слишком большой сравнительно с другими.

Теорема Муавра-Лапласа - один из первых вариантов центральной предельной теоремы, относится к биномиальным величинам.

### *13.1. Теорема Муавра-Лапласа*

Пусть мы имеем дело со схемой Бернулли: p - вероятность успеха в единичном испытании, n - число таких испытаний, $X_n$ - число успехов в n единичных испытаниях -биномиальная величина с параметрами n и p>0.

Нормировкой получим $X_n^* = \dfrac{X_n - MX_n}{\sqrt{DX_n}} = \dfrac{X_n - np}{\sqrt{npq}}$.

Тогда при неограниченном увеличении n функция распределения величины X* неограниченно приближается к функции распределения стандартного нормального закона N(0,1) и, следовательно, $\lim P\{a<X^*_n<b\}=\Phi(b)-\Phi(a)$.

На практике это позволяет при больших значениях n вычислять вероятность для биномиальной величины.

Так, пусть при большом n надо вычислить для числа успехов $X_n$ вероятность попадания в отрезок $[m_1,m_2]$, где $m_1,m_2$ - заданные натуральные числа.

$$P\{m_1 \le X_n \le m_2\} \approx P\left\{\frac{m_1 - np}{\sqrt{npq}} \le \frac{X_n - np}{\sqrt{npq}} \le \frac{m_2 - np}{\sqrt{npq}}\right\} = P\{a \le X^* \le b\} =$$

$$= \Phi(b) - \Phi(b)$$

где $a=\frac{m_1 - np}{\sqrt{npq}}$, $b=\frac{m_2 - np}{\sqrt{npq}}$.

Это соотношение строго верно в пределе. При конечном n лучше ввести поправку $a=\frac{m_1 - np - \frac{1}{2}}{\sqrt{npq}}$, $b=\frac{m_2 - np + \frac{1}{2}}{\sqrt{npq}}$.

**Пример 15.** В автопарке 100 автобусов. Из опыта известно, что вероятность отказа в течении суток мотора автобуса равна 0,1. Найти вероятность того, что за сутки неисправных моторов будет больше 5.

Решение. n=100, p=0,1 - вероятность отказа мотора, np=10, X - число автобусов с отказавшим за день мотором.

1-ый путь строгий, но утомительный:

$P\{X>5\}=P_{6,100}+P_{7,100}+...+P_{100,100}$

где $P_{m,100} = C_{100}^m p^m (1-p)^{100-m}$ - совершенно не практичный путь, неприемлемо.

Лучше $P\{X>5\}=1-P\{X\le5\}=1-P_{0,100}-P_{1,100}-P_{2,100}-...-P_{5,100}$ - лучше, но тяжело.

2-ой путь. С помощью теоремы Муавра-Лапласа

$$P\{X>5\}=P\{6\le X\le 100\}=\Phi\left(\frac{100-np+\frac{1}{2}}{\sqrt{npq}}\right)-\Phi\left(\frac{6-np-\frac{1}{2}}{\sqrt{npq}}\right)=$$

$$=\Phi\left(\frac{100{,}5-10}{\sqrt{9}}\right)-\Phi\left(\frac{5{,}5-10}{\sqrt{9}}\right)=\Phi(30)-\Phi(-1{,}5)=1-0{,}1=0{,}9$$

- в среднем в 90% всех дней года потребуется ремонт больше пяти моторов из 100 имеющихся. •

**Пример 16.** Работница обслуживает больше 1000 веретен. Вероятность обрыва нити на веретене за смену равна 0,005. Найти вероятность того, что число X обрывов не превысит 2.

Решение.

n=1000, p=0,005, q=1-p=0,995, np=5, $npq=2{,}23$

*1-й способ.*

$$P\{X\le 2\}=P_{0,1000}+P_{1,1000}+P_{2,1000}=0{,}995^{1000}+1000*0{,}095^{999}*0{,}005+$$

$$+\frac{1000*999}{2}0{,}995^{998}*0{,}005^{2}=0{,}12$$

- строгое, но утомительное решение.

*2-й способ.*

Формула Муавра-Лапласа.

$$P(X \le 2) = P(0 \le X \le 2) = \Phi\left(\frac{2 - np + 0,5}{\sqrt{npq}}\right) - \Phi\left(\frac{2 - np - 0,5}{\sqrt{npq}}\right) =$$

$$\Phi(\frac{-2,5}{2,23}) - \Phi\left(\frac{-5,5}{2,23}\right) = \Phi(-1,12) - \Phi(-2,47) = 0,13 - 0,007 = 0,124$$

.

*3-й способ.*

Биномиальное распределение при большом n и малом p хорошо приближается пуассоновским распределением с $\lambda$=np=5.

$$P\{X \le 2\} = P_0 + P_1 + P_2 = e^{-5} + \frac{\lambda}{1!}e^{-5} + \frac{\lambda^2 e^{-5}}{2!} = e^{-5}\left(1 + 5 + \frac{25}{2}\right) = \frac{18,5}{148} = 0,125$$

Как видим, все три способа дают близкие результаты.

## Задачи для домашних заданий.

### *Вариант 1.*

**Задача 1.** Вероятность попадания в мишень P=0,8. Стрелок имеет 4 патрона и ведет стрельбу до первого попадания в мишень. Написать закон распределения случайной величины X, равной числу израсходованных патронов.
Вычислить: 1) M(x); 2) Д(x); 3) $P(-2 \le X \le 1{,}3)$; 4) $P(3 \le X \le 8)$. Построить функцию распределения F(x).

**Задача 2.** В коробке находится 6 красных и 4 черных шара. Достают из коробки первый раз 3 шара, а затем еще 2. Написать закон распределения системы случайных величин (X,Y), где X - количество черных шаров среди трех, а Y - количество черных среди двух шаров, вынутых из коробки.

**Задача 3.**

| X | -1 | 2 | 4 |
|---|---|---|---|
| вер. | 0.2 | ? | 0.4 |

| Y | 0 | 1 | 2 |
|---|---|---|---|
| вер. | ? | 0.5 | 0.1 |

Написать закон распределения для Z=2X+Y, вычислить M(Z), Д(Z). Чему равны $M(X^2)$ и Д(X-Y)?

**Задача 4.** Плотность вероятности задана в виде:

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

Найти: 1) А, 2) М(х), 3) Д(х), 4) $P(0 \le X \le 2M)$, 5) вычислить вероятность $P_4$(более двух раз $x \in [M, M+Д]$) при четырех испытаниях.

**Задача 5.** Случайная величина X распределена по нормальному закону N (1;3). Вычислить вероятности случайных событий: 1) X более 6, 2) произвели 2 испытания, в первом $X \in [M, Д\ \ ]$, а во втором $X \in [0,2]$, 3) произвели 10 испытаний, хотя бы 2 раза $X \in [M, M+ Д\ ]$.

## *Вариант 2.*

**Задача 1.** Бросаем игральную кость до первого появления шестерки. Случайная величина X равна количеству бросков кости. Найти закон распределения случайной величины X, вычислить M(X), если произвели не более 3 бросков; вычислить Д(X), функцию распределения F(x) при тех же условиях; вычислить вероятности событий
$X \in [-2,1,3]$и $X \in [3,8]$.

**Задача 2.** В коробке лежат 5 белых и 3 желтых конверта. Вынимаем 3 конверта и X- количество желтых среди них, затем вынимаем два конверта и У - количество желтых среди двух вынутых. Написать закон распределения системы случайных величин (X,У).

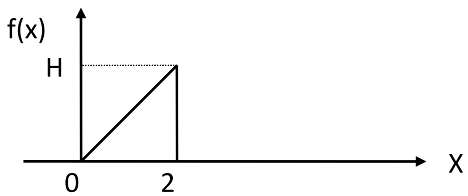
**Задача 3.**

| X | 2 | 0 | 4 |
|---|---|---|---|
| вер. | 0.7 | ? | 0.1 |

| Y | -2 | 0 | 1 |
|---|---|---|---|
| вер. | 0.6 | 0.2 | ? |

Написать закон распределения Z=X+2Y и вычислить M(3X+6-Y) и Д(3Y+4), M(Z), Д(Z), $M(X^2)$.

**Задача 4.** Плотность вероятности задана в виде графика:




Определить H; вероятность события $X \in [M, Д+M]$; вероятность того, что при 4 испытаниях более двух раз X попала в интервал [M, Д+M].

**Задача 5.** Случайная величина распределена по нормальному закону N(0;4). Вычислить: 1) вероятность того, что
$X \in [-6;1]$; 2) вероятность того, что при 11 испытаниях 5 раз X попадет в интервал [M,Д]; 3) вероятность того, что в первом испытании $X \in [-6;1]$, а во втором $X \in [0;2]$.

**Задача 6.** Вычислить вероятность того, что из четырех испытаний хотя бы один раз X попадет в интервал [-1;6], если распределено по равномерному закону R[0;10].

## 14. Математическая статистика. Введение.

При исследовании экономических процессов используются методы и модели математической статистики, позволяющие получить необходимые знания об исследуемом объекте.

Одним из основных методов математической статистики является выборочный метод. При этом используются понятия генеральной совокупности и выборочной совокупности.

Генеральная совокупность – это случайная величина $X$, распределенная по неизвестному закону и имеющая такое достаточно большое количество элементов, что сплошное обследование трудоемко или невозможно.

Выборочная совокупность или просто выборка – это совокупность случайно отобранных объектов из генеральной совокупности. К выборочной совокупности предъявляются определенные математические требования: выборка должна обладать всеми свойствами генеральной совокупности, т. е. должна быть представительной (репрезентативной); на основании закона больших чисел теории вероятностей отбор объектов для исследования должен носить случайный характер; каждый объект генеральной совокупности должен иметь одну и ту же вероятность того, что он попадет в выборку.

Таким образом, задача математической статистики состоит в следующем:

1) Осуществление математически правильного отбора данных для обследования;
2) Получение оценки функции распределения вероятностей и ее параметров;
3) Проверка статистических гипотез о виде законов распределения и величине параметров распределения, вид которых известен.

По результатам статистического исследования получают научные и практические выводы

## 15. Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка. При этом $x_1$ наблюдалось $n_1$ раз, $x_2$ наблюдалось $n_2$ раз, … , $x_к$ наблюдалось $n_к$ раз. Сумма $\sum_{i=1}^{k} n_i = n$ называется объемом выборки. Наблюдаемые значения называют вариантами. Последовательность вариант, записанных в возрастающем порядке, называют вариационным рядом.

Статистическим распределением выборки называют таблицу частот:

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | … | $x_к$ | варианты |
| $n_i$ | $n_1$ | $n_2$ | … | $n_к$ | частоты |

$$\sum_{i=1}^{k} n_i = n$$

или относительных частот:

| $x_i$ | $x_1$ | $x_2$ | … | $x_к$ | варианты |
|---|---|---|---|---|---|
| $\overline{n}$ | $\overline{n}$ | $\overline{n}$ | … | $\overline{n}$ | относительные частоты |

$$\sum_{i=1}^{k}\frac{n_i}{n}=1$$

Если объем выборки достаточно большой, то весь массив значений разбивается на $k$ интервалов и подсчитывается число элементов выборки, попавших в соответствующий интервал. Элемент, попадающий на границу двух интервалов, учитывается в левом интервале.

Статистическое распределение в этом случае имеет вид:

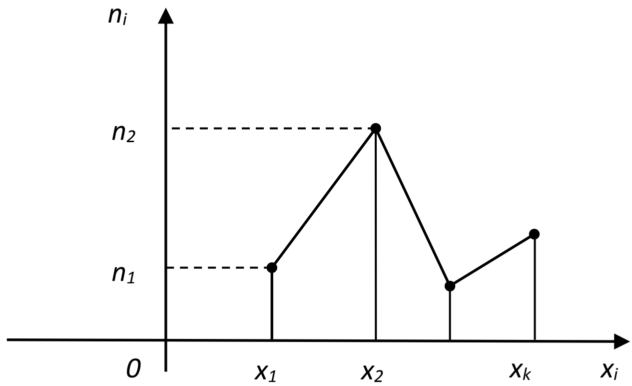
| $(x_{i-1};x_i)$ | $(x_0;x_1)$ | $(x_1;x_2)$ | … | $(x_{к-1};x_к)$ | варианты |
|---|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | … | $n_к$ | частоты |
| $\frac{n_i}{n}$ | $\frac{n_1}{n}$ | $\frac{n_2}{n}$ | … | $\frac{n_k}{n}$ | относительные частоты |

## 16. Полигон и гистограмма

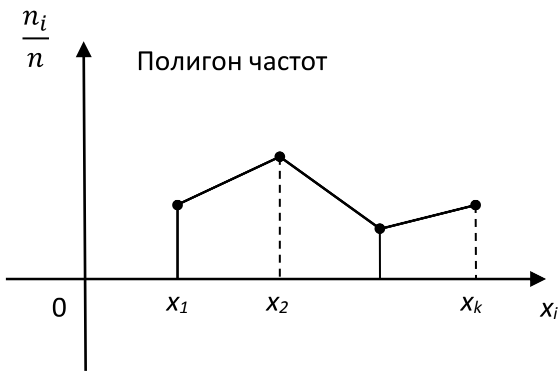
Для наглядности строят различные графики статистических распределений.

Полигоном частот называют ломаную, соединяющую точки $(x_i;n_i)$. По оси абсцисс откладывают $x_i$, а по оси ординат $n_i$.

Полигоном относительных частот называют ломаную, соединяющую точки $(x_i; \frac{n_i}{n})$.

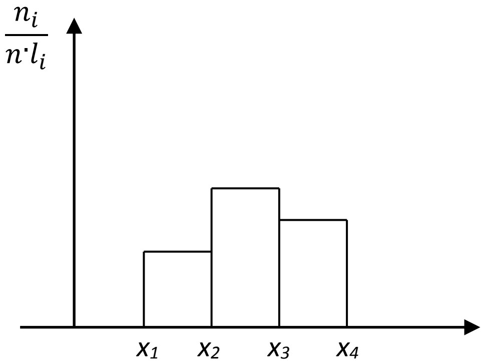



Полигон частот




Полигон относительных частот

Гистограмма строится по данным интервального статистического распределения и представляет собой набор прямоугольников, основаниями которых являются интервалы, а высоты равны: $h_i = \frac{n_i}{n \cdot l_i}$, где $l_i$ -длина $i$-го интервала, то есть площадь $i$-го прямоугольника равна относительной частоте попадания элементов выборки в $i$-й интервал. Площадь всей гистограммы равна 1.




Таким образом, гистограмма относительных частот является аналогом плотности вероятностей.

Замечание. Возможно построение гистограммы частот. В этом случае площадь $i$-го прямоугольника равна частоте попадания элементов выборки в $i$-й интервал, а площадь всей гистограммы равна объему выборки $n$.

Гистограмма частот и гистограмма относительных частот отражают один и тот же характер плотности вероятностей генеральной совокупности. Разница только в выборе масштаба.

**Пример 1.** Дана выборка: 1, 3, 2, 1, 0, 4, 3, 1, 0, 2. Составить статистическое распределение (таблицу частот) и построить полигон.

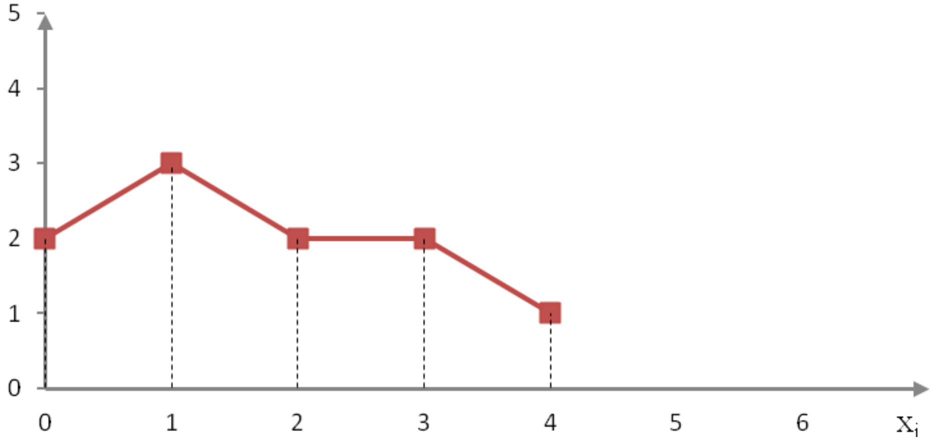
Решение.

Таблица частот:

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 0 | 1 | 2 | 3 | 4 |
| $n_i$ | 2 | 3 | 2 | 2 | 1 |

$n$=10

Полигон частот:




Пример 2. Дана выборка: 13, 15, 20, 12, 11, 35, 17, 38, 23, 27, 24, 39, 22, 25, 36, 18, 16, 23, 14, 27, 24, 26, 18, 37, 23, 14, 39, 27, 13, 22, 33, 31, 18, 35, 24, 14, 23, 34, 37, 12, 11, 11, 23, 13, 34, 25, 15, 23, 13, 17.

Требуется составить интервальное статистическое распределение и построить гистограмму относительных частот, разбив выборку на 5 интервалов.

Решение. Выберем сначала длину интервала. Так как размах выборки $R=x_{наиб}-x_{наим}$=39-11=28, то длина интервала должна быть больше, чем $\frac{\ }{5}$ =5,6. Поэтому возьмем длину интервала $l$=6. Начало первого интервала нужно выбрать так, чтобы $x_1 \le x_{min}$. В примере возьмем

$x_1$=10<11. Записав интервалы, определим количество чисел, попавших в каждый интервал. Если число совпадает с границей интервала, то запишем его в левый интервал.

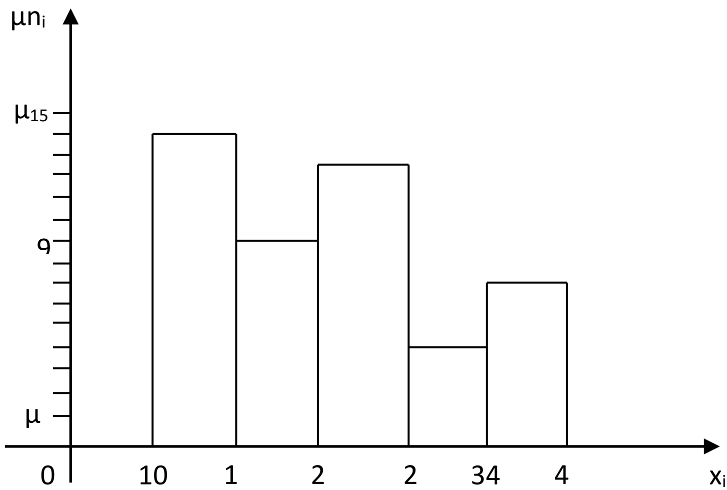
Таким образом, получим интервальное статистическое распределение:

| $(x_i\text{-}x_{i+1})$ | 10-16 | 16-22 | 22-28 | 28-34 | 34-40 | |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 9 | 14 | 4 | 8 | n=50 |
| $\frac{n_i}{n}$ | 0,30 | 0,18 | 0,28 | 0,08 | 0,16 | $\sum\frac{n_i}{n}=1$ |

Для построения гистограммы высоту *i*-го прямоугольника вычисляем по формуле:

$$h_i = \frac{n_i}{n \cdot l_i} = \frac{n_i}{50 \cdot 6} = \frac{1}{300} n_i = \mu n_i \,.$$

Число $\mu = \dfrac{1}{300}$ будем считать масштабной единицей на оси ординат.



На полученном рисунке гистограмма интервального статистического распределения данной выборочной совокупности.

## 17. Эмпирическая функция распределения

Напомним, что в теории вероятностей функцией распределения $F(x)$ называется $P(X<x)$ – вероятность того, что случайная величина $X$ (дискретная или непрерывная) расположена слева от фиксированного значения $x$, то есть $F(x)=P(X<x)$.

В математической статистике случайная величина X-это генеральная совокупность, которую мы изучаем по выборке. И по выборке строим аналогичную функцию:

$F^*(x)=\frac{\sum n_i}{n}$ для всех $x_i<x$, где $x$-фиксированное значение из выборки, в числителе $\sum n_i$ означает количество элементов выборки с учетом частоты по величине меньших выбранного элемента x.

Другими словами теоретическая функция $F(x)$ определяет вероятность события $X<x$, а эмпирическая $F^*(x)$ определяет относительную частоту этого же события.

Доказано (см. т. Бернулли в законе больших чисел), что относительная частота и вероятность события при больших $n$ мало отличаются одно от другого, то есть выполняется:

$$\lim_{n\to\infty} P(|F(x) - F^*(x)| < \varepsilon) = 1,\ \varepsilon>0.$$

Из определения функции $F^*(x)$ вытекают ее свойства, аналогичные свойствам $F(x)$:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x)$-неубывающая функция;
3. Если $x_1$ наименьшая варианта, то $F^*(x)=0$ для всех $x \le x_1$;
4. Если $x_к$ наибольшая варианта, то $F^*(x)=1$ для всех $x > x_к$.

Таким образом, эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

**Пример.** Составить эмпирическую функцию распределения для выборки из предыдущего параграфа:

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 0 | 1 | 2 | 3 | 4 |
| $n_i$ | 2 | 3 | 2 | 2 | 1 |

$n$=10

Решение. Из определения $F^*$следует:

$F^*(0)=0,\ x \le 0$;

$F^*(1)=\frac{2}{10}=0{,}2$;

$F^*(2)=\frac{2+3}{10}=0{,}5$;

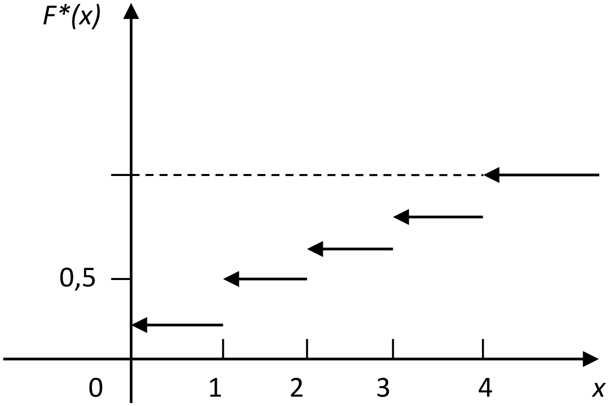
$F^*(3)=\frac{2+3+2}{10}=0{,}7$;

$F^*(4)=\frac{2+3+2+2}{10}=0{,}9$;

$F^*(x)=1$, x>4.

Таким образом,

График $F^*(x)$




$$F^*(x)=\begin{cases}0, x\le 0,\\ 0{,}2, 0<x\le 1,\\ 0{,}5, 1<x\le 2,\\ 0{,}7, 2<x\le 3,\\ 0{,}9, 3<x\le 4,\\ 1, x>4,\end{cases}$$

## 18. Статистические оценки параметров распределения

Параметры генеральной совокупности это $\bar{x}_г$ - средняя генеральная величина, $D_г$-дисперсия генеральная и $\sigma_г= \sqrt{D_г}$ -среднее квадратическое отклонение. Эти параметры неизвестны.

По случайной выборочной совокупности находят среднюю выборочную $\bar{x}_в$, дисперсию выборочную $D_в$ и среднее выборочное отклонение $\sigma_в = \sqrt{D_в}$ . Выборочные параметры являются случайными числами, их называют оценками неизвестных параметров генеральной совокупности.

К оценкам предъявляются определенные статистические требования: несмещенность, состоятельность и эффективность.

Введем обозначения: пусть $\theta$-неизвестный параметр генеральной совокупности, $\theta^*$-оценка этого параметра, вычисленная по выборочной совокупности.

1. Несмещенность. Оценка $\theta^*$ называется несмещенной, если $M\theta^*=\theta$, то есть математическое ожидание несмещенной оценки равно ее истинному значению.
2. Состоятельность. Оценка $\theta^*$ называется состоятельной, если с ростом объема выборки значение $\theta^* \to \theta$ по вероятности, то есть выполняется:

   $lim_{n\to\infty} P(|\theta^* - \theta| < \varepsilon) = 1$, $\varepsilon>0$.

   При этом достаточным условием состоятельности является стремление дисперсии $D\theta^* \to 0$ при $n \to \infty$.
3. Эффективность. Оценка $\theta^*$ называется эффективной, если дисперсия $D\theta^*$ принимает минимальное значение.

Говорят $\theta^*_1$ эффективнее $\theta^*_2$, если $D\theta^*_1 < D\theta^*_2$.

## 19. Оценка генеральной средней по выборочной средней

Пусть для изучения генеральной совокупности $X$ извлечена выборка объема $n$. Различные значения $x_1$, $x_2$, … $x_k$ и соответствующие частоты $n_1$, $n_2$, … $n_k$ помещены в таблицу:

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | … | $x_к$ | варианты |
| $n_i$ | $n_1$ | $n_2$ | … | $n_к$ | частоты |

$\sum n_i = n$ – объем выборки

Выборочная средняя $\overline{x}_в = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{n} = \frac{1}{n}\sum x_i n_i$.

Так как при повторении опыта получается другая выборка и другое значение $\overline{x}_в$, то можно говорить о математическом ожидании средней выборочной $M\overline{x}_в$ и дисперсии $D\overline{x}_в$.

Покажем, что $\overline{x}_в$ является несмещенной оценкой генеральной средней $\overline{x}_г$. Заметим, что $\overline{x}_г$ совпадает с математическим ожиданием генеральной совокупности $X$, то есть

$\overline{x}_г = MX$.

Выборки, полученные в результате независимых испытаний, можно считать случайными величинами $X_1$, $X_2, \ldots, X_n$, имеющими одинаковое распределение, совпадающее с распределением генеральной совокупности

$X$, то есть числовые характеристики случайных величин $X_i$ $(i=1, 2, 3,\ldots,n)$ одинаковые:

$MX_1= MX_2=\ldots=MX_n=a=MX=\bar{x}_г$,

$DX_1= DX_2=\ldots=DX_n=DX$.

Рассмотрим математическое ожидание средней выборочной:

$M\bar{x}_в=M\frac{X_1+X_2+\cdots+X_n}{n}=\frac{1}{n}(MX_1+MX_2+\cdots+MX_n)=\frac{na}{n}=a=MX.$

Поэтому имеем $M\,\bar{x}_в=\bar{x}_г$, следовательно, $\bar{x}_в$ является несмещенной оценкой генеральной средней.

Состоятельность средней выборочной $\bar{x}_в$ доказывает теорема Чебышева из закона больших чисел:

$$\lim_{n\to\infty}P(|\bar{x}_в-\bar{x}_г|<\varepsilon)=1,\ \varepsilon>0.$$

То есть $\bar{x}_в\to\bar{x}_г$ по вероятности при $n\to\infty$.

Таким образом, средняя выборочная является несмещенной состоятельной и, можно доказать, эффективной оценкой генеральной средней.

## 20. Среднее квадратическое отклонение

Так же, как в предыдущем параграфе 7, будем обозначать генеральную совокупность через дискретную случайную величину $X$ достаточно большого объема. Тогда генеральная средняя $\bar{x}_г=MX$ и $D_г=DX$ и по определению:

$$DX=M(X-MX)^2=MX^2-(MX)^2.$$

Эта формула выведена при изучении случайных величин (см. раздел «Теория вероятностей»). Выборочная дисперсия находится аналогично по выборочной совокупности:

$$D^*=\frac{1}{n}\sum(x_i-\bar{x}_в)^2 n_i,$$

среднее выборочное отклонение $\sigma^*=\sqrt{D^*}$.

Обратим внимание на то, что сумма произведений отклонений $x_i-\bar{x}$ на соответствующие частоты $n_i$ равна нулю.

Действительно:

$$\sum(x_i-\bar{x})n_i=\sum x_i n_i-\sum\bar{x}\,n_i=n\bar{x}-\bar{x}\sum n_i=n\bar{x}-n\bar{x}=0.$$

Пример. Имеется выборка:

| | | | | |
|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | |
| $n_i$ | 2 | 3 | 5 | $n$=10 |

Вычислить:

1) Среднюю величину отклонения;
2) Среднее квадратическое отклонение.

Решение:

$$\bar{x}=\frac{1\cdot2+2\cdot3+3\cdot5}{10}=2{,}3.$$

1. $$\frac{1}{n}\sum\left(x_i-\bar{x}\right)n_i=\frac{1}{10}\left[(1-2{,}3)\cdot2+(2-2{,}3)\cdot3+(3-2{,}3)\cdot5\right]=$$

$$=\frac{1}{10}(-1{,}3\cdot2-0{,}3\cdot3+0{,}7\cdot5)=\frac{1}{10}(-2{,}6-0{,}9+3{,}5)=$$

$$=\frac{1}{10}(-3{,}5+3{,}5)=0$$

2. $D^* = \frac{1}{10}\left[(1-2{,}3)^2 \cdot 2 + (2-2{,}3)^2 \cdot 3 + (3-2{,}3)^2 \cdot 5\right] =$

$$= \frac{1}{10}\left[(-1{,}3)^2 \cdot 2 + (-0{,}3)^2 \cdot 3 + (0{,}7)^2 \cdot 5\right] =$$

$$\frac{1}{10}(1{,}69 \cdot 2 + 0{,}09 \cdot 3 + 0{,}49 \cdot 5) = \frac{1}{10}(3{,}38 + 0{,}27 + 2{,}45) =$$

$$= \frac{6{,}1}{10} = 0{,}61$$

$$\sigma^* = \sqrt{0{,}61} = 0{,}78.$$

Ответ: 1. Среднее отклонение равно нулю.

2. Среднее квадратическое отклонение равно 0,78.

Из примеров видно, что вычисление дисперсии по определению достаточно громоздко. Выведем формулу, которой обычно пользуются на практике:

$$D^* = \frac{1}{n}\sum n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum n_i\left(x_i{}^2 - 2x_i\overline{x} + \left(\overline{x}\right)^2\right) =$$

$$= \frac{1}{n}\left(\sum n_i x_i{}^2 - 2\sum n_i x_i \overline{x} + \sum n_i\left(\overline{x}\right)^2\right) =$$

$$\left(\sum\frac{n_i x_i{}^2}{n} - 2\overline{x}\sum\frac{n_i x_i}{n} + \left(\overline{x}\right)^2\sum\frac{n_i}{n}\right) = \overline{x^2} - 2\left(\overline{x}\right)^2 + \left(\overline{x}\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2,$$

то есть $D^* = \overline{x^2} - \left(\overline{x}\right)^2$

Выборочная дисперсия равна усредненному квадрату выборочных значений минус средняя выборочная величина в квадрате.

Пример. Возьмем выборку из предыдущего примера:

| | | | | |
|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | |
| $n_i$ | 2 | 3 | 5 | $n$=10 |

Средняя величина $\bar{x}$=2,3.

Средний квадрат выборочных значений:

$\overline{x^2} = \frac{1}{10}(1^2 \cdot 2 + 2^2 \cdot 3 + 3^2 \cdot 5) = \frac{1}{10}(2 + 12 + 45) = 5{,}9$;

$D^*$=5,9-(2,3)$^2$=5,9-5,29=0,61.

## 21. Оценка генеральной дисперсии по исправленной выборочной

Покажем, что выборочная дисперсия $D^*$ является смещенной оценкой генеральной дисперсии $DX$. Так же, как и в пункте 6, будем считать выборочные совокупности одинаково распределенными случайными величинами, имеющими одинаковые числовые характеристики.

Выведено в 7, что для средней $\bar{X}$ одинаково распределенных величин выполняется:

$$M\bar{X}=MX,$$

$$M\overline{X^2}=MX^2,$$

$$D\bar{X} = \frac{DX}{n}. \qquad (1)$$

Рассмотрим

$$M(D^*)=M(\overline{X^2} - (\bar{X})^2)=M(\overline{X^2})-M(\bar{X})^2; \qquad (2)$$

По определению D($\bar{X}$)= $M(\bar{X})^2$-$(M\bar{X})^2$, поэтому получим равенство:

$$\frac{DX}{n} = M(\bar{X})^2 - (M\bar{X})^2 \qquad (3)$$

Из (3) следует, учитывая (1)

$$M(\bar{X})^2 = \frac{DX}{n} + (MX)^2 \qquad (4)$$

Подставим теперь в (4) в (2):

$$M(D^*) = M(X^2) - \frac{DX}{n} - (MX)^2 = DX - \frac{DX}{n} = \frac{n-1}{n} DX.$$

Таким образом, $M(D^*) \neq DX$ и, следовательно, дисперсия выборочная является смещенной оценкой генеральной дисперсии. Однако можно заметить, что $\frac{n-1}{n} \to 1$ при $n \to \infty$, то есть для больших объемов выборки $D^*$ фактически будет несмещенной.

При $n \leq 30$ используют исправленную дисперсию:

$$S^2 = \frac{n}{n-1} D^*,$$

которая является несмещенной оценкой $DX$.

Действительно:

$$M(S^2) = M\left(\frac{n}{n-1} D^*\right) = \frac{n}{n-1} MD^* = \frac{n}{n-1} \cdot \frac{n-1}{n} DX = DX.$$

Вычисляется $S^2$ по формуле:

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 n_i.$$

$S = \sqrt{S^2}$ называется стандартным отклонением.

## 36. Статистические (точечные) оценки параметров основных законов распределения

1. Нормальное распределение: $X=N(m,\sigma)$. Теоретические параметры: $m=MX$, $\sigma^2=DX$. Оценки параметров $m^*=\bar{X}$ -средняя арифметическая величина, определяемая по выборке и $S^2=D^*$-исправленная выборочная дисперсия.

Асимметрия $A_s^* = \frac{\overline{(X-\bar{X})^3}}{(\sigma^*)^3}$ и эксцесс $E_x^* = \frac{\overline{(X-\bar{X})^4}}{(\sigma^*)^4} - 3$.

Оценка асимметрии закона вычисляется по формуле:

$A_s^* = \frac{\overline{(X-\bar{X})^3}}{(\sigma^*)^3} = \frac{1}{(\sigma^*)^3}\left(\overline{x^3} - 3\overline{x^2}\cdot\bar{x} + 2(\bar{x})^3\right)$.

Точечная оценка коэффициента эксцесса находится по формуле:

$E_x^* = \frac{1}{(\sigma^*)^4}\overline{(x-\bar{x})^4} - 3$.

2. Показательное распределение: $X=E(\lambda)$

Параметр закона λ, $M(X)=\frac{1}{\lambda}$; $D(X)=\frac{1}{\lambda^2}$.

Оценка параметра $\lambda^*=\frac{1}{\bar{X}}$; $\bar{X}$-средняя выборочная.

3. Равномерное распределение: $X=R[a,b]$

Теоретические параметры, $MX=\frac{a+b}{2}$, $DX=\frac{(b-a)^2}{12}$.

Оценки параметров:

средняя выборочная $\bar{X} = \frac{a^*+b^*}{2}$ и исправленная выборочная дисперсия $S^2=\frac{(b^*-a^*)^2}{12}$.

## 22. Метод наибольшего правдоподобия для точечной оценки параметров распределения

Пусть $X$-дискретная случайная величина и выборка из нее: $x_1$, $x_2,\dots,x_n$. Допустим, что предполагаемый закон распределения вероятностей выбран и требуется найти точечную оценку параметра $\theta$ закона. Обозначим через $p(x_i;\theta)$ вероятность того, что величина X примет значение $x_i$ $(i=1, 2,\dots,n)$

Функцией правдоподобия дискретной случайной величины называется функция:

$L(x_1, x_2,\dots,x_n, \theta)=p(x_1,\theta)\, p(x_2,\theta)\dots p(x_n,\theta)\dots$

Находят такое значение $\theta^*$, при котором функция достигает максимума и $\theta^*$ называют оценкой наибольшего правдоподобия.

Замечено, что функции $L$ и $\ln L$ достигают максимума при одном и том же значении θ, поэтому исследуем на экстремум логарифмическую функцию правдоподобия:

$\ln L=\ln p(x_1,\theta)+\ln p(x_2,\theta)+\dots+\ln p(x_n,\theta)$ .

Для этого по известному правилу математического анализа нужно:

1. Найти производную $\frac{d(\ln L)}{d\theta}$;
2. Приравнять производную к нулю и, решая уравнение $\frac{d(\ln L)}{d\theta}=0$, вычислить $\theta^*$;
3. Найти производную второго порядка $\frac{d^2(\ln\theta)}{d\theta^2}$ ; подставив в нее значение $\theta=\theta^*$, убедиться в том, что вторая производная отрицательна, и, следовательно, $\theta^*$-точка максимума, которую принимают в качестве оценки наибольшего правдоподобия параметра $\theta$.

Метод наиболее применим для малых выборок. Недостаток метода состоит в том, что приводит к сложным вычислениям.

Пример. Найти методом наибольшего правдоподобия оценку параметра $\lambda$ в распределении вероятностей Пуассона:

$$P_n(m)=\frac{\lambda^m}{m!}e^{-\lambda}$$

Решение. Пусть проводится $k$ опытов, каждый из которых состоит из $n$ испытаний. В каждом опыте фиксируется число появлений рассматриваемого события. Обозначим число появлений события в $i$-м опыте через $m_i$ и подставим в формулу Пуассона:

$$P_n(m_i)=\frac{\lambda^{m_i}}{m_i!}e^{-\lambda},\ i=1,\ 2,\ldots,k.$$

Составим теперь функцию наибольшего правдоподобия:

$$L=P_n(m_1,\lambda)\,P_n(m_2,\lambda)\ \ldots\ P_n(m_k,\lambda)=$$

$$=\frac{\lambda^{m_1}}{m_1!}e^{-\lambda}\frac{\lambda^{m_2}}{m_2!}e^{-\lambda}\ldots\frac{\lambda^{m_k}}{m_k!}e^{-\lambda}=\frac{\lambda^{(m_1+m_2+\ldots+m_k)}}{m_1!m_2!\ldots m_k!}e^{-k\lambda}=\frac{\lambda^{\sum_{i=1}^{k}m_i}}{m_1!m_2!\ldots m_k!}e^{-k\lambda}$$

.

Запишем логарифмическую функцию наибольшего правдоподобия:

$$\ln L=\left(\sum_{i=1}^{k}m_i\right)\ln\lambda-\ln(m_1!m_2!\ldots m_k!)-k\lambda.$$

Исследуем функцию на экстремум.

1) Возьмем производную по $\lambda$ и приравняем ее к нулю:
$$\frac{d\ln L}{d\lambda}=\frac{1}{\lambda}\sum_{i=1}^{k}m_i-k=0.$$

Получим $\lambda^* = \frac{1}{k}\sum_{i=1}^{k} m_i = \overline{m_в}$, то есть $\lambda^*$ равна среднему выборочному значению числа появлений события в $k$ опытах.

2) Возьмем производную второго порядка:
$\frac{d^2 \ln L}{d\lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{k} m_i$.

Очевидно, что она отрицательна, следовательно, $\lambda^* = \overline{m_в}$ - точка максимума. Таким образом, в качестве оценки наибольшего правдоподобия параметра $\lambda$ распределения Пуассона принимается выборочная средняя числа появлений события в k опытах.

## 23. Число степеней свободы

Статистические оценки параметров предполагаемого теоретического распределения вычисляются по случайным выборкам и поэтому являются случайными числами. Они подвергаются дальнейшему анализу с помощью различных критериев согласия: например, хи-квадрат *($\chi^2$)*, Фишера-Снедекора, Стьюдента и другие.

Для применения этих стандартных законов распределения вводится понятие числа степеней свободы изучаемого статистического материала.

Различные независимые измерения $X_1$, $X_2, \ldots, X_k$ одной и той же величины можно рассматривать как различные случайные величины, имеющие одно и то же распределение и имеющие $k$ степеней свободы. Сумма $X_1+X_2+\ldots+X_k$ и сумма квадратов также будет иметь число степеней свободы, равное $k$.

Если для рассмотренных случайных величин используется какая-либо связь, то число степеней свободы уменьшается. Например, если по результатам выборки вычислена средняя величина:

$\overline{x_{в}} = \frac{x_1+x_2+\cdots+x_k}{k}$,

и зафиксирована для всех $X_1, X_2,\ldots,X_k$, то есть $\overline{x_{в}} = \frac{X_1+X_2+\cdots+X_k}{k}$, то тогда одну из величин можно выразить через остальные. Значит, система оказалась связанной и потеряла одну степень свободы.

Если по результатам выборки вычислена выборочная дисперсия и присвоена любой из случайных величин $X_i$ $(i=1, 2,\ldots,k)$, то число степеней свободы станет равно $k$-2.

В некоторых статистических исследованиях фиксируются не только общая средняя выборочная и дисперсия выборочная, но и групповые средние и групповые дисперсии. Тогда число степеней свободы у системы случайных величин уменьшается еще и на число таких связей.

## 24. Основные стандартные законы распределения, их таблицы и применение этих таблиц

Для практического применения статистических расчетов служат таблицы основных стандартных законов распределения: нормального, $\chi^2$, Стьюдента и Фишера.

Введем понятие квантиля и критической точки. Пусть $X$-СВ, имеющая стандартное распределение. $P$-Квантилем распределения $X$ называется такое значение $x_p$, что $F(x_p)=p$, где $F(x)$-функция распределения СВ $X$. Часто вместо квантиля используется критическая точка. Критической точкой уровня $\alpha$ распределения $X$ называется такое значение $x_{кр(\alpha)}$, что $P(X>x_{кр(\alpha)})=\alpha$. Ясно, что квантиль уровня $p$ равен критической точке уровня $1-p$. Если распределение $X$ симметрично относительно оси ординат, то $x_\alpha=-x_{1-\alpha}$(см. рис.1)

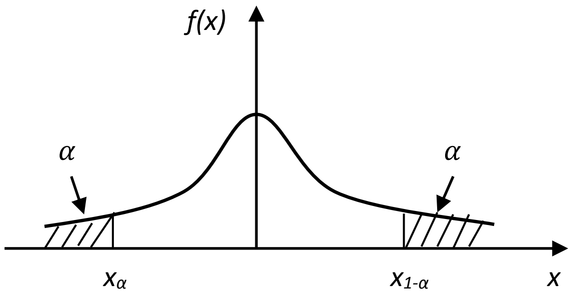



Рис. 1

### 24.1. Работа с таблицами стандартного нормального распределения

Стандартным нормальным распределением называется нормальное распределение $N[0,1]$ с параметрами $m=0$, $\sigma=1$, $p$-Квантиль нормального распределения обозначается $u_p$. Обычно в приложениях для отыскания $u_p$ используются таблицы значений функции Лапласа:

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

Функция Лапласа связана с функцией распределения стандартной нормальной СВ равенством: $F(u)=\Phi(u)+0,5$ (см. рис.2)

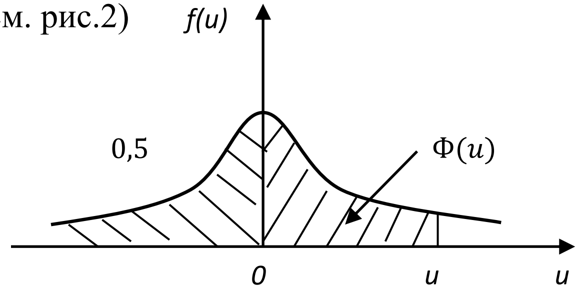



Рис. 2

Если $p>0,5$, то, используя таблицы функции Лапласа, находим квантиль $u_p$ из равенства $\Phi(u_p)=p-0,5$.

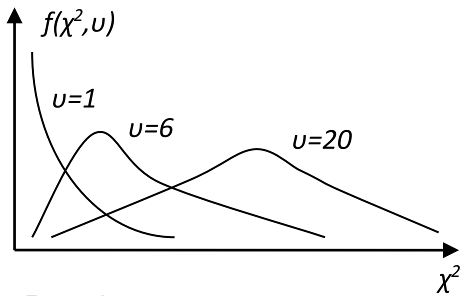
Например, найдем $u_{0,95}$. Из таблиц имеем $\Phi(1,64)=0,4495$, $\Phi(1,65)=0,4505$. Полагаем $\Phi(1,645)=0,45=0,95-0,5$, т. е. $u_{0,95}=1,645$. Если же $p<0,5$, то $1-p>0,5$. Находим $u_{1-p}$. Тогда $u_p=-u_{1-p}$.

Пример. Найти $u_{0,05}$. Тогда $1-0,05=0,95$; $u_{0,95}=1,645$. Значит, $u_{0,05}=-1,645$.

<u>Замечание.</u> Если в приложениях даются таблицы функции $F(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{t^2}{2}}dt$, то тогда сразу по заданному уровню $p$ находим $u_p$.

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

Пусть $U_1, U_2, \ldots, U_n$ – взаимно независимые случайные величины, имеющие стандартные нормальные распределения. Тогда $\chi^2 = \sum_{i=1}^{n} U_i^2$ имеет $\chi^2(n)$ распределение с $n$ степенями свободы. Отметим, что число степеней свободы $(\upsilon)$ определяется числом случайных величин, ее составляющих, уменьшенному на число линейных связей между ними. Случайная величина, имеющая $\chi^2$ распределение, может принимать только неотрицательные значения, ее график плотности имеет асимметрический вид с вытянутым правым «хвостом». Однако, с увеличением числа степеней свободы распределение $\chi^2(n)$ постепенно приближается к нормальному.




$M(\chi^2)=\upsilon$

$D(\chi^2)=2\upsilon$

Рис. 3

Распределение $\chi^2$ применяется для нахождения интервальных оценок и проверки статистических гипотез. В приложениях обычно для данного уровня $p$ и числа степеней свободы $\upsilon$ даются значения критических точек $\chi^2_{кр}(p,\upsilon)$.

Например, $\chi^2_{кр}(0{,}995,4)=0{,}207$, т. е. $P(\chi^2(4)>0{,}207)=0{,}995$.

### *24.2.1. Распределение Стьюдента*

Пусть СВ $U$ имеет стандартное нормальное распределение, а СВ $V$ не зависит от $U$ и распределена по закону $\chi^2$ с $n$ степенями свободы.

Тогда СВ $T=\frac{U}{\sqrt{V/n}}$ имеет распределение Стьюдента ($t$ распределение) с $n$ степенями свободы $T=St(n)$. Распределение Стьюдента имеет один параметр, график функции плотности симметричен относительно оси ординат.


$$M(T)=0;\ D(T)=\frac{n}{n-2}$$

Рис. 4

С увеличением числа степеней свободы распределения Стьюдента приближается к нормальному и при $n$>30 его практически можно заменить нормальным. Так же, как и распределение $\chi^2$ его применяют для построения интервальных оценок и проверки статистических гипотез.

При этом используют либо таблицы квантилей, либо критических точек. Поскольку распределение Стьюдента симметрично, то $t_{1-p}=-t_p$. Значения квантиля находятся на пересечении строки с числом степеней свободы и столбцавероятности. Например, $t_{0,95}(5)=2{,}015$, т. е. $P(St(5)<2{,}015)=0{,}95$. Это же значение $t=2{,}015$ является критической точкой для уровня 0,05 $t_{кр}(0{,}05;5)=2{,}015$, т. е. $P(St(5)>2{,}015)=0{,}05$.

### *24.2.2. Распределение Фишера-Снедекора*

Пусть $V=\chi^2(m)$, $W=\chi^2(n)$ независимые СВ, распределенные по закону $\chi^2$ со степенями свободы $\upsilon_1=m$ и $\upsilon_2=n$. Тогда СВ $F=\frac{V/m}{W/n}$ имеет распределение Фишера со степенями свободы $\upsilon_1=m$ и $\upsilon_2=n$, $F=F(m,n)$. Таким образом, распределение Фишера ($F$-распределение) определяется двумя параметрами, причем $M(F)=\frac{n}{n-2}$ ($n>2$); $D(F)=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $n>4$.

СВ, распределенное по закону Фишера, принимает только неотрицательные значения, применяется для проверки статистических гипотез в дисперсионном и регрессионном анализе. Таблицы квантилей (критических точек) должны иметь три входа, поэтому в приложениях приводятся таблицы для некоторых наиболее употребительных значений уровней значимости: $\alpha=0{,}01$; $\alpha=0{,}05$; $\alpha=0{,}1$. Для каждого из этих значений приводятся значения критических точек (или квантилей) для распределения Фишера с числом степеней свободы $\upsilon_1$ и $\upsilon_2$. Например, для $\alpha=0{,}05$ $F_{кр}(0{,}05;\ (4{,}2))=19{,}2$. Это значит, что $P(F(4{,}2)>19{,}2)=0{,}05$. Если же $\alpha=0{,}1$, то $F_{кр}(0{,}1;\ (4{,}2))=9{,}24$, т. е. $P(F(4{,}2)>9{,}24)=0{,}1$. В то же время эти критические точки являются квантилями уровней 0,95 и 0,9 соответственно. $F_{0,95}(4,2)=19,2$ и $F_{0,9}(4,2)=9,24$, т. е. $P(F(4,2)<19,2)=0,95$ и $P(F(4,2)<9,24)=0,9$.

График плотности распределения Фишера имеет вид:

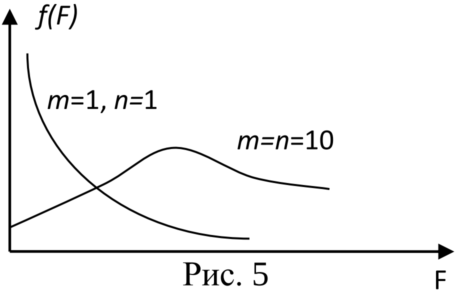



Рис. 5

## 25. Интервальные оценки

Точечная оценка $\theta^*$ параметра $\theta$ является функцией выборки и зависит в том числе от ее объема. Поскольку обычно оценки берутся несмещенные и состоятельные, то если объем выборки достаточно большой, оценка дает хорошее приближение оцениваемому параметру. Однако нередко, особенно, если невозможно в силу условий построить выборку большого объема, возникает необходимость задать некоторые границы, в которых заключается истинное значение параметра с некоторой вероятностью (надежностью).

Это приводит нас к задаче построения интервальных оценок или доверительных интервалов. Дадим точное определение.

Пусть $\theta$-неизветсный параметр и $\gamma$-вероятность, тогда интервал $[\theta_1(n);\theta_2(n)]$, здесь n – объем выборки, покрывающий параметр $\theta$ с вероятностью $\gamma$, называется доверительным интервалом, или интервальной оценкой параметра $\theta$, а $\gamma$-надежностью или доверительной вероятностью. Мы говорим «интервал покрывает параметр $\theta$», поскольку $\theta$-неслучайная величина, а случайными являются концы интервала. Очевидно, интервал зависит от оценки $\theta^*$ параметра. Иногда, но не всегда, он имеет вид $\theta^*$ . В этом случае $\Delta$-предельная абсолютная погрешность, а величина $\Delta$ или $2\Delta$ называется <u>точностью</u> оценки. Ясно, что при одном и том же объеме выборки, если мы хотим увеличить точность оценки (т. е. уменьшить длину интервала), то это приведет к уменьшению ее надежности и наоборот. Погрешность $\Delta$ (ошибка $\Delta$) возникает вследствие того, что вместо всей генеральной совокупности исследуется только ее часть – случайная выборка. Она называется ошибкой <u>репрезентативности</u>.

Алгоритм построения доверительного интервала:

1. Задается уровень значимости $\alpha=1-\gamma$, где $\gamma$-надежность оценки. Обычно для $\alpha$ используется значение 0,1; 0,05; 0,01; 0,001.
2. Пусть $\theta$-неизвестный параметр распределения СВ $X$. Обычно предполагается, что $X$ имеет нормальное

распределение. В силу Ц. П. Т. для больших объемов n выборки эти же формулы используются и для других распределений.

3. Строится СВ $Z$, являющаяся функцией выборки, зависящая от $\theta$ и ее точечной оценки $\theta^*$ и имеющая стандартное распределение. Она называется статистикой.
4. С помощью таблиц, имеющихся для этих стандартных распределений, находятся такие $z_1$ и $z_2$, что $P(z_1<Z<z_2)=\gamma=1-\alpha$.
5. Используя связь между $Z$ и $\theta$, строится равносильное неравенство $\theta_1<\theta<\theta_2$, так что $P(\theta_1<\theta<\theta_2)=\gamma$. Если распределение $Z$ симметрично, то интервал имеет вид $\theta^* \pm\Delta$, где $\Delta$-предельная абсолютная погрешность.

Схема построения доверительных интервалов

| Оцениваемый параметр | Статистика и ее распределение | Доверительный интервал |
|---|---|---|
| 1. Математическое ожидание нормальной СВ $X=N[m,\sigma]$ $MX=m$, $\sigma$ известно | $U=\dfrac{(\bar{x}-m)\sqrt{n}}{\sigma}=N[0,1]$ | $\Delta=u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ $P(\bar{x}-\Delta<m<\bar{x}+\Delta)=$ $=1-\alpha$ |
| 2. Математическое ожидание нормальной СВ $X=N[m,\sigma]$ $MX=m$, $\sigma$ неизвестно | $T=\dfrac{(\bar{x}-m)\sqrt{n}}{s}=St[n-1]$ $s$-исправленное среднее квадратическое | $\Delta=t_{1-\alpha/2}(n-1)\dfrac{s}{\sqrt{n}}$ $P(\bar{x}-\Delta<m<\bar{x}+\Delta)=$ $=1-\alpha$ |

| | | |
|---|---|---|
| | отклонение | |
| 3. Дисперсия нормальной СВ $X=N[m,\sigma]$ $DX=\sigma^2$, $m$ неизвестно | $\chi 2=\frac{(n-1)s^2}{\sigma^2}=\chi 2[n-1$ | $P\left(\frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}<\sigma^2<\frac{s^2(n-1)}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right)$ $=1-\alpha$ $P\left(s\sqrt{\frac{(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}<\sigma<s\sqrt{\frac{(n-1)}{\chi^2_{\frac{\alpha}{2}}(n-1)}}\right)$ $=1-\alpha$ |
| 4. Вероятность $p$ биномиального распределения (доля), $n$ велико | $U=\frac{(P-W)}{\sqrt{\frac{W(1-W)}{n}}}=N[0,1]$ | $\Delta=u_{1-\alpha/2}\sqrt{\frac{W(1-W)}{n}}$ $P(w-\Delta<p<w+\Delta)$ $=1-\alpha$ |

Здесь $u_p$, $t_p(k)$, $\chi^2_p(k)$ - квантили соответственно стандартного нормального распределения, распределения Стьюдента и $\chi^2$ с $k$ степенями свободы.

$W=\frac{m}{n}$ - относительная частота (доля).

**Задача 1.** В результате предварительного опроса больших групп населения по 50 избирательным округам установлено, что в среднем 35,1% готовы проголосовать за кандидата данной партии.

а) Считая, что СВ $X$-процент проголосовавших в каждом округе за данного кандидата имеет нормальное распределение со средним квадратическим отклонением $\sigma$=1,5%, указать с 90% уверенностью, в каких пределах будет заключаться процент проголосовавших на предстоящих выборах.

b) Решить ту же задачу, заменяя заданное $\sigma$ стандартным отклонением, полученным в данной выборке объема 50 и составляющим 2,1%.
c) На уровне значимости $\alpha$=0,1 проверить, не является ли стандартное отклонение в 1,5% заниженной оценкой?
Решение. a) $\gamma$=0,9, 1-$\alpha$/2=0,95; $u_{0,95}$=1,645, $\Delta=\frac{1{,}645\cdot 1{,}5}{\sqrt{50}}=0{,}35$.
Интервал 35,1±0,35. Итак, в 90% случаев можно быть уверенным, что процент проголосовавших за данного кандидата будет заключен в интервале (34,75%;35,45%).
b) Поскольку σ неизвестно, пользуемся распределением Стьюдента $t_{0,95}(49)$=1,675,
$\Delta=\frac{1{,}675\cdot 2{,}1}{\sqrt{50}}=0{,}5$. Интервал (34,6%;35,6%) оказался шире.
c) Построим доверительный интервал для $\sigma$.
$\chi^2_{0,95}(49)=67{,}5$, $\chi^2_{0,05}(49)=34{,}8$,
$s\sqrt{\frac{49}{67{,}5}}=2{,}1\cdot 0{,}85=1{,}79$, $s\sqrt{\frac{49}{34{,}8}}=2{,}1\cdot 1{,}9=2{,}49$,
$P(1{,}79<\sigma<2{,}49)=0{,}9$.

Итак, на уровне значимости $\alpha$=0,1 можно утверждать, что оценка стандартного отклонения в 1,5% занижена, т. к. 1,5 не принадлежит интервалу (1,79;2,49).

**Задача 2.** Из большой партии деталей для проверки было отобрано 100 деталей, среди которых оказалось 96 стандартных. Найти: *а)* вероятность того, что доля нестандартных деталей во всей партии отличается от полученной доли в выборке по абсолютной величине не более, чем на 0,01; *в)* границы, в которых с надежностью 0,95 заключена доля нестандартных деталей во всей партии; *c)* определить число деталей, которое надо отобрать в выборку, чтобы с вероятностью 0,9 доля нестандартных деталей в выборке отличалась от генеральной доли (по абсолютной величине) не более, чем на 0,02.

Решение. Поскольку n велико, статистика $\frac{(P-W)\sqrt{n}}{\sqrt{W(1-W)}}$ имеет нормальное стандартное распределение.

a) $W=\frac{100-96}{100}=0{,}04$; $\sigma_w=\sqrt{\frac{W(1-W)}{n}}=\sqrt{\frac{0{,}04\cdot 0{,}96}{100}}=\frac{0{,}196}{10}=0{,}0196$;

$P(|W-p|<0{,}01)=2\Phi(\frac{0{,}01}{0{,}0196})=2\Phi(0{,}51)=2\cdot 0{,}0196=0{,}39$.

b) $\Delta=u_{1-\alpha/2}\cdot\sqrt{\frac{W(1-W)}{n}}$; $\alpha/2=0{,}025$; $1-\alpha/2=0{,}975$; $U_{0{,}975}=1{,}96$

$\Delta=1{,}96\cdot 0{,}0196=0{,}038$; $p\in W\pm\Delta=0{,}04\pm 0{,}038$;

$P(0{,}002<p<0{,}078)=0{,}95$;

c) $P(|W-p|<0{,}02)=2\Phi\left(\frac{0{,}02}{\sigma_w'}\right)=0{,}9$; $\Phi(1{,}645)=0{,}45$;

$0{,}012=\frac{0{,}02}{1{,}645}=\sigma_w'$, т. к. $\frac{0{,}02}{\sigma_w'}=1{,}645$; откуда $\sqrt{\frac{0{,}04\cdot 0{,}96}{n}}=0{,}012$

Значит, $n>\frac{0{,}04\cdot 0{,}96}{0{,}012^2}=1361$.

Замечание. Для решения задачи *c)* мы предположили, что доля нестандартных деталей не изменилась. Иногда ее полагают равной средней доли данного производства. Если о доли ничего, даже приблизительно, неизвестно, то полагают: $W(1-W)=W(1-W)_{max}=0{,}25$, что увеличивает объем выборки.

**Вывод:** с надежностью 0,39 можно полагать, что доля нестандартных деталей всей партии лежит в границах

(3%,5%) и с надежностью 0,95 лежит в границах (0,2%,7,8%). Объем выборки должен быть не менее, чем 1361, чтобы с вероятностью 0,9 гарантировать отклонение доли нестандартных деталей от генеральной доли не более, чем на 0,02.

## 26. Проверка статистических гипотез

Статистической называют гипотезу о виде закона распределения или о параметре известного распределения. В первом случае гипотеза называется непараматрической, а во втором – параметрической.

Рассмотрим сначала случай параметрических гипотез. Пусть $\theta$-неизвестный параметр. Выдвигается гипотеза $H_0{:}\theta=\theta_0$. Эта гипотеза называется нулевой или основной. В качестве конкурирующих или альтернативных гипотез рассматриваются гипотезы $\theta\neq\theta_0$, $\theta>\theta_0$, $\theta<\theta_0$. С логической точки зрения альтернативной является только $\theta\neq\theta_0$. Но часто, исходя из смысла параметра $\theta$, оказывается, что одна из них, например, $\theta>\theta_0$ не имеет смысла, и тогда $\theta<\theta_0$ является альтернативной гипотезой. Заметим, что обычно в качестве нулевой гипотезы берется та, которую собираются отвергнуть. Причина в том, что для доказательства некоторого утверждения требуется рассмотреть все случаи, а для его отрицания достаточно одного опровергающего примера.

Алгоритм проверки параметрической гипотезы

1. Выбирается уровень значимости $\alpha$;

2. Выдвигаются основная $H_0$ гипотеза и альтернативная $H_1$;
3. Строится случайная величина $Z$, которая при условии, что гипотеза $H_0$ справедлива, имеет стандартное распределение;
4. Используя стандартное распределение и уровень значимости $\alpha$, область изменения СВ $Z$ разбивается на две области: критическую, вероятность попадания в которую равна $\alpha$, и область принятия гипотезы, вероятность попадания в которую СВ $Z$ равна $\gamma=1-\alpha$;
5. Вычисляется выборочное значение критерия $Z$. Если оно попадает в критическую область, то гипотеза $H_0$ отвергается. В противном случае говорят, что гипотеза не противоречит эксперименту и может быть принята.

Следующий пример позволяет лучше понять задачу проверки гипотезы.

Пусть известно, что во время эпидемии гриппа заболевает в среднем 30%. Была разработана вакцина от гриппа. Разработчики утверждают, что в результате прививки данной вакцины человек либо не заболевает, либо вероятность заболеть резко уменьшается. Для подтверждения этого было отобрано 10 групп добровольцев, подвергшихся вакцинации. В результате в пяти из них никто не заболел, в двух группах процент заболевших был равен двум и в двух группах – 4% заболевших, но в одной группе оказалось 27% заболевших.

Можно ли утверждать, что вакцина действует, и какой на самом деле средний процент заболевших?

Для решения этой задачи была выдвинута гипотеза $H_0$: $MX=30$, где $X$-процент заболевших. В качестве альтернативной гипотезы принята гипотеза $H_1$: $MX<30$. В результате проверки гипотеза $H_o$, означающая, что вакцина не работает, была отвергнута. Значит процент заболевших уменьшился, но как? Последовательно проверялись гипотезы: $MX=20$, $MX=10$, $MX=5$. В результате проверки была принята гипотеза $MX<5$, что означает, что после вакцинации на уровне значимости 0,05 можно утверждать, что процент вакцинированных заболевших не превышает 5.

Статистическая проверка гипотез основывается на выбранных данных, а потому не исключена возможность принятия неправильного решения. При этом возможны ошибки двух родов: ошибка первого рода состоит в том, что будет отвергнута правильная нулевая гипотеза, ошибка второго рода заключается в том, что будет принята ошибочная нулевая гипотеза.

Результаты статистических выводов представлены следующей таблицей:

| Результаты проверки гипотезы | Возможные состояния гипотезы | |
|---|---|---|
| | Верна $H_0$ | Верна $H_1$ |
| Гипотеза $H_0$ отклоняется, принимается $H_1$ | Ошибка первого рода $P(H_1/H_0)=\alpha$ | Правильный вывод $P(H_1/H_1)=1-\beta$ |
| Гипотеза $H_1$ отклоняется, принимается $H_0$ | Правильный вывод $P(H_0/H_0)=1-\alpha$ | Ошибка второго рода $P(H_0/H_1)=\beta$ |

$P(H_1/H_1)=1-\beta$ называется мощностью критерия, т. е. вероятностью отвергнуть неверное решение.
Вероятность $\alpha$ совершить ошибку первого рода называют также уровнем значимости.

Невозможно исключить ошибки первого и второго рода, поскольку проверка осуществляется на основании выборки, но можно попытаться уменьшить их вероятности.
Однако одновременное уменьшение $\alpha$ и $\beta$ достигается только увеличением объема выборки.

Вероятность $\alpha$ ошибки первого рода называют также риском продавца, а $\beta$ – вероятность ошибки второго рода – риском покупателя.
В самом деле, пусть продавец продает оптом большую партию ящиков с апельсинами. Так как невозможно проверить все ящики, покупатель выбирает несколько из них. Если апельсины в них оказались второго сорта $(H_1)$, хотя фактически подавляющее большинство апельсинов партии первого сорта $(H_o)$, то вся партия считается вторым сортом и проигрывает продавец. Его риск равен α. Если, напротив, в выбранных ящиках апельсины оказались первого сорта $(H_o)$, хотя в действительности почти все они второго сорта $(H_1)$, то покупатель платит за них по цене первого сорта и проигрывает. Его риск равен β.

Итак, пусть нулевая гипотеза $H_o$: $\theta=\theta_0$, $\alpha$ – уровень значимости, $Z$ – статистика, $K_p$ – квантиль этой статистики уровня $p$, $V$ – область принятия гипотезы $H_0$. следующая таблица устанавливает связь между конкурирующей гипотезой $H_1$ и областью $V$.

| Конкурирующая гипотеза | Область $V$ принятия гипотезы |
|---|---|
| $H_1$: $\theta \neq \theta_0$ | $K_{\alpha/2} < Z_{наб} < K_{1-\alpha/2}$ |
| $H_1$: $\theta > \theta_0$ | $Z_{наб} < K_{1-\alpha}$ |
| $H_1$: $\theta < \theta_0$ | $Z_{наб} > K_{\alpha}$ |

<u>Примеры проверки гипотез</u>

1. Проверка гипотезы о значении математического ожидания нормальной СВ при известной и неизвестной дисперсии.

$\sigma^2$ известно $H_0$: $MX=m_0$ , $Z=\frac{(\bar{x}-m_0)\sqrt{n}}{\sigma}=N[0,1]$ , $K_p=u_p$

$\sigma^2$ неизвестно $H_0$: $MX=m_0$ , $Z=\frac{(\bar{x}-m_0)\sqrt{n}}{s}=St[n-1]$ , $K_p=t_p(n-1)$

$s$ - исправленное среднее квадратическое отклонение

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

*Замечание 1*

$u_p$ – квантиль нормального распределения, $t_p(n-1)$ – квантиль распределения Стьюдента с $(n-1)$ степенью свободы. Поскольку стандартное нормальное распределение и распределение Стьюдента симметричны относительно $x=0$, то $u_p=-u_{1-p}$ и тоже самое $t_p=-t_{1-p}$.

*Замечание 2*

Если альтернативная гипотеза $\theta \neq \theta_0$, то проверка гипотезы может быть проведена следующим образом: строится доверительный интервал параметра $\theta$ с надежностью $\gamma=1-\alpha$, а затем, если значение $\theta_0$ попадает в данный интервал, то гипотеза $H_0$ не отвергается, если нет – то отвергается.

2. Проверка гипотезы о величине дисперсии нормальной СВ.
   $H_0$: $\sigma^2 = \sigma_o^2$
   $Z = \frac{(n-1)s^2}{\sigma_o^2} = \chi 2[n-1]$; $\chi_p^2[n-1] = K_p$.

Здесь снова $s^2$-исправленная дисперсия $\chi_p^2[n-1]$ – квантиль распределения $\chi^2$ с $(n-1)$ степенью свободы.

3. Проверка гипотезы о значении вероятности в независимых испытаниях Бернулли.
   $H_0$: $p = p_o$

$Z=\frac{(W-p_o)\sqrt{n}}{\sqrt{p_o(1-p_o)}} = N[0,1]$ $K_p=u_p$

Здесь $W=\frac{m}{n}$-относительная частота. Из теоремы Бернулли следует, что при достаточно больших значениях n относительная частота W сходится по вероятности к вероятности события. При этом $M(W)=p$, $\sigma(W)=\sqrt{\frac{pq}{n}}$.

Величина $U=\frac{(W-p)\sqrt{n}}{\sqrt{pq}}$ имеет приближенно стандартное нормальное распределение.

4. Проверка гипотезы о равенстве математических ожиданий.

Пусть генеральные совокупности $X$ и $Y$ распределены нормально и их дисперсии известны: $\sigma_x^2$ и $\sigma_y^2$ . По независимым выборкам, объемы которых равны $n$ и $m$ соответственно найдены выборочные средние $\bar{X}$ и $\bar{Y}$.

*$H_0$: MX=MY*

а) $\sigma_x^2$ и $\sigma_y^2$ известны $Z=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}}=N[0,1]$ $K_p=u_p$

*в) $H_0$: MX=MY*

$\sigma_x^2$ и $\sigma_y^2$ неизвестны,

$X$ и $Y$ имеют нормальное распределение или

$\sigma_x^2$ и $\sigma_y^2$ неизвестны, $n$ и $m>30$ и выборки независимы

$$Z=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}}=N[0,1] \quad K_p=u_p$$

*с) $H_0$: MX=MY*

$\sigma_x^2$ и $\sigma_y^2$ неизвестны,

$X$ и $Y$ нормально распределены

$m$, $n$ малы, $\sigma_x^2=\sigma_y^2$

$$Z=\frac{\bar{X}-\bar{Y}}{\sqrt{(n-1)s_x^2+(m-1)s_y^2}}\sqrt{\frac{nm(n+m-2)}{m+n}}=St(n+m-2)$$

$K_p=t_p(n+m-2)$

5. Проверка гипотезы о равенстве двух дисперсий нормальных распределений.

Пусть генеральные совокупности $X$ и $Y$ распределены нормально. По независимым выборкам с объемами соответственно n и m найдены исправленные выборочные дисперсии $s_x^2$ и $s_y^2$ и пусть для определенности $s_x^2 > s_y^2$.

$H_0$: $DX=DY$ $Z=s_x^2/s_y^2=F(n-1, m-1)$; $K_p=F_p(n-1, m-1)$

Замечание. Если наоборот $s_x^2 < s_y^2$, то $Z=s_y^2/s_x^2=F(m-1, n-1)$.

6. Проверка равенства двух вероятностей биномиальных распределений (равенстве долей).

Пусть в двух генеральных совокупностях проведены независимые испытания: $n_1$ – в первой и $n_2$ – во второй. При этом частота появления события A в первой равна $W_1=\frac{m_1}{n_1}$, а во второй - $W_2=\frac{m_2}{n_2}$. Обозначим неизвестные вероятности появления события A в первой совокупности через $p_1$, а во второй - $p_2$.

$H_0$: $p_1=p_2=p$; $Z=\frac{W_1-W_2}{\sqrt{\frac{m_1+m_2}{n_1+n_2}\left(1-\frac{m_1+m_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = N[0,1]$ ; $K_p=u_p$

$n_1$, $n_2$ велики

Замечание. В качестве наилучшего значения берут $p_0=\frac{m_1+m_2}{n_1+n_2}$, если две выборки смешать в одну.

**Задача №3.** Анализируется доход $X$ фирм в некоторой отрасли. Предполагается, что СВ $X$ имеет нормальное распределение и при этом средний доход в отрасли не менее 1 млн. а среднее квадратическое отклонение не более 0,1 млн (в $). По выборке из 50 фирм получены следующие данные $\overline{X}=0,9$ млн, $s=0,15$ млн. На уровне значимости $\alpha=0,01$ проверить эти предположения.

Проверим, что средний доход не менее 1 млн при известном и неизвестном $\sigma$.

$H_0$: $MX=m_0=1$. Поскольку $\overline{X}=0{,}9<1$, то в качестве гипотезы $H_1$: $MX<1$

$\sigma$ известно $Z_{набл}=\dfrac{(0{,}9-1)\sqrt{50}}{0{,}1}=-7{,}07;$

Область левосторонняя. $u_\alpha=u_{0.01}=-u_{0.99}$.

$\Phi(2{,}3)=0{,}49=0{,}99-0{,}5$, т. е. $u_{0.01}=-2{,}3$.

Поскольку $Z_{набл}=-7{,}07<-2{,}3$, то гипотеза $H_0$ отвергается, принимается $H_1$, т. е. средний доход в отрасли меньше 1 млн $.

σ неизвестно $Z_{набл}=\dfrac{(\overline{X}-m_0)\sqrt{n}}{s}=\dfrac{(0{,}9-1)\sqrt{50}}{0{,}15}=-4{,}67;$

$V=t_{кр}(0{,}01;49)=-2{,}4$. Снова $Z_{набл}=-4{,}67<t_{кр}=-2{,}4$ и делается тот же вывод.

Проверим теперь, что среднее квадратическое не более, чем 0,1; $\sigma_{выб}=0{,}15;$, поэтому

$H_0$: $\sigma^2=\sigma_0^2=0{,}1^2$

$H_1$: $\sigma^2>\sigma_0^2=0{,}1^2$ $\qquad Z_{набл}=\dfrac{(n-1)s^2}{\sigma_0^2}=\dfrac{49\cdot 0{,}15^2}{0{,}1^2}=110{,}25$

Область правосторонняя:

$\chi^2_{кр}(0{,}01;49)=60=\chi^2_{0{,}99}(49)<Z_{набл}=110{,}25$, то

принимается гипотеза $H_1$, т. е. $\sigma>0{,}1$

<u>Вывод</u>. Средний доход фирм в этой отрасли меньше 1 млн$, а среднее квадратическое отклонение $\sigma>0{,}1$.

**Задача №4.** В рекламе фирмы *A* утверждается, что месячный доход по ее акциям в среднем равен 0,75% и превышает доход по акциям фирмы *B* более, чем на 0,3%. И при этом ее риски меньше. В течение годичного периода средний месячный доход по акциям *B* составил 0,4%, а по акциям *A* – 0,65%, а их средние квадратические отклонения 1,9% для *B* и 2% для *A*. Полагая распределение доходности по каждой акции нормальными, на уровне значимости $\alpha$=0,05 проверить все три утверждения, содержащиеся в рекламе.

Пусть *X* и *Y* месячный доход по акциям фирм *A* и *B* соответственно и пусть $Y_1=Y+0,3$. Тогда $\overline{Y_1}=\overline{Y}+0,3$;

$$s_{y1}^2=s_y^2.$$

а) $H_0$: $MX$=0,75%=$m_o$ $\qquad Z_{набл}=\frac{(0,65-0,75)\sqrt{12}}{2}=-0,17;$

$s$=2%

$H_1$: $MX$<0,75%,

$t_{0,05}(11)$=-1,796; -0,17>-1,796, принимается гипотеза $H_0$, т. е. $MX$=0,75%.

в) $H_0$: $MX=MY+0,3\%$;

$$Z_{набл}=\frac{0,65-0,4-0,3}{\sqrt{\frac{1,9^2}{2}}+\sqrt{\frac{2^2}{12}}}=\frac{-0,05\cdot\sqrt{12}}{\sqrt{7,61}}=0,06$$

$H_1$: $MX>MY+0,3\%$; $\qquad Z_{набл}=-0,06<u_{0,95}=1,645$

$S_x$=2%, $S_y$=1,9%,

$u_{1-\alpha}=u_{0,95}=1,645$

Принимается гипотеза $H_0$: $MX=MY+0,3\%$

с) $H_0$: $\sigma_x^2=\sigma_y^2$ $\qquad Z_{набл}=\frac{2^2}{1,9^2}=1,1$

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

$F_{кр}(11,11)= F_{0,05}(11,11)=2,82> Z_{набл}=1,1$

Принимается гипотеза $H_0$.

Вывод: Действительно, средний доход по акциям в фирме *A* равен 0,75% и он на 0,3% больше, чем в фирме *B*, но риски в обеих фирмах примерно одинаковые.

**Задача №5.** В университете из 102 студентов специальности *A* 62 успешно сдали все экзамены, а из 50 студентов специальности *B* сдали все экзамены 32 студента. На уровне значимости $\alpha$=0,05 проверить гипотезу об отсутствии существенных различий в успеваемости студентов этих двух специальностей.

Решение:

$$p_0 = \frac{32+62}{50+102} = 0,625; W_1 = \frac{62}{102} = 0,608; W_2 = \frac{32}{50} = 0,64$$

$$Z_{набл} = \frac{0,64-0,608}{\sqrt{0,625(1-0,625)\left(\frac{1}{102}+\frac{1}{50}\right)}} = \frac{0,032}{0,0836} = 0,38$$

$H_0$: $p_1=p_2$ $\alpha$=0,05; $\alpha$/2=0,025; 1-$\alpha$/2=0,975;

$H_1$: $p_1 \neq p_2$ $u_{0,025}$=-1,96; $u_{0,975}$=1,96

$-1,96<Z_{набл}=0,38<1,96$.

Следовательно, гипотеза $H_0$_принимается, т. е. не существует значимого различия в успеваемости студентов специальностей *A* и *B*.

## 27. Непараметрические гипотезы. Критерий Пирсона

Пусть по виду гистограммы или на основании теоретического анализа выдвигается гипотеза о виде закона распределения СВ *X*. Для ответа на вопрос, объясняются ли расхождения между гипотетическим и эмпирическим распределениями случайными обстоятельствами или гипотеза неверна, служат критерии согласия. На практике чаще всего используется критерий $\chi^2$-Пирсона.

Схема применения критерия Пирсона. Пусть *X* – непрерывная СВ. Полученная выборка объема *n* сформирована в виде интервального вариационного ряда.

1. Выбирается уровень значимости $\alpha$.
2. Число интервалов. Необходимо, чтобы в каждый интервал попадало не менее 5 наблюдений, т. е. $n_i \geq 5$. Если это условие не выполнено, то объединяют соседние интервалы. Пусть число интервалов равно *k*.
3. Параметры распределения. Число степеней свободы.

Параметры гипотетического распределения могут быть известными и неизвестными. Если они неизвестны, то их заменяют точечными оценками. Оценки чаще всего находят методом моментов, т. е. приравнивают теоретические и эмпирические моменты. Число степеней свободы $\nu=k-1$, если параметры известны и $\nu=k-s-1$, если приходится оценивать *s* параметров.

Заметим, что всегда должно быть выполнено соотношение: $\sum_{i=1}^{k} n_i' = n$

4. Теоретические частоты

На основании гипотетического распределения находятся вероятности $p_i = P(x \in \Delta_i)$ и теоретические частоты: $n_i' = np_i$

5. Выдвигается гипотеза $H_0$: исследуемая СВ величина распределена по данному закону. Строится статистика:

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

Эта статистика имеет $\chi^2$ распределение с числом степеней свободы $v=k\text{-}s\text{-}1$ при достаточно большом $n$. Для данного $\alpha$ находим $^{\Lambda кр}(\alpha,\ k\text{-}s\text{-}1)$

6. Если $Z_{наб} > \chi^2_{кр}(\alpha,\ k\text{-}s\text{-}1)$, то гипотезу $H_o$ отвергают, в противном случае считают, что расхождение между теоретическим и эмпирическим распределениями несущественно, и гипотезу $H_o$ можно считать правдоподобной.

<u>Комментарии</u>

1. При проверке гипотезы о нормальном распределении СВ $X$, параметры $m$ и $\sigma$ могут быть как известными, так и неизвестными. В этом случае полагают $m = \bar{X}$, $\sigma = s$. Тогда $n_i' = np_i = n\left(\Phi\left(\frac{x_{i+1}-\bar{x}}{s}\right) - \Phi\left(\frac{x_i-\bar{x}}{s}\right)\right)$.

Так как $x\epsilon(-\infty,+\infty)$, то $x_0=-\infty$, а $x_k=+\infty$, поскольку должно быть выполнено равенство $\sum_{i=1}^{k} p_i = 1$, или иначе $\sum_1^k n_i' = n$. Возможна небольшая погрешность, вызванная ошибками округления

2. если проверяется гипотеза о равномерном распределении СВ $X=R[a,b]$ и концы интервала известны, то $n_i' = \frac{n}{k}$, где $k$ – число интервалов одинаковой длины, или же $n_i' = n\frac{\Delta_i}{b-a}$, если пришлось объединять интервалы. В случае, когда концы интервала неизвестны, их заменяют оценками $a^*=\bar{X}-s\sqrt{3}$, $b^*=\bar{X}+s\sqrt{3}$.

Тогда $n_1' = \frac{x_1-a^*}{b^*-a^*}\cdot n$; $n_k' = \frac{b^*-x_{k-1}}{b^*-a^*}\cdot n$.

Для внутренних интервалов $n_i' = \frac{x_{i+1}-x_i}{b^*-a^*}\cdot n$

7. Показательное распределение $X\sim E(\lambda)$ имеет один параметр $\lambda$. Если он неизвестен, его заменяют оценкой $\lambda^*= \frac{1}{\bar{\bar{x}}}$. Так как $x\epsilon[0,+\infty]$, то полагают $x_0=0$ и $x_k= +\infty$, чтобы $\sum_1^k n_i'$ была равна $n$. Частоты $n_i' = n(e^{-\lambda^* x_i} - e^{-\lambda^* x_{i+1}})$

**Пример 1.** Для оценки месячного дохода на душу населения $X$ (тыс.руб.) была произведена случайная выборка объема $n$=100 и сформирован интервальный вариационный ряд.

| $x_i$ | Менее 7,3 | 7,3-8,5 | 8,5-9,7 | 9,7-10,9 | 10,9-12,1 | 12,1-13,3 | 13,3-14,5 | Свыше 14,5 |
|---|---|---|---|---|---|---|---|---|

| $n_i$ | 10 | 10 | 22 | 18 | 20 | 6 | 8 | 6 |
|---|---|---|---|---|---|---|---|---|

На уровне значимости $\alpha$=0,05 проверить гипотезу о нормальном распределении СВ $X$.

Решение.

При наличии открытых интервалов значений $X$ типа «менее $x_0$» или «свыше $x_k$» для проведения расчетов их условно заменяют интервалами той же длины. Интервал «менее 7,3» заменяем интервалом [6,1;7,3], а интервал «свыше 14,5» - интервалом [14,5;15,1]. Беря в качестве представителя интервала его середину, получаем $\overline{X} = 10{,}396 \approx 10{,}4$; $D_B = 5{,}242$, $s$=2,3.

Вычисления оформляем в виде таблицы:

| № интервала | $x_i$ | $x_{i+1}$ | $n_i$ | $a_i$ | $a_{i+1}$ | $\Phi(a_i)$ | $\Phi(a_{i+1})$ | $n_i'$ | $\dfrac{(ni - n_i')^2}{n_i'}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $-\infty$ | 7,3 | 10 | $-\infty$ | -1,35 | -0,5 | -0,4115 | 8,85 | 0,15 |
| 2 | 7,3 | 8,5 | 10 | -1,35 | -0,83 | -0,4115 | -0,2967 | 11,48 | 0,19 |
| 3 | 8,5 | 9,7 | 22 | -0,83 | -0,3 | -0,2967 | -0,1179 | 17,88 | 0,95 |
| 4 | 9,7 | 10,9 | 18 | -0,3 | 0,22 | -0,1179 | 0,0871 | 20,5 | 0,3 |
| 5 | 10,9 | 12,1 | 20 | 0,22 | 0,74 | 0,0871 | 0,2703 | 18,32 | 0,15 |
| 6 | 12,1 | 13,3 | 6 | 0,74 | 1,26 | 0,2703 | 0,3962 | 12,59 | 3,45 |
| 7 | 13,3 | 14,5 | 8 | 1,26 | 1,78 | 0,3962 | 0,4625 | 6,63 | 0,28 |
| 8 | 14,5 | $+\infty$ | 6 | 1,78 | $+\infty$ | 0,4625 | 0,5 | 3,75 | 1,35 |
| | | | | | | | | | 6,82 |

Здесь $a_i = \dfrac{x_i - 10{,}4}{2{,}3}$; $\Phi(x)$ – функция Лапласа

$\chi^2_{наб}$=6,82; $V$=8-1-2=5; $\chi^2_{кр}$(0,05;5)=11,1> $\chi^2_{наб}$=6,82.

Следовательно, гипотеза о нормальном распределении месячного дохода на душу населения не противоречит эксперименту и может быть принята.

Рассмотрим случай дискретной случайной величины.

**Пример 2.** Анализируется количество договоров, заключенных страховыми агентами компании в течение недели, для чего была произведена выборка объема $n$=50. Результат оформлен в виде дискретного вариационного ряда, где $x_i$ – количество заключенных договоров, а $n_i$ – число агентов, заключивших $x_i$ договоров.

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 10 | 15 | 9 | 6 | 5 | 4 |

На уровне значимости $\alpha$=0,05 проверить гипотезу о том, что СВ $X$ – количество заключенных договоров страховым агентом, распределено по закону Пуассона.

Поскольку $n_i$ должно быть ≥5, то объединяем два значения $x$=0 и $x$=1, а также $x$=5 и $x\geq 6$. Распределение Пуассона имеет один параметр $\lambda$. В качестве его оценки берем $\lambda^* = \overline{X}$.

Вычислим $\overline{X} = \dfrac{0\cdot 1 + 1\cdot 10 + 2\cdot 15 + 3\cdot 9 + 4\cdot 6 + 5\cdot 5 + 6\cdot 4}{50} = 2{,}8$

$\lambda^*=2{,}8=\lambda$.

Для распределения Пуассона $P(m) = \dfrac{\lambda^m e^{-\lambda}}{m!}$; $m=0,\ 1, \ldots n, \ldots$

Вычисления оформим в виде таблицы, полагая $e^{-2,8}$=0,06

| № варианта | $x_i$ | $n_i$ | $n_i'$ | $\dfrac{(ni - n_i')^2}{n_i'}$ |
|---|---|---|---|---|
| 1 | 0<br>1 | 11 | 11,4 | 0,014 |
| 2 | 2 | 15 | 11,76 | 0,893 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 9 | 10,98 | 0,357 |
| 4 | 4 | 6 | 7,68 | 0,368 |
| 5 | 5<br>≥6 | 9 | 8,18 | 0,082 |

$\sum=1{,}714.$

$\chi^2_{наб}$=1,714

Пояснения

$P(X=0)+P(X=1)=\frac{\lambda^0 e^{-\lambda}}{0!}+\frac{\lambda e^{-\lambda}}{1!}$=0,06(1+$\lambda$)=0,06·3,8=0,228;

$n'_1$=50·$p_1$=50·0,228=11,4;

$P(X=2)=\frac{2{,}8^2\cdot 0{,}06}{2!}$=0,2352; $n'_2$=0,2352·50=11,76

$P(X=3)=\frac{2{,}8^3\cdot 0{,}06}{3!}$=0,2196; $n'_3$=10,98

$P(X=4)=\frac{2{,}8^4\cdot 0{,}06}{4!}$=0,1536; $n'_4$=7,68

$P(X\geq 5)=1-P(X<5)$=1-(0,228+0,2352+0,2196+0,1537)=0,1635

$n'_5$=8,18

$V$=5-1-1=3; $\chi^2_{кр}$(0,05;3)=7,8> $\chi^2_{наб}$ =1,714.

Следовательно, на уровне значимости $\alpha$=0,05 можно принять гипотезу $H_0$ о пуассоновском распределении числа страховых договоров, заключенных страховым агентом.

## Задачи для домашних заданий.

Для исследования доходов сотрудников предприятия было отобрано по 15 человек в каждом подразделении. В результате получены 2 выборки, где *X* и *Y* доходы подразделений (в тыс. руб.). Объединив обе выборки в одну:

1. Составить интервальный вариационный ряд, разбив значения варианты на 6 интервалов.
2. Построить гистограмму, и эмпирическую функцию распределения.
3. Найти средний доход работников предприятия, медиану *Me*, моду *Mo*, дисперсию $s^2$ и среднее квадратическое отклонение.

### *Вариант 1.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *X:* | 8,5 | 19,2 | 23,1 | 15,1 | 10,8 | 19,5 | 15,2 | 28,4 | 12,6 | 20,4 |
| *Y:* | 18,5 | 32,1 | 16,7 | 35,6 | 30,4 | 27,8 | 52,4 | 40,8 | 18,5 | 13,6 |

| | | | | | |
|---|---|---|---|---|---|
| *X:* | 18,5 | 32,1 | 16,4 | 25,4 | 30,7 |
| *Y:* | 9,4 | 12,1 | 15,7 | 20,1 | 18,5 |

### *Вариант 2.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *X:* | 30,4 | 17,5 | 15,2 | 18,1 | 23,4 | 24,3 | 52,1 | 32,4 | 10,8 | 33,5 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Y:* | 52,1 | 32,4 | 10,8 | 32,5 | 20,5 | 15,7 | 10,2 | 52,1 | 50,4 | 32,4 |

| | | | | | |
|---|---|---|---|---|---|
| *X:* | 20,5 | 15,7 | 12,6 | 14,8 | 20,7 |
| *Y:* | 26,4 | 18,1 | 35,2 | 27,2 | 40,1 |

## *Вариант 3.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *X:* | 10,2 | 52,1 | 50,4 | 32,4 | 26,4 | 18,1 | 14,6 | 16,7 | 26,4 | 26,8 |
| *Y:* | 15,7 | 42,3 | 30,5 | 18,5 | 50,2 | 17,5 | 44,6 | 16,7 | 36,4 | 28,6 |

| | | | | | |
|---|---|---|---|---|---|
| *X:* | 50,2 | 13,8 | 10,5 | 14,6 | 12,2 |
| *Y:* | 40,2 | 50,3 | 12,4 | 13,8 | 30,5 |

## *Вариант 4.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *X:* | 8,5 | 15,2 | 18,5 | 30,4 | 52,1 | 10,2 | 18,1 | 20,3 | 15,7 | 14,6 |
| *Y:* | 52,1 | 40,7 | 45,1 | 42,3 | 16,7 | 23,1 | 25,2 | 16,4 | 15,2 | 10,8 |

| | | | | | |
|---|---|---|---|---|---|
| *X:* | 19,2 | 20,4 | 32,1 | 17,5 | 32,4 |
| *Y:* | 50,4 | 46,2 | 39,7 | 30,5 | 16,7 |

## *Вариант 5.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *X:* | 15,1 | 28,4 | 35,6 | 18,1 | 33,5 | 32,4 | 35,8 | 40,1 | 18,5 | 16,4 |
| *Y:* | 26,4 | 27,1 | 45,2 | 50,2 | 16,8 | 19,5 | 20,4 | 27,8 | 24,3 | 15,7 |

| | | | | | |
|---|---|---|---|---|---|
| *X:* | 10,8 | 12,6 | 30,4 | 23,4 | 20,5 |
| *Y:* | 18,1 | 26,4 | 15,8 | 17,5 | 50,2 |

**Приложение**

## *Таблицы квантилей*

Квантили **$u_p$** стандартного нормального закона распределения **N(1,0).**

| p | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 | 0,999 | 0,9995 |
|---|---|---|---|---|---|---|---|
| $u_p$ | 1,282 | 1,645 | 1,960 | 2,325 | 2,576 | 3,090 | 3,291 |

Квантили $\chi_p^2(k)$ закона распределения $\chi^2(k)$,

| k \ p | 0,01 | 0,025 | 0,05 | 0,1 | 0,9 | 0,95 | 0,975 | 0,99 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0,0002 | 0,001 | 0,004 | 0,0158 | 2,71 | 3,84 | 5,02 | 6,63 |
| 2 | 0,02 | 0,05 | 0,103 | 0,211 | 4,61 | 5,99 | 7,38 | 9,21 |
| 3 | 0,115 | 0,216 | 0,352 | 0,584 | 6,25 | 7,81 | 9,35 | 11,3 |
| 4 | 0,297 | 0,484 | 0,711 | 1,06 | 7,78 | 9,49 | 11,1 | 13,3 |
| 5 | 0,554 | 0,831 | 1,15 | 1,61 | 9,24 | 11,1 | 12,8 | 15,1 |
| 10 | 2,56 | 3,25 | 3,94 | 4,87 | 16 | 18,3 | 20,5 | 23,2 |
| 20 | 7,63 | 9,59 | 10,9 | 12,4 | 28,4 | 31,4 | 34,2 | 37,6 |
| 30 | 14,3 | 16,8 | 18,5 | 20,6 | 40,3 | 43,8 | 47,0 | 50,9 |
| 40 | 22,2 | 24,4 | 26,5 | 29,1 | 51,8 | 55,8 | 59,3 | 63,7 |
| 50 | 29,7 | 32,4 | 34,8 | 37,7 | 63,2 | 67,5 | 71,4 | 76,2 |
| 75 | 49,5 | 53 | 56,1 | 59,8 | 91,1 | 96,2 | 100,8 | 106,4 |
| 100 | 70,1 | 74,2 | 77,9 | 82,4 | 118,5 | 124,3 | 129,6 | 135,6 |

Квантили $\mathbf{t_p}$ **(k)** закона распределения Стьюдента **T (k)**,

| k \ p | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 |
|---|---|---|---|---|---|
| 1 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 |
| 2 | 1,886 | 2,92 | 4,303 | 6,965 | 9,925 |
| 3 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 |
| 4 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 |
| 5 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 |
| 10 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 |
| 20 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 |
| 30 | 1,312 | 1,697 | 2,042 | 2,457 | 2,750 |
| 40 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 |
| 120 | 1,289 | 1,658 | 2,98 | 2,358 | 2,617 |

Квантили $\mathbf{F_{0,9}}$ **($n_1$,$n_2$)** закона распределения Фишера

| $n_1$ \ $n_2$ | 1 | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 30 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 40 | 8,53 | 5,54 | 4,54 | 4,06 | 3,29 | 3,07 | 2,97 | 2,88 | 2,75 |
| 2 | 49,5 | 9 | 5,46 | 4,32 | 3,78 | 2,92 | 2,7 | 2,59 | 2,49 | 2,35 |
| 3 | 53,6 | 9,16 | 5,39 | 4,19 | 3,62 | 2,73 | 2,49 | 2,38 | 2,28 | 2,13 |
| 4 | 55,8 | 9,24 | 5,34 | 4,11 | 3,52 | 2,61 | 2,36 | 2,25 | 2,14 | 1,99 |
| 30 | 62,2 | 9,46 | 5,17 | 3,82 | 3,17 | 2,16 | 1,87 | 1,74 | 1,61 | 1,41 |

Квантили $\mathbf{F_{0,95}\ (n_1,n_2)}$ закона распределения Фишера

| $n_1 \setminus n_2$ | 1 | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 30 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161 | 18,5 | 10,13 | 7,71 | 6,61 | 4,96 | 4,54 | 4,35 | 4,24 | 3,92 |
| 2 | 199 | 19 | 9,55 | 6,94 | 5,79 | 4,1 | 3,68 | 3,49 | 3,39 | 3,07 |
| 3 | 216 | 19,16 | 9,28 | 6,59 | 5,41 | 3,71 | 3,29 | 3,1 | 2,99 | 2,68 |
| 4 | 225 | 19,25 | 9,12 | 6,39 | 5,19 | 3,48 | 3,05 | 2,87 | 2,76 | 2,45 |
| 30 | 250 | 19,46 | 8,62 | 5,75 | 4,5 | 2,7 | 2,25 | 2,04 | 1,84 | 1,55 |