МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БРЕСТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

КАФЕДРА ВЫСШЕЙ МАТЕМАТИКИ

Elements of Probability Theory And Mathematical Statistics

методические указания на английском языке
по дисциплине **«Теория вероятностей и математическая статистика»**

Брест 2020

УДК 519.2(076)=111

Данные методические указания адресованы преподавателям и студентам технических ВУЗов для проведения аудиторных занятий и организации самостоятельной работы студентов при изучении материала из рассматриваемых разделов. Методические указания на английском языке «Elements of Probability Theory And Mathematical Statistics» содержат необходимый материал по темам «Теория вероятностей», «Математическая статистика», изучаемым студентами БрГТУ технических специальностей в курсе дисциплины «Теория вероятностей и математическая статистика». Теоретический материал сопровождается рассмотрением достаточного количества примеров и задач, при необходимости приводятся соответствующие иллюстрации. Для удобства пользования каждая тема разделена на три части: краткие теоретические сведения (определения, основные теоремы, формулы для расчетов); задания для аудиторной работы и задания для индивидуальной работы.

Данные методические указания являются продолжением серии методических разработок на английском языке коллектива авторов [1]-[9]. Практика использования разработок данной серии показала целесообразность её применения в процессе обучения студентов не только технических, но и экономических специальностей. Также были получены положительные отзывы об упомянутой серии от иностранных студентов.

Составители: Дворниченко А.В., м.э.н., старший преподаватель
Лебедь С.Ф., к.ф.-м.н., доцент
Бань О.В., старший преподаватель кафедры иностранных языков

Рецензент: Мирская Е.И., доцент кафедры алгебры, геометрии и математического моделирования УО «Брестский государственный университет им. А.С. Пушкина», к.ф.-м.н., доцент

CONTENT

I PROBABILITY THEORY

1.1 Elements of Combinatorics

Theorem 1 (fundamental principles of combinatorics) Let an action A_1 be able to be done by n_1 ways, an action A_2 by n_2 ways,…, an action A_k by n_k ways. Then all these actions can be done together [or simultaneously] by $n_1 \cdot n_2 \cdot n_3 \cdots n_k$ ways and any of these actions can be done by $n_1 + n_2 + n_3 + \cdots + n_k$ ways if these ways are not the same.

Example 1 Let's suppose that one has a coins and b dice. Then he can take a coin and a die by $a \cdot b$ ways. Indeed, each coin generates $1 \cdot b = b$ pairs “coin-die”. Therefore, a coins generate $a \cdot b$ pairs.

Example 2 One has 2 coins, 3 ties and 5 books. He can take one coin, one tie and one book by 2 · 3 · 5 = 30 ways.

Main notions of combinatorics

Let there be given some set M containing n elements.

Definition An ***arrangement*** of n elements (taken) k at a time (k-fold arrangement of n elements) is called any ordered k-fold subset of the n-fold set M.

Various arrangements differ by at least one element or by the order of their elements.

Definition A ***permutation*** of n elements is called any arrangement of all n elements of the n-fold set M.

Distinct permutations differ by the order of (the same) elements. One can say that permutation of n elements is the ordered set of all elements of the set M.

Definition A ***combination*** of n elements (taken) k at a time (k-fold combination of n elements) is called any k-fold subset of the n-fold set M.

Every combination differs from another one by at least one element.

Theorem 2 Numbers of all k-fold arrangements, of all permutations, of all k-fold combinations of n elements, without repetitions, are respectively equal

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!} \tag{1}$$

$$P_n = n! \tag{2}$$

$$C_n^k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \tag{3}$$

Theorem 3 Numbers of all k-fold arrangements, of all permutations, of all k-fold combinations of n elements, with repetitions, are respectively equal

$$\left(A_n^k\right)^* = n^k \tag{4}$$

$$P^{*}_{n_1,n_2,n_3,\ldots n_p}(n)=\frac{n!}{n_1!\cdot n_2!\cdots n_p!} \qquad (5)$$

$$\left(C_n^k\right)^{*}=C_{n+k-1}^{k}=\frac{(n+k-1)!}{k!\cdot(n-1)!} \qquad (6)$$

Example 3 There are 3 elements a, b, c. Use them to make:

a) permutations (without repetition) $P_3=3!=\{abc,bca,bac,acb,cab,cba\}$

b) permutations with repetition, where „ a“ occurs 2 times,“ b“ one time and „c“ one time

$$P_4(2,1,1)=12=\begin{Bmatrix} aabc & aacb & acba & abca \\ bcaa & cbaa & caba & baca \\ baac & caab & acab & abac \end{Bmatrix}$$

Example 4 A group containing 25 students can elect the leader and their assistant by

$$A_{25}^{2}=\frac{25!}{(25-2)!}=\frac{25!}{23!}=24\cdot 25=600$$

ways, because these two students form 2-fold arrangement of 25 elements .

Example 5 One can invite any 4 students of the same group to do some work by

$$C_{25}^{4}=\frac{25!}{4!(25-4)!}=\frac{25!}{4!\cdot 21!}=\frac{22\cdot 23\cdot 24\cdot 25}{2\cdot 3\cdot 4}=12650$$

ways, because these 4 students form 4-fold combinations of 25 elements.

Example 6 8 books can be placed in a bookshelf by

$$P_8=8!=1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8=40320$$

ways, because they form a permutation of 8 elements.

Example 7 How many numbers can 240 arrangements be made from if the number of elements to be selected is 2?

Solution $A_x^2=240,\ \frac{x!}{(x-2)!}=240,\ x(x-1)=240,\ x^2-x-240=0 \ \Rightarrow\ x=16$

16 numbers are needed.

Example 8 Number of arrangements without repetition with $k=3$ from x members is lower than number of arrangements with repetition with $k=3$ from x members by 225. How many members are there?

Solution $\left(A_x^3\right)^{*}-A_x^3=225 \Rightarrow x^3-\frac{x!}{(x-3)!}=225 \Rightarrow x^3-x(x-1)(x-2)=225$

$$x^3-x^3+3x^2-2x=225,\ 3x^2-2x-225=0\ ,\ x_1=9\ \ x_2=-\frac{25}{3}<0.$$

There are 9 members.

Example 9 A password of 6 digits is made of digits 926002.How many possible passwords are there? How long would it take to try all the possible passwords if trying one password takes 5 seconds?

Solution $n_1 = 2, n_2 = 2,\ n_3 = 1,\ n_4 = 1,\ n = 6$

$$P^*_{2,2,1,1}(6) = \frac{6!}{2!\cdot 2!\cdot 1!\cdot 1!} = 180$$

$t = 180 \cdot 5 = 900 \text{ sec} = 15 \text{min}$. There are 180 possible passwords; trying all of them would take 15 minutes.

Example 10 a) Find out a formula for counting the number of diagonals in a convex n-gon. b) How many diagonals does a 10-gon have?

Solution:

a) $N = C_n^2 - n = \dfrac{n!}{2!\cdot(n-2)!} - n = \dfrac{n(n-1)}{2} - n = \dfrac{n^2 - 3n}{2} = \dfrac{n(n-3)}{2}$.

b) $N = C_{10}^2 - 10 = \dfrac{10!}{2!\cdot(10-2)!} - 10 = \dfrac{10(10-3)}{2} = 35$.

Convex 10-gon has 35 diagonals.

Example 11 5 different ice-creams are sold in the confectioners. A father would like to buy 15 caps of ice-cream for his family. In how many ways can he buy the ice-cream?

$$\left(C_5^{15}\right)^* = C_{15+5-1}^{15} = \frac{19!}{15!\cdot 4!} = \frac{16\cdot 17\cdot 18\cdot 19}{4!} = 3876$$

The father can buy the ice-cream in 3876 different ways.

Example 12 A student has to make 4 university entry exams. For passing each exam he gets either 2, 3 or 4 points. He needs to reach at least 13 points to get to the university. In how many ways can he do the exams to be successful?

Solution He needs to get 13, 14, 15 or 16 points.

13 points:	*(3,3,3,4)*	$P^*_{3,1}(4) = 4$
	(4,4,3,2)	$P^*_{2,1,1}(4) = 12$
14 points:	*(4,4,4,2)*	$P^*_{3,1}(4) = 4$
	(4,4,3,3)	$P^*_{2,2}(4) = 6$
15 points:	*(4,4,4,3)*	$P^*_{3,1}(4) = 4$
16 points:	*(4,4,4,4)*	$P^*_{4}(4) = 1$

N = 4 + 12 + 4 + 6 + 4 + 1 = 31

There are 31 possible ways of reaching at least 13 points.

Exercise Set 1.1

1. If there are 2 intern positions open at a firm with 5 applicants, how many different combinations of applicants could be hired?

2. If a class of 10 students has five men, in how many ways can the men and women be arranged in a circle so that no two men sit next to each other?

3. You have 6 different tickets in your pocket marked with numbers 1-6. How many ways are there to choose 3 of them (considering the order), if

a) The selected ticket is not returned to the pocket.

b) The selected ticket is returned to the pocket.

4. Eight students should be accommodated in two 3-bed and one 2-bed rooms. In how many ways can they be accommodated?

5. In how many ways can you throw a sum of 11 by 3 playing cubes?

6. You need to put your reindeer, Gloopin, Quentin, Ezekiel, and Lancer, in a single-file line to pull your sleigh. However, Quentin and Gloopin are best friends, so you have to put them next to each other, or they won't fly. In how many ways can you arrange your reindeer?

7. There are 4 Czech and 3 Slovak books on the bookshelf. Czech books should be placed on the left side of the bookshelf and Slovak books on the right side of the bookshelf. How many ways are there to arrange the books?

8. In how many ways can you choose 8 of 32 playing cards not considering their order?

9. Count all possible arrangements of 15 people on the photo.

a) How many photos will be made?

b) How long would the photographing last if you need 10 seconds for one photo?

10. On a circle there are 9 points selected. How many triangles with edges in these points exist?

11. If the number of elements rise by 2, the number of arrangements without repetition would rise 42 times. How many elements are there?

12. A teacher has prepared 20 arithmetic tasks and 30 geometry tasks. He would like to use: a) 3 arithmetic and 2 geometry tasks; b) 1 arithmetic and 2 geometry tasks for a test. How many ways are there to build the test?

Individual Tasks 1.1

I.

1. How many unique ways are there to arrange the letters in the word SONG?

2. Four names are drawn from 24 members of a club for the offices of President, Vice-President, Treasurer, and Secretary. In how many different ways can this be done?

3. Suppose a box contains 4 blue, 5 white, 6 red and 7 green balls. In how many of all possible samples of size 5, chosen without any replacement, will every color be represented?

4. In a particular softball league each team consists of 5 women and 5 men. Determining a betting order for 10 players, a woman must bet first, and successive betters must be of the opposite sex. How many different betting orders are possible for a team?

5. How many positive integers of 5 digits may be made from the ciphers 1,2,3,4,5, if each cipher may be used just once? How many of them will be even?

6. Two groups consist of 26 elements and 160 combinations without repetition for $k=2$ together. How many elements are there in the first and how many in the second group?

II.

1. How many different unique combinations of letters can be created by rearranging the letters in *mathematics*?

2. How many committees of two chemists and one physicist can be formed from 4 chemists and 3 physicists?

3. The Mathematics Department of the University of Louisville consists of 8 professors, 6 associate professors, 13 assistant professors. In how many of all possible samples of size 4, chosen without any replacement, will every type of a professor be represented?

4. Eight students promised to send a postcard to each other. How many postcards did they send together?

5. How many positive integers of 5 digits may be made from the ciphers 1,2,3,4,5, if a cipher may be used a few times? How many of them will begin with 5?

6. If the number of members increments by 2, the number of possible variations of arrangements with $k=3$ increments by 384. How many members are there?

1.2 Trial and Event. Definition of Probability

Definition A ***trial*** is a realization of some complex of conditions. It is supposed that a trial can be arbitrary realized many times.

Definition An ***outcome*** is the result of a single trial of an experiment.

Definition An ***event*** is one or more outcomes of an experiment.

Example 1 (see the table).

Trial	***Events***
1. Coin flip [coin tossing]	"head" (occurrence of a head), "tail"
2. Dice toss(ing), fair dice rolling	"1", "2", "3", "4", "5", "6"
3. Drawing a ball from an urn containing *a* white and *b* black balls	"white ball", "black ball"

Events are usually denoted by capitals ($A, B, C, \ldots$). There are impossible, certain and random events.

Definition An event is called ***impossible*** if it can not occur in any trial (denoted by $\varnothing$).

Definition An event is called ***certain*** if it necessarily occurs in any trial (denoted by Ω).

The occurrence of a head *or* a tail in one coin tossing and the occurrence of at least one of the digits 1, 2, 3, 4, 5, 6 in one dice rolling are the examples of certain events.

Definition An event is called ***random*** if it can or can not occur in a trial.

Definition Events A, B are called ***joint*** (*compatible*) if they can occur together [or simultaneously] in a trial.

Example 2 "head", "head"; "tail", "tail"; "head", "tail"; "tail", "head" if a trial implies double coin tossing.

Definition Events A, B are called ***disjoint*** (*incompatible, non-compatible*) if they can not occur together (or simultaneously) in a trial.

Example 3 "head", "tail" in one coin toss.

Example 4 The events “1”, “2”, “3”, “4”, “5”, “6” are pairwise disjoint in one dice rolling.

Definition Two events A, B are called ***independent*** if the occurrence of one of them does not depend on the occurrence (or non-occurrence) of the other one.

Definition n events (for $n \geq 2$) are called ***mutually independent*** if the occurrence of one of them does not depend on the occurrence or non-occurrence of any group of the other one.

Definition One says that events $A, B..., C$ form ***a total [complete] group*** (of events) ($A, B..., C$ are only possible events, $A, B..., C$ are exhaustive events) if at least one of them occurs in any trial.

Example 5 Events “head” and “tail” in one coin toss form the total group of events. All the events “1”, “2”, “3”, “4”, “5”, “6” in one dice rolling form the total group of events.

Definition Two events A and $\overline{A}$ (non A) are called ***opposite*** if they are disjoint and form a total group.

Example 6 If A is “head”, then $\overline{A}$ (non A) is “tail” (in one coin toss). If A is “1”,then $\overline{A}$ (non A) is the occurrence of at least one of events “2”, “3”, “4”, “5”, “6”, A = {“2” or “3”, or “4”, or “5”, or “6”} (in one fair dice rolling).

Definition The sum $A + B$ $(A \text{ or } B)$ of two events A and B is called an event which consists in occurrence of at least one of them [which means that at least one of these events occurs] (A but not B or B but not A or A and B together).

Example 7 The sum of an event A and its opposite one $\overline{A}$ is a certain event.

Example 8 If an event A is “1” in one dice rolling, then the opposite event $\overline{A}$ is the sum $\overline{A} = "2" + "3" + "4" + "5" + "6"$.

Definition The product AB (A *and* B) of two events A and B is called an event consisting in occurrence of both these events [an event which means that both these events occur] together.

Example 9 The product of an event A and its opposite one $\overline{A}$ is an impossible event.

Example 10 (***Euler circles***). Let M and N be two circles having the non-empty intersection $V = M \cap N$, also let $U = M \setminus N$, $W = N \setminus M$, and so $M = U \cup V$, $N = V \cup W$ (see fig. 1). If an event A means that a point P belongs to M, and B means that P belongs to N, then $A + B = \{P \in (U \cup V \cup W)\}$, $A \cdot B = \{P \in V\}$.

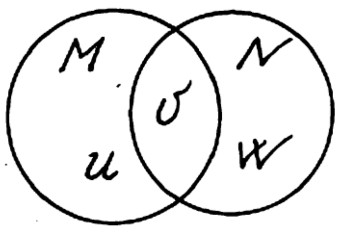


Figure 1

When solving probabilistic problems, it is sometimes useful to represent an event in question as a sum of the product of other events (with pairwise disjoint summands).

Example 11 Let events A, B, C mean that the first, the second, the third device (correspondingly) work. In this case the events

$$D = ABC, \; E = \overline{ABC}, \; F = A\overline{BC} + \overline{A}B\overline{C} + \overline{AB}C, \; G = \overline{A}BC + A\overline{B}C + AB\overline{C},$$

$$F + G + D = \overline{E}$$

mean respectively that all the three devices work, none of devices works, only one device works (and the other two do not work), two devices work (and one does not work), at least one device works.

There are events for which we can subtract a set of ***elementary events*** (*chances, possibilities*) that is a total group of pairwise disjoint and equally possible events. A chance is called ***favorable*** for an event A if A occurs when this chance occurs.

Probability is the measure of how likely an event is.

Definition Let n be the number of all chances [of all elementary events, of all possibilities] and m be the number of those favorable for some event A. In this case the probability of this event is expressed by the next ratio:

$$P(A) = \frac{m}{n} \qquad (1)$$

The probability of an event A is the number of ways the event A can occur divided by the total number of possible outcomes.

Example 12 Find the probability of occurrence of the head in one coin-tossing.

Solution Let A be an event which means that a head occurs. We can subtract the next $n = 2$ chances [elementary events, possibilities]: “head”, “tail”. There is $m = 1$ favorable chance, namely “head”. By the formula (1)

$$P(A) = \frac{m}{n} = \frac{1}{2} = 0.5.$$

Example 13 Find the probability of occurrence of an even number in one dice-rolling.

Solution Let A be an event which consists in occurrence of an even number in one dice-rolling. The chances [elementary events, possibilities] connected with the event A are “1”, “2”, “3”, “4”, “5”, “6”, $n = 6$. The favorable chances are “2”, “4”, “6”, $m = 3$. By the formula (1)

$$P(A) = \frac{m}{n} = \frac{3}{6} = 0.5.$$

Example 14 There are 6 white and 14 black balls in some urn. One takes 10 balls at randomly. Find the probability of drawing of 4 white and 6 black balls.

Solution Let A be an event consisting of drawing of 4 white and 6 black balls. The chances [elementary events, possibilities] for the event A are various sets of

10 balls, that is 10-fold combinations of 20 elements. Therefore, the number of all chances is equal to $n = C_{20}^{10}$ that is the number of all 10-fold combinations of 20 elements.

To determine the number m of favorable chances we must take into account that one can take 4 white balls (4-fold combination of 6 elements) by C_6^4 ways and 6 black balls (6-fold combination of 14 elements) by C_{14}^6 ways. Therefore, he can take 4 white and 6 black balls together by virtue of the fundamental principle of combinatorics by $C_6^4 \cdot C_{14}^6$ ways. It means that $m = C_6^4 \cdot C_{14}^6$.

Hence,

$$P(A) = \frac{m}{n} = \frac{C_6^4 \cdot C_{14}^6}{C_{20}^{10}} = \frac{\frac{6!}{4!\cdot 2!} \cdot \frac{14!}{6!\cdot 12!}}{\frac{20!}{10!\cdot 10!}} \approx 0.24.$$

Example 15 It is necessary to place 8 books on a bookshelf. Find the probability for two certain books *A*, *B* to stand side by side.

Solution Let an event C be the required position of our books. The chances [elementary events, possibilities] are their various locations which are permutations of 8 elements. Therefore, the number of all chances is that of all possible permutations of 8 elements,

$$n = P_8 = 8! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 = 40320.$$

To find the number of favorable chances (that is that of required positions of books) we will introduce the following table:

Actions to place A, B side by side	*Number of ways to do these actions*
1. Finding places for *A*, *B*	7
2. Location of *A*, *B* on these places (permutation of 2 elements)	$P_2 = 2!$
3. Disposition of the other 6 books (permutation of 6 elements)	$P_6 = 6!$
4.Getting a required disposition of all 8 books	$7 \cdot P_2 \cdot P_6$ (by virtue of the main combinatorial principle)

On the base of the classical definition of probability

$$P(A) = \frac{m}{n} = \frac{7 \cdot P_2 \cdot P_6}{P_8} = \frac{1}{4} = 0.25.$$

Statistic Probability

Definition Let some event A be studied and there very large number N of independent trials be fulfilled on A. Let is denote the number of occurrences of A as $N(A)$ in these trials. The ratio

$$p^* = P_N^*(A) = \frac{N(A)}{N} \tag{2}$$

is called a ***relative frequency*** (or sometimes frequency) of the event A.

Let us fulfill the series of very large numbers $N_1, N_2, ...$ of the independent trials on A and denote them by

$$p_1^* = P_{N_1}^*(A),\, p_2^* = P_{N_2}^*(A), \dots$$

corresponding relative frequencies of A.

There are many events for which relative frequencies possess a property of ***statistic stability*** that is they are approximately equal to some number p

$$p_1^* \approx p,\, p_2^* \approx p,\, \dots$$

Definition If event A possesses such a stability property, we say that it has a probability (so-called ***statistic probability***), and this probability equals

$$P(A) = p \qquad (3)$$

Geometric Probability

Random events that take place in *continuous sample space* may invoke geometric imagery due to for at least two reasons: due to the nature of the problem or the nature of the solution.

We may think of *geometric probabilities* as of non-negative quantities (not exceeding 1) being assigned to subregions of a given domain subject to certain rules. If a function μ is an expression of this assignment defined on a domain D, then, for example, we require

$$0 \le \mu(A) \le 1 \text{ , } A \subset D \text{ and } \mu(D) = 1.$$

The function μ is usually not defined for all $A \subset D$. Those subsets of D, for which μ is defined are the *random events* that form particular sample spaces. Very often μ is defined by means of the ratio of areas so that, if $\sigma(A)$ is defined as the "area" of set A, then one may set $\mu(A) = \dfrac{\sigma(A)}{\sigma(D)}$.

Example 16 Two friends who take metro to their jobs from the same station arrive to the station uniformly randomly between 7 and 7:20 in the morning. They are willing to wait for one another for 5 minutes, after which they take a train whether together or alone. What is the probability of their meeting at the station?

Solution In a Cartesian system of coordinates (s,t) a square of side 20 (minutes) represents all the possibilities of the morning arrivals of the two friends at the metro station.

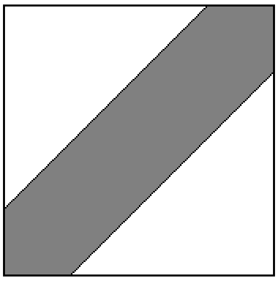


Figure 2

The grey area A is bounded by two straight lines, $t = s + 5$ and $t = s - 5$, so that inside A, $|s - t| \leq 5$. It follows that the two friends will meet only provided their arrivals s and t fall into region A. The probability of this happening is given by the ratio of the area of A to the area of the square:

[400 - (15×15/2 + 15×15/2)] / 400 = 175/400 = 7/16.

Exercise Set 1.2

1. If you randomly pull a single card from a standard deck, what is the probability that the card is anything other than a king?

2. We roll two six-sided dices. Find the probability that the sum of the numbers of dices is less than 8.

3. What is the probability of not pulling a face card from a standard deck?

4. What is the probability that a single roll of a 10-sided die will not land on a 7?

5. What is the probability that a single roll of a standard die will be 1, 2, 3, 4, or 5?

6. When rolling two six-sided dice, find the classical probability that: a) the sum of cast numbers is less than 5, b) we roll 5 by any dice, c) we roll only one 5, d) the sum of rolled numbers is divided by 3, e) the module of the difference of rolled numbers is equal to 3.

7. We have 5 segments with lengths 1, 3, 4, 7 and 9. Find the classical probability that by accidently choosing 3 segments we can construct a triangle.

8. There is a lottery with 10000 tickets, 120 monetary prizes and 80 material prizes. What is the probability of winning any prize for a participant having: a) 1 ticket; b) 2 tickets?

9. A candy jar contains 6 red, 7 blue, 8, green, and 9 yellow candies. What is the probability that choosing a single candy at random will result in a piece that is either red, blue, or green?

10. What is the probability that a single choice from the jar in Q 9 will result in a piece that is either red or yellow? What is the probability that a single choice from the same jar will not be blue?

11. You have a bag of crazy-flavor jellybeans. You know that the flavors are distributed as: 12 earwax, 14 belly-button lint. What is the probability that it will be only 2 earwax if you choose 10 jellybeans at random?

12. There are 450 songs on the .mp3 player you share with your father and sister. If your dad has 125 80's songs, and your sister has twice that many country music songs, what is the percent probability that a randomly chosen song will not be one of yours?

13. What is the percent probability that a random roll of a fair die will not result in an even or prime number?

14. The probability that a student has called in sick and that it is Monday is 12%. The probability that it is Monday and not another day of the school week is 20%

(there are only five days in the school week). What is the probability that a student has not called in sick, given that it is Monday?

15. The cube whose faces are painted is then cut into 1000 pieces. The resulting cubes are thoroughly shuffled. What is the probability that a randomly chosen cube will have: a) one; b) two; c) three; d) four colored faces?

16. If we randomly pick two television sets in succession from a shipment of 240 television sets of which 15 are defective, what is the probability that they both will be defective?

17. Mr. Flowers plants 10 rose bushes in a row. Eight of the bushes are white and two are red, and he plants them in a random order. What is the probability that he will consecutively plant seven or more white bushes?

18. Ten books are arranged on the bookshelf. What is the probability that some 3 of them will be near each other?

19. The point is placed randomly into the circle. What is the probability that the point will get inside of: a) a square inscribed in the circle; b) a regular triangle inscribed in the circle; c) a regular hexagon inscribed in the circle?

20. Real numbers p, q are chosen at random between 0 and 1. What is the probability equation $x^2 + px + q = 0$ has real roots?

21. We choose a point from a square with an inscribed circle. Find the geometrical probability that the chosen point does not belong to the circle.

22. The telephone line is damaged by storm between 160 and 290 kilometers. Find the probability that this line is damaged between 200 and 240 kilometers.

23. We accidentally choose a point in a tetrahedron, in which a ball is inscribed. Find the geometrical probability that an accidentally chosen point does not belong to the ball.

24. We accidentally choose three points A, B and C on the circumference of a circle with radius R. Find the probability that ΔABC will be an acute triangle.

Individual Tasks 1.2

I.

1. There are 5 white and 10 black balls in the box. Find the probability that the accidentally chosen ball would be black.

2. There are 17 students in the group. 8 of them are boys. There are staged 7 tickets to be drawn. Find the probability that there are 4 boys among the owners of tickets.

3. You have \$39 in cash, composed of the largest bills possible. What is the probability that a randomly chosen bill from the \$39 will not be a \$1 bill?

4. A neighborhood wanted to improve its parks so it surveyed kids to find out whether they rode bikes or skateboards or not. Out of 2300 children in the neighborhood that ride something, 1800 rode bikes, and 500 rode skateboards, while 200 of those ride both a bike and a skateboard. What is the probability that a student does not ride a skateboard, given that he or she rides a bike?

5. One has cards with the letters *m,a,t,h,e,m,a,t,i,c,s*. Take out one card for another randomly. What is the probability that you get a word: a) “mathematics”; b) ”team”; c) ”cat”?

6. An urn contains 3 red balls, 2 green balls and 1 yellow ball. Three balls are selected at random and without replacement from the urn. What is the probability that at least 1 color is not drawn?

7. We accidentally choose a point in a cube, in which a ball is inscribed. Find the geometrical probability that an accidentally chosen point does not belong to the ball.

8. Two friends must meet at the particular place in the interval of time [12, 13]. The friend which arrived first waits no longer than 20 minutes. Find the probability that the meeting between friends will happen within the mentioned interval.

II.

1. There are 7 white and 13 red balls in the box. Find the probability that among accidentally chosen 3 balls 2 balls would be red.

2. A group of 10 girls and 10 boys is accidentally divided into two subgroups. Find the classical probability that in both subgroups the numbers of girls and boys are equal.

3. A box contains four $10 bills, six $5 bills and two $1 bills. Two bills are taken at random from the box without replacement. What is the probability that both bills will be of the same denomination?

4. A movie theatre is curious about how many of its patrons buy food, how many buy a drink, and how many buy both products. They track 300 people through the concessions stand one evening, out of the 300, 78 buy food only, 113 buy a drink only and the remainder buy both. What is the probability that a patron does not buy a drink if they have already bought food?

5. A randomly taken phone number contains 6 digits. What is the probability that all digits are: a) different b) odd?

6. A box contains five green balls, three black balls, and seven red balls. Two balls are selected at random without replacement from the box. What is the probability that both balls are of the same color?

7. We accidentally choose a point in a ball, in which a cube is inscribed. Find the geometrical probability that an accidentally chosen point does not belong to the cube.

8. A student has planned to take money out of the bank. It is possible that he comes to the bank in the interval of timefrom 14:15 to 14:25. It is also known that the robbery of this bank is planned in the same interval of time and it will continue for 4 minutes. Find the probability that the student will be in the bank at the moment of the robbery.

1.3 Axioms of Probability Theory. Corollaries

We state axioms of the probability theory on the base of statistic definition of probability ($P(A) \approx P_N^*(A)$ for a large number N of trials).

1. If A is an *impossible* event, then its probability equals zero, $P(\varnothing) = 0$ ($\varnothing$ is impossible event).

2. If A is a *certain* event, then its probability equals unity, $P(\Omega) = 1$ (Ω is certain event).

3. If A is a *random* event, then its probability is contained between zero and unity, $0 \le P(A) \le 1$ (A is random event).

4. If A and B are two *disjoint* events, then the probability of their sum is equal to

the sum of probabilities of these events, $P(A+B)=P(A)+P(B)$ (A, B are disjoint events).

Definition A ***conditional probability*** of an event, namely $P(B/A)$ is the probability of an event B by condition that an event A occurs. Analogous is the probability of A if B occurs.

5. Probability of a product of two events equals the product of the probability of one event and the condition probability of the other, $P(AB)=P(A)\cdot P(B/A)=P(B)\cdot P(A/B)$.

Example 1 An urn contains 3 white and 2 black balls. One takes two balls successively and at random. Find the probability that they are white.

Solution Let us denote an event which means that two drawn balls are white by A. Also let us denote the events which mean that the first and the second drawn balls are white by B and C respectively. It is evident that $A=BC$, hence, by virtue of the fifth axiom (and classic definition of probability)

$$P(BC)=P(B)\cdot P(C/B)=P(C)\cdot P(B/C)=\frac{3}{5}\cdot\frac{2}{4}=0.3.$$

Some corollaries

1. If some events A, B, C are *pairwise disjoint events*, then

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

If, moreover, they form *a total group*, then

$$P(A+B+C)=P(A)+P(B)+P(C)=1.$$

2. The sum of probabilities of two *opposite events* equals 1,

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

3. Probabilities of a product of three, four etc. events are equal to

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

4. For two arbitrary events A and B the probability of their sum equals

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

5. If A, B are *independent events*, then $P(B/A)=P(B)$, $P(A/B)=P(A)$ and so

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

that is the probability of a product of two *independent events* is equal to the product of their probabilities.

6. If A, B, C are *mutually independent events*, then

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

Example 2 To pass an exam successfully a student has to know the proofs of 50 theorems but he knows only 40 of them. What is the probability for him to pass an exam if exam tasks contain 3 theorems?

Solution Let A be an event which means that a student will pass an exam. Let us introduce the following three auxiliary events: B_1 that is a student knows the proof of the first theorem, B_2 - of the second, B_3 - of the third. Then by the fourth corollary (and classic definition of probability)

$$A = B_1B_2B_3 \Rightarrow P(A) = P(B_1B_2B_3) = P(B_1)\cdot P(B_2 / B_1)\cdot P(B_3 / B_1B_2);$$

$$P(A) = \frac{40}{50}\cdot\frac{39}{49}\cdot\frac{38}{48} \approx 0.5.$$

Example 3 Three independently working engines are installed in a workshop. Probabilities to work at a given time equal for them 0.6, 0.9, 0.7 respectively. Find probabilities of the following events: a) only one engine works; b) at least one engine works.

Solution Let an event A means that only one engine works and an event B means that at least one engine works. Our problem is to find the probabilities of these events.

Let us introduce three auxiliary events, namely C_1 which means that the first engine works, C_2 - the second engine works and C_3 - the third engine works. By conditions of the problem

$$P(C_1) = 0.6, P(\overline{C_1}) = 1 - P(C_1) = 1 - 0.6 = 0.4;$$

$$P(C_2) = 0.9, P(\overline{C_2}) = 1 - P(C_2) = 1 - 0.9 = 0.1;$$

$$P(C_3) = 0.7, P(\overline{C_3}) = 1 - P(C_3) = 1 - 0.7 = 0.3.$$

a) The event A can be represented as the sum of products

$$A = C_1\overline{C_2C_3} + \overline{C_1C_2}C_3 + \overline{C_1}C_2\overline{C_3}$$

with pairwise disjoint summands and independent factors in every summand. Hence the probability of the event A equals

$$P(A) = P(C_1\overline{C_2C_3} + \overline{C_1C_2}C_3 + \overline{C_1}C_2\overline{C_3}) = P(C_1\overline{C_2C_3}) + P(\overline{C_1C_2}C_3) + P(\overline{C_1}C_2\overline{C_3});$$

$$P(A) = 0.6\cdot 0.1\cdot 0.3 + 0.4\cdot 0.9\cdot 0.3 + 0.4\cdot 0.1\cdot 0.7 = 0.154.$$

b) To find the probability of the event B we will evaluate at first the probability of its opposite one $\overline{B}$ (which means that all three engines do not work, $\overline{B} = \overline{C_1}\,\overline{C_2}\,\overline{C_3}$).

We will obtain $P(\overline{B}) = P(\overline{C_1}\,\overline{C_2}\,\overline{C_3}) = 0.4\cdot 0.1\cdot 0.3 = 0.012$ whence it follows that

$$P(B) = 1 - P(\overline{B}) = 1 - 0.012 = 0.988.$$

Exercise Set 1.3

1. The probability that the student will pass the examinations is equal to 0.8 for the first one, 0.7 - for the second one, 0.65 - for the third one. Find the probability that the student will pass: a) two examinations; b) not less than two examinations; c) at least one examination.

2. A shooter fires at the target three times. The probabilities of hitting the target are respectively equal to p_1, p_2, p_3. Find the probabilities of the following events: 1) hitting the target three times; 2) hitting the target not less than two times; 3) hitting the target at least once.

	p_1	p_2	p_3
A	0.8	0.85	0.9
B	0.7	0.8	0.9
C	0.9	0.75	0.8

3. How many times is it necessary to flip a coin so that the probability of at least one occurrence of a "tail" is greater then 0.625?

4. A drawer contains 4 black, 6 brown, and 8 olive socks. Two socks are selected at random from the drawer. (a) What is the probability that both socks are of the same color? (b) What is the probability that both socks are olive if it is known that they are of the same color?

5. Flip a coin and then independently cast a die. What is the probability of observing heads on the coin and a 2 or 3 on the die?

6. Three shooters alternately fire at the same target. Each shooter has two rounds; at the first hit, the shooting stops. With one shot, the probability of hitting a target is 0.2 for the first shooter, 0.3 for the second, and 0.4 for the third. Find the probability that all three shooters will use up all their ammunition.

7. There are 10 parts in the box, of which four are painted. The collector took three details at random. Find the probability that at least one of the taken parts is colored.

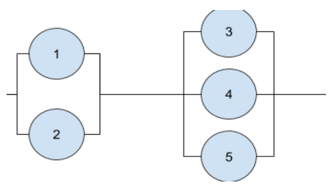
8. The probability of one hitting the target with one volley of two guns is 0.38. Find the probability of hitting the target with one shot the first of the guns, if for the second gun this probability is equal to 0.8.

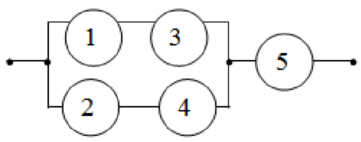
9. A card is drawn at random from an ordinary deck of 52 cards and replaced. This is done a total of 5 independent times. What is the conditional probability of drawing the ace of spades exactly 4 times, given that this ace is drawn at least 4 times?

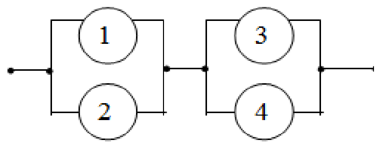
10. An urn contains 4 balls numbered 0 through 3. One ball is selected at random and removed from the urn and not replaced. All balls with nonzero numbers less than that of the selected ball are also removed from the urn. Then a second ball is selected at random from those remaining in the urn. What is the probability that the second ball selected is numbered 3?

11. A box contains 2 green and 3 white balls. A ball is selected at random from the box. If the ball is green, a card is drawn from a deck of 52 cards. If the ball is white, a card is drawn from the deck consisting of just 16 pictures. (a) What is the probability of drawing a king? (b) What is the probability of a white ball was selected given that a king was drawn?

12. A circuit path includes 5 elements working independently from each other with corresponding probabilities of failure: $p_1 = 0.3$; $p_2 = 0.6$; $p_3 = 0.5$; $p_4 = 0.4$; $p_5 = 0.2$. Calculate the probability of circuit failure.





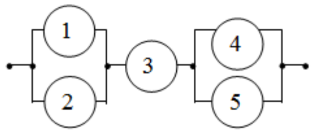


Individual Tasks 1.3

I.

1. The first worker makes 40% of the second-class parts, and the second makes 30%. Two parts are taken from each worker at random. Find the probability that: a) all the four parts are second-class; b) at least three parts are second-class; c) less than three parts are second-class.

2. A circuit path includes 5 elements working independently from each other with corresponding probabilities of failure: $p_1 = 0.3$; $p_2 = 0.6$; $p_3 = 0.5$; $p_4 = 0.4$; $p_5 = 0.2$. Calculate the probability of a circuit failure.



3. How many times is it necessary to flip a coin so that the probability of at least one occurrence of a "head" is greater then 0.875?

4. If we randomly pick two television sets in succession from a shipment of 240 television sets of which 15 are defective, what is the probability that they both will be defective?

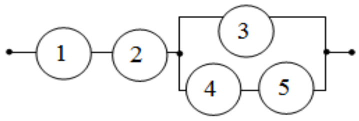
5. An urn contains 3 red, 2 white and 4 yellow balls. An ordered sample of size 3 is drawn from the urn. If the balls are drawn with replacement so that one outcome does not change the probabilities of others, then what is the probability of drawing a sample that has the balls of each color? Also, find the probability of drawing a sample that has two yellow balls and a red ball or a red ball and two white balls?

6. A family has five children. Assuming that the probability of a girl on each birth was 0.5 and that the five births were independent, what is the probability the family has at least one girl, given that they have at least one boy?

II.

1. A worker services three machine tools. The probability that during his shift the machine tools will claim his attention is equal to 0.7 for the first one, 0.65 for the second one, 0.55 for the third one. Find the probability that during his shift his attention will be claimed by: a) two machine tools; b) not less than two machine tools; c) at least one machine tool.

2. A circuit path includes 5 elements working independently from each other with corresponding probabilities of failure: $p_1 = 0.3$; $p_2 = 0.6$; $p_3 = 0.5$; $p_4 = 0.4$; $p_5 = 0.2$. Calculate the probability of a circuit failure.



3. The probability of hitting a plane with one shot from a rifle 0.04. How many shooters should shoot at the same time so that the probability of hitting the plane is more than 70%?

4. A box of fuses contains 20 fuses, of which 5 are defective. If 3 of the fuses are selected at random and removed from the box in succession without replacement, what is the probability that all three fuses are defective?

5. An urn contains 6 red balls and 3 blue balls. One ball is selected at random and is replaced by a ball of the other color. A second ball is then chosen. What is the conditional probability that the first ball selected is red, given that the second ball was red?

6. A small grocery store had 10 cartons of milk, 2 of which were sour. If you are going to buy the 6th carton of milk sold that day at random, find the probability of selecting a carton of sour milk.

1.4 Formulae of Total Probability and Bayes

In practice we often deal with the following situation. An event A can occur only together with one of pairwise disjoint event $H_1, H_2, ..., H_n$, which form a total group. Let us call these events ***hypotheses*.** Their probabilities and corresponding conditional probabilities of the event A are known. In this case the probability of the event A can be found with the help of the following formula (***the formula of total probability***):

$$P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+...+P(H_n)P(A/H_n) \qquad (1)$$

$$(P(H_1)+P(H_2)+...+P(H_n)=1)$$

Bayes formulae

Let an event A, which can occur only together with one of given hypotheses $H_1, H_2, ..., H_n$, occurs. In this case the following probabilities $P(H_k / A)$ of its occurrence together with each of these hypotheses can be evaluated with the help of the known Bayes formulas

$$P(H_k / A)=\frac{P(H_k)P(A/H_k)}{P(A)},\ k=\overline{1,n} \qquad (2)$$

Bayes formulae (2) state the probability that namely the k-th hypothesis has occurred together with the event A in question.

Example 1 There are 6 white and 2 black balls in the first urn and 8 white and 3 black balls in the second urn. One moves a ball from the first urn to the second one at random, and then he takes a ball from the second urn (also at random).

1. Find the probability for him to take a white ball from the second urn.

2. Let a white ball be taken from the second urn. A ball of which color was most probably moved from the first urn?

Solution

1. Let an event A means that one will take a white ball from the second urn. We can introduce the following two hypotheses: H_1 means that one has moved a white ball from the first urn; H_2 that he has moved a black ball from there. Their probabilities equal

$$P(H_1)=\frac{6}{8}=\frac{3}{4};\ P(H_2)=\frac{2}{8}=\frac{1}{4}$$

by condition, and corresponding conditional probabilities of the event A equal

$$P(A/H_1)=\frac{9}{12}=\frac{3}{4};\ P(A/H_2)=\frac{8}{12}=\frac{2}{3}.$$

On the base of the formula (1) of total probability the probability of the event A equals

$$P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)=\frac{3}{4}\cdot\frac{3}{4}+\frac{1}{4}\cdot\frac{2}{3}=0.73.$$

2. To solve the second problem we must find and compare the following conditional probabilities $P(H_1/A)$, $P(H_2/A)$. On the base of Bayes formulae (2)

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{P(A)}=\frac{\frac{3}{4}\cdot\frac{3}{4}}{0.73}\approx 0.77;$$

$$P(H_2/A)=\frac{P(H_2)P(A/H_2)}{P(A)}=\frac{\frac{1}{4}\cdot\frac{2}{3}}{0.73}\approx 0.23.$$

We see that 0.77 > 0.23, therefore one has the most probably moved a white ball from the first urn to the second one.

Exercise Set 1.4

1. Articles from three conveyors enter for assembling. The number of articles given for assembling is $\alpha\%$ for the first one, $\beta\%$ for the second one, $\gamma\%$ for the third one. On the average, the number of defective articles is $\delta_1\%$ from the first conveyor; $\delta_2\%$ from the second conveyor; $\delta_3\%$ from the third conveyor. Find the probability that a defective part has entered for assembling. What is the probability that the part from the i-th of the conveyor is defective?

	$\alpha\%$	$\beta\%$	$\gamma\%$	$\delta_1\%$	$\delta_2\%$	$\delta_3\%$	i
A	30	15	55	2	2	3	3
B	40	40	20	1	3	2	1
C	50	30	20	2	4	3	2
D	20	45	35	3	5	2	3

2. The number of trucks passing by a gas station is refered to the number of cars as 3:2. The probability that a truck will refuel equals 0.1; the probability that a car will refuel equals 0.2. A car drove up for refueling to the gas station. Find the probability that it is the truck.

3. A cable message consists of the signals “dot” and “dash”. The number of their occurrences is refered as 5:3. On average 2/5 message of the signals “dot” and 1/3 message of the signals “dash” are distorted. Find the probability that: a) the transmitted signal is accepted; b) the accepted signal is “dash”.

4. The first urn contains 10 balls, among them 8 are white; the second urn contains 20 balls, among them 4 white. We pick one ball from each urn at random.

Then we picked one from these two balls randomly. Find the probability that the chosen ball is white.

5. In a certain village, 20% of the population has some disease. A test is administered which has the property that if a person is sick, the test will be positive 90% of the time and if the person is not sick, then the test will still be positive 30% of the time. All people tested positive are prescribed a drug which always cures the disease but produces a rash 25% of the time. Given that a random person has the rash, what is the probability that this person had the disease to start with?

6. An insurance company considers that people can be split in two groups: those who are likely to have accidents and those who are not. Statistics show that a person who is likely to have an accident has probability 0.4 to have one over a year; this probability is only 0.2 for a person who is not likely to have an accident. We assume that 30% of the population is likely to have an accident. (a) What is the probability that a new customer has an accident over the first year of his contract? (b) A new customer has an accident during the first year of his contract. What is the probability that he belongs to the group likely to have an accident?

7. The probability of shooting the game for the first hunter is equal to 0.8. The same probability for the second hunter is 0.7. The beast was shot with simultaneous shots. The mass of the game was 190 kg. It was found that the game was killed with one bullet. How should the game be divided between hunters?

8. Two shots shoot a target. The probability that the first shot will hit the shooting mark is equal to 0.9. The analogous probability for the second shot is 0.7. Find the probability that the target will be hit by both shots.

9. 100 and 200 details are produced in plants I and II, respectively. The probabilities of the producing of a standard detail in plants I and II are equal to 0.9 and 0.8, respectively. a) A damage caused by the realization of non-standard details made up $3000. Find a fine which must be payed by the administration of plant II caused by the realization of its non-standard details. b) The profit received by the realization of standard details made up $5000. Find a portion of the profit due to plant I.

Individual Tasks 1.4

I.

1. 3 white and 3 black balls are placed in urn I, 3 white and 4 black balls - in urn II and 4 white and 1 black balls - in urn III. We accidentally choose a box and further accidentally choose a ball from this urn. What is the probability that an accidentally chosen ball will be white if the probability of a choice of any urn is equal to $1/3$?

2. The entire output of a factory is produced on three machines. The three machines account for 40%, 35%, and 25% of the output, respectively. The fraction of defective items produced is this: for the first machine, 0.2%; for the second machine, 0.3%; for the third machine, 0.5%. What is the probability that: a) the randomly chosen item being defective b) the defective item was produced by the second machine?

3. A transmitting system transmits 0’s and 1’s. The probability of a correct transmission of a 0 is 0.8, and it is 0.9 for a 1. We know that 45% of the transmitted symbols are 0's. (a) What is the probability that the receiver gets a 0? (b) If the receiver gets a 0, what is the probability the transmitting system actually sent a 0?

4. Sixty percent of new drivers have had a driver's education. During their first year, new drivers without driver education have probability 0.08 of having an accident, but new drivers with driver education have only 0.05 probability of an accident. What is the probability a new driver has had driver education, given that the driver has had no accident the first year?

5. A diagnostic test for a certain disease is said to be 90% accurate in that, if a person has the disease, the test will detect with probability 0.9. Also, if a person does not have the disease, the test will report that he or she does not have it with probability 0.9. Only 1% of the population has the disease in question. If the diagnostic test reports that a person chosen at random from the population has the disease, what is the conditional probability that the person, in fact, has the disease?

II.

1. Two machines produce the items that go into the overall conveyor. The probability of an item produced being defective are 0.075 and 0.08 correspondingly for the first machine and for the second. The productivity of the second machine is triple times greater than the first. Find the probability of the item being defective.

2. There are 20 shooters in a group. Among them there are 5 excellent, 9 good and 6 satisfactory shooters. The probability of the excellent shooter hitting the target with one shot equals 0.9; of the good shooter - 0.8 and of the satisfactory shooter - 0.7. A random shooter fired twice. One hit and one miss are observed. Who shot with the most probability?

3. 46% of the electors of a town consider themselves as independent, whereas 30% consider themselves democrats and 24% republicans. In a recent election, 35% of the independents, 62% of the democrats and 58% of the republicans voted. (a) What proportion of the total population actually voted? (b) A random voter is picked. Given that he voted, what is the probability that he is independent? democrat? republican?

4. One-half percent of the population has AIDS. There is a test to detect AIDS. A positive test result is supposed to mean that you have AIDS but the test is not perfect. For people with AIDS, the test misses the diagnosis 2% of the times. And for the people without AIDS, the test incorrectly tells 3% of them that they have AIDS. (a) What is the probability that a person picked at random will have the positive test? (b) What is the probability that you have AIDS given that your test comes back positive?

5. English and American spelling are *rigour* and *rigor*, respectively. A man staying at Al Rashid hotel writes this word, and a letter taken at random from his spelling is found to be a vowel. If 40 percent of the English-speaking men at the hotel are English and 60 percent are American, what is the probability that the writer is an Englishman?

1.5 Scheme of Independent Trials

Definition A ***binomial experiment*** is an experiment that has the following properties:

- The experiment consists of n repeated trials.
- Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other - a failure.
- The probability of success, denoted by p is the same on every trial.
- The trials are independent; that is, the outcome on of trial does not affect the outcome on other trials.

Binomial Formula (Bernoulli Scheme) Suppose a binomial experiment consists of n trials and results in m successes. If the probability of success on an individual trial is p, then the binomial probability is:

$$P_n(m) = C_n^m p^m q^{n-m} \tag{1}$$

Formula (1) is called ***Bernoulli Formula.***

Example 1 Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

$$P_5(2) = C_5^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0.161.$$

Note 1 For $m = 0,1,2,...$ the value of the probability (1) first increases and then decreases, and therefore there exists the so-called ***most probable number*** m_0 of successes. It is defined by the next double inequality

$$np - q \le m_0 \le np + p \ , \ P_n(m_0) = \max \tag{2}$$

The number m that satisfies (2) is known as the *most probable* (*most likely*) number of successes in n Bernoulli trials. It is also always different from the *average number* of successes.

Note 2 Probability of no less than k_1 and no greater than k_2 successes equals

$$P_n(k_1 \le m \le k_2) = P_n(k_1) + P_n(k_1 + 1) + P_n(k_1 + 2) + ... + P_n(k_2) \tag{3}$$

Example 2 The probability that a student is accepted to a prestigious college is 0.3. If 5 students from the same school apply, what is the probability that at most 2 are accepted?

Solution To solve this problem, we compute 3 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,

$$P_5(k \le 2) = P_5(0) + P_5(1) + P_5(2) = 0.8369.$$

Example 3 6 independently working engines are installed in a shop. Probability for any engine to work at a given moment is 0.8. Find the probabilities of at least one engine to work, of no less than 2 and no greater than 5 engines to work and the most probable number of working engines at this moment.

Solution We can consider setting of an engine as a trial. So we have n = 6 independent trials. Let a success A mean that an engine works.

$$p = P(A) = 0.8, P(\overline{A}) = 1 - p = q = 0.2.$$

Let a success B means that at least one engine works. Using the axioms of probability theory gives us the following formula to calculate the probability of B

$$P(B) = 1 - P_6(0) = 1 - q^6 = 0.99994.$$

The most probable a number of working engines at this moment can be evaluated by the formula (2)

$$np - q \le m_0 \le np + p \ \Rightarrow \ 6 \cdot 0.8 - 0.2 \le m_0 \le 6 \cdot 0.8 + 0.8$$

whence it follows that

$$4.6 \le m_0 \le 5.6 \Rightarrow m_0 = 5.$$

Probability of no less than 2 and no greater than 5 engines to work on the base of the formula (3) equals

$$P_6(2 \le m \le 5) = P_6(2) + P_6(3) + P_6(4) + P_6(5) =$$
$$= C_6^2 p^2 q^4 + C_6^3 p^3 q^3 + C_6^4 p^4 q^2 + C_6^5 p^5 q^1 =$$
$$= 0.01536 + 0.08192 + 0.24576 + 0.39321 + 0.73625 \approx 0.74.$$

Bernoulli formula (1) is not convenient for a large number n of trials. There are some approximate formulas.

Poisson Formula

Let us suppose that the number n of trials tends to infinity, the probability p of a success A goes to zero, but a product np retains constant,

$$n \to \infty, p \to 0, np = const = \lambda.$$

Given the mean number of successes λ that occur in a specified region ($np \le 10$), we can compute the Poisson probability based on the following formula.

Poisson Formula. Suppose we conduct a Poisson experiment, in which the average number of successes within a given region is λ. Then, the Poisson probability is:

$$P_n(m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!}, \ \ \lambda = np \qquad (4)$$

where m is the actual number of successes that result from the experiment.

Example 4 The average number of houses sold by the Acme RealtyCompany is 2 houses per day. What is the probability that exactly 3 houses will be sold tomorrow?

Solution This is a Poisson experiment in which we know the following:

- $\lambda = 2$; since 2 houses are sold per day, on average.
- $m = 3$; since we want to find the likelihood that 3 houses will be sold tomorrow.

We plug these values into the Poisson formula as follows: $P_n(3) = \dfrac{2^3 \cdot e^{-2}}{3!} = 0.18$.

Thus, the probability of selling 3 houses tomorrow is 0.180.

A ***cumulative Poisson probability*** refers to the probability that the Poisson random variable is greater than some specified lower limit and less than some specified upper limit.

Example 5 Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that tourists will see fewer than four lions on the next 1-day safari?

Solution This is a Poisson experiment in which we know the following:

- $\lambda = 5$; since 5 lions seen on a 1-day safari, on average.
- $m = 0,1,2,3$; since we want to see fewer than four lions on the next 1-day safari.

To solve this problem, we need to find the probability that tourists will see 0, 1, 2, or 3 lions. Thus, we need to calculate the sum of four probabilities: $P(m=0)+P(m=1)+P(m=2)+P(m=3)$. To compute this sum, we use the Poisson formula:

$$P_n(m \le 3) = P(0)+P(1)+P(2)+P(3) = 0.265.$$

Thus, the probability of seeing at no more than 3 lions is 0.265.

Laplace Local and Integral Theorems

Laplace local theorem gives an approximate value of the probability $P_n(m)$ of m successes in n independent trials (with constant probability p of the success in any trial). Given the mean number of successes np that occur in a specified region $np > 10$, we can compute the probability based on the following formula. Namely, for large n we can substitute Bernoulli formula (1) by the next approximate one:

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x) \qquad (5)$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{and} \quad x = \frac{m-np}{\sqrt{npq}} \qquad (6)$$

is so-called *small Laplace function*. The function is even one, that is $\varphi(-x)=\varphi(x)$ and therefore its graph is symmetric with respect to the *Oy*-axis. The function is tabulated one (see Appendix table 1). We can suppose $\varphi(x)=0,\ |x|\ge 4$.

Integral Laplace Theorem gives an approximate value of the probability of hitting of values of Bernoulli scheme in a segment $[m_1, m_2]$. Namely for large n we can substitute the exact formula (5) by the following approximate formula:

$$P_n(m_1 \le m \le m_2) = \Phi(x_2) - \Phi(x_1) \qquad (7)$$

where $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$, $x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$ and a function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \qquad (8)$$

is the so-called ***Laplace function*** (or ***normed Laplace function***). The function is an odd one, that is $\Phi(-x) = -\Phi(x)$. Laplace function is tabulated (see Appendix table 2). We can suppose

$$\Phi(x) = 0.5 \ , \ x \ge 5 .$$

Example 6 The probability of the appearance of an event in each of 245 independent trials is constant and equal to 0.25. Find the probability that an event will begin exactly 50 times.

Solution Using Laplace local and integral theorems we have

$n = 245, m = 50, p = 0.25, q = 1 - p = 0.75$

$$x = \frac{50 - 245 \cdot 0.25}{\sqrt{245 \cdot 0.75 \cdot 0.25}} = \frac{-11.25}{\sqrt{45.9375}} = -\frac{11.25}{6.778} \approx -1.66 .$$

The function is an even one $\varphi(-1.66) = \varphi(1.66)$.

$$P_{245}(50) \approx \frac{1}{\sqrt{npq}} \varphi(1.66) = \frac{0.1006}{6.778} \approx 0.0148 .$$

Example 7 The probability of the appearance of an event in each of 245 independent trials is constant and equal to 0.25. Find the probability that an event will begin not less than 45 times and not more than 60 times.

Solution Using integral Laplace theorem we have

$n = 245, m_1 = 45, m_2 = 60, p = 0.25, q = 1 - p = 0.75$

$$x_1 = \frac{45 - 245 \cdot 0.25}{\sqrt{245 \cdot 0.75 \cdot 0.25}} = \frac{-16.25}{\sqrt{45.9375}} \approx -2.40 ,$$

$$x_2 = \frac{60 - 245 \cdot 0.25}{\sqrt{245 \cdot 0.75 \cdot 0.25}} = \frac{-1.25}{\sqrt{45.9375}} \approx -0.18 .$$

The function is an odd one. Using the table of Laplace function, we get

$$\Phi(-2.40) = -\Phi(2.40) \approx -0.4918 ,$$

$$\Phi(-0.18) = -\Phi(0.18) \approx -0.0714 ,$$

$$P_{245}(45 \le m \le 60) \approx \Phi(-0.18) - \Phi(-2.40) = -0.0714 + 0.4918 = 0.4204 .$$

The probability of the deviation of the relative frequency of an event A ***from its probability*** $p = P(A)$ (in n independent trails with constant probability $p = P(A) = const$ of the event) can be found by the following formula

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

Exercise Set 1.5

1. Dice are thrown five times. Find the probability that a) three points fall four times; b) three points will appear at least once.

2. In the new district 10,000 code locks were installed at the entrance doors of the houses. The probability of failure of one lock during the month is 0.0002. Find the probability that two locks will fail during in a month.

3. The probability of manufacturing a product of highest quality is 0.9. 50 products were manufactured. Find the most probable number of products of the highest quality and its probability.

4. The machine produces of 10,000 parts per shift. The probability of manufacturing a defective part is $p = 0.0001$. Find the probability that a) three; b) from four to six; c) at least one the defective part will be manufactured for a shift.

5. What is the probability that from 1500 to 1600 lamps will be lit from 2450 lamps illuminating the street by the end of the year? Assume that each lamp will be lit during the year with the probability of 0.64.

6. The probability of manufacturing a defective part is 0.004. Find the probability that among 1000 parts 5 will be defective.

7. The probability of the birth of a girl equals 0.485. Find the probability that of 600 children born: a) 300 will be girls; b) there will be more girls than boys.

8. The probability that a part is defective equals 0.1. How many parts need to be selected so that with a probability of 0.9544 it can be argued that the relative frequency of occurrence of defective parts deviates from the probability in absolute value by no more than 0.03?

9. Industrial television installation contains 2000 transistors. The probability of failure of each transistor is 0.0006. Find the probability of failure of two to five transistors.

10. The probability that a product of the highest quality is taken is 0.5. Determine the probability that 500 out of 1000 products are of the highest quality.

11. The probability of an event occurring in each of the 2000 independent trials is 0.7. Find the probability that an event will occur from 1350 to 1500 times.

12. The probability of the appearance of an event in each of n independent trials is constant and equal to p. Find the probability that an event will begin exactly m times. Find the probability that an event will begin not less than m_1 times and not more than m_2 times.

	n	m	m_1	m_2	p
A	*144*	*120*	*115*	*125*	*0.8*
B	*110*	*18*	*15*	*20*	*0.15*
C	*220*	*140*	*130*	*145*	*0.6*

Individual Tasks 1.5

I.

1. The worker serves 12 machines of the same type. The probability that the machine will require the attention of a worker during a shift is 1/3. Find the probability that during the shift from 3 to 6 machines will require the attention of the worker.

2. The probability of production of a defective part is equal to 0.008. Find the probability that out of 1000 parts taken for inspection there will be 10 defective ones.

3. The manufacturer sent to the base 12000 good-quality products. The number of products damaged during transportation averages 0.05%. Find the probability that: a) no more than three damaged products; b) at least two damaged products will go to the base.

4. The probability of failure of each device during the testing is 0.2. Instruments are tested independently of each other. Find the probability of failure of 10 devices during the testing 80 of them.

5. Find the probability of simultaneous shutdown of 30 machines out of 100 working, if the probability of stopping for each machine is 0.2.

6. How many experiments need to be made so that with a probability of 0.9 one could argue that the frequency of event A will differ from the absolute value by no more than 0.1?

7. The probability of hitting the target with one shot is 0.9. Find the probability that at 100 shots the target will be hit from 84 to 96 times.

II.

1. Coin is thrown 5 times. Find the probability that a "tail" will fall: a) 2 times; b) less than two times; c) at least three times.

2. The machine consists of 2,000 independently operating units. The probability of failure of one node during the year is 0.0005. Find the probability of failure within a year from two to four operating units.

3. 1000 seeds were sowed. The probability of not germinating for each seed is 0.002. Find the probability that: a) 10 seeds do not germinate; b) all seeds germinate.

4. The probability of an event occurring in each of the independent trials is 0.25. Find the probability that an event will occur 50 times in 243 trials.

5. Find the probability of hitting the target 75 times with 90 shots, if the probability of hitting the target with one shot is 0.8.

6. The probability of an event occurring in each of the 900 independent tests is 0.5. Find the probability that the relative frequency of occurrence of an event deviates from its probability in absolute value by no more than 0.02.

7. Find the probability that at 400 trials an event will appear from 90 to 180 times, if the probability of occurrence in one trial is 0.4.

1.6 Random Variables

A Random Variable

Definition A ***random variable*** is a variable which takes on some value in any trial and this value is not known beforehand [in advance].

We will denote random variables by letters $X, Y, Z, ...$ and their possible values by $x, y, z, \ldots$. We will study *discrete* and *continuous* random variables.

Definition A random variable is called ***discrete*** if it can take on only separate isolated possible values (with some probabilities).

Example 1 A number X of occurrences of an event A in one trial: $X = 1$ if A occurs, and $X = 0$, if A does not occur (that is an opposite event $\overline{A}$ occurs);

$$P(X = 1) = P(A), P(X = 0) = P(\overline{A}) = 1 - P(A).$$

Example 2 The number of students of the lecture and the daily production of some factory (in items) are examples of discrete random variables.

Definition of a ***continuous*** random variable will be given below. Now we will only say that its possible values fill some interval completely.

Example 3 The human height and weight, the size of an item, the error of measurement are examples of continuous random variables.

Certain frequently occurring random variables and their distributions have special names (*Bernoulli Random Variables, Binomial Random Variables, Discrete Uniform Random Variables, Geometric Random Variables*).

Definition The ***distribution*** (*the distribution law, the distribution function, the law of distribution, the law*) of a random variable is a rule which sets a correspondence between its possible values and corresponding probabilities.

Definition The distribution table of a discrete random variable X (with finite number n of possible values) is called table which first row contains all possible values of the random variable, and the second row contains corresponding probabilities of these values. The distribution table has the following form:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

The notation $X = x_i$ means that the random variable X takes on a value x_i. Events $(X = x_1)$, $(X = x_2)$,…,$(X = x_n)$ are pairwise disjoint and form a total group. Therefore, the sum of their probabilities $\sum p_i = 1$.

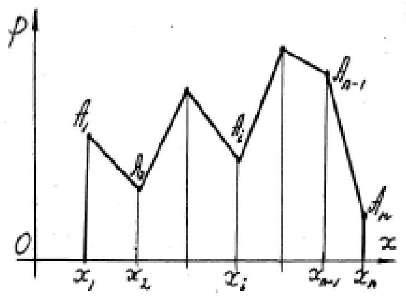


Figure 3

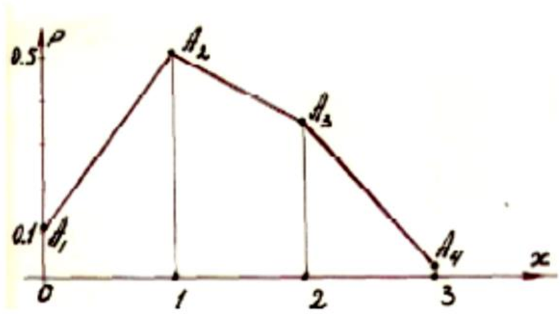


Figure 4

Definition The ***distribution polygon*** of a discrete random variable is a broken line (a polygonal line, an open polygon) which is generated by the successive joining of the points

$$A_1(x_1, p_1),\ A_2(x_2, p_2),\ \dots,\ A_n(x_n, p_n) \text{ (See Figure 3).}$$

Example 4 An urn contains 7 balls (namely 3 white and 4 black balls). One draws 3 balls at random. Find the distribution law of the number X of white balls which can be taken from the urn.

Solution Possible values of the random variable X are 0, 1, 2, 3. We determine corresponding probabilities

$$p_1 = P(X = 0),\ p_2 = P(X = 1),\ p_3 = P(X = 2),\ p_4 = P(X = 3)$$

with the help of the classical definition of probability. Elementary events (chances) for every of these four cases are sets of 3 balls that is 3-fold combinations of 7 elements. Hence the general number of chances equals

$$n = C_7^3 = \frac{7!}{3!\cdot 4!} = 35.$$

Numbers of favorable chances are represented in the table

Event	*Number of favorable chances*	*Explication*
$X = 0$	$m_1 = C_4^3 = C_4^1 = 4$	One can take 0 white (and so 3 black) balls by the number of 3-fold combinations of 4 elements
$X = 1$	$m_2 = 3 \cdot C_4^2 = 18$	One can take 1 white ball by 3 ways and 2 black balls by C_4^2 ways
$X = 2$	$m_3 = 4 \cdot C_3^2 = 12$	One can take 2 white balls by C_3^2 ways and 1 black ball by 4 ways
$X = 3$	$m_4 = 1$	One can take 3 white balls in one unique way

Numbers m_2, m_3 are calculated with the help of the main principle of combinatorics. The distribution law of the random variable X is represented by the following distribution table:

X	0	1	2	3
p_i	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$

The sum of obtained probabilities equals 1: $\sum_{i=1}^{4} p_i = \frac{4+18+12+1}{35} = 1$.

The most probable value of the random variable is $X = 1$. The distribution polygon of the random variable X is shown on Figure 4.

The Distribution Function of a Random Variable

The distribution function is the most general form of the distribution law of a random variable X.

Definition The ***distribution function*** of a random variable X is called a function:

$$F(x) = P(X < x) = P(-\infty < X < x) \qquad (1)$$

The distribution function $F(x)$ is the probability for the random variable X to take on values which are less than x or the probability of hitting of the random variable in the infinite interval (the half-axis) $(-\infty, x)$.

Properties of the distribution function

1. The distribution function, being a probability, ranges from 0 to 1: $0 \le F(x) \le 1$.

2. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$.

3. $F(x_1) \le F(x_2)$, if $x_1 < x_2$.

4. $P(\alpha < X < \beta) = F(\beta) - F(\alpha)$ (2).

Example 5 From the urn, which contains 3 white and 5 black spheres, 3 spheres are extracted. Let random variable X be the number of taken out black spheres. Find the distribution law. Plot the graph of the function of distribution.

Solution Possible values of the random variable X are 0, 1, 2, 3. We determine corresponding probabilities

$$p_1 = P(X=0),\ p_2 = P(X=1),\ p_3 = P(X=2),\ p_4 = P(X=3)$$

with the help of the classical definition of probability.

$$p_1 = P(X=0) = \frac{C_3^3}{C_8^3} = \frac{1\cdot 2\cdot 3}{8\cdot 7\cdot 6} = \frac{1}{56},\ p_2 = P(X=1) = \frac{C_5^1\cdot C_3^2}{C_8^3} = \frac{5\cdot 3}{56} = \frac{15}{56},$$

$$p_3 = P(X=2) = \frac{C_5^2\cdot C_3^1}{C_8^3} = \frac{5\cdot 4\cdot 3}{2\cdot 56} = \frac{15\cdot 2}{56} = \frac{30}{56},$$

$$p_4 = P(X=3) = \frac{C_5^3}{C_8^3} = \frac{5\cdot 4\cdot 3}{2\cdot 3}\cdot\frac{1}{56} = \frac{10}{56}.$$

The distribution law of the random variable X is represented by the following distribution table:

X	0	1	2	3
P	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

The distribution function of the random variable X, namely of the number of taken out black spheres is given as follows.

If $x\in(-\infty;1]$, then $F(x) = P\left(X=0\right) = 1/56$.

If $x\in(1;2]$, then $F(x) = P(X=0)+P(X=1) = \frac{1}{56}+\frac{15}{56} = \frac{16}{56}$.

If $x\in(2;3]$, then $F(x) = P\left(X=0\right)+P\left(X=1\right)+P\left(X=2\right) = \frac{16}{56}+\frac{30}{56} = \frac{46}{56}$.

If $x\in(3;+\infty)$, then $F(x) = P\left(X=0\right)+P\left(X=1\right)+P\left(X=2\right)+P\left(X=3\right) = 1$.

$$F(x) = \begin{cases} 0, & if\quad x\le 0,\\ \frac{1}{56} = 0.018, & if\quad 0<x\le 1,\\ \frac{16}{56} = 0.286, & if\quad 1<x\le 2,\\ \frac{46}{56} = 0.821, & if\quad 2<x\le 3,\\ 1, & if\quad 3<x<+\infty. \end{cases}$$

The graph of this function is shown in Figure 5.

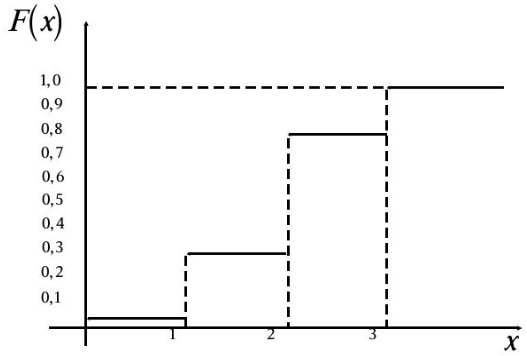


Figure 5

Figure 6

The Distribution Density of a Random Variable

Definition A random variable X is called a ***continuous*** random variable, if its possible values constitute a finite or infinite interval and a distribution function is not a step function, but a continuous function.

Definition The ***distribution density*** (the density of probability) of a continuous random variable X is called the derivative of its distribution function,

$$f(x) = F'(x) \qquad (3)$$

Properties of the distribution density

1.	The distribution density is a non-negative function, $f(x) \geq 0$.	
2.	$P(a < X < b) = \int_a^b f(x)\,dx$	(4)
3.	$F(x) = \int_{-\infty}^{x} f(t)\,dt$	(5)
4.	$\int_{-\infty}^{\infty} f(x)\,dx = 1$	(6)

Example 6 Let there be given a function $f(x) = ae^{-|x|}$. Find the value of the parameter a so that the function can be the distribution density of some continuous random variable.

Solution The function in question is a non-negative one. To be the distribution density it must satisfy the condition (6). We find the value of a

$$\int_{-\infty}^{\infty} ae^{-|x|}dx = 1, \quad a\int_{-\infty}^{0} e^{x}dx + a\int_{0}^{\infty} e^{-x}dx = 1$$

$$ae^{x}\Big|_{-\infty}^{0} - ae^{-x}\Big|_{0}^{\infty} = a(1-0) - a(0-1) = 2a = 1, \; a = \frac{1}{2}.$$

Example 7 Let there be given a function $f(x) = \begin{cases} 4/x^5, & if \;\; x \geq 1, \\ 0, & if \;\; x < 1. \end{cases}$. Find its distribution function.

Solution By virtue of the formula (5) the distribution function of the random variable is

$$F(x) = \int_1^x f(t)\,dt = \int_1^x \frac{4}{t^5}\,dt = 4\cdot\frac{t^{-4}}{-4}\Big|_1^x = \left(-\frac{1}{t^4}\right)\Big|_1^x = 1 - \frac{1}{x^4},$$

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

Exercise Set 1.6

1. In the lot of 6 components 4 are standard. 2 components are selected at random. Compose the law of distribution of the random variable *X*, which is the number of standard parts among those selected.

2. The probabilities of the first, second and third shooters to hit the target are respectively equal to 0.4; 0.3 and 0.6. The random variable X is the number of shots on the target. Find the distribution law of the random variable X .

3. The probability of hitting the target with one shot is equal to 0.6. The random variable *X* is the number of hitting the target with 5 shots. Find the distribution law of the random variable *X*.

4. Let X be the number of heads minus the number of tails obtained in four independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

5. Suppose we perform independent Bernoulli trials with parameter p, until we obtain two consecutive successes or two consecutive failures. Draw a tree diagram and find the probability function of the number of trials.

6. Let X be the number obtained in a single roll of a fair die. Draw a histogram for its probability function and a graph for its distribution function.

7. We roll two fair dice independently of each other. Let X be the absolute value of the difference of the numbers obtained on them. Draw a histogram for its probability function and a graph for its distribution function.

Let us suppose that n independent trials on a random variable *X* are fulfilled, and the obtained results are represented by the following tables:

8.

X	-2	0	1	5
P	0.5	0.2	0.1	0.2

9.

X	-3	-1	2	4
P	0.3	0.2	0.4	0.1

Build the plotted function of distribution.

Let there be given functions.

10. $f(x)=\begin{cases}0, & x\le 0,\\ A(x+2), & 0<x\le 2,\\ 0, & x>2.\end{cases}$ *11.* $f(x)=\begin{cases}0, & x\le 1,\\ A(x-2), & 1<x\le 3,\\ 0, & x>3.\end{cases}$

12. $f(x)=\begin{cases}0, & x\le 0,\\ Ax^2, & 0<x\le 2,\\ 0, & x>2.\end{cases}$ *13.* $f(x)=\begin{cases}A/x^4, & if\ \ x\ge 1,\\ 0, & if\ \ x<1.\end{cases}$

14. $F(x)=\begin{cases}0, & if\ \ x<0,\\ A(x^2+x), & if\ \ 0\le x\le 1,\\ 1, & if\ \ x>1.\end{cases}$ *15.* $F(x)=\begin{cases}0, & if\ \ x<0,\\ A(x^2+2x), & if\ \ 0\le x\le 3,\\ 1, & if\ \ x>3.\end{cases}$

Find the value of the parameter A so that the function $f(x)$ can be the distribution density of some continuous random variable. Find the probability $P(0.5<X<1)$, $P(X<2)$. In exercises 14 to 17 find its distribution function. Draw the graph for its distribution function.

Individual Tasks 1.6

I.

1. Let X be the number of hearts in a randomly dealt poker hand of five cards. Draw a histogram for its probability function and a graph for its distribution function.

2. Let X be the larger of the numbers of heads obtained in five independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

3. Find the value of the parameter A so that the function can be the distribution density of some continuous random variable. Find its distribution function. Draw the graph for its distribution function. Find the probability $P(0.5<X<1)$, $P(X<0)$.

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

4. Find the value of the parameter A so that the function $f(x)$ can be the distribution density of some continuous random variable. Find the probability $P(0.5<X<1.5)$, $P(X<1)$.

$$F(x)=\begin{cases}0, & if\ \ x<0,\\ A(x^3+3x), & if\ 0\le x\le 2,\\ 1, & if\ \ x>2.\end{cases}$$

II.

1. Let X be the number of heads obtained in five independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

2. Let X be the absolute value of the difference between the number of heads and the number of tails obtained in four independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

3. Find the value of the parameter A so that the function can be the distribution density of some continuous random variable. Find its distribution function. Draw the graph for its distribution function. Find the probability $P(0.5 < X < 1)$, $P(X < 2)$.

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

4. Find the value of the parameter A so that the function $f(x)$ can be the distribution density of some continuous random variable. Find the probability $P(-0.5 < X < 1)$, $P(X < 0)$.

$$F(x) = \begin{cases} 0, & if \ \ x < -1, \\ A(x^3+1), & if \ -1 \le x \le 2, \\ 1, & if \ \ x > 2. \end{cases}$$

1.7 Numerical Characteristics of a Random Variable

A random variable X is characterized by its probability density function, which defines the relative likelihood of assuming one value over the others. The density function can be used to infer a number of characteristics of the underlying random variable. The two most important attributes are measures of location and dispersion.

The mathematical expectation of a random variable

Example 1 (Average of Dice Rolls). Suppose that we roll a die $n = 18$ times, and observe the following outcomes: 2, 4, 2, 1, 5, 5, 4, 3, 4, 2, 6, 6, 3, 4, 1, 2, 5, 6. The average of these numbers can be computed as

$$average = \frac{2\cdot 1 + 4\cdot 2 + 2\cdot 3 + 4\cdot 4 + 3\cdot 5 + 3\cdot 6}{18} =$$

$$= \frac{2}{18}\cdot 1 + \frac{4}{18}\cdot 2 + \frac{2}{18}\cdot 3 + \frac{4}{18}\cdot 4 + \frac{3}{18}\cdot 5 + \frac{3}{18}\cdot 6 = \sum_{i=1}^{6} f_i \cdot i = \frac{65}{18} \approx 3.611$$

where f_i stands for the relative frequency of the outcome i. Now ideally, since for a fair die the six outcomes are equally likely, we should have obtained each number 3 times, but that is not what usually happens. For large n, however, the relative frequencies are approximately equal to the corresponding probabilities $p_i = 1/6$ and the average becomes close to

$$\sum_{i=1}^{6} p_i \cdot i = \sum_{i=1}^{6} \frac{1}{6}\cdot i = \frac{1}{6}\sum_{i=1}^{6} i = \frac{21}{6} = 3.5$$

Let us suppose that n independent trials on a random variable X are fulfilled and the obtained results are represented by the following table:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Definition The mathematical expectation of a discrete random variable X is defined by the expression

$$E(X)=\sum_{i=1}^{n} x_i p_i \qquad (1)$$

which is the sum of products of its possible values and corresponding probabilities of these values.

Let X be a continuous random variable with the distribution density $f(x)$. We get the mathematical expectation of a continuous random variable in the form of the improper integral

$$E(X)=\int_{-\infty}^{\infty} x\, f(x)\,dx, \text{ if } X\in(-\infty,\infty) \qquad (2)$$

$$E(X)=\int_{a}^{b} x\, f(x)\,dx, \text{ if } X\in(a;b). \qquad (3)$$

Probability [probabilistic] sense of the mathematical expectation of a random variable: it is its mean [average] value.

Remarks

1. Because of the occurrence of infinite sums and integrals, $E(X)$ does not exist for some random variables. These cases are rare, however, in real-life applications.

2. The expected value of a random variable X, is not necessarily a possible value of X despite its name; see, for instance, Example 1, but in many cases it can be used to predict, before the experiment is performed, that a value of X close to $E(X)$ can be expected.

The expected value is a measure of the center of a probability distribution.

Properties of the mathematical expectation

1. $E(C)=C$.
2. $E(CX)=C\,E(X),\quad C=const$.
3. $E(X+Y)=E(X)+E(Y),\quad X \text{ and } Y \text{ are}$ random variable;
4. $E(X\cdot Y)=E(X)\cdot E(Y),\quad X \text{ and } Y \text{ are}$ independent random variables.

Example 2 The mathematical expectation of the following discrete random variable

X	0	1	2	3
P	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

on the base of the same formula (1) equals

$$E(X)=0\cdot\frac{1}{56}+1\cdot\frac{15}{56}+2\cdot\frac{30}{56}+3\cdot\frac{10}{56}=\frac{15+60+30}{56}=\frac{105}{56}=1.875.$$

Example 3 The distribution density of a continuous random variable X equals

$$f(x)=\begin{cases}4/x^5, & if \quad x\ge 1,\\ 0, & if \quad x<1.\end{cases}$$

In this case the mathematical expectation must be calculated by the formula (2),

$$E(X)=\int_1^{\infty}x\,f(x)dx=\int_1^{\infty}\frac{4}{x^4}dx=4\cdot\left.\frac{x^{-3}}{-3}\right|_1^{\infty}=-\left.\frac{4}{3x^3}\right|_1^{\infty}=\frac{4}{3}.$$

The variance (dispersion) and root-mean-square (standard) deviation

Definition The deviation of a random variable X from its mathematical expectation is the following random variable

$$X-E(X) \qquad (4)$$

However, $E\left(X-E(X)\right)=0$ for every random variable that has an expectation, and so this is a useless definition.

Definition The ***variance (dispersion) of a random variable*** X is the mathematical expectation of the square of its deviation (4) from its mathematical expectation

$$D(X)=E\left((X-E(X))^2\right) \qquad (5)$$

Probability [probabilistic] sense of the variance: the variance characterizes [describes] a dissipation of a random variable about its mathematical expectation, that is about its mean [or average] value.

Definition The root-mean-square deviation of a random variable X is the square root of its variance, that is

$$\sqrt{D(X)}=\sigma(X) \qquad (6)$$

Properties of the variance

1. The variance of a constant quantity C equals zero, $D(C)=0$.

2. $D(CX)=C^2D(X), \quad C=const$.

3. $D(X+Y)=D(X)+D(Y)$, where X and Y are independent random variables.

4. $D(X)=E(X^2)-E^2(X)$.

Let X be a continuous random variable with the distribution density $f(x)$. We get the variance of a continuous random variable in the form of the improper integral

$$D(X)=E\left((X-E(X))^2\right)=\int_{-\infty}^{\infty}\left(x-E(X)\right)^2 f(x)dx;$$

$$D(X)=E\left((X-E(X))^2\right)=\int_{-\infty}^{\infty}x^2 f(x)dx-E^2(X) \qquad (7)$$

The variance of a discrete random variable X is the following expression

$$D(X)=E(X^2)-E^2(X)=\sum_{i=1}^{n} x_i^2 p_i - E^2(X) \tag{8}$$

Example 4 Calculate the variance and root-mean-square deviation of the random variable X of Example 2.

Solution The distribution tables of X (see Example 2) are

X	0	1	2	3
P	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

By the formulas (8) and (6) we obtain

$$E(X^2)=1\cdot\frac{15}{56}+4\cdot\frac{30}{56}+9\cdot\frac{10}{56}=\frac{15+120+90}{56}=\frac{225}{56}=4.018;$$

$$D(X)=4.018-1.875^2=0.5024;\sigma(X)=\sqrt{0.5024}=0.71.$$

Example 5 Calculate the variance and root-mean-square deviation of the random variable X of Example 3.

Solution The distribution density of the random variable is

$$f(x)=\begin{cases} 4/x^5, & if \quad x\ge 1, \\ 0, & if \quad x<1. \end{cases}$$

The integral of the formula (7) equals

$$D(X)=E(X^2)-\left(\frac{4}{3}\right)^2=\int_1^{\infty} x^2 f(x)\,dx-\frac{16}{9}=\int_1^{\infty}\frac{4}{x^3}dx-\frac{16}{9}=4\cdot\left.\frac{x^{-2}}{-2}\right|_1^{\infty}-\frac{16}{9}=$$

$$=-\left.\frac{4}{2x^2}\right|_1^{\infty}-\frac{16}{9}=2-\frac{16}{9}=\frac{2}{9}.$$

Therefore, the root-mean-square deviation of the random variable X equals

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{2}{9}}\approx 0.471.$$

Moments of a random variable

Definition The n***-th order initial moment of a random variable*** X is the mathematical expectation of its n-th power,

$$\alpha_n=E\left(X^n\right) \tag{9}$$

Definition The n***-th order central moment of a random variable*** X is the mathematical expectation of the n-th power of its centered random variable,

$$\mu_n=E\left(\left(X-E(X)\right)^n\right) \tag{10}$$

Theorem If the distribution of a random variable is symmetric about its mathematical expectation, then all its odd-order central moments are equal to zero.

Central moments can be expressed in terms of initial moments:

$$\mu_2 = \alpha_2 - 2\alpha_1\alpha_1 + \alpha_1^2 = \alpha_2 - \alpha_1^2,$$

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 3\alpha_1\alpha_1^2 - \alpha_1^3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3,$$

Definition Let X be a random variable whose probability density function is $f(x)$. A real valued function $M(t)$ defined by

$$M(t) = E\left(e^{tX}\right) \qquad (11)$$

is called the ***moment generating function of*** X if this expected value exists for all t in the interval $-h < t < h$ for some $h > 0$.

In general, not every random variable has a moment generating function. But if the moment generating function of a random variable exists, then it is unique. Using the definition of expected value of a random variable, we obtain the explicit representation for $M(t)$ as

$$M(t) = \begin{cases} \sum\limits_{k=0}^{\infty} e^{kt} p_k, \text{ if } X \text{ is descrete,} \\ \int\limits_{-\infty}^{+\infty} e^{xt} f(x)dx, \text{ if } X \text{ is continuous.} \end{cases} \qquad (12)$$

All the moments of X can be generated by taking derivatives of the moment generating function and then evaluating them at zero.

$$E(X) = M'(0),\ D(X) = \mu_2 = \alpha_2 - \alpha_1^2 = M''(0) - \left(M'(0)\right)^2.$$

Exercise Set 1.7

1. In the lot of 6 components 4 are standard. 2 components are selected at random. Compose the law of distribution of the random variable X, which is the number of standard parts among those selected. Find number characteristics of a random variable.

2. Probabilities of hitting the target of the first, second and third shooters are respectively equal to 0.4; 0.3 and 0.6. Each shooter makes one shot. The random variable X is the number of hits on the target. Find number characteristics of a random variable.

3. The probability of hitting the target with one shot is equal to 0.6. The random variable X is the number of hits the target with 5 shots. Find number characteristics of a random variable.

4. We roll two fair dice independently of each other. Let X be the absolute value of the difference of the numbers obtained on them. Find number characteristics of a random variable.

5. From a regular deck of 52 playing cards we pick one at random. Let the random variable X equal the number on the card if it is a numbered one (ace counts as 1) and 10 if it is a face card. Find number characteristics of a random variable.

6. 4 indistinguishable balls are distributed randomly into 3 distinguishable boxes. Let X denote the number of balls that end up in the first box. Find number characteristics of a random variable.

Let us suppose that n independent trials on a random variable X are fulfilled, and the obtained results are represented by the following tables:

7.

X	-2	0	1	5
P	0.5	0.2	0.1	0.2

8.

X	-3	-1	2	4
P	0.3	0.2	0.4	0.1

Find number characteristics of a random variable.

Let there be given functions. Find number characteristics of a random variable.

9. $f(x) = \begin{cases} 0, & x \le 0, \\ A(x+2), & 0 < x \le 2, \\ 0, & x > 2. \end{cases}$

10. $f(x) = \begin{cases} 0, & x \le 2, \\ A(x-2), & 2 < x \le 4, \\ 0, & x > 4. \end{cases}$

11. $f(x) = a e^{-|x|}$

12. $f(x) = \begin{cases} 0, & x \le 0, \\ Ax^3, & 0 < x \le 1, \\ 0, & x > 1. \end{cases}$

13. $f(x) = \begin{cases} ax^2, & if \;\; 0 \le x \le 2, \\ 0, & if \;\; x \notin (0;\, 2). \end{cases}$

14. $f(x) = \begin{cases} a/x^3, & if \;\; x \ge 1, \\ 0, & if \;\; x < 1. \end{cases}$

15. Let X be an exponential random variable with parameter λ

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

Find $E(X^2)$.

16. Find $E(X)$ for a random variable X with density

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & if \quad x \ge 0, \\ 0, & if \quad x < 0. \end{cases}$$

(This is the density of the sum of two independent exponential random variables with parameter $\lambda > 0$.)

Individual Tasks 1.7

I.

1. The probabilities of hitting the target of the first, second and third shooters are respectively equal to 0.5; 0.7 and 0.6. Each shooter makes one shot. The random variable X is the number of hits on the target. Find number characteristics of a random variable.

2. The worker serves 8 machines of the same type. The probability that the machine will require the attention of a worker during a shift is 1/3. The random variable X is the number of machine will require the attention of a worker during a shift. Find number characteristics of a random variable.

3. An urn contains 5 balls numbered 1 through 5. Two balls are selected at random without replacement from the urn. If the random variable X denotes the sum of the numbers on the 2 balls, then find the distribution law and number characteristics of a random variable.

4. Let us suppose that n independent trials on a random variable X are fulfilled, and the obtained results are represented by the following table. Find number characteristics of a random variable.

X	–5	2	3	4
p	0.4	0.3	0.1	0.2

5. Let there be given functions. Find number characteristics of a random variable. Find the probability $P(0.5 < X < 1)$.

$$F(x) = \begin{cases} 0, & if \;\; x < -1, \\ a + b \arcsin x, & if \;\; -1 \le x \le 1, \\ 1, & if \;\; x > 1. \end{cases}$$

II.

1. A fair coin is tossed 4 times. Let the random variable X denote the number of heads in 4 tosses of the coin. Find number characteristics of a random variable.

2. The probability that the student will pass the examinations is equal to 0.8 for the first one, 0.7 for the second one, 0.65 for the third one. The random variable X is the number of passing examinations. Find number characteristics of a random variable.

3. A pair of six-sided dice is rolled and the sum is determined. If the random variable X denotes the sum of the numbers rolled, then find the distribution law and number characteristics of a random variable.

4. Let us suppose that n independent trials on a random variable X are fulfilled, and the obtained results are represented by the following table. Find number characteristics of a random variable.

X	–3	–1	0	2	4
p	0.1	0.2	0.3	0.3	0.1

5. Let there be given functions

$$F(x) = \begin{cases} 0, & if \;\; x < -1, \\ \frac{1}{9}(x^3 + 1), & if \;\; -1 \le x \le 2, \\ 1, & if \;\; x > 2. \end{cases}$$

Find number characteristics of a random variable. Find the probability

$P(0.5 < X < 1)$.

1.8 Some Remarkable Distributions of Discrete Random Variable

Certain frequently occurring random variables and their distributions have special names.

Bernoulli [Binomial] Distribution

Definition A random variable X is called a ***Bernoulli random variable with parameter*** p, if it has two possible values, 0 and 1, with $P(X=1)=p$ and $P(X=0)=1-p=q$, where p is any number from the interval $[0,1]$. An experiment whose outcome is a Bernoulli random variable is called a *Bernoulli trial*.

Let a random variable X be the number of successes (the number of occurrences of some event A) in n independent trials with constant probability of the success A in any trial

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

Let us find a probability $P(X=m)=P_n(m)$ that is the probability of m successes. We will get the so-called Bernoulli formula:

$$P_n(m) = C_n^m p^m q^{n-m}, \quad q = 1 - p \tag{1}$$

Definition A random variable X is called a ***binomial random variable with parameters*** n and p, if it has the binomial distribution with probability function

$$f(k) = C_n^k p^k q^{n-k}, \quad k = 0,1,...,n \tag{2}$$

The distribution function of a binomial random variable is given by

$$F(x) = \begin{cases} 0, \ if \quad x < 0, \\ \sum_{k=0}^{[x]} C_n^k p^k q^{n-k} \ , if \ 0 \le x < n \\ 1, \quad\quad if \quad x \ge n. \end{cases}$$

Here $[x]$ denotes the floor or greatest integer function, that is, $[x]$ = the greatest integer $\le x$.

One says that X is distributed binomially (by Bernoulli [binomial] law) or simply: X is Bernoulli (binomial) distribution (briefly“ X *distributed B*”).

If a random variable X is a Bernoulli random variable with parameter p, then *the mathematical expectation, the variance, the root-mean-square deviation are represented by the following formulas:*

$$E(X) = n\,p, \quad D(X) = n\,p\,q, \quad \sigma(X) = \sqrt{n\,p\,q} \tag{3}$$

Example 1 6 independently working engines are installed in a shop. The probability for any engine to work at a given moment is 0.8. Find number characteristics of a random variable X, if the random variable X is a number of working engines at this moment.

Solution We can consider setting of an engine as a trial. So we have n = 6 independent trials. Let an event A means that an engine works.

$p = P(A) = 0.8, P(\overline{A}) = 1 - p = q = 0.2$.

The random variable X has Bernoulli distribution (briefly "X distributed B"), it can take on the values 0,1,2,3,4,5,6, which one calculates by Bernoulli formula (1). For example

$P(X = 0) = P_6(0) = C_6^0 p^0 q^6 = 0.2^6 = 0.00006$;

$P(X = 1) = P_6(1) = C_6^1 p^1 q^5 = 6 \cdot 0.8 \cdot 0.2^5 = 0.00154$.

Using formulas (3), we get

$E(X) = n\,p = 6 \cdot 0.8 = 4.8, \quad D(X) = n\,p\,q = 6 \cdot 0.8 \cdot 0.2 = 0.96,$

$\sigma(X) = \sqrt{n\,p\,q} = \sqrt{0.96} = 0.979$.

Poisson distribution

Poisson random variables are used to model the number of occurrences of certain events that come from a large number of independent sources, such as the number of calls to an office telephone during business hours, the number of atoms decaying in a sample of some radioactive substance, the number of visits to a web site, or the number of customers entering a store.

Let a random variable X be distributed B. Let us suppose that the number n of trials tends to infinity, the probability p of a success A goes to zero, but a product np retains constant,

$$n \to \infty, p \to 0, np = const = \lambda .$$

In this case the limit of the probability $P(X = k) = P_n(k)$, which is defined by Bernoulli formula (1), equals

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \qquad (4)$$

Definition One says that a discrete random variable X (with non-negative integer possible values) has Poisson distribution with a parameter λ if its distribution law is given by the following formula (Poisson formula):

$$P(X = k) = P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (5)$$

The histogram of a typical Poisson probability function (for $\lambda = 4$) is shown in Figure 7.

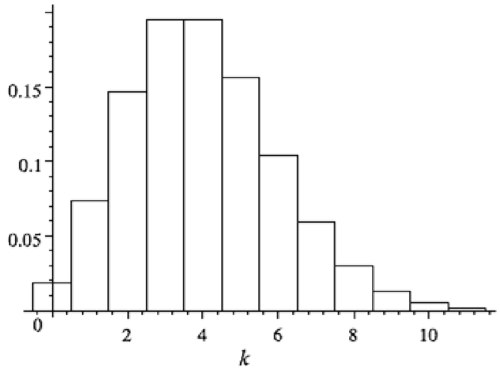


Figure 7

The mathematical expectation, the variance, the root-mean-square deviation are represented by the following formulas:

$$E(X) = D(X) = \lambda,\ \sigma(x) = \sqrt{\lambda} \qquad (6)$$

Example 2 Suppose that on average a restaurant has on average 50 diners per night. What is the probability that on a certain night 40 or fewer will show up?

Solution Suppose that the diners come from a large pool of potential customers, who show up independently with the same small probability for each one. Then their number X may be taken to be Poisson, and so

$$P(X \le 40) = \sum_{k=0}^{40} \frac{50^k}{k!} e^{-50} \approx 0.086\,.$$

On the other hand, if we assume that the customers come in independent pairs rather than individually, and denote the number of pairs by Y, then the corresponding probability is

$$P(Y \le 20) = \sum_{k=0}^{20} \frac{20^k}{k!} e^{-20} \approx 0.185\,.$$

These numbers show that, in order to estimate the probability of a slow night, it is not enough to know how many people show up on average, but we need to know the sizes of the groups that decide, independently from one another, whether to come or not.

In most applications, we are interested not just in one Poisson random variable but in a whole family of Poisson random variables. For instance, in the previous example we may ask for the probabilities of the number of diners in a week or a month.

In general, a family of random variables $X(t)$ depended on a parameter t is called a *stochastic or random process*. The parameter t is time in most applications, but not always. Here we are concerned with the particular stochastic process called the Poisson process.

Definition A family of random variables $X(t)$ depended on a parameter t is called a ***Poisson process with rate*** λ, for any $\lambda > 0$, if $X(t)$, the number of occurrences of some kind in any interval of length t, has a Poisson distribution with parameter λt for any $t > 0$, that is,

$$P(X(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \textit{for} \;\; t > 0, \;\; k = 0,1,2,... \tag{7}$$

and the numbers of occurrences in nonoverlapping time intervals are independent of each other.

Example 3 Suppose the pages of a book contain misprinted characters, independently of each other, with a rate of $\lambda = 0.1$ misprints per page. Assume that the numbers $X(t)$ of misprints on any t pages constitute a Poisson process. Find the probabilities of having (a) no misprint on the first three pages, (b) at least two misprints on the first two pages.

Solution

(a) In this case $t = 3$ and $\lambda t = 0.3$. Thus, $P(X(3) = 0) = \frac{(0.3)^0}{0!} e^{-0.3} \approx 0.74$.

(b) Now $t = 2$ and $\lambda t = 0.2$, and so

$$P(X(2) \geq 2) = 1 - \left(P(X(2) = 0) + P(X(2) = 1)\right) = 1 - \left(\frac{(0.2)^0}{0!} e^{-0.2} + \frac{(0.2)^1}{1!} e^{-0.2} \right) \approx$$

≈ 0.0175.

Geometric Random Variable

Definition Suppose we perform independent Bernoulli trials with parameter p, with $0<p<1$, *until we obtain a success*. The number X of trials is called a ***geometric random variable with parameter*** p. It has the probability function

$$f(k)=P(X=k)=pq^{k-1}, \quad for \quad k=1,2,... \tag{8}$$

The name “geometric” comes from the fact that the $f(k)$ values are the terms of a geometric series. Using the formula for the sum of a geometric series, we can confirm that they form a probability distribution:

$$\sum_{k=1}^{\infty} f(k)=\sum_{k=1}^{\infty} pq^{k-1}=\frac{p}{1-q}=1$$

Let X be geometric with parameter p.

The mathematical expectation, the variance, the root-mean-square deviation of a geometric random variable are represented by the following formulas:

$$E(X)=\frac{1}{p}, \quad D(X)=\frac{q}{p^2}, \quad \sigma(X)=\frac{\sqrt{q}}{p} \tag{9}$$

Hypergeometric Random Variable

Definition One says that a discrete random variable X has a ***hypergeometric distribution*** if it counts the number of successes, that is, the number of standard items picked, if we select a sample of size n without replacement from a mixture of N standard and nonstandard items. If p stands for the fraction of standard items in the lot and $q=1-p$ the fraction of nonstandard items, then the probability function of X is

$$f(x;n,N,p)=\frac{C_{Np}^{x}\cdot C_{Nq}^{n-x}}{C_N^n}, \quad for \quad \max(0,n-Nq)\le x \le \min(n,Np)$$

The mathematical expectation, the variance, the root-mean-square deviation of a hypergeometric random variable are represented by the following formulas:

$$E(X)=np, \quad D(X)=n\cdot\frac{np}{N-1}\cdot\frac{(N-np)(N-n)}{N^2}, \quad \sigma(X)=\sqrt{D(X)} \tag{10}$$

Exercise Set 1.8

1. Five independent experiments are being produced, in each of them the event A can occur with probability 0.6. Let X denotes the number of occurrences of the event A in 5 experiments. Find the distribution law of the random variable X. Find $E(X), D(X), \sigma(X)$.

2. Find the constant probability p of hitting the target in every shot and the number of shots being fired if the average number of hits equals 240 and the standard deviation of hits equals 12.

3. The probability that the product will be broken is equal to 0.0004. Find the probability that no less than two from 1000 products will be broken.

4. The average number of taxi reservations in 1 minute equals 5. Find the probability that during 3 minute a taxi will be called: a) 7 times; b) no less than 4 times; c) at least one time.

5. The probability of hitting the target with one shot is equal to 0.6. The random variable X is a number of striking the target with 5 shots. Find number characteristics of a random variable.

6. A basketball player makes three penalty throws. The probability of hit with each throw is equal to 0.7. Find number characteristics of a random variable X, if the random variable X is a number of shots at the basket.

7. From a regular deck of 52 playing cards we pick one at random. Let the random variable X equals the number on the card if it is a numbered one (ace counts as 1) and 10 if it is a face card. Find $E(X)$.

8. 4 indistinguishable balls are distributed randomly into 3 distinguishable boxes. Let X denotes the number of balls that end up in the first box. Find $E(X)$.

9. Customers enter a store at a mean rate of 1 per minute. Find the probabilities that: 1) more than one will enter during the first minute, 2) more than two will enter during the first two minutes, 3) more than one will enter during each of the first two minutes, 4) two will enter during the first minute and two in the second minute if four have entered during the first two minutes.

10. The probability that a machine produces a defective item is 0.02. Each item is checked as it is produced. Assuming that these are independent trials, what is the probability that at least 100 items must be checked to find one that is defective?

11. A random sample of 5 students is drawn without replacement from among 300 seniors, and each of these 5 seniors is asked if she/he has tried a certain drug. Suppose 50% of the seniors actually have tried the drug. Let X denotes the number of the students interviewed who have tried the drug. What is the probability that two of the students interviewed have tried the drug? Find $E(X), D(X), \sigma(X)$.

12. Suppose an experiment consists of tossing a fair coin until three heads occur. What is the probability that the experiment ends after exactly six flips of the coin with a head on the fifth toss as well as on the sixth?

13. The number of hits, X, per baseball game, has a Poisson distribution. If the probability of a no-hit game is 1/3, what is the probability of having 2 or more hits in a specified game?

14. A textile plant turns out cloth that has 1 defect per 20 square yards. Assume that 2 square yards of this material are in a pair of pants and 3 square yards in a coat. 1) What percentage of the pants will be defective? 2) About what percentage of the coats will be defective?

15. Let the probability that the birth weight (in grams) of babies in America is less than 2547 grams be 0.1. If X equals the number of babies that weigh less than 2547 grams at birth among 20 of these babies selected at random, then what is $P(X \le 3)$?

Individual Tasks 1.8

I.

1. What is the probability of rolling at most two sixes in 5 independent casts of a fair die?

2. Nine independent experiments are being produced, in each of them the event A can occur with probability 0.8. Let X denotes the number of occurrences of the event A in 9 experiments. Find the distribution law of the random variable X. Find $E(X), D(X), \sigma(X)$.

3. Find the constant probability p of hitting the target in every shot and the number of shots being fired if the average number of hits equals 210 and the standard deviation of hits equals 14.

4. The probability that a machine produces a defective item is 0.05. Each item is checked as it is produced. Assuming that these are independent trials, what is the probability that at least 120 items must be checked to find one that is defective?

5. The number of traffic accidents per week in a small city has a Poisson distribution with the mean equal to 3. What is the probability that exactly 2 accidents occur in 2 weeks?

6. The probability that the product will be broken is equal to 0,006. Find the probability that at least three from 1000 products will be broken.

7. Suppose there are 3 defective items in a lot of 50 items. A sample of size 10 is taken at random and without replacement. Let X denote the number of defective items in the sample. What is the probability that the sample contains at most one defective item? Find $E(X), D(X), \sigma(X)$.

II.

1. What is the probability of rolling two sixes and three nonsixes in 5 independent casts of a fair die?

2. Eight independent experiments are being hold, in each of them the event A can occur with probability 0.7. Let X denotes the number of occurrences of the event A in 8 experiments. Find the distribution law of the random variable X. Find $E(X), D(X), \sigma(X)$.

3. Find the constant probability p of hitting the target in every shot and the number of shots being fired if the average number of hits equals 310 and the deviation of hits equals 15.

4. The probability that a machine produces a defective item is 0.03. Each item is checked as it is produced. Assuming that these are independent trials, what is the probability that at least 80 items must be checked to find one that is defective?

5. A random variable X has a Poisson distribution with a mean of 3. What is the probability that X is bounded by 1 and 3, that is, $P(1 \le X \le 3)$?

6. The probability that the product will be broken is equal to 0.003. Find the probability that at least two from 2000 products will be broken.

7. A radio supply house has 200 transistor radios, of which 3 are improperly soldered and 197 are properly soldered. The supply house randomly draws 4 radios without replacement and sends them to a customer. Let X denotes the number of improperly soldered radios. What is the probability that the supply house sends 2 improperly soldered radios to its customer? Find $E(X), D(X), \sigma(X)$.

1.9 Some Remarkable Distributions of Continuous Random Variable

The uniform distribution

Definition One says that a random variable X has a ***uniform distribution*** over an interval (a,b) [X is uniformly distributed or simply X is the uniform distribution over an interval (a,b)], if its distribution density is constant inside and equals zero outside this interval,

$$f(x)=\begin{cases} c, & if \quad x\in[a;b], \\ 0, & if \quad x\notin[a;b]. \end{cases} \qquad (1)$$

On the base of property 4 of the distribution density we must have

$$\int_{-\infty}^{\infty} f(x)dx=1 \Rightarrow \int_{-\infty}^{\infty} cdx=\int_{a}^{b} cdx=c(b-a)=1 \Rightarrow c=\frac{1}{b-a},$$

and so the distribution density of the uniform distribution is the next one:

$$f(x)=\begin{cases} \dfrac{1}{b-a}, & if \ x\in[a;b], \\ 0, & if \ x\notin[a;b]. \end{cases} \qquad (2)$$

Using property 3 of the distribution density, we must study three cases.

$$F(x)=\begin{cases} 0, & if \quad x\le a, \\ \dfrac{x-a}{b-a}, & if \quad a<x\le b, \\ 1, & if \ b<x<+\infty. \end{cases} \qquad (3)$$

The graph of distribution density and distribution function of the uniform distribution is shown in Figures 8, 9.

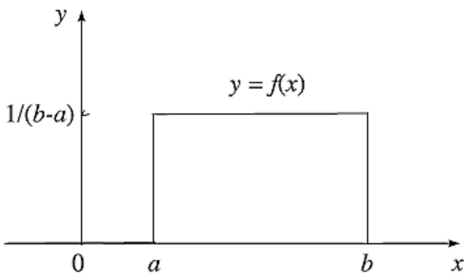


Figure 8

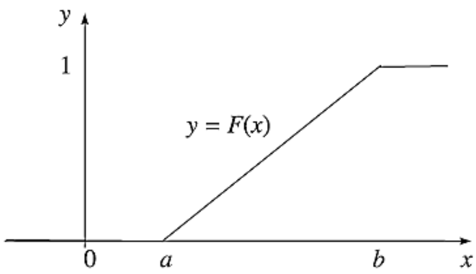


Figure 9

The mathematical expectation, the variance, the root-mean-square deviation of a uniformly distributed random variable are represented by the following formulas:

$$E(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12},\ \sigma(X)=\frac{b-a}{\sqrt{12}},\ P(c<X<d)=\frac{d-c}{b-a} \qquad (4)$$

Example 1 The time interval of the trolleybus service equals 5 minutes. Find the probability that one will wait a trolleybus no longer then 2 minutes.

Solution The waiting time T is a random variable uniformly distributed over the interval $(0,5)$, and we have to find the probability $P(0<X<2)$:

$$P(0,2)=\frac{2-0}{5-0}=0.4.$$

The normal distribution

Definition One says that a random variable X has a ***normal distribution with parameters*** a, σ ($\sigma>0$) (or that X is distributed $N(a,\sigma)$) if its distribution density is the following function:

$$p(x;a,\sigma)=f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (5)$$

The graph of distribution density of the normal distribution is shown in Figure 10. It is called a ***normal curve***. The normal distribution is often called Gauss distribution, and the corresponding normal curve is called *Gauss curve*.

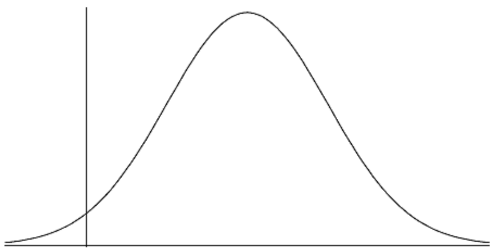


Figure 10

Let the parameter σ of the normal distribution tend to 0. For $\sigma\to 0$ the normal curve stretches along the straight line $x=a$ and simultaneously presses to the *Ox* axis $(\sigma_1<\sigma<\sigma_2)$. (See figure 11)

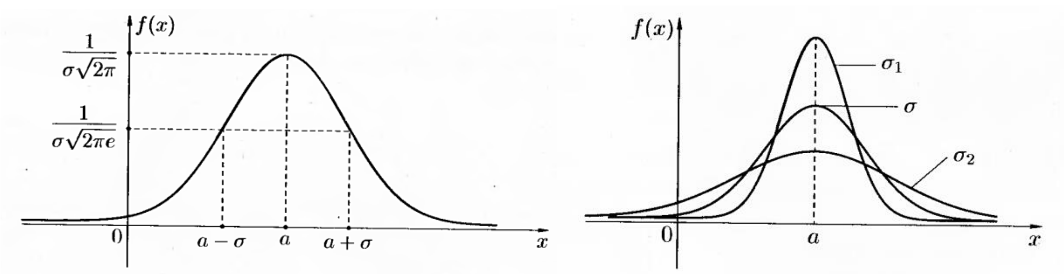


Figure 11

Let a random variable X be normally distributed with parameters $a,\sigma,\sigma>0$ (X distributed $N(a,\sigma)$). *The mathematical expectation, the variance, the root-mean-square deviation of a normally distributed random variable are represented by the following formulas:*

$$a = E(X), \quad \sigma^2 = D(X), \quad \sigma(X) = \sigma \tag{6}$$

Let a random variable X be distributed $N(a,\sigma)$. The probability of its hitting on an interval (α,β) can be calculated by the following formula:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) \tag{7}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ is known as Laplace function.

The probability of the deviation of a random variable X, which is distributed $N(a,\sigma)$, from its mathematical expectation a is given by the following formula:

$$P\left(|X - E(X)| < \delta\right) = P\left(|X - a| < \delta\right) = 2\Phi\left(\frac{\delta}{\sigma}\right) \tag{8}$$

For example, let $\delta = 3\sigma$. The formula (8) gives

$$P\left(|X - a| < 3\sigma\right) = 2\Phi(3) = 0.9973.$$

We have got the so-called ***"3σ – rule"***: with a probability 0.9973 all values of the normal distribution are concentrated in the interval $(a - 3\sigma;\ a + 3\sigma)$.

Example 2 A plant makes balls for the bearings. The nominal diameter of the balls is equal to 6 (mm). As a result of an inaccuracy in the production of the balls its actual diameter is a random variable, distributed according to the normal law with an average value of 6 (mm) and a mean-square deviation of 0.04 (mm). The balls, whose diameter varies from the nominal by more than 0.1 (mm), are inspected out. Find: 1) what percentage of balls will be rejected on average; 2) the probability that the actual diameter of balls will be contained in the range from 5.97 to 6.05 (mm).

Solution Let the random variable X be an actual diameter of a ball. It is distributed according to the normal law, i.e. $X \in N(a,\sigma)$. If $a = d_0 = 6$ and $\sigma = 0.04$, then $X \in N(6;0.04)$.

Since according to the condition of the task the balls whose diameter differs from the nominal one by more than 0.1 (mm) are inspected out, then let us examine the event $|X - 0.6| > 0.1$. To find the probability of this event we will use the opposite event $|X - 0.6| \le 0.1$. Since the random variable X is continuous, then

$$P\left(|X - 6| \le 0.1\right) = P\left(|X - 6| < 0.1\right);$$

$$P\left(|X - a| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) \Rightarrow P\left(|X - 6| < 0{,}1\right) = 2\Phi\left(\frac{0.1}{0.04}\right) = 2\Phi(2.5).$$

Using the table of the values of the function of Laplace (see Appendix table 2), let us find that $\Phi(2.5) \approx 0.4938$ and $P\left(|X - 6| < 0.1\right) \approx 2 \cdot \Phi(2.5) = 0.9876$.

If $P\left(|X - 6| < 0{,}1\right) + P\left(|X - 6| > 0{,}1\right) = 1$, then

$$P\left(|X - 6| > 0.1\right) = 1 - P\left(|X - 6| < 0.1\right) = 1 - 0.9876 = 0.0124.$$

Consequently, $1,24\%$ of the balls will be rejected on average.

The probability of its hitting on an interval (α,β) can be calculated by the following formula:

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

If $X \in N(6;0.04)$ and $\alpha = 5.97$, $\beta = 6.05$, then

$$P(5.97 < X < 6.05) = \Phi\left(\frac{6.05-6}{0.04}\right) - \Phi\left(\frac{5.97-6}{0.04}\right) = \Phi(1.25) + \Phi(0.75).$$

Using the table of the values of the function of Laplace, let us find that

$$\Phi(1.25) \approx 0.3944 \text{ and } \Phi(0.75) \approx 0.2734,$$

$$P(5.97 < X < 6.05) \approx 0.3944 + 0.2734 = 0.6678.$$

The exponential distribution

Definition One says that X has the ***exponential distribution with parameter*** λ, if its distribution density is the following function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & if \quad x \ge 0, \\ 0, & if \quad x < 0. \end{cases} \tag{9}$$

The distribution function of the random variable X for $x > 0$ equals

$$F(x) = 1 - e^{-\lambda x},\ x \ge 0 \tag{10}$$

The expression $1 - e^{-\lambda x}$ tends to zero with x, and therefore the distribution function can be defined as follows:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & if \quad x \ge 0, \\ 0, & if \quad x < 0. \end{cases} \tag{11}$$

The graphs of both functions $F(x)$, $f(x)$ are represented on Figure 12.

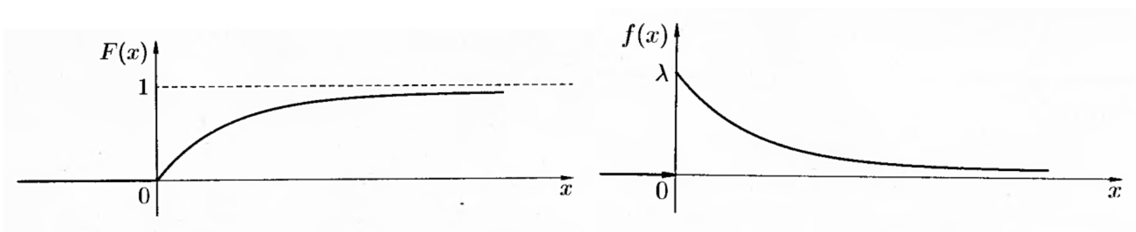


Figure 12

The mathematical expectation, the variance, the root-mean-square deviation of a random variable distributed according to the exponential law are represented by the following formulas:

$$E(X) = \frac{1}{\lambda},\quad D(X) = \frac{1}{\lambda^2},\quad \sigma(X) = \frac{1}{\lambda},\quad \lambda > 0 \tag{12}$$

The probability of its hitting on an interval (a,b) can be calculated by the following formula:

$$P(a < X < b) = e^{-\lambda a} - e^{-\lambda b} \tag{13}$$

Example 3 Time t of the reliable work of a radio-technical system is distributed according to the exponential law. The failure rate of the system is $\lambda = 0.02$. Find the mean time of the failure-free operation and the probability of the failure-free operation in 80 hours.

Solution The density of probability of this distribution takes the form of

$$f(t) = \begin{cases} 0.02e^{-0.02t}, & if \quad t \geq 0, \\ 0, & if \quad t < 0. \end{cases}$$

Mathematical expectation is this mean time of the reliable work of a system.

$$E(T) = \frac{1}{\lambda} = \frac{1}{0.02} = \frac{100}{2} = 50 \text{ (hours)}.$$

Let us determine the probability of the failure-free operation for 80 hours with the help of the function of reliability $R(t) = e^{-\lambda t} = P(0 < X < t)$:

$$R(80) = e^{-0.02 \cdot 80} = e^{-1.6} \approx 0.2019.$$

Exercise Set 1.9

1. The random variable X has the uniform distribution with $E(X) = 4$ and $D(X) = 12$. Find $F(x)$, $f(x)$ and $P(0 < X < 2)$.

2. Someone expects a telephone call between 19.00 and 20.00. The waiting time of the ring is the random variable X, which has a uniform distribution in section [19; 20]. Find the probability that the telephone will ring in the period from 19.22 to 19.46.

3. If X has a uniform distribution on the interval from 0 to 10, then what is $P\left(X + \frac{10}{X} \geq 7\right)$?

4. If the random variable X has a uniform distribution on the interval $[0, a]$, what is the probability that the random variable is greater than its square, that is $P(X > X^2)$?

5. A box is to be constructed so that its height is 10 inches and its base is X inches by X inches. If X has a uniform distribution over the interval (2, 8), then what is the expected volume of the box in cubic inches?

6. The radio equipment for 1000 hours of work goes out of order on average one time. Find the probability of failure of the radio equipment for 200 hours of work, if the period of failure-free operation is a random variable, distributed according to the exponential law.

7. Find the time of the failure-free operation of a radio lamp (the probability of the failure-free operation of the radio lamp is 0.8) if the average time of the failure-free operation is equal to 700 hours.

8. The mean time of operation of each of the three elements, entering the technical device, is equal to T hours. For the reliable work of the device the failure-free operation of at least one of these three elements is necessary. Find the probability that the device will work from t_1 to t_2 hours, if the time of operation of each of the three elements independently is distributed according to the exponential law.

Exercise	*T*	t_1	t_2
A	*800*	*650*	*700*
B	*1000*	*800*	*900*
C	*850*	*750*	*820*
D	*1200*	*900*	*1000*

9. If the random variable X is normal with mean 1 and standard deviation 2, then what is $P\left(X^2-2X\le 8\right)$?

10. If X is normal with mean 0 and variance 4, then what is the probability of the event $X-\dfrac{4}{X}\ge 0$?

11. If the waiting time at Rally's drive-in-window is normally distributed with mean 13 minutes and standard deviation 2 minutes, then what percentage of customers wait longer than 10 minutes but less than 15 minutes?

12. A random variable X is normally distributed with the mathematical expectation $E(X)$ and the variance $D(X)$. Find the density of probability distribution $f(x)$ and build the graph of this function. Write down the interval of practically probable values of a random variable. Which is more probable $X\in(\alpha;\beta)$ or $X\in(\gamma;\delta)$?

Exercise	$E(X)$	$D(X)$	α	β	γ	δ
A	-3	9	-5	-4	-2	-1
B	4	1	0	3	2	5
C	-5	16	-10	-6	0	4
D	3	4	0	5	6	7

Individual Tasks 1.9

I.

1. The random variable X has the uniform distribution with $E(X)=6$ and $D(X)=10$. Find $F(x)$, $f(x)$ and $P(0<X<4)$.

2. If X is uniform on the interval from −5 to 5, what is the probability that the quadratic equation $100t^2+20tX+2X+3=0$ has complex solutions?

3. Find the time of the failure-free operation of the radio lamp (the probability of the failure-free operation of the radio lamp is 0.6), if the average time of failure-free operation is equal to 500 hours.

4. The mean time of operation of each of the three elements, entering the technical device, is equal to 900 hours. For the reliable work of the device the failure-free operation of at least one of these three elements is necessary. Find the probability that the device will work from 700 to 900 hours, if the time of operation of each of the three elements independently is distributed according to the exponential law.

5. The random variable X has the normal distribution with $E(X)=2$ and $D(X)=9$. Find $F(x)$, $f(x)$. Write down the interval of practically probable values of a random variable. Which is more probable $X\in(-2;2)$ or $X\in(1;5)$?

6. If the random variable X is normal with mean 2 and standard deviation $\sigma=1$, then what is $P\left(X^2-X\le 6\right)$?

II.

1. The random variable X has the uniform distribution with $E(X)=4$ and $D(X)=8$. Find $F(x)$, $f(x)$ and $P(0<X<3)$.

2. If X has a uniform distribution on the interval from 0 to 3, what is the probability that the quadratic equation $4t^2+4tX+X+2=0$ has real solutions?

3. Find the time of the failure-free operation of the radio lamp (the probability of the failure-free operation of the radio lamp is 0.7), if the average time of failure-free operation is equal to 450 hours.

4. The mean time of operation of each of the three elements, entering the technical device, is equal to 950 hours. For the reliable work of the device the failure-free operation of at least one of these three elements is necessary. Find the probability that the device will work from 720 to 850 hours, if the time of operation of each of the three elements independently is distributed according to the exponential law.

5. The random variable X has the normal distribution with $E(X)=0$ and $D(X)=16$. Find $F(x)$, $f(x)$. Write down the interval of practically probable values of a random variable. Which is more probable $X\in(-2;2)$ or $X\in(0;4)$?

6. If the random variable X is normal with mean 3 and standard deviation $\sigma=4$, then what is $P\left(X^2-X\ge 12\right)$?

1.10 Laws of Large Numbers. The Central Limit Theorem

Theorem (Chebyshev inequality) For any random variable X with the mathematical expectation $E(X)$ and the variance $D(X)$, the next inequality holds

$$P\left(|X-E(X)|\ge\varepsilon\right)\le\frac{D(X)}{\varepsilon^2} \qquad (1)$$

or, as a result, the inequality

$$P\left(|X-E(X)|<\varepsilon\right)\ge 1-\frac{D(X)}{\varepsilon^2} \qquad (2)$$

In the case of *Bernoulli distribution* X (here $E(X)=np$, $D(X)=npq$) the inequality (2) takes the form

$$P\left(\,|\,X-np\,|<\varepsilon\,\right)\ge 1-\frac{npq}{\varepsilon^2} \qquad (3)$$

and in the case of *Poisson distribution* X ($E(X)=D(X)=\lambda$) the form

$$P\left(\,|\,X-\lambda\,|<\varepsilon\,\right)\ge 1-\frac{\lambda}{\varepsilon^2} \qquad (4)$$

Example 1 A device consists of 200 independent working elements. The probability of the failure of each element during the time *T* equals 0.01. With the help of Chebyshev inequality estimate the probability that the number of failed elements during the time *T* differs from the mean number of failures: a) less than by 5; b) not less than by 5.

Solution Let a random variable X be the number of failed elements during the time *T*. It is obvious that X distributed *B* with

$$n=200,\, p=0.01,\, q=0.99,\, E(X)=n\,p=2,\; D(X)=n\,p\,q=1.98.$$

By the inequality (3) with $\varepsilon=5$ we have

$$P\left(\,|\,X-2\,|<5\,\right)\ge 1-\frac{1.98}{5^2}\approx 0.92.$$

On the base of the inequality (1) we have

$$P\left(\,|\,X-2\,|\ge 5\,\right)\le \frac{1.98}{5^2}\approx 0.08.$$

Let $X_1,\, X_2,\, \ldots,\, X_n$ be a sequence of (mutually) independent random variables. Chebyshev inequality (2) gives us the estimate of the deviation of the arithmetical average of n first independent random variables from the arithmetical average of their mathematical expectations, namely

$$P\left(\,\left|\frac{1}{n}\sum X_i-\frac{1}{n}\sum E(X_i)\right|<\varepsilon\,\right)\ge 1-\frac{D(Y)}{n\varepsilon^2} \qquad (5)$$

Example 2 How many measurements must one fulfill to assert that with the probability 0.99 the error of the arithmetic mean of results of these measurements is less than 0.01 if the root-mean-square deviation of each measurement equals 0.03?

Solution We can consider the results of measurements as independent random variables X_i with the same mathematical expectation a (the exact value of a quantity to be measured) and variance $D(X_i)=0.03^2=0.0009$.

Supposing $D(X_i)=0.03^2=0.0009, \varepsilon=0.01$ in the inequality (5), we have to find n from the following relation

$$P\left(\,\left|\frac{1}{n}\sum X_i-a\right|<0.01\right)\ge 1-\frac{0.0009}{n\cdot 0.01^2}=0.99$$

or

$$1-\frac{0.0009}{n\cdot 0.01^2}=0.99.$$

This last equation gives $n = 900$, and therefore it's sufficient to fulfill not less than 900 measurements.

Chebyshev inequality permits to estimate the deviation of the relative frequency $P_n^*(A)$ of some event A from its probability $p = P(A)$. Let us suppose that we fulfill n independent trials with a constant probability $p = P(A)$ of an event A. If a random variable X is the number of occurrences of the event A, then the relative frequency of A equals $P_n^*(A) = \dfrac{X}{n}$. It is obvious that X has Bernoulli distribution for which $E(X) = n\,p,\ D(X) = n\,pq$. Therefore

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) = P\left(|X - np| < \varepsilon \cdot n\right) \ge 1 - \frac{pq}{n\varepsilon^2} \qquad (6)$$

Example 3 The probability for an item to be nonstandard is 0.03. How many nonstandard items are contained in a batch of 100 items with probability 0.9? Solve the problem with the help of Chebyshev inequality.

Solution Let a random variable X be a number of nonstandard items in a batch, and an event A means that an item, which is taken at random, is not standard one. By conditions of the problem $p = P(A) = 0.03,\ q = 0.97$, and by virtue of the formula (6) we can write

$$P\left(\left|\frac{X}{100} - 0.03\right| < \varepsilon\right) \ge 1 - \frac{0.03 \cdot 0.97q}{100 \cdot \varepsilon^2} = 0.9 .$$

The equality $1 - \dfrac{0.03 \cdot 0.97q}{100 \cdot \varepsilon^2} = 0.9$ gives $\varepsilon \approx 0.05$, and so we must find possible nonnegative values of X from the relation

$$P\left(\left|\frac{X}{100} - 0.03\right| < 0.05\right) = 0.9 .$$

$$\left|\frac{X}{100} - 0.03\right| < 0.05 \Leftrightarrow |X - 3| < 5 \Leftrightarrow -5 < X - 3 < 5 \Leftrightarrow 0 \le X < 8 .$$

Definition One says that a sequence of random variables $X_1, X_2, \dots, X_n$ converges to some value A ***in probability*** $X_n \underset{prob}{\rightarrow} A$, if for any positive however small number ε the next limit takes place**:**

$$\lim_{n \to \infty} P\left(|X_n - A| < \varepsilon\right) = 1 \qquad (7)$$

Law of Large Numbers

Theorem (Chebyshev small theorem) Let $X_1, X_2, \dots, X_n$ be independent random variables with the same mathematical expectation $E(X) = a$ and whose variances don't surpass some number C. In this case the arithmetic mean $\dfrac{1}{n}\sum_{i=1}^{n} X_i$ of these random variables converges in the probability to their common mathematical expectation.

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n}\sum X_i - \frac{1}{n}\sum E(X_i) \right| < \varepsilon \right) = 1 \qquad (8)$$

Chebyshev small theorem is often applicable in a measurement practice.

Let us we fulfill independent measurements of some quantity. We can regard results of measurements as independent random variables $X_1, X_2, \dots, X_n$ with the same mathematical expectation $E(X_i) = a$ (which is the exact value of a quantity to be measured). If systematic biases are absent that is the measurement dispersions are uniformly bounded by some number C, $D(X_i) < C$, then by Chebyshev small theorem the arithmetical mean of the results of measurements converges in probability to a.

Theorem (Chebyshew large theorem) Let the random variables $X_1, X_2, \dots, X_n$ have different mathematical expectations in conditions of Chebyshev small theorem. In this case the difference $\frac{1}{n}\sum X_i - \frac{1}{n}\sum E(X_i)$ that is the difference between the arithmetical average $\frac{1}{n}\sum_{i=1}^{n} X_i$ of the first n random variables and the arithmetical average of their mathematical expectations, converges in probability to zero (if $n \to \infty$),

$$\frac{1}{n}\sum X_i - \frac{1}{n}\sum E(X_i) \underset{prob}{\to} 0 \qquad (9)$$

Theorem (Bernoulli) A relative frequency of every event A converges in probability to its probability $P(A)$ (if $n \to \infty$),

$$P_n^*(A) \underset{prob}{\to} P(A) \qquad (10)$$

Theorem (De Moivre–Laplace Limit Theorem) The binomial probabilities $P_n(m) = C_n^m p^m q^{n-m}$ can be approximated by the corresponding values of the $N(a,\sigma)$ distribution with matching parameters, that is, with $E(X) = np,\ D(X) = npq$. More precisely,

$$P_n(m) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (11)$$

where the symbol $\sim$ means that the ratio of the two sides tends to 1 as $n \to \infty$.

The central limit theorem (CLT) states that even though the population distribution may be far from being normal, yet for a large sample size n, the distribution of the standardized sample mean is approximately standard normal with better approximations obtained with the larger sample size.

Theorem (The Central Limit Theorem) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean a and variance $\sigma^2 < +\infty$, then the limiting distribution of

$$Z_n = \frac{\bar{X} - a}{\left(\sigma / \sqrt{n}\right)} \tag{12}$$

is standard normal, that is Z_n converges in distribution to Z where Z denotes a standard normal random variable.

$$\lim_{n\to\infty} P\left(Z_n < x\right) = \Phi(x) \tag{13}$$

Example 4 (Total Weight of People in a Sample) Assume that the weight of adults in a population has a mean $a = 150$ pounds and standard deviation $\sigma = 30$ pounds. Find (approximately) the probability that the total weight of a random sample of 36 such people exceeds 5700 pounds.

Solution The weight of any individual is not a normal random variable, but a mixture (that is, the probability density function is a weighted average) of two, approximately normal, random variables: one for the women and one for the men. This fact is, however, immaterial because by the CLT, the total weight W is approximately normal with a mean $n \cdot a = 36 \cdot 150 = 5400$ and $SD = \sqrt{n} \cdot \sigma = \sqrt{36} \cdot 30 = 180$. Thus,

$$P\left(W > 5700\right) = P\left(\frac{W - 5400}{180} > \frac{5700 - 5400}{180}\right) \approx P(Z > 1.667) = 1 - \Phi(1.667) \approx 0.048.$$

Example 5 (Determining Sample Size) Suppose that in a public opinion poll the proportion p of voters who favor a certain proposition is to be determined. In other words, we want to estimate the unknown probability p of a randomly selected voter being in favor of the proposition. We take a random sample, with the responses being independent and identically distributed Bernoulli random variables X_i with parameter p and use $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$ to estimate p. Approximately how large must a random sample be taken to ensure that $P\left(\left|\bar{X}_n - p\right| < 0.1\right) \geq 0.95$?

Solution By the CLT

$$P\left(\left|\bar{X}_n - p\right| < 0.1\right) = P\left(\frac{\left|\bar{X}_n - p\right|}{\sqrt{p(1-p)/n}} < \frac{0.1}{\sqrt{p(1-p)/n}}\right) \approx P\left(|Z| < \frac{0.1}{\sqrt{p(1-p)/n}}\right) =$$

$$= 2\Phi\left(\frac{0.1}{\sqrt{p(1-p)/n}}\right) - 1.$$

Now, this quantity is greater than 0.95 if $\Phi\left(\frac{0.1}{\sqrt{p(1-p)/n}}\right) \geq 0.975$ or, equivalently, if

$$\frac{0.1}{\sqrt{p(1-p)/n}} \geq \Phi^{-1}(0.975) \approx 1.96 \Rightarrow \sqrt{n} \geq \frac{1.96}{0.1}\sqrt{p(1-p)} \Rightarrow n \geq 384.16 \cdot p(1-p).$$

Here $p(1-p)$ has its maximum at $p = 1/2$, and then $p(1-p) = 1/4$. Thus $n \geq 97$ ensures that $P\left(\left|\bar{X}_n - p\right| < 0.1\right) \geq 0.95$ for any value of p.

Example 6 Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n = 25$ from a population that has the mean $a = 71.43$ and the variance $\sigma^2 = 56.25$. Let $\bar{X}$ be a sample mean. What is the probability that the sample mean is between 68.91 and 71.97?

Solution The mean of $\bar{X}$ is given by $E(\bar{X}) = 71.43$. The variance of $\bar{X}$ is given by

$$D(\bar{X}) = \frac{\sigma^2}{n} = \frac{56.25}{25} = 2.25.$$

In order to find the probability that the sample mean is between 68.91 and 71.97, we need the distribution of the population. However, the population distribution is unknown. Therefore, we use the central limit theorem. The central limit theorem says that $\frac{\bar{X} - a}{\left(\sigma / \sqrt{n}\right)} \approx N(0,1)$ as n approaches infinity.

Therefore

$$P\left(68.91 \leq \bar{X} \leq 71.97\right) = P\left(\frac{68.91 - 71.43}{\sqrt{2.25}} \leq \frac{\bar{X} - 71.43}{\sqrt{2.25}} \leq \frac{71.97 - 71.43}{\sqrt{2.25}}\right) =$$

$$= P\left(-0.68 \leq W \leq 0.36\right) = P(W \leq 0.36) + P(W \leq 0.68) - 1 = 0.5941.$$

Example 7 Light bulbs are installed successively into a socket. If we assume that each light bulb has a mean life of 2 months with a standard deviation of 0.25 months, what is the probability that 40 bulbs last at least 7 years?

Solution Let X_i denote the life time of the i-th bulb installed. The 40 light bulbs last a total time of $S_{40} = X_1 + X_2 + \ldots + X_{40}$.

By the central limit theorem $\frac{\sum_{i=1}^{40} X_i - na}{\left(\sqrt{n\sigma^2}\right)} \approx N(0,1)$.

Thus $\frac{S_{40} - 40 \cdot 2}{\left(\sqrt{40 \cdot (0.25)^2}\right)} \approx N(0,1)$. That is $\frac{S_{40} - 80}{1.581} \approx N(0,1)$.

Therefore $P\left(S_{40} \ge 7(12)\right) = P\left(\frac{S_{40}-80}{1.581} \ge \frac{84-80}{1.581} \right) = P(Z \ge 2.530) = 0.0057$.

Exercise Set 1.10

1. The height of 100 persons is measured to the nearest inch. What is the approximate probability that the average of these rounded numbers differs from the true average by less than 1%?

2. Light bulbs are installed into a socket. Assume that each has a mean life of 2 months with standard deviation of 0.25 month. How many bulbs n should be bought so that one can be 95% sure that the supply of n bulbs will last 5 years?

3. The mathematical expectation of the amount of precipitation in a given region during a year is 60 cm. Determine the probability that precipitation in this region is not less than 180 cm.

4. The daily water flow in a given region is a random variable X, for which $\sigma(X) = 10000$ liters. Estimate the probability that during the day the water consumption in the given region deviates from the expectation in absolute value by more than 25,000 liters.

5. The shooter shoots the target 300 times, and the probability of hitting the target with each shot is 2/3. Determine the probability that the shooter will hit the target from 185 to 215 times.

6. As a result of a medical examination of 900 draftees, it was established that their average weight is 1.2 kg more than the average weight of draftees for one of the preceding periods. What is the probability of this deviation, if the standard deviation of the weight of recruits is 8 kg?

7. The variance of each of the 2500 independent random variables does not exceed five. Estimate the probability that the deviation of the average mean of these random variables from the average mean of their mathematical expectations does not exceed 0.4.

8. The probability of occurrence of event A in a separate test is 0.6. Applying the Bernoulli theorem, determine the number of independent tests, starting from which the probability of the event frequency deviation from its probability in absolute value less than 0.1, will be greater than 0.97.

Individual Tasks 1.10

I.

1. The average value of the wind speed at the earth at this point is equal to 16 km / h. Estimate the likelihood that at this point the wind speed will not exceed 80 km / h.

2. The standard deviation of each of the 2134 independent random variables does not exceed 4. Estimate the probability that the deviation of the average mean of these random variables from the average mean of their mathematical expectations does not exceed 0.5.

3. The number of sunny days per year for this region is a random variable X, for which $E(X)=75$ days. Estimate the probability that during the year there will be more than 200 sunny days in the region.

II.

1. The average value of water flow in a village is 50,000 l / d. Assess the likelihood that water consumption in this locality will not exceed 150,000 l / d.

2. The probability of occurrence of an event in one experiment is 0.5. Is it possible with a probability greater than 0.97 to assert that the number of occurrences of an event in 1000 experiments is in the range from 400 to 600?

3. Taking the probability of ripening of a corn stalk with three cobs equal to 0.75, estimate with the help of Chebyshev inequality the probability that among 3000 stalks of the experimental section of such stems there will be from 2190 to 2310 inclusive ones.

1.11 Two Dimensional Random Variables

Definition A system of n random variables $X=\left(X_1, X_2, \dots, X_n\right)$ is called the ***n-dimensional random variable*** or n-dimensional random vector.

A random variable $X=\left(X_1, X_2, \dots, X_n\right)$ is called a discrete [continuous] one if all its components X_i are discrete [*resp.* continuous] random variables.

For simplicity we will consider two-dimensional random variables $\left(X,Y\right)$.

Distribution law of a discrete two-dimensional random variable

Let $\left(X,Y\right)$ be a discrete two-dimensional random variable. It means that its components X,Y are those discrete. Let the random variable X have n possible values $x_1, x_2, \dots, x_n$, and Y have m possible values $y_1, y_2, \dots, y_m$. In this case we can represent the distribution law of $\left(X,Y\right)$ by the following distribution table (table 1.1).

Table 1.1

$Y \backslash X$	x_1	x_2	$\dots$	x_n
y_1	p_{11}	p_{12}	$\dots$	p_{1n}
y_2	p_{21}	p_{22}	$\dots$	p_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	p_{m1}	p_{m2}	$\dots$	p_{mn}

Here $P(X=x_i,\ Y=y_i)=p_{ij},\ \ i=\overline{1,n},\ \ j=\overline{1,m},\ \ \sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}=1.$

Definition The next function is called the distribution function of a two-dimensional random variable $\left(X,Y\right)$:

$$F(x,y)=P(X<x,\ Y<y)=P\left(\left(X,Y\right)\in R_{xy}\right) \qquad (1)$$

Here R_{xy} is the infinite open rectangle generated by two rays starting from the point $M(x,y)$ to the left and downwards in parallel to the coordinate axes (see figure 13). It is the infinite open quadrant with the right upper vertex at the point $M(x,y)$. The distribution function in question is the probability of hitting of the random variable (X,Y) in the rectangle (quadrant) R_{xy}.

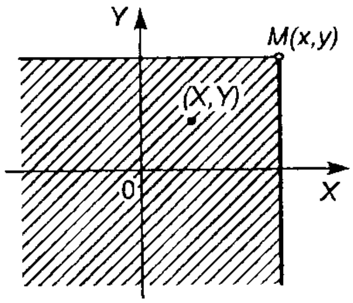


Figure 13

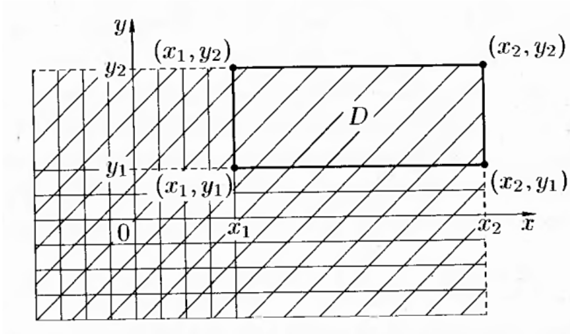


Figure 14

Properties of the distribution function

1. If $y \to +\infty$, the rectangle R_{xy} transforms into the left half-plane $R_{x,\infty}$ (with respect to the straight line passing through the point (x,y) in parallel to the *Oy*-axis). So we can take

$$F(x,+\infty) = P(X < x) = F_X(x)$$

by the definition, where $F_X(x)$ is the distribution function of the random variable X. If now $x \to +\infty$, we obtain

$$F(+\infty, y) = P(Y < y) = F_Y(y)$$

by analogy way, where $F_Y(y)$ is the distribution function of the random variable Y.

2. Directing $x \to +\infty$ and $y \to +\infty$, we rush the rectangle R_{xy} to the *xOy*-plane and the event $(X,Y) \in R_{xy}$ to a certain one. As a result we obtain at the property

$$F(+\infty,+\infty) = F_X(+\infty) = F_Y(+\infty) = 1.$$

3. Directing $x \to -\infty$ and (or) $y \to -\infty$, the rectangle R_{xy} vanishes, the event $(X,Y) \in R_{xy}$ rushes to the impossible one. As a result we obtain at the property

$$F(-\infty,-\infty) = F_X(-\infty) = F_Y(-\infty) = 0.$$

4. The probability of hitting of a two-dimensional random variable (X,Y) in a finite half-open rectangle (see figure 14) is determined by the formula

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \qquad (2)$$

5. The distribution function never decreases with respect to its each argument: if $x_1 < x_2, y_1 < y_2$, then $F(x_1,y) \le F(x_2,y), F(x,y_1) \le F(x,y_2)$.

6. The components X,Y of a two-dimensional random variable (X,Y) are independent if and only if the distribution function $F(x,y)$ of (X,Y) can be represented in the form of a product of the distribution functions of these components, $F(x,y)=F_X(x)\cdot F_Y(y)$.

Remark Knowing the distribution function of a two-dimensional random variable (X,Y) we always can find the distribution functions of its components X and Y (see the property 1). The inverse fact is not true in general. But the property 6 indicates that knowing the distribution functions of two independent random variables we can find the distribution function of their system (X,Y).

Let (X,Y) be a continuous two-dimensional random variable.

Definition The ***distribution density*** (or the probability density) of a two-dimensional continuous random variable (X,Y) is called the function $f(x,y)$ defined by the formula

$$f(x,y)=F''_{xy}(x,y) \tag{3}$$

that is the second mixed partial derivative of the distribution function $F(x,y)$ of this random variable.

Properties of the distribution density

1. The distribution density is a non-negative function, $f(x,y)\geq 0$.

2. The probability of hitting of a random variable (X,Y) in some domain D of the *xOy*-plane equals the double integral of its distribution density over the domain D,

$$P\big((X,Y)\in D\big)=\iint_D f(x,y)\,dx\,dy \tag{4}$$

3. As a result we can find the distribution function of a random variable (X,Y) that is the double integral of its distribution density over the rectangle R_{xy},

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv \tag{5}$$

4. On the base of the property (1) of the distribution function of a random variable (X,Y) we find the distribution functions of its components X and Y,

$$\begin{aligned} F_X(x)&=F(x,+\infty)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(u,y)\,du\,dy \\ F_Y(y)&=F(+\infty,y)=\int_{-\infty}^{y}\int_{-\infty}^{+\infty} f(x,v)\,dx\,dv \end{aligned} \tag{6}$$

5. On the supposition of satisfaction of corresponding conditions we find the distribution densities of the random variables X and Y by differentiating the obtained integrals (6) with respect to their variable upper bounds [limits] x and y,

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)\,dy; \quad f_2(y)=\int_{-\infty}^{\infty} f(x,y)\,dx \tag{7}$$

6. On the base of the property 2 of the distribution function $F(x,y)$ we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1 \qquad (8)$$

7. The components X, Y of a two-dimensional continuous random variable (X,Y) are independent if and only if the distribution density $f(x,y)$ of (X,Y) can be represented in the form of a product of the distribution densities of these components, $f(x,y) = f_1(x)\cdot f_2(y)$.

Example 1 A two-dimensional continuous random variable (X,Y) is given by the distribution density

$$f(x,y) = \begin{cases} Axy, & if\ (x,y)\in D, \\ 0, & if\ x\notin D. \end{cases}$$

where D is a rectangle: $D = \{(x,y): 0\le x\le a, 0\le y\le b\}$. Find the value of the parameter A, the distribution densities of the components X,Y and the distribution function of the given random variable (X,Y).

Solution We find the value of A on the base of the property 6 of the distribution density, namely

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \iint_D Axydxdy = A\int_0^a xdx\int_0^b ydy = \frac{A}{4}a^2b^2 = 1 \Rightarrow A = \frac{4}{a^2b^2}.$$

$$f(x,y) = \begin{cases} \dfrac{4}{a^2b^2}xy, & if\ (x,y)\in D, \\ 0, & if\ x\notin D. \end{cases}$$

To find the distribution densities of the components X,Y we will make use of the formulas (7).

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} \dfrac{4}{a^2b^2}\displaystyle\int_0^b xydy = \dfrac{4x}{a^2b^2}\int_0^b ydy = \dfrac{2x}{a^2}, & if\ (x,y)\in D, \\ 0, & if\ x\notin D. \end{cases}$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} \dfrac{4}{a^2b^2}\displaystyle\int_0^a xydx = \dfrac{4y}{a^2b^2}\int_0^a xdx = \dfrac{2y}{b^2}, & if\ (x,y)\in D, \\ 0, & if\ x\notin D. \end{cases}$$

Definition Let X and Y be any two random variables with the distribution density $f(x,y)$ and the distribution densities of the components $f_1(x)$ and $f_2(y)$. ***The conditional probability density function*** g of X, given (the event) $Y = y$, is defined as

$$g\left(x/y\right)=\frac{f(x,y)}{f_2(y)} \qquad f_2(y)>0 \tag{9}$$

Similarly, the conditional probability density function h of Y, given (the event) $X=x$, is defined as

$$h\left(y/x\right)=\frac{f(x,y)}{f_1(x)} \qquad f_1(x)>0 \tag{10}$$

Definition If X and Y are discrete random variables, then a point $\left(E(X),E(Y)\right)=(m_x, m_y)$ is called its mathematical expectation, whose coordinates are determined by the formulas

$$m_x=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij}, \quad m_y=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij} \tag{11}$$

Definition If X and Y are continuous random variables, then a point $\left(E(X),E(Y)\right)=(m_x, m_y)$ is called its mathematical expectation, whose coordinates are determined by the formulas

$$m_x=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\, f(x,y)\,dx\,dy, \quad m_y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\, f(x,y)\,dx\,dy \tag{12}$$

Definition If $\left(X,Y\right)$ is a two-dimensional random variable, then a point $\left(D(X),D(Y)\right)$ is called its variance, whose coordinates are determined by the formulas

$$D(X)=E\left((X-m_x)^2\right),\ D(Y)=E\left((Y-m_y)^2\right) \tag{13}$$

The expected value and the variance provided useful summary information about random variables. The new notions of covariance and correlation provide the information about the relationship between two random variables.

Definition Given random variables X and Y with expected values (m_x, m_y), their ***covariance*** is defined as

$$\operatorname{cov}\left(X,Y\right)=\mu_{xy}=E\left((X-m_x)(Y-m_y)\right) \tag{14}$$

whenever the expected value on the right-hand side exists.

Example 2 Let $\left(X,Y\right)$ be uniform on the triangle $D=\left\{(x,y):0\le x\le y\le 1\right\}$. Find the value of the covariance of the given random variable $\left(X,Y\right)$.

Solution A two-dimensional continuous random variable $\left(X,Y\right)$ is given by the distribution density

$$f(x,y)=\begin{cases}A, & if\ (x,y)\in D,\\ 0, & if\ x\notin D.\end{cases}$$

We find the value of A on the base of the property 6 of the distribution density, namely

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \iint_D A\,dx\,dy = A\int_0^1 dy\int_0^y dx = A\int_0^1 y\,dy = \frac{A}{2} = 1 \Rightarrow A = 2.$$

$$f(x,y) = \begin{cases} 2, & if\ (x,y)\in D, \\ 0, & if\ x \notin D. \end{cases}$$

If $f(x,y) = 2$ on D, and

$$m_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2x\,dx\,dy = 2\int_0^1\int_0^y x\,dx\,dy = \int_0^1 y^2 dy = \frac{1}{3},$$

$$m_y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2y\,dx\,dy = 2\int_0^1\int_0^y y\,dx\,dy = \int_0^1 2y^2 dy = \frac{2}{3},$$

$$\operatorname{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2\left(x-\frac{1}{3}\right)\left(y-\frac{2}{3}\right)dx\,dy = \int_0^1\left(y^2-\frac{2}{3}y\right)\left(y-\frac{2}{3}\right)dy = \frac{1}{36}.$$

Theorem If X and Y are random variables such that $E(X)$, $E(Y)$ and $E(XY)$ exist, then

$$\operatorname{cov}(X,Y) = \mu_{xy} = E\big((X-m_x)(Y-m_y)\big) = E(XY) - m_x\cdot m_y \qquad (15)$$

Theorem For independent random variables X and Y whose expectations exist, $\operatorname{cov}(X,Y) = 0$.

Definition We define the ***correlation coefficient*** of any random variables X and Y with nonzero variances and existing covariance as

$$r_{xy} = E\left(\frac{X-m_x}{\sigma_x}\cdot\frac{Y-m_y}{\sigma_y}\right) = \frac{\mu_{xy}}{\sigma_x\sigma_y} \qquad (16)$$

Properties of the covariance and correlation coefficient

1. $\mu_{xy} = \mu_{yx}$.

2. $\mu_{xx} = {\sigma_x}^2$, $\mu_{yy} = {\sigma_y}^2$.

3. For any random variables X and Y such that r_{xy} exists, $|r_{xy}| \le 1$.

4. $r_{xx} = r_{yy} = 1$.

5. For independent random variables X and Y whose expectations exist, $r_{xy} = 0$.

Definition Two random variables are called ***those correlated*** if their correlation coefficient does not equal zero and non-correlated otherwise.

Example 3 *(Correlation Between Two Exams)*. Suppose five students take two exams. Let X and Y denote the grades of a randomly selected student, as given in the X and Y columns of Table 1.2. Find the correlation coefficient of two random variables.

Solution The rest of the table is included for the computation of r_{xy}.

Table 1.2

Student	X	Y	X^2	Y^2	XY
A	40	50	1600	2500	2000
B	60	55	3600	3025	3300
C	80	75	6400	5625	6000
D	90	80	8100	6400	7200
E	80	90	6400	8100	7200
Ave.	70	70	5220	5130	5140

Hence $m_x = m_y = 70,\ \sigma_x = \sqrt{5220 - 70^2} \approx 17.889,\ \sigma_y = \sqrt{5130 - 70^2} \approx 15.166$

and $r_{xy} = \dfrac{5140 - 70^2}{17.889 \cdot 15.166} \approx 0.88$.

Exercise Set 1.11

1. A two-dimensional continuous random variable (X,Y) is given by the following table

$Y \backslash X$	0	1	2	3
- 1	0.01	0.06	0.05	0.04
0	0.04	0.24	0.05	0.17
1	0.05	0.10	0.10	0.09

Find the distribution law and mathematical expectations of the components X, Y, the correlation coefficient of two random variables.

2. A two-dimensional continuous random variable (X,Y) is given by the distribution density

$$f(x,y) = \begin{cases} 0, & \text{if} \quad x < 0,\ y < 0, \\ 6e^{-(3x+2y)}, & \text{if} \quad x \ge 0,\ y \ge 0. \end{cases}$$

Find the correlation coefficient of the given random variable (X,Y).

3. Let (X,Y) be uniform on the triangle $D = \{(x,y): 0 \le x \le 2,\ 0 \le y \le x\}$. Find the value of the covariance of the given random variable (X,Y).

4. Let (X,Y) be uniform on the half disc $D = \{(x,y): y > 0,\ x^2 + y^2 \le 1\}$. Find the correlation coefficient of two random variables.

5. A two-dimensional continuous random variable (X,Y) is given by the distribution density

$$f(x,y) = \begin{cases} \dfrac{1}{6\pi}, & \text{if} \quad \dfrac{x^2}{4} + \dfrac{y^2}{9} \le 1, \\ 0, & \text{if} \quad \dfrac{x^2}{4} + \dfrac{y^2}{9} > 1. \end{cases}$$

Find the correlation coefficient of the given random variable (X,Y).

Individual Tasks 1.11

I.

1. A two-dimensional continuous random variable (X,Y) is given by the following table

Y \ X	2	4	6	8
- 1	0.08	0.05	0.02	0.05
2	0.15	0.25	0.03	0.07
3	0.07	0.09	0.12	0.02

Find the distribution law and mathematical expectations of the components X,Y, the correlation coefficient of two random variables.

2. A two-dimensional continuous random variable (X,Y) is given by the distribution density

$$f(x,y)=\begin{cases}36xye^{-3(x^2+y^2)}, & if \quad x\ge 0,\ y\ge 0,\\ 0, & if \quad x<0,\ y<0.\end{cases}$$

Find the correlation coefficient of the given random variable (X,Y).

3. Let (X,Y) be uniform on the triangle $D=\{(x,y):\ 0\le y\le 3,\ 0\le x\le y\}$. Find the value of the covariance of the given random variable (X,Y).

II.

1. A two-dimensional continuous random variable (X,Y) is given by the following table

Y \ X	- 1	0	1
0	1/12	1/2	1/12
2	1/12	1/6	1/12

Find the distribution law and mathematical expectations of the components X,Y, the correlation coefficient of two random variables.

2. A two-dimensional continuous random variable (X,Y) is given by the distribution density

$$f(x,y)=\begin{cases}x+y, & if \quad x<0,\ y<1,\\ 0, & if \quad x\ge 0,\ y\ge 1.\end{cases}$$

Find the correlation coefficient of the given random variable (X,Y).

3. Let (X,Y) be uniform on the half disc $D=\{(x,y):\ x>0,\ x^2+y^2\le 4\}$. Find the correlation coefficient of two random variables.

II ELEMENTS OF MATHEMATICAL STATISTICS

2.1. General Remarks. Sampling Method. Variation Series

Let us suppose that we study some random variable X .

We will dwell upon three typical problems of the mathematical statistics.

1. *Exact or approximate determination of the distribution law of a random variable* (for example, it can be stated or hypothesized that a random variable X is distributed normally).

2. *Estimation (approximate calculation) of parameters of the distribution law of a random variable* (for example, estimation of $E(X)$; $\sigma(X)$; $D(X)$; $As(X)$ a random variable X).

3. *Testing statistical hypotheses* (for example, testing a hypothesis that a given random variable X is distributed normally).

There are various methods of solving such problems. One of the most widespread is ***the sampling method***. Suppose that we have some population, consisting of a great number N of things (the population of the size N) which must be studied with respect to some random variable X . We take randomly things $n \le N$ from the population, fulfill their allround testing with respect to X and extend obtained results on the whole population.

Using the language of mathematical statistics we do a sampling of the size n ($n \le N$) getting the sample (of the size n) which is subjected to thorough investigation with respect to a random variable in question. A sample must be representative, that is it must certainly represent the population. To be representative the sample must be a random one.

We study a sample of the size n, which we obtain by a random sampling from the population with the help of the so-called *variation (or statistical) series*. There are variation series of two types, namely those discrete and interval ones.

Table 2.1 A discrete variation series

X , x_i	x_1	x_2	…	x_k
m_i	m_1	m_2	…	m_k
$p_i^* = m_i / n$	$p_1^* = m_1 / n$	$p_2^* = m_2 / n$	…	$p_k^* = m_k / n$

1. A discrete variation series contains the row of observed values x_i of a random variable X (as the rule in increasing order), the row of numbers m_i of occurrences of these values and the row or their relative frequencies $p_i^* = m_i / n$ (table 2.1). It must be $\sum_{i=1}^{k} m_i = n$ and $\sum_{i=1}^{k} p_i^* = 1$.

Such a discrete variation series can be represented geometrically with the help of a polygon of frequencies or a polygon of relative frequencies. The polygon of relative frequencies is a broken line which successively joins the next points:

$$A_1(x_1, p_1^*),\ A_2(x_2, p_2^*), ...,\ A_n(x_n, p_n^*) \text{ (see figure 15).}$$

The polygon of frequencies is a broken line which successively joins the other points, namely those with ordinates (frequencies) $m_1, m_2, ..., m_k$

$$B_1(x_1, p_1),\ B_2(x_2, p_2), ...,\ B_n(x_n, p_n)\ .$$

If there are a lot of distinct observed values of a random variable X , they can be united in intervals which generate the so-called ***interval variation series***.

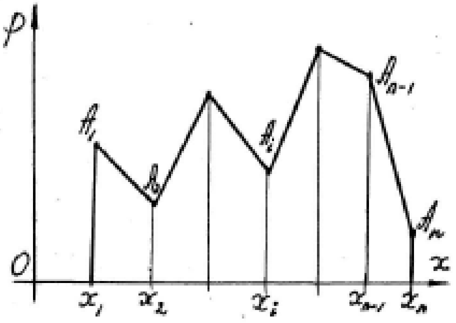


Figure 15

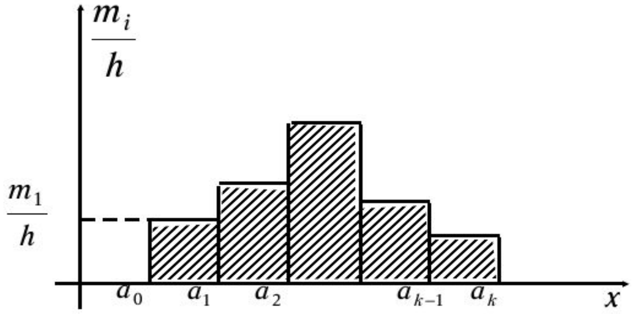


Figure 16

2. An interval variation series contains the row of intervals with all observed values of a random variable X in question, the row of numbers (frequencies) m_i of hitting of these values in the i-th interval and the row of corresponding relative frequencies $p_i^* = m_i / n$ (table 2.2).

Table 2.2 An interval variation series

Intervals	$a_0 - a_1$	$a_1 - a_2$	…	$a_{k-1} - a_k$
Frequencies m_i	m_1	m_2	…	m_k
$p_i^* = m_i / n$	$p_1^* = m_1 / n$	$p_2^* = m_2 / n$		$p_k^* = m_k / n$

Just as in the case of a discrete variation series it must be $\sum_{i=1}^{k} m_i = n$ and $\sum_{i=1}^{k} p_i^* = 1$.

The interval variation series can be represented geometrically by a histogram of frequencies or relative frequencies. The histogram of relative frequencies is the set of rectangles with the bases $(a_{k-1}; a_k]$, of the lengths $\Delta x_i = a_{i+1} - a_i$ and the areas $S_i = p_i^*$ (figure 16). The altitude of the i-th rectangle of such a histogram equals m_k / h .

The histogram of frequencies contains the rectangles of the areas $S_i = m_i$ with the same bases.

Often we form the intervals of the same length Δx (in particular on figure 16).

If one wants to generate k intervals, he can put

$$h = \frac{x_{\max} - x_{\min}}{k} \tag{1}$$

where $x_{\max}$ is the greatest and $x_{\min}$ is the least of the observed values of a random variable X . In practice it is useful sometimes to take some greater number instead of $x_{\max}$ and some less number instead of $x_{\min}$ in the formula (1).

One can preset the length of intervals instead of their number. For this purpose he can use the foiiowing known approximate formula:

$$h \cong \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg n} \tag{2}$$

He can take

$$a_0 = x_{\min} - h/2 \tag{3}$$

as the left point of the first interval and then get the other points as follows

$$a_i = x_{i-1} + h \tag{4}$$

The number k of the intervals is determined by the condition

$$a_k \geq x_{\max} \tag{5}$$

which means that the last k-th interval must contain the value $x_{\max}$ of the random variable X in question.

Having an interval variation series we must often compile a corresponding discrete variation series by taking some inner point x_i in the *i*-th interval. For example we can take

$$x_i = \frac{a_i + a_{i+1}}{2} \tag{6}$$

and obtain the discrete variation series which is given by the table 2.3. We conditionally suppose that the point x_i represents the *i*-th interval and therefore that the values $X = x_i$ were observed m_i times (with the relative frequency $p_i^* = m_i / n$). It means that the tables 2.2 and 2.3 contain the same second and third rows.

Table 2.3 The discrete variation series which corresponds to that interval

x_i	x_1	x_2	…	x_k
m_i	m_1	m_2	…	m_k
$p_i^* = m_i / n$	$p_1^* = m_1 / n$	$p_2^* = m_2 / n$	…	$p_k^* = m_k / n$

Example (***the Basic example***). To study a random variable X a sampling is fulfilled and the sample of the size n = 100 is obtained (see the table 2.4). Study the random variable.

Table 2.4 The sample with respect to the random variable X of the Basic example

24.8	26.2	25.6	24.0	26.4	25.2	26.7	25.4	25.3	26.1
24.3	25.3	25.6	26.7	24.5	26.0	25.7	25.0	26.4	25.9
24.4	25.4	26.1	23.4	26.5	25.9	23.9	25.7	27.1	24.9
23.8	25.6	25.2	26.4	24.2	26.5	25.7	24.7	26.0	25.8
24.3	25.5	26.7	24.9	26.2	26.7	24.6	26.0	25.4	25.0
25.4	25.3	24.1	26.6	24.8	25.6	23.7	26.8	25.2	26.1
24.5	25.4	25.1	26.2	24.2	26.4	25.7	23.9	27.2	25.0
23.9	25.6	24.9	24.5	26.2	26.7	24.3	26.1	27.7	25.8
25.6	25.2	24.2	26.0	24.7	26.5	23.5	25.4	27.1	24.0
26.2	24.2	25.5	26.0	25.7	26.4	24.6	27.0	25.2	26.9

$$x_{max} - x_{min} = 27.7 - 23.4 = 4.3$$

We have $x_{min} = 23.4$ and $x_{max} = 27.7$ here. Let us compile intervals of the length

$$h = \frac{x_{max} - x_{min}}{1 + 3.322 \cdot \lg n} = \frac{4.3}{1 + 3.322 \cdot \lg 100} = \frac{4.3}{7.644} \approx 0.56,$$

$$a_1 = x_{min} - \frac{h}{2} = 23.4 - 0.28 = 23.12,$$

$$a_1 = 23.12; \quad a_2 = 23.68; \quad a_3 = 24.24; \quad a_4 = 24.8; \quad a_5 = 25.36;$$

$$a_6 = 25.92; \quad a_7 = 26.48; \quad a_8 = 27.04; \quad a_9 = 27.6; \quad a_{10} = 28.16$$

Table 2.5 represents the interval variation series, inner points x_i for corresponding discrete variation series and altitudes h_i of rectangles for plotting the histogram of relative frequencies.

Table 2.5 The variation series for the Basic example

Intervals	x_k	Frequencies m_k	$p_i^* = m_i / n$	*Altitudes* $h_i = \frac{p_i^*}{\Delta x}$
23.12-23.68	23.40	2	0.02	0.04
23.68-24.24	23.96	11	0.11	0.20
24.24-24.80	24.52	14	0.14	0.25
24.80-25.36	25.08	14	0.14	0.25
25.36-25.92	25.64	23	0.23	0.41
25.92-26.48	26.20	20	0.20	0.36
26.48-27.04	26.76	12	0.12	0.21
27.04-27.60	27.32	3	0.03	0.05
27.60-28.16	27.88	1	0.01	0.02
$\sum$		100	1.00	

The histogram of relative frequencies for the interval and corresponding discrete variation series are represented on figure 17. For the sake of geometric visualization we draw two different scales along the axes.

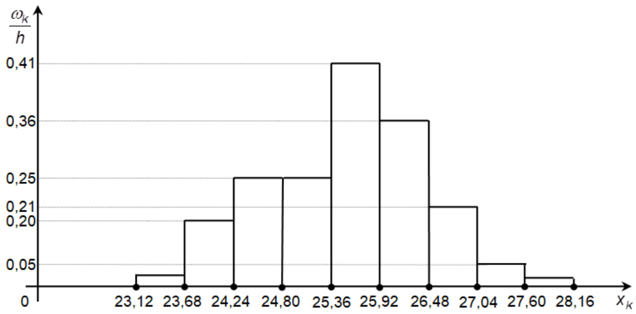


Figure 17

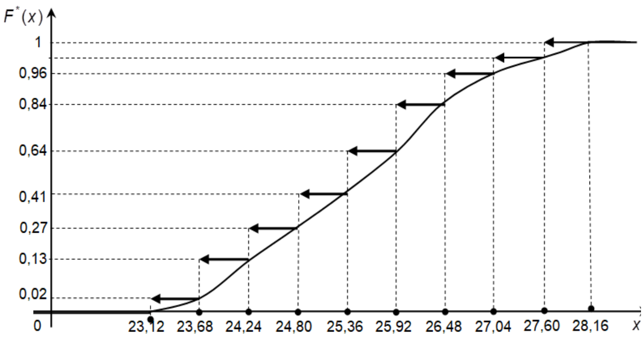


Figure 18

2.2 Estimation of Parameters of the Distribution Law

The statistic distribution function

Definition The statistic distribution function of a random variable X is called the relative frequency of hitting of its observed values on the interval $(-\infty, x)$.

$$F^*(x) = W(X < x) = \frac{n_x}{n} \qquad (1)$$

It is obvious that the statistic distribution function $F^*(x)$ equals the sum of relative frequencies of those observed values of X which are less than *x*.

$$F^*(x) = \sum_{x_i < x} p_i^* \qquad (2)$$

Example Find the statistic distribution function of the random variable X which is studied in the ***Basic example***.

Solution Proceeding from the discrete variation series (see inner points x_i and corresponding relative frequencies $p_i^* = m_i / n$ in the table 2.5) we obtain $F^*(x)$.

Using the interval variation series we get approximate values of the statistic distribution function which represent corresponding points on the *xOy*-plane and plot the approximate graph of the statistic distribution function in the form of a continuous line (figure 18).

$$F^*(x) = \begin{cases} 0, & x \le 23.12; \\ 0.02, & 23.12 < x \le 23.68; \\ 0.13, & 23.68 < x \le 24.24; \\ 0.27, & 24.24 < x \le 24.80; \\ 0.41, & 24.80 < x \le 25.36; \\ 0.64, & 25.36 < x \le 25.92; \\ 0.84, & 25.92 < x \le 26.48; \\ 0.96, & 26.48 < x \le 27.04; \\ 0.99, & 27.04 < x \le 27.60; \\ 1, & 27.60 < x \le 28.16; \\ 1, & x > 28.16. \end{cases}$$

Let θ be a parameter to be estimated and $\tilde{\theta}$ be its estimate [or *estimator*]. This latter is a function of the results of *n* trials on X.

In theory we consider the results of trials as random variables $X_1, X_2, ..., X_n$ with the same distribution law as X. In particular with the same mathematical expectation and variance

$$E(X_i) = E(X),\ D(X_i) = D(X), i = \overline{1,n} \ .$$

Respectively we suppose the estimate $\tilde{\theta}$ to be a function of these random variables:

$$\tilde{\theta} = f(X_1, X_2, ..., X_n).$$

In practice we express the estimate $\tilde{\theta}$ with the help of the results of trials on the random variable X represented by variation series. For discrete variation series (with k observed various values x_i of the random variable X) we can write $\tilde{\theta} = f(x_1, x_2, ..., x_n)$. For interval variation series (with k intervals) we use the inner points x_i^* of the intervals and so $\tilde{\theta} = f(x_1^*, x_2^*, ..., x_n^*)$.

There exist three necessary requirements which must be laid to every estimate: consistency, unbiasedness and efficiency [effectiveness].

1. An estimate $\tilde{\theta}$ is called *a consistent estimator of the parameter* θ if it converges to θ in probability ($n \to \infty \quad : P\left(\left| \tilde{\theta} - \theta \right| < \varepsilon \right) \to 1$)

2. An estimate $\tilde{\theta}$ is called *an unbiased estimator of the parameter* θ if its mathematical expectation equals θ ($E(\tilde{\theta}) = \theta$).

3. An estimate $\tilde{\theta}$ is called *an efficient [effective] estimator of* θ if it has minimal variance in comparison with all other estimates.

There are pointwise and interval estimates of the parameters of distribution laws of random variables.

In mathematical statistics we try at first to estimate corresponding parameters by analogous formulas.

1. The estimate [estimator] of the mathematical expectation [estimation of expectation].

$$E(\overline{x}_s) = \overline{x}_g \tag{3}$$

$$\overline{x_s} = \overline{x} = \frac{1}{n} \sum_{k=1}^{m} n_k x_k \tag{4}$$

2. The estimate [estimator] of the variance [variance estimate].

$$D_s = \frac{1}{n} \sum_{k=1}^{m} n_k \left(x_k - \overline{x} \right)^2 = \frac{1}{n} \sum_{k=1}^{m} n_k x_k^{\ 2} - \left(\overline{x} \right)^2 = \overline{x^2} - \left(\overline{x} \right)^2 \tag{5}$$

Therefore the sample dispersion D_s is a consistent but biased [shifted] estimate of the dispersion. To have an unbiased consistent estimator one introduces the so-called *corrected dispersion*

$$S^2 = \frac{n}{n-1} \cdot D_s, \quad E(S^2) = D_g \tag{6}$$

3. The root-mean-square deviation of the random variable X is estimated by the sample root-mean-square deviation

$$\sigma_s = \sqrt{D_s} \tag{7}$$

and the corrected root-mean-square deviation (or the standard deviation)

$$s = \sqrt{S^2} \tag{8}$$

Example Estimate the mathematical expectation, variance, root-mean-square deviation of the random variable X of the ***Basic example***.

Corresponding calculations are represented in table 2.6.

Table 2.6

$(a_k;a_{k+1})$	x_k	m_k	$x_k m_k$	$x_k{}^2 m_k$
23.12-23.68	23.40	2	46.8	1095.12
23.68-24.24	23.96	11	263.56	6314.898
24.24-24.80	24.52	14	343.28	8417.226
24.80-25.36	25.08	14	351.12	8806.09
25.36-25.92	25.64	23	589.72	15120.42
25.92-26.48	26.20	20	524	13728.8
26.48-27.04	26.76	12	321.12	8593.171
27.04-27.60	27.32	3	81.96	2239.147
27.60-28.16	27.88	1	27.88	777.2944
$\sum$		100	2549.44	65092.17

Thus the sample means of the random variable X and its square are equal to

$$\bar{x}_s = \frac{2549.44}{100} = 25.4944 \approx 25.49\,;\ \overline{x^2} = 650.9217$$

and therefore by (5) and (7) the sample variance and standard deviation are equal to

$$D_S = 650.9217 - 25.4944^2 = 650.9217 - 649.9644 \approx 0.9573;$$

$$\sigma_s = \sqrt{0.9573} \approx 0.9784 \approx 0.98\,.$$

The corrected dispersion and root-mean-square deviation are equal by the virtue of formulas (12), (14)

$$S^2 = \frac{100}{99}\cdot 0.9573 \approx 0.967 \approx 0.97;\ s = \sqrt{0.97} \approx 0.98\,.$$

Interval Estimates [Estimation by Confidence Interval]

Let θ be a parameter to be estimated and $\tilde{\theta}$ its estimate. Let δ be some small positive number (the so-called *accuracy*) and γ is some large probability (the *reliability*).

The next relation $P\left(\left|\theta-\tilde{\theta}\right|<\delta\right)=\gamma$ means that with the reliability γ: $\theta \in \left(\tilde{\theta}-\delta, \tilde{\theta}+\delta\right)$, it is therefore equivalent to the next one:

$$P\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)=P\left(\left|\theta-\tilde{\theta}\right|<\delta\right)=\gamma \qquad (9)$$

This last equality shows that with the reliability γ the estimating parameter θ is covered by a random interval $\left(\tilde{\theta}-\delta, \tilde{\theta}+\delta\right)$.

Definition An interval $\left(\tilde{\theta}-\delta, \tilde{\theta}+\delta\right)$ which covers an estimated parameter θ with the reliability γ is called a ***confidence*** one.

Let a random variable X be normally distributed and it is necessary to find the confidence interval for its mathematical expectation $E(X)=a$. We can suppose that the sample mean $\overline{x}_s$ and corrected root-mean-square deviation σ_s of X are found.

Let us study an auxiliary random variable

$$t=\sqrt{n}\cdot\frac{\overline{x}_S-a}{\sigma_s} \qquad (10)$$

It can be proved that t has Student distribution (or t-distribution) with $k=n-1$ degrees of freedom. Therefore for the given reliability γ we can find a number t_γ such that

$$P\left(|t|<t_\gamma\right)<\gamma \qquad (11)$$

There is a corresponding table for finding t_γ by the known γ and n (see Appendix table 4).

The formulas (10) and (11) determine the confidence interval in question.

1. If σ is known, then

$$\overline{x}_S-\frac{t\sigma}{\sqrt{n}}<a<\overline{x}_S+\frac{t\sigma}{\sqrt{n}} \qquad (12)$$

where $2\Phi(t)=\gamma$ or $P\left(\left|a-\overline{x}_B\right|<\frac{t\sigma}{\sqrt{n}}\right)=\gamma$.

2. If σ is unknown, then

$$\overline{x}_S-\frac{t_\gamma\cdot S}{\sqrt{n}}<a<\overline{x}_S+\frac{t_\gamma\cdot S}{\sqrt{n}} \qquad (13)$$

where $S^2=\frac{n}{n-1}\cdot D_s$, $s=\sqrt{S^2}$ and $t_\gamma=t(\gamma,n)$.

3. Confidence interval for the root-mean-square deviation σ is

$$S(1-q)<\sigma<S(1+q)\text{, if } q<1$$

$$0<\sigma<S(1+q)\text{, if } q>1$$

where we find the number $q=q(\gamma,n)$ according to table 5 of Appendix.

Example Supposing that the random variable X of the ***Basic example*** has a normal distribution, find the confidence intervals for its mathematical expectation a with the reliabilities $\gamma=0.95$.

For the reliability $\gamma=0.95$ and $n=100$ we have $t_\gamma=t(\gamma,n)=t(0.95;100)=1.984$ (see Appendix table 4). Hence the corresponding accuracy equals

$$\delta = \frac{s}{\sqrt{n}} t_\gamma, \; s = 0.98, \; \delta = \frac{0.98}{10} \cdot 1.984 \approx 0.1944 \approx 0.19$$

and the confidence interval is

$$\left(\bar{x} - \delta; \bar{x} + \delta\right) = \left(25.49 - 0.19; 25.49 + 0.19\right) = \left(25.30; 25.68\right).$$

The confidence interval for its root-mean-square deviation is calculated by the formula

$$s(1-q) < \sigma < s(1+q), \;\; s - \delta < \sigma < s + \delta.$$

$$q = q(\gamma, n) = q(0.95; 100) = 0.143, \; \delta = s \cdot q = 0.93 \cdot 0.143 = 0.13299 \approx 0.13.$$

The confidence interval for its root-mean-square deviation is

$$\left(s - \delta; s + \delta\right) = \left(0.98 - 0.13; 0.98 + 0.13\right) = \left(0.85; 1.11\right).$$

2.3 Testing Statistic Hypotheses

We will limit ourselves to hypotheses about the distribution law of a random variable which is investigated. For instance, our ***Basic example*** has generated the hypothesis that X has a normal distribution.

Let us test a hypothesis that a random variable X has a certain distribution law. For this purpose we introduce some non-negative random variable (***goodness-of-fit test***) K which is the measure of deviation of the theoretical assumption (based on the hypothesis) and the results of trials on the random variable. It is supposed that we know the exact or approximate distribution law of K. On the base of the result of trials on X (for example, on the base of the fulfilled sample) we find the so-called calculated value K_{calc} of K and compare it with some well defined critical value K_{crit} of the same K.

Let, for example, the critical value K_{crit} of the goodness-of-fit test K be defined by the relation

$$P(K > K_{crit}) > \alpha$$

where α is some small probability. This probability is often called the ***significance level*** in such a sense that the event $K > K_{crit}$ can be considered as highly improbable or even practically impossible. Correspondingly, we consider the occurrence of the result $K_{calc} > K_{crit}$ as a low-probability [unlikely] outcome.

The comparison of the calculated value K_{calc} of the test K with its critical value K_{crit} can give a rise to two cases: $K_{calc} > K_{crit}$ or $K_{calc} \le K_{crit}$.

In the first case we say that the results of trials *contradict the hypothesis* because of the occurrence of a low-probability event. Therefore we can reject the hypothesis in question. In the second case we say that the results of trials *do not contradict the hypothesis* and we can accept it.

Remark 1 It is necessary to understand that we ascertain the contradiction or noncontradiction of the results of trials to the hypothesis but *we do not state its validity or invalidity*.

Henceforth, we will study some frequently used goodness-of-fit tests.

Pearson χ^2-goodness-of-fit test

Let us introduce the next random variable (so-called Pearson distribution; it is often called Pearson χ^2-goodness-of-fit test or simply χ^2-goodness-of-fit test)

$$\chi^2{}_{calc} = \sum \frac{(m_i - m_i')^2}{m_i'}, \; m_i' = n \cdot P_i, \; i = \overline{1,k} \tag{1}$$

where n is the number of trials on the random variable X (for example, the size of a sample if the trials consist of sampling); the sense of the other quantities, namely the sense of m_i', P_i, k, depends on the form of a variation series which represents the results of trials.

a) In case of a *discrete variation series*, k is the number of different observed values of the random variable X and m_i is the number of occurrences of the value x_i of X and P_i is the probability of occurrence of this value $P_i = P(X = x_i)$. This latter is calculated on the base of the advanced hypothesis. If, for example, we have hypothesized that a random variable X has Poisson distribution, then

$$P(X = m) = P_n(m) = \frac{\lambda_s{}^m}{m!} e^{-\lambda_s} \tag{2}$$

where λ_s is the point estimate of the parameter λ (sample λ).

b) In the case of an *interval variation series*, k is the number of intervals, m_i is the number of values of X which have hit in the *i*-th interval (that is the frequency of hitting of observed values of X in this interval) and P_i is the probability of hitting of X in this interval $P_i = P(a_{i-1} < X < a_i)$, calculated by the virtue of the hypothesis.

For example, in the hypothesis of the normal distribution of a random variable X we can write

$$P_i = P \;\; (a_{i-1} < X < a_i) = \Phi \left(\frac{a_i - \overline{x_s}}{\sigma_s} \right) - \Phi \left(\frac{a_{i-1} - \overline{x_s}}{\sigma_s} \right) \tag{3}$$

The law of Pearson distribution (21) is known: it approximately coincides with the χ^2- distribution. The number ν of the degrees of freedom of this distribution is proved to be the following:

$$\nu = k - r - 1 \tag{4}$$

where r is the number of independent parameters which we estimate as to our random variable X. Values of the parameter r for some known distributions are represented in table 8. For the known number of degrees of freedom ν and a small probability (a significance level) α there exists the ***critical value*** χ^2_{crit} of χ^2 such that

$$P(\chi^2 > \chi^2_{crit}) = \alpha \qquad (5)$$

It can be found with the help of a corresponding table if we preset ourselves the significance level (see Appendix table 6).

	Hypothesis	*Parameters to be estimated*	*Value of the parameter*
1	X has a normal distribution	a, σ	2
2	X has Bernoulli distribution	$E(X) = np$	1
3	X has Poisson distribution with the parameter λ	$E(X) = \lambda$	1
4	X has an exponential distribution with the parameter λ	$E(X) = 1/\lambda$	1
5	X has a uniform distribution on $[a,b]$	There aren't such the parameters	0

The application of the goodness-of-fit test χ^2-is fulfilled according to following plan.

1. We find approximate values of all the probabilities $P_i = P(a_{i-1} < X < a_i)$ at first on the base of advanced hypothesis and define the calculated value χ^2_{calc}.

2. Knowing the number of degrees of freedom ν and specifying some significance level (a small probability) α, we find the critical value χ^2_{crit} from the table.

3. Now we compare χ^2_{calc} with χ^2_{crit}.

a) If $\chi^2_{calc} \le \chi^2_{crit}$, then we say that the results of trials do not contradict the hypothesis.

b) If $\chi^2_{calc} > \chi^2_{crit}$, then we say that the results of trials contradict the hypothesis (because of the event $\chi^2 > \chi^2_{crit}$, which we regarded as practically impossible, has occurred).

In the case (a) we can accept the hypothesis and in the case (b) we can reject it.

Example 1 Test the hypothesis that the random variable X of the ***Basic example*** has a normal distribution.

1. By virtue of the hypothesis we find approximate values of the probabilities of hitting of the random variable in all the intervals of the interval variation series using the formulas (3), (4). Corresponding evaluations up to finding the calculated value $\chi^2_{calc} \approx 4.1218$ are represented in tables 2.7, 2.8.

Table 2.7

a_k	$\frac{a_k - \overline{x}_s}{\sigma_s}$	$\Phi\left(\frac{a_k - \overline{x}_s}{\sigma_s}\right)$	$\Phi\left(\frac{a_{k+1} - \overline{x}_s}{\sigma_s}\right) - \Phi\left(\frac{a_k - \overline{x}_s}{\sigma_s}\right)$
23.12	-2.41	-0.4921	P_1=0.0243
23.68	-1.85	-0.4678	P_2=0.0681
24.24	-1.28	-0.3997	P_3=0.1417
24.80	-0.70	-0.2580	P_4=0.2063
25.36	-0.13	-0.0517	P_5=0.2217
25.92	0.44	0.1700	P_6=0.1738
26.48	1.01	0.3438	P_7=0.0991
27.04	1.58	0.4429	P_8=0.041
27.60	2.15	0.4839	P_9=0.0128
28.16	2.72	0.4967	$\sum$ =0.9888

Table 2.8

№	x_k	P_k	m_k	$m'_k = 100P_k$	$(m_k - m'_k)^2$	$\frac{(m_k - m'_k)^2}{m'_k}$
1	23.40	0.0243	2 } 13	2.43 } 9.24	14.1376	1.5300
2	23.96	0.0681	11	6.81		
3	24.52	0.1417	14	14.17	0.0289	0.0020
4	25.08	0.2063	14	20.63	43.9569	2.1307
5	25.64	0.2217	23	22.17	0.6889	0.0311
6	26.20	0.1738	20	17.38	6.8644	0.3950
7	26.76	0.0991	12 } 16	9.91 } 15.29	0.5041	0.0330
8	27.32	0.041	3	4.1		
9	27.88	0.0128	1	1.28		
$\sum$		0.9888	100	98.88		$\chi^2_{calc} \approx 4.1218$

2. Let us choose the significance level (a small probability) $\alpha = 0.05$. Finding the number of degrees of freedom by the formula (4)

$$\nu = k - r - 1,\ k = 6,\ r = 2,\ \nu = 3$$

we find the critical value $\chi^2_{crit}(\nu;\ \alpha) = \chi^2_{crit}(3;\ 0.05) = 7.815$ (see Appendix table 6).

3. We have obtained $\chi^2_{calc} = 4.1218 < \chi^2_{crit}$. It means that the results of trials on the random variable X in question do not contradict the hypothesis that its distribution law is a normal one.

Kolmogorov Goodness-of-Fit Test

Let X be a continuous random variable and we generate the following hypothesis: the given function $F(X)$ is the distribution function of this random variable.

Let a number D be the greatest value of the absolute value of a difference of the function $F(X)$ and the statistical distribution function $F^*(X)$ that is

$$D_{calc.} = \sqrt{n} \cdot \max_{1 \le i \le k} \left| F^*(x_i) - F(x_i) \right| \quad (6)$$

Table 2.9 gives some values of the D_{crit}.

Table 2.9 The values of the D_{crit}.

α	0.20	0.10	0.05	0.02	0.01	0.001
D_{crit}	1.073	1.224	1.358	1.520	1.627	1.950

a) If $D_{calc} \le D_{crit}$, then we say that the results of trials do not contradict the hypothesis.

b) If $D_{calc} > D_{crit}$, then we say that the results of trials contradict the hypothesis (because of the event $D > D_{crit}$, which we regarded as practically impossible, has occurred).

In the case (a) we can accept the hypothesis and in the case (b) we can reject it.

Example 2 Test the hypothesis that the random variable X of the ***Basic example*** has a normal distribution with the help of Kolmogorov goodness-of-fit test.

We represent all necessary calculations for finding the calculated value D_{calc} of D in the table 2.10 using the results obtained above (tables 2.5, 2.7 and 2.8). The values of the statistical distribution function $F^*(X)$ at the right end-points of the intervals were found in the example 2. By the same way we find approximate values of the distribution function $F(X)$.

Table 2.10

Intervals	p_i	$F(X)$	p_i^*	$F^*(X)$	$\left\lvert F^*(x_i) - F(x_i) \right\rvert$
23.12-23.68	0.0243	0.0243	0.02	0.02	0.0043
23.68-24.24	0.0681	0.0924	0.11	0.13	***0.0376***
24.24-24.80	0.1417	0.2341	0.14	0.27	0.0359
24.80-25.36	0.2063	0.4404	0.14	0.41	0.0304
25.36-25.92	0.2217	0.6621	0.23	0.64	0.0221
25.92-26.48	0.1738	0.8359	0.20	0.84	0.0041
26.48-27.04	0.0991	0.935	0.12	0.96	0.025
27.04-27.60	0.041	0.976	0.03	0.99	0.014
27.60-28.16	0.0128	0.9888	0.01	1	0.012
$\sum$		100	1.00		

1. If we choose the significance level $\alpha = 0.05$, we will get $D_{crit} = 1.358$ from the table 2.9.

2. Comparing $D_{calc} = \sqrt{100} \cdot 0.0376 = 0.376$ and $D_{crit} = 1.358$, we see that $D_{calc} \le D_{crit}$. Therefore the results of trials do not contradict the stated hypothesis.

Tasks for individual work on mathematical statistics

PROBLEM The independent trials are fulfilled on a random variable X and the results of trials are represented by the corresponding sample of the size *n*.

1. Compile the interval variation series for the random variable X. Plot the histogram of relative frequencies and the approximate graph of the distribution density. Find the statistical distribution function for the interval variation series and construct its approximate graph.

2. Form the discreet variation series on the base of the interval variation series by taking inner points in each interval. Construct the polygon of relative frequencies. Form the statistical distribution function for the discreet variation series and plot its graph.

3. Calculate the sample mean [the sample average], variance, root-mean-square deviation of the random variable.

4. Put forward a hypothesis about the type of distribution of the random variable X. Using the points 1, 2 and 3 substantiate the choice of the type of distribution. Find the corrected dispersion and standard deviation of the random variable. Compare their values with corresponding sample estimators.

5. Search out the confidence intervals for the mathematical expectation of the random variable X with reliabilities 0.95 and 0.99, basing on the hypothesis of its normal distribution (for task 2).

6. Find approximate values of the probabilities of hitting of the random variable X on all the intervals of its variation series proceeding from the hypothesis of the normal distribution of X (for task 2).

7. Test the hypothesis on the selected distribution of the random variable X with the significance levels $\alpha = 0.05$ making use of Pearson χ^2-goodness-of-fit tests and Kolmogorov goodness-of-fit test.

Individual Tasks 2.1

1.1

2 4 2 4 3 3 3 2 0 6 1 2 3 2 2 4 3 3 5 1 3 6 4 1 3 2
0 2 4 3 2 2 3 3 1 3 3 3 1 1 2 3 1 4 3 1 4 1 3 1 0 0 7
4 3 4 2 3 2 3 3 1 4 3 1 4 5 3 4 2 4 5 4 6 4 7 4 1 3

1.2

26.65	26.55	26.25	26.20	26.15	26.00	26.00	25.95	25.85	26.25
25.80	25.75	25.70	25.60	25.50	25.50	25.35	25.10	25.10	25.65
25.35	25.50	25.55	25.65	25.70	25.70	25.75	25.75	25.85	25.60
25.85	25.95	26.00	26.15	26.20	26.25	26.45	26.55	26.65	26.00
26.65	26.55	26.45	26.25	26.15	26.10	26.00	26.00	25.85	26.25
25.75	25.70	25.65	25.55	25.50	25.40	25.10	26.85	25.70	25.60
25.30	25.45	25.50	25.65	26.70	25.70	25.75	25.85	25.85	25.55
26.00	26.00	26.10	26.25	25.55	26.55	26.65	25.65	26.25	26.20
26.65	26.55	26.25	26.00	25.85	25.75	25.65	25.00	25.30	26.10
26.65	26.25	26.20	25.90	25.80	25.70	25.65	25.00	25.50	26.00

1.3

Intervals	0-50	50-100	100-150	150-200	200-250	250-300	300-350
m_i	45	23	17	8	4	3	2

2.1

0 4 2 0 5 1 1 3 0 2 2 4 3 2 3 3 0 4 5 1
3 1 5 2 0 2 2 3 2 2 2 6 2 1 3 1 3 1 5 4
5 5 3 2 2 0 2 1 1 3 2 3 5 3 5 2 5 2 1 1
2 3 4 3 2 3 2 4 2

2.2

2.54	0.69	2.59	2.60	0.89	1.85	0.95	2.89	1.07	2.92
2.23	1.13	2.27	2.30	1.22	1.75	1.25	2.44	1.29	2.22
2.08	1.38	2.10	2.11	1.43	1.64	1.46	2.20	1.49	1.48
1.91	1.58	1.93	1.95	1.62	1.44	1.66	2.04	1.69	2.07
1.80	2.58	1.82	1.83	1.75	1.25	1.76	1.87	1.78	1.90
1.71	2.24	1.73	1.19	1.84	0.90	2.84	1.77	1.88	1.09
1.54	2.09	1.60	1.41	1.97	1.99	2.34	1.68	2.05	1.31
1.34	1.91	1.40	1.61	2.13	2.15	2.18	1.28	2.90	1.51
0.67	1.81	1.16	1.74	2.31	2.33	2.03	1.28	2.46	1.31
1.12	1.72	0.73	0.84	2.69	2.75	1.86	1.02	2.21	1.09

2.3

Intervals	0-60	60-120	120-180	180-240	240-300	300-360	360-420
m_i	35	23	17	7	12	1	2

3.1

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

3.2

25.45	25.50	26.00	25.85	25.65	25.70	26.00	26.00	25.65	25.85
25.65	25.50	25.70	26.00	25.85	25.75	26.10	26.35	25.90	25.95
25.50	25.85	25.50	25.40	25.75	26.10	25.50	25.75	25.25	25.70
25.80	26.25	25.75	25.75	26.15	26.10	26.00	27.00	25.85	26.00
25.00	25.65	26.25	26.10	25.85	26.65	26.35	25.50	25.65	26.30
26.15	25.85	25.50	26.10	25.25	25.75	25.55	26.75	25.50	25.50
25.50	25.75	25.90	26.10	25.75	25.85	26.10	25.55	26.35	26.50
25.65	26.25	25.70	25.05	25.75	26.35	25.75	25.75	25.85	25.65
25.75	25.70	26.25	25.90	26.10	25.50	26.15	25.50	25.50	25.40
25.85	25.75	25.50	26.10	25.75	26.00	25.50	26.05	26.05	25.60

3.3

Intervals	0-70	70-140	140-210	210-280	280-350	350-420	420-490
m_i	33	27	13	7	5	3	1

4.1

4 6 0 2 1 3 3 1 2 5 3 1 2 2 4 4 4 3 2 5
2 5 1 2 3 0 3 0 5 1 2 1 3 0 4 0 2 2 1 0
5 1 4 2 4 2 1 3 1 0 6 1 2 1 4 2 2 0 2 4
2 2 1 2 2

4.2

2.74	3.01	2.99	3.11	3.00	3.00	3.02	3.02	2.93	3.18
3.13	2.93	3.11	3.01	3.00	3.05	2.99	3.07	2.91	3.04
3.15	3.06	3.03	3.07	3.21	3.13	3.09	3.01	3.07	3.04
2.99	2.93	3.02	3.05	3.03	2.97	2.93	3.09	3.11	3.00
3.09	3.09	3.01	3.04	3.07	3.10	3.19	2.92	3.03	3.05
2.78	3.15	3.09	3.06	3.03	3.08	2.96	3.16	2.95	3.00
3.00	3.32	3.12	3.05	3.01	3.02	2.89	3.02	2.99	3.14
3.01	3.12	2.98	3.03	2.95	3.03	3.12	3.11	3.10	3.01
3.02	3.16	3.08	3.08	2.97	3.08	2.95	2.98	3.02	2.97
3.01	3.10	3.12	3.02	3.11	2.92	2.99	3.02	3.04	3.05

4.3

Intervals	0-40	40-80	80-120	120-160	160-200	200-240	240-280
m_i	41	30	20	10	4	3	2

5.1

0 0 3 1 0 0 1 3 1 4 1 0 3 0 2 0 0 0 0 1
1 1 1 3 2 0 0 1 4 1 0 5 0 2 0 1 2 1 2 0
1 2 1 0 0 1 0 1 1 0 2 1 4 2 0 1 5 0 0 2
1 2 0 1 1 1 2 6 0 2 2 1 2 2 0 0 0 2 0 0
0 1 0 4

5.2

66.3	66.0	55.5	57.9	59.3	61.7	47.3	49.7	63.5	65.9
54.0	56.4	15.3	17.7	26.3	38.7	26.8	29.2	51.3	53.7
39.8	42.2	80.3	82.7	68.3	70.7	60.3	62.7	61.8	54.1
62.5	64.9	55.2	57.6	54.1	56.5	74.3	76.7	35.3	37.7
61.9	64.3	29.1	31.5	55.2	57.7	28.3	30.7	53.1	55.5
52.7	55.1	58.5	59.0	62.9	66.3	37.7	50.1	68.7	62.1
42.9	45.3	38.8	41.2	55.3	57.7	45.0	47.4	62.8	65.2
51.6	76.8	54.8	57.2	32.0	34.4	73.9	76.3	36.4	37.0
35.3	70.7	82.0	54.0	42.4	88.4	44.6	76.9	35.3	47.0
49.3	44.3	49.1	37.7	70.1	72.5	44.1	46.5	47.6	50.0

5.3

Intervals	0-56	56-112	112-168	168-224	224-280	280-336	336-448
m_i	35	24	17	7	4	2	1

6.1

2 0 0 3 1 2 2 2 3 4 1 2 3 3 2 1 1 3 3 0
4 1 3 3 0 1 0 0 1 2 1 1 3 2 3 0 1 0 4 2
3 1 2 1 1 1 1 2 1 2 5 2 1 3 2 3 1 1 1 1
2 1 1 1 3 1 3 1 2 1 2 1 1 0 0 3 3 1 2 3

6.2

122	172	208	187	208	194	115	201	172	93
129	115	194	43	108	86	122	122	108	201
136	180	100	158	151	129	144	187	172	144
115	79	158	129	100	136	122	108	158	172
122	165	100	158	151	108	93	172	158	72
151	64	129	136	108	151	144	100	57	129
165	151	144	165	151	136	165	100	129	108
100	136	158	129	50	122	151	151	86	129
172	108	100	129	115	100	180	136	93	122
172	136	115	122	151	129	172	144	86	144

6.3

Intervals	0-66	66-132	132-198	198-264	264-330	330-396	396-462
m_i	39	27	16	8	5	3	2

7.1

1 0 3 1 1 2 1 2 1 3 0 1 0 3 3 1 1 2 0 1 6
1 5 0 2 0 1 1 2 1 0 0 2 1 0 2 5 1 1 2 1 0
1 2 3 0 0 0 3 0 3 1 1 0 2 3 2 2 1 0 2 1 1
1 0 2 0 5 0 0 2 1 0 2 1 1 4 0 1 0 2 3 2 2
2 1 0 1 0 1 3 1 1 2 1 2 1 0 1 1 3 3 1 2 1

7.2

66.3	66.0	55.5	57.9	59.3	61.7	46.3	48.7	47.3	49.7
54.0	56.4	15.3	17.7	26.3	28.7	50.3	52.7	26.8	29.2
39.8	42.2	80.3	82.7	68.3	70.7	20.6	24.0	60.3	62.7
62.5	64.9	55.2	57.6	54.1	56.5	55.0	57.4	74.3	76.7
61.9	64.3	29.1	31.5	55.2	57.7	76.5	78.9	28.3	30.7
52.7	55.1	58.5	59.0	62.9	66.3	33.7	46.1	47.7	50.1
42.9	45.3	38.8	41.2	55.3	57.7	32.3	34.7	45.0	47.4
51.6	76.8	54.8	57.2	32.0	34.4	59.3	61.7	73.9	76.3
35.3	70.7	82.0	54.0	42.4	88.4	74.5	44.8	44.6	76.9
49.3	44.3	49.1	37.7	70.1	72.5	27.3	29.7	44.1	46.5

7.3

Intervals	0-76	76-152	152-228	228-304	304-380	380-456	456-532
m_i	45	32	20	10	6	4	3

8.1

1 1 2 0 1 1 0 3 1 1 2 1 2 1 3 0 1 0 3 3
2 0 1 1 2 0 1 5 0 0 0 2 1 0 2 1 1 2 1 0
1 1 2 1 0 1 1 3 3 1 2 3 0 0 0 3 0 3 1 1
0 1 0 2 1 1 1 2 1 2 1 0 2 1 1 0 1 1 4 0
0 5 0 0 2 1 2 1 0 2 3 2 0 2 3 2 2 0 2 1

8.2

59.3	61.7	46.3	48.7	90.1	92.5	47.3	49.7	81.8	84.2
26.3	28.7	50.3	52.7	47.8	50.2	60.3	62.7	54.3	56.7
68.3	70.7	20.6	24.0	68.1	70.5	83.8	81.2	72.3	74.7
54.1	56.5	55.0	57.4	27.7	30.1	57.3	59.7	75.3	78.1
55.2	57.7	76.5	78.9	56.1	58.5	83.6	86.0	32.3	34.7
62.9	66.3	33.7	46.1	54.1	56.5	54.1	56.5	54.7	65.2
55.3	57.7	32.3	34.7	23.7	26.1	40.9	40.3	33.7	36.1
32.0	34.4	59.3	61.7	36.5	38.9	43.1	51.3	74.4	51.7
42.4	88.4	74.5	44.8	51.4	50.7	93.5	53.8	68.3	97.7
70.1	72.5	27.3	29.7	46.3	48.7	23.1	25.5	35.3	37.7

8.3

Intervals	0-80	80-160	160-240	240-320	320-400	400-480	480-560
m_i	48	30	21	9	7	3	2

9.1

2 0 1 4 2 0 1 5 2 0 0 2 1 0 2 1 1 2 1 0
1 1 2 1 0 1 1 3 3 1 2 1 0 0 0 3 0 3 1 1
0 2 1 1 4 0 6 0 2 1 1 0 2 3 2 0 2 3 2 2
2 1 3 0 1 0 3 3 1 1 2 0 1 1 0 3 1 1 2 1
2 1 0 1 0 1 0 2 1 4 1 2 1 2 0 5 0 0 0 1

9.2

55.5	57.9	59.3	61.7	46.3	48.7	47.3	49.7	63.5	65.9
15.3	17.7	26.3	28.7	50.3	52.7	26.8	29.2	51.3	53.7
80.3	82.7	68.3	70.7	20.6	24.0	60.3	62.7	61.8	54.1
55.2	57.6	54.1	56.5	55.0	57.4	74.3	76.7	35.3	37.7
29.1	31.5	55.2	57.7	76.5	78.9	28.3	30.7	53.1	55.5
58.5	59.0	62.9	66.3	33.7	46.1	47.7	50.1	68.7	62.1
38.8	41.2	55.3	57.7	32.3	34.7	45.0	47.4	62.8	65.2
54.8	57.2	32.0	34.4	59.3	61.7	73.9	76.3	36.4	37.0
82.0	54.0	42.4	88.4	74.5	44.8	44.6	76.9	35.3	47.0
49.1	37.7	70.1	72.5	27.3	29.7	44.1	46.5	47.6	50.0

9.3

Intervals	0-90	90-180	180-270	270-360	360-450	450-540	540-630
m_i	50	33	21	8	4	2	2

10.1

1 0 2 3 2 0 2 3 2 2 0 5 0 0 0 1 2 1 0 1
1 1 2 0 1 5 0 3 1 6 2 1 2 1 3 0 1 0 3 3
2 0 1 1 2 0 1 5 0 0 0 2 1 0 2 1 1 2 4 0
0 1 0 2 1 1 1 2 1 2 1 3 3 1 0 2 1 1 4 0
0 0 2 1 1 0 1 2 1 0 0 0 3 0 3 1 1 1 1 2

10.2

81.8	54.3	72.3	75.3	32.3	62.8	33.7	74.4	68.3	35.3
84.2	56.7	74.7	78.1	34.7	65.2	36.1	51.7	97.7	37.7
64.6	49.3	59.3	37.3	52.7	50.0	50.1	72.5	68.0	48.3
70.1	67.0	51.7	60.7	39.7	55.1	59.4	54.4	70.4	37.7
63.5	51.3	61.8	35.3	53.1	68.7	62.8	36.4	35.3	47.6
65.9	53.7	54.1	37.7	55.5	62.1	65.2	37.0	47.0	50.0
46.3	50.3	20.6	55.0	76.5	33.7	32.3	59.3	74.5	27.3
48.7	52.7	24.0	57.4	78.9	46.1	34.7	61.7	44.8	29.7
47.3	60.3	83.8	57.3	83.6	54.7	40.9	43.1	93.5	23.1
25.5	53.8	51.3	40.3	57.1	86.0	59.7	81.2	62.7	49.7

10.3

Intervals	0-100	100-200	200-300	300-400	400-500	500-600	600-700
m_i	44	29	19	7	3	1	2

11.1

6 1 0 2 3 1 1 2 1 2 1 1 2 1 0 4 1 3 3 0
2 1 1 4 0 0 0 2 1 1 0 2 3 2 0 2 3 2 2 2
1 1 2 0 1 1 0 3 1 1 0 5 6 0 0 1 2 1 0 1
1 2 1 0 0 0 3 0 3 1 1 1 5 0 0 1 2 1 3 0
0 2 1 0 2 1 1 2 1 0 2 0 1 1 0 3 3 1 2 0

11.2

4.03	4.05	4.13	4.26	4.25	4.13	4.12	4.25	4.14	4.05
4.07	4.14	4.13	4.22	4.14	4.15	4.19	4.29	4.30	4.16
4.16	4.27	4.05	4.25	4.22	4.14	4.28	4.08	4.17	4.25
4.25	4.30	4.03	4.03	4.15	4.19	4.30	4.17	4.13	4.30
4.17	4.15	4.14	4.14	4.15	4.16	4.26	4.15	4.16	4.07
4.06	4.08	4.13	4.05	4.13	4.13	4.13	4.12	4.14	4.15
4.19	4.15	4.04	4.11	4.11	4.19	4.17	4.16	4.26	4.21
4.20	4.25	4.12	4.26	4.25	4.28	4.28	4.14	4.09	4.03
4.28	4.25	4.23	4.24	4.18	4.22	4.18	4.20	5.14	4.14
4.10	4.11	4.30	4.17	4.20	4.10	4.22	4.17	4.11	4.15

11.3

Intervals	0-44	44-88	88-132	132-176	176-220	220-264	264-308
m_i	46	24	18	10	5	7	4

12.1

0 1 0 1 0 2 0 1 1 2 1 2 1 1 2 0 1 1 0 1 0
3 1 1 2 1 2 1 3 0 1 0 3 3 0 5 0 0 0 1 2 1
1 5 0 0 0 2 1 3 2 1 0 1 1 3 3 1 2 0 2 1 1
3 1 1 2 1 0 2 0 2 3 2 2 0 0 0 2 1 1 0 0 0
0 1 1 2 0 1 1 1 0 2 3 3 0 2 4 3

12.2

5.50	5.53	5.55	5.60	5.53	5.57	5.55	5.49	5.60	5.57
5.50	5.50	5.45	5.53	5.55	5.62	5.57	5.65	5.62	5.55
5.49	5.60	5.50	5.57	5.60	5.40	5.53	5.55	5.65	5.33
5.60	5.45	5.57	5.50	5.62	5.68	5.50	5.55	5.60	5.57
5.62	5.45	5.50	5.53	560	5.40	5.55	5.57	5.53	5.49
5.68	5.55	5.60	5.68	5.57	5.70	5.55	5.45	5.57	5.55
5.60	5.55	5.60	5.49	5.50	5.53	5.57	5.55	5.62	5.53
5.60	5.70	5.53	5.55	5.60	5.49	5.50	5.62	5.53	5.55
5.60	5.47	5.57	5.55	5.55	5.49	5.53	5.57	5.60	5.68
5.57	5.60	5.49	5.53	5.60	5.62	5.49	5.50	5.67	5.50

12.3

Intervals	0-54	54-108	108-162	162-216	216-270	270-324	324-378
m_i	51	26	21	8	6	5	3

13.1

1 2 1 2 1 3 4 1 0 3 3 0 5 5 0 0 1 2 1 0 7 0 0 0 2
1 1 2 1 6 1 1 1 2 5 2 1 1 2 0 1 1 0 3 1 1 1 1 0 2
2 0 1 1 2 0 1 5 0 0 0 2 1 0 2 0 0 3 0 3 2 2 1 0 1
1 0 1 0 2 1 0 2 3 2 0 2 3 0 2 1 1 4 1 2 1 3 3 1 1

13.2

226	113	307	356	259	307	291	291	324	221
437	421	405	421	405	421	437	405	437	405
388	307	372	388	324	374	372	307	291	307
372	359	307	226	324	243	197	194	259	340
324	324	340	259	372	275	388	324	307	340
372	324	324	340	356	291	327	307	243	324
243	243	356	275	453	486	291	226	340	291
307	372	356	356	307	178	145	129	259	324
372	388	324	259	259	275	210	162	162	178
275	243	388	243	340	226	275	243	240	291

13.3

Intervals	0-64	64-128	128-192	192-256	256-320	320-384	384-448
m_i	52	27	20	9	5	4	3

14.1

2 1 1 1 2 1 2 2 0 5 1 2 7 1 5 0 0 0 2 1
1 1 2 6 1 1 0 3 1 1 2 1 2 1 3 0 1 0 3 3
1 0 2 0 1 1 3 3 1 2 1 6 0 0 3 0 3 1 1 0
1 1 2 0 1 0 2 3 2 2 1 0 2 1 1 3 1 0 1 1
0 0 2 1 1 0 2 3 2 0 4 0 5 0 0 0 1 2 1 0

14.2

43.9	43.7	68.7	43.6	17.5	45.0	27.2	43.2	40.0	23.7
47.7	14.7	56.7	42.5	43.7	52.3	43.7	20.4	70.4	37.5
34.7	38.7	39.0	43.4	64.9	22.1	20.7	47.7	30.8	58.5
35.7	15.2	48.7	15.7	69.9	62.3	33.4	36.1	16.7	32.7
33.0	32.5	23.0	51.2	48.1	23.7	41.5	40.2	51.9	39.7
37.7	53.0	46.7	25.7	41.1	45.4	41.5	56.4	23.7	36.0
58.4	36.7	24.9	12.1	42.5	44.5	16.1	56.6	36.2	78.5
34.7	39.8	37.5	43.1	72.0	45.6	72.1	48.7	35.7	29.3
70.2	42.7	40.7	64.1	20.7	51.2	22.1	37.7	83.7	11.5
23.6	56.7	62.8	31.3	41.1	50.3	50.9	28.2	52.0	42.4

14.3

Intervals	0-74	74-148	148-222	222-296	296-370	370-444	444-518
m_i	51	26	17	6	4	5	4

15.1

1 1 4 3 1 4 2 1 2 1 3 0 1 0 3 3 0 1 0 1
2 0 1 1 2 0 1 5 0 0 0 2 1 0 2 1 1 2 1 0
1 1 2 1 0 1 1 3 3 1 2 1 0 0 0 3 0 3 1 1
0 2 1 1 4 0 0 6 2 1 1 0 2 3 2 0 2 3 2 2
0 2 1 1 2 0 1 2 1 1 1 1 2 0 2 0 5 0 0 1

15.2

66.3	66.0	29.7	27.3	63.5	65.9	90.1	48.7	47.3	49.7
54.0	56.4	44.8	74.5	51.3	53.7	47.8	50.7	26.8	29.2
39.8	42.2	61.7	59.3	61.8	54.1	68.1	38.9	60.3	62.7
62.5	64.9	34.7	32.3	35.3	37.7	27.7	26.1	74.3	76.7
61.9	64.3	46.1	33.7	53.1	55.5	56.1	56.5	28.3	30.7
52.7	55.1	78.9	76.5	68.7	62.1	54.1	58.5	47.7	50.1
42.9	45.3	57.4	55.0	62.8	65.2	23.7	30.1	45.0	47.4
51.6	76.8	24.0	20.6	36.4	37.0	36.5	70.5	73.9	76.3
35.3	70.7	52.7	50.3	35.3	47.0	51.4	50.2	44.6	76.9
49.3	44.3	48.7	46.3	47.6	50.0	46.3	92.5	44.1	46.5

15.3

Intervals	0-50	50-100	100-150	150-200	200-250	250-300	300-350
m_i	45	23	17	8	4	3	2

2.4 Elements of Correlation Theory

There is *deterministic* (*functional, stiff*) dependence between random variables X, Y (for example, linear dependence $Y = aX + b$). And there is *undetermined* (*non-functional, non-stiff, statistic, correlation*) dependence between X, Y, for example, the dependence between labor productivity and living standard, between a state of health and the productivity of a worker, between the height and weight of a man.

Definition *Correlation dependence* between random variables X and Y is that between values of one variable and a corresponding mean value (average value, distribution center, mathematical expectation) of the other. Such dependence is defined by introducing conditional distributions of random variables and conditional mathematical expectations.

Definition The ***conditional mathematical expectation*** of a random variable Y, that is $f(x)$, is called the *regression function* of Y on X, its graph is called the *regression curve* (regression line) of Y on X and the equation

$$\overline{y}_x = f(x) \qquad (1)$$

is called the ***regression equation*** of Y on X.

By analogy, the regression function $\varphi(y)$, the regression curve (regression line) and the regression equation of X on Y are defined

$$\overline{x}_y = \varphi(y) \qquad (2)$$

Definition A ***correlation dependence*** between random variables X, Y is called a functional dependence between possible values of one random variable and corresponding regression function (average [mean] value) of the other.

Main problems of correlation theory (for the case of two random variables)

1. Determine the *form of correlation dependence* between random variables.

If regression functions of random variables X, Y are linear, then one says about a linear correlation (or linear correlation dependence) between these random variables. Otherwise one says about non-linear (or curvilinear) correlation.

2. Determine the *closeness of relation* between random variables X, Y. The closeness problem is resolved with the help of the correlation coefficient r_s, which is the measure of a linear dependence between X, Y, and the correlation ratios ρ_{xy}, ρ_{yx}, which are the measure of a functional (not necessary linear) dependence between X, Y.

Definition The ***correlation coefficient*** of random variables X, Y is defined by the formula

$$r_s = \frac{\sum xym_{xy} - n \cdot \overline{x} \cdot \overline{y}}{n\sigma_x\sigma_y} \qquad (3)$$

where $\sum xym_{xy} - n\cdot\bar{x}\cdot\bar{y}$ is the *correlation moment* of X, Y.

If random variables X, Y are independent, then $r_s = 0$. The converse is not true in general: there are dependent random variables with $r_s = 0$.

Definition Random variables X, Y are called *those correlated* if their correlation coefficient does not equal zero ($r_s \neq 0$), and *non-correlated* otherwise ($r_s = 0$).

It is known that $|r_s| \leq 1$, and $|r_s| = 1$ if X, Y are connected by the linear (functional) dependence $Y = aX + b$.

Linear Correlation

Let the regression functions of random variables X, Y be linear, that is there is a linear correlation between X, Y. It can be proved that the regression functions are given by the following formulas

$$\bar{y}_x - \bar{y} = r_s\cdot\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right),\ \bar{x}_y - \bar{x} = r_s\cdot\frac{\sigma_x}{\sigma_y}\left(y - \bar{y}\right) \qquad (4)$$

Corresponding regression straight lines $\bar{y}_x = f(x)$ and $\bar{x}_y = \varphi(y)$:

a) intersect at the point $(\bar{x}, \bar{y})$;

b) have the slopes $k_1 = r_s\cdot\frac{\sigma_y}{\sigma_x}$, $k_2 = r_s\cdot\frac{\sigma_x}{\sigma_y}$;

c) coincide if $|r_s| = 1$ that is if there is a linear functional dependence between the random variables X and Y;

d) are perpendicular respectively to the *Ox*-axis and *Oy*-axis if $r_s = 0$, that is if X and Y are not correlated.

The simplest case (every pair of random variables was observed only one time)

Let each pair $\left(X = x_i, Y = y_i\right)$ of the random variables X, Y occurs only one time. In this case we will use the following formulas

$$\bar{x}_s = \frac{\sum x_i}{n},\ \overline{x^2} = \frac{\sum x^2}{n},\ \bar{y}_s = \frac{\sum y_i}{n},\ \overline{y^2} = \frac{\sum y^2}{n},\ \overline{xy} = \frac{\sum x_i y_i}{n},$$

$$\sigma_s^{\,2}(y) = \sqrt{y^2 - \bar{y}_s^{\,2}},\ \sigma_s^{\,2}(x) = \overline{x^2} - (\bar{x}_s)^2,\ r_s = \frac{\sum xym_{xy} - n\,\bar{x}\,\bar{y}}{n\sigma_x\sigma_y}.$$

Example 10 enterprises collected data on the cost of repairing equipment Y (rubles) depending on the time of its use X (years). Estimate number characteristics of X, Y, regression functions of X on Y and Y on X and correlation coefficient.

X	4	5	5	6	8	10	8	7	11	6
Y	1.5	2.0	1.4	2.3	2.7	4.0	2.3	2.5	6.6	1.7

Solution We insert preliminary calculations into a table.

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
4	1.5	16	2.25	6.0
5	2.0	25	4.0	10.0
5	1.4	25	1.96	7.0
6	2.3	36	5.29	13.8
8	2.7	64	7.29	21.6
10	4.0	100	16.00	40.0
8	2.3	64	5.29	18.4
7	2.5	49	6.25	17.5
11	6.6	121	43.56	72.6
6	1.7	36	2.89	10.2
$\sum$ 70	27.0	536	94.78	217.1

$$\bar{x}_s = \frac{70}{10} = 7; \quad \overline{x^2} = \frac{536}{10} = 53.6; \quad \sigma_s^2(x) = 53.6 - 49 = 4.6;$$

$$\sigma_s(x) = 2.145; \quad \bar{y}_s = \frac{27}{10} = 2.7; \quad \overline{y^2} = \frac{94.78}{10} = 9.478;$$

$$\sigma_s^2(y) = 9.478 - 7.29 = 2.188; \quad \sigma_s(y) = 1.479.$$

$$\overline{xy} = \frac{217.1}{10} = 21.71; \quad r_s = \frac{21.71 - 7 \cdot 2.7}{2.145 \cdot 1.479} = 0.886.$$

This latter is sufficiently large, so we can say that X and Y are connected by the essential linear dependence. By the formula (4) we find the estimates of values of the regression function X on Y: $\bar{y}_x - \bar{y} = r_s \cdot \frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right)$,

$$\bar{y}_x - 2.7 = 0.886 \cdot \frac{1.479}{2.145} \cdot (x - 7), \quad \bar{y}_x = 0.61x - 1.58,$$

and Y on X: $\bar{x}_y - \bar{x} = r_s \cdot \frac{\sigma_x}{\sigma_y}\left(y - \bar{y}\right)$,

$$\bar{x}_y - 7 = 0.886 \cdot \frac{2.145}{1.479} \cdot (y - 2.7), \quad \bar{x}_y = 1.28y + 3.53.$$

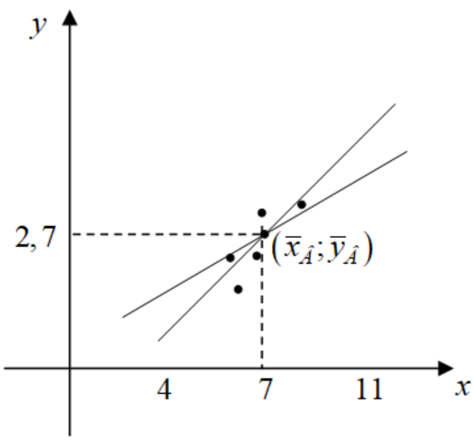


Figure 19

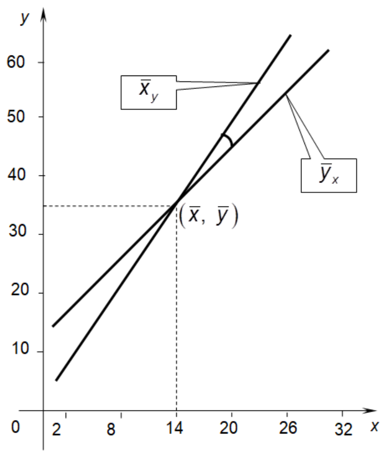


Figure 20

Let us plot the graphs of the obtained straight lines on one drawing (see Figure 19). The nearer to zero the acute angle between them (marked by the arc), the closer the connection between the features.

General case

Now we will consider the *general case* when as a rule several values of a random variable Y correspond to an arbitrary value of a random variable X or when an arbitrary pair of values $\left(X = x_i, Y = y_i\right)$ of random variables X, Y appears n_{ij} times. We study the correlative dependence between random variables X, Y with the help of the following table, which is called the ***correlation table*** (table 2.10).

Table 2.10

$X \backslash Y$	y_1	y_2	…	y_p	m_x
x_1	n_{11}	n_{12}	…	n_{1p}	m_{x1}
x_2	n_{21}	n_{22}	…	n_{2p}	m_{x2}
…	…	…	…	…	…
x_k	n_{k1}	n_{k2}	…	n_{kp}	m_{xk}
m_y	m_{y1}	m_{y2}	…	m_{yp}	n

In this case we will use the following formulas

$$\bar{x}_s = \frac{\sum x_i m_x}{n}, \quad \overline{x^2} = \frac{\sum x_i^2 m_x}{n}, \quad \sigma_s^2(x) = \overline{x^2} - (\bar{x}_s)^2, \ \sigma_s^2(y) = \overline{y^2} - (\bar{y}_s)^2$$

$$\bar{y}_s = \frac{\sum y_j m_y}{n}, \quad \overline{y^2} = \frac{\sum y_j^2 m_y}{n}, \quad \overline{xy} = \frac{\sum\sum x_i y_j m_{xy}}{n}, \quad r_s = \frac{\sum xy m_{xy} - n\bar{x}\,\bar{y}}{n\sigma_x\sigma_y}.$$

Example The results of n = 100 trials on random variables X and Y are represented in the table. Estimate number characteristics of X, Y, regression functions of X on Y and Y on X, the correlation coefficient.

$X \backslash Y$	10	20	30	40	50	60	m_x
2	2	4					6
8		3	7				10
14		1	48	10	2		61
20			2	7	5		14
26				1	2	2	5
32					2	2	4
m_y	2	8	57	18	11	4	$n = 100$

Solution We insert preliminary calculations into an “extended” table.

X \ Y	10	20	30	40	50	60	m_x	xm_x	x^2m_x
2	2	4					6	12	24
8		3	7				10	80	640
14		1	48	10	2		61	854	11956
20			2	7	5		14	280	5600
26				1	2	2	5	130	3380
32					2	2	4	128	4096
m_y	2	8	57	18	11	4	*n = 100*	1484	25696
ym_y	20	160	1710	720	550	240	3400		
y^2m_y	200	3200	51300	28800	27500	14400	125400		

Estimate the number characteristics of the random variables X and Y.

$$\bar{x}=\frac{\sum xm_x}{n}=\frac{1484}{100}=14.84;\ \bar{y}=\frac{\sum ym_y}{n}=\frac{3400}{100}=34;$$

$$D_x=\frac{\sum x^2m_x}{n}-\bar{x}^2=\frac{25696}{100}-14.84^2\approx 36.73;\quad \sigma_x=\sqrt{36.73}\approx 6.06;$$

$$D_y=\frac{\sum y^2m_y}{n}-\bar{y}^2=\frac{125400}{100}-34^2=98,\quad \sigma_y=\sqrt{98}\approx 9.90.$$

$$\sum xym_{xy}=2(10\cdot 2+20\cdot 4)+8(20\cdot 3+30\cdot 7)+14(20\cdot 1+30\cdot 48+40\cdot 10+50\cdot 2)+$$
$$+20(30\cdot 2+40\cdot 7+50\cdot 5)+26(40\cdot 1+50\cdot 2+60\cdot 2)+32(50\cdot 2+60\cdot 3)=55400.$$

Let us pass to the estimation of the correlation coefficient of random variables X and Y.

$$r_s=\frac{\sum xym_{xy}-n\cdot\bar{x}\cdot\bar{y}}{n\cdot\sigma_x\sigma_y}=\frac{55400-100\cdot 14.84\cdot 34}{100\cdot 6.06\cdot 9.9}\approx 0.824.$$

This latter is sufficiently large, so we can say that X and Y are connected by the essential linear dependence. By the formula (4) we find the estimates of values of the regression function X on Y: $\bar{y}_x-\bar{y}=r_s\cdot\frac{\sigma_y}{\sigma_x}(x-\bar{x})$,

$$\bar{y}_x-34=0.824\cdot\frac{9.9}{6.06}(x-14.84),\ \bar{y}_x=1.35x+13.97,$$

and Y on X: $\bar{x}_y-\bar{x}=r_s\cdot\frac{\sigma_x}{\sigma_y}(y-\bar{y})$,

$$\bar{x}_y-14.84=0.824\cdot\frac{6.06}{9.9}(y-34),\ \bar{x}_y=0.5y-2.16.$$

Let us plot the graphs of the obtained straight lines on one drawing (see Figure 20).

The nearer to zero the acute angle between them (marked by the arc), the closer the connection between the features. But if this angle is close to 90º, this indicates that the connection is weak or absent at all.

Individual Tasks 2.2

The results of trials on random variables X and Y are represented in the table. Estimate number characteristics of X, Y regression functions of X on Y and Y on X, the correlation coefficient. Plot the graphs of the obtained regression straight lines.

1.1

x_i	10	12	12.5	13	14	14.5	15.2	15.8	16	16.5
y_i	7	6	7	8	7	8	10	11	11	12

1.2

X \ Y	2	5	8	11	14	17	m_x
15	5	3	6				14
25		7	8	11			26
35			9	10	12		31
45				9	9		18
55					7	4	11
m_y	5	10	23	30	28	4	$n = 100$

2.1

x_i	6	6.3	7	7.3	8	8.7	9	9.5	10.7	11
y_i	12	12	12	11	13	13	14	14	15	16

2.2

X \ Y	3	5	7	9	11	13	m_x
12	4	3	5				12
14	6	7	8				21
16		10	12	11			33
18			8	8	5		21
20				4	5	4	13
m_y	10	20	33	23	10	4	$n = 100$

3.1

x_i	7	7.3	8	8.3	9	9.7	10	11	11.7	12
y_i	2	2	3	2	4	4	5	5	6	6

3.2

Y \ X	4	8	12	16	20	24	m_x
1	3	2	9				14
4		7	10	9			26
7			12	10	5		27
10				9	8	5	22
13					6	5	13
m_y	3	9	31	28	19	10	$n = 100$

4.1

x_i	2	2.5	3	3.4	3.6	4	4.5	5	5.2	6.8
y_i	3	2	2	3	3	4	4	4.5	4.5	5

4.2

Y \ X	11	14	17	20	23	26	m_x
10	4	6	3				13
15		7	9	10			26
20			13	9	7		29
25				12	6	3	21
30					6	5	11
m_y	4	13	25	31	19	8	$n = 100$

5.1

x_i	4	4.5	5.5	6	6.5	7	7.2	7.8	8	10
y_i	2	3	3	4	4	5	5	4	5	6

5.2

Y \ X	4	12	20	28	36	44	m_x
1				5	4	6	15
5				11	6		17
9			5	14	8		27
13		9	8	7			24
17	4	6	7				17
m_y	4	15	20	37	18	6	$n = 100$

6.1

x_i	77	96	86	92	98	63	80	53	64	66
y_i	81	77	76	86	53	36	40	47	49	60

6.2

Y / X	6	8	10	12	14	16	m_x
11				6	5	3	14
16			10	8	6		24
21		7	12	9			28
26	6	8	10				24
31	3	7					10
m_y	9	22	32	23	11	3	$n = 100$

7.1

x_i	81	57	86	80	87	163	153	133	159	134
y_i	54	40	61	68	88	145	136	129	126	96

7.2

Y / X	2	8	14	20	26	32	m_x
10					7	3	10
20			6	9	7	2	24
30		8	12	10			30
40	5	7	12				24
50	4	6	2				12
m_y	9	21	32	19	14	5	$n = 100$

8.1

x_i	129	145	142	120	95	107	133	140	149	147
y_i	100	95	206	118	109	107	120	114	113	123

8.2

Y / X	10	15	20	25	30	35	m_x
5				5	5	3	13
12			5	9	7		21
19		9	13	6			28
26	9	8	10				27
33	4	7					11
m_y	13	24	28	20	12	3	$n = 100$

9.1

x_i	104	108	93	124	112	113	95	112	116	93
y_i	94	84	73	107	94	107	99	100	104	88

9.2

Y \ X	2	4	6	8	10	12	m_x
13				5	4	3	12
17			5	10	6		21
21			9	14	5		28
25		6	12	6			24
29	5	4	6				15
m_y	5	10	32	35	15	3	*n = 100*

10.1

x_i	96	112	136	104	103	115	123	111	127	129
y_i	84	94	162	98	77	88	94	76	84	73

10.2

Y \ X	14	21	28	35	42	49	m_x
3				3	6	5	14
4				9	5	3	17
5			16	8	4		28
6		7	10	5			22
7	6	13					19
m_y	6	20	26	25	15	8	*n = 100*

11.1

x_i	60	62	68	70	73	75	82	84	87	88
y_i	175	171	160	151	150	141	133	131	125	120

11.2

Y \ X	50-62	62-74	74-86	86-98	98-110	110-122	122-134	134-146	m_x
0.7-1.1	2	3	5						10
1.1-1.5		6	3	5					14
1.5-1.9			5	8	15				28
1.9-2.3				6	9	10			25
2.3-2.7					1	6	8		15
2.7-3.1						3	4	1	8
m_y	2	9	13	19	25	19	12	1	*n = 100*

12.1

x_i	80	56	85	79	86	162	152	132	158	133
y_i	55	41	62	69	89	146	137	130	126	97

12.2

X \ Y	10-30	30-50	50-70	70-90	90-110	110-130	130-150	150-170	m_x
50-150						3	7	2	12
150-200					5	4	6		15
200-250				7	9	8			24
250-300			5	14	7				26
300-350		4	7	5					16
350-400	3	4							7
m_y	3	8	12	26	21	15	13	2	$n = 100$

13.1

x_i	73	92	82	88	98	59	76	49	60	62
y_i	77	73	72	82	49	32	36	43	45	56

13.2

X \ Y	200-400	400-600	600-800	800-1000	1000-1200	1200-1400	1400-1600	1600-1800	m_x
2.5-7.5	1	2	5						8
7.5-12.5		2	7	4					13
12.5-17.5			9	6	4				19
17.5-22.5				14	6	7			27
22.5-27.5					1	8	9		18
27.5-32.5						4	5	6	15
m_y	1	4	21	24	11	19	14	6	$n = 100$

14.1

x_i	92	108	132	100	99	114	119	107	123	125
y_i	80	90	158	94	73	84	90	72	80	69

14.2

X \ Y	210-310	310-410	410-510	510-610	610-710	710-810	810-910	910-1010	m_x
1-5							7	2	9
5-9						7	8		15
9-13				15	5	9			29
13-17			6	6	7				19
17-21		5	9	2					16
21-25	2	4	6						12
m_y	2	9	21	23	12	16	15	2	$n = 100$

15.1

x_i	6.7	6.9	7.2	8.4	8.8	9.1	9.8	10.6	10.7	11.1
y_i	2.8	2.2	3.0	3.2	3.7	4.0	4.8	6.0	5.4	5.2

15.2

Y / X	80-160	160-240	240-320	320-400	400-480	480-560	560-640	640-720	m_x
8.5-12.5	4	5	2						11
12.5-16.5		6	7	5					18
16.5-20.5			6	8	14				28
20.5-24.5					12	9	2		23
24.5-28.5					6	4			10
28.5-32.5						5	3	2	10
m_y	4	11	15	13	32	18	5	2	$n = 100$

Appendix

Table 1 The values of the function $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0.4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0.2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2179	2155	2331	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2.0	0.0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0.0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3.5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3.6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3.8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3.9	0002	0002	0002	0002	0002	0002	0002	0001	0001	0001

Table 2 The values of the function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.0000	0.45	0.1736	0.90	0.3159	1.35	0.4115	1.80	0.4641	2.50	0.4938
0.01	0.0040	0.46	0.1772	0.91	0.3186	1.36	0.4131	1.81	0.4649	2.52	0.4941
0.02	0.0080	0.47	0.1808	0.92	0.3212	1.37	0.4147	1.82	0.4656	2.54	0.4945
0.03	0.0120	0.48	0.1844	0.93	0.3238	1.38	0.4162	1.83	0.4664	2.56	0.4948
0.04	0.0160	0.49	0.1879	0.94	0.3264	1.39	0.4177	1.84	0.4671	2.58	0.4951
0.05	0.0199	0.50	0.1915	0.95	0.3289	1.40	0.4192	1.85	0.4678	2.60	0.4953
0.06	0.0239	0.51	0.1950	0.96	0.3315	1.41	0.4207	1.86	0.4686	2.62	0.4956
0.07	0.0279	0.52	0.1985	0.97	0.3340	1.42	0.4222	1.87	0.4693	2.64	0.4959
0.08	0.0319	0.53	0.2019	0.98	0.3365	1.43	0.4236	1.88	0.4699	2.66	0.4961
0.09	0.0359	0.54	0.2054	0.99	0.3389	1.44	0.4251	1.89	0.4706	2.68	0.4963
0.10	0.0398	0.55	0.2088	1.00	0.3413	1.45	0.4265	1.90	0.4713	2.70	0.4965
0.11	0.0438	0.56	0.2123	1.01	0.3438	1.46	0.4279	1.91	0.4719	2.72	0.4967
0.12	0.0478	0.57	0.2157	1.02	0.3461	1.47	0.4292	1.92	0.4726	2.74	0.4969
0.13	0.0517	0.58	0.2190	1.03	0.3485	1.48	0.4306	1.93	0.4732	2.76	0.4971
0.14	0.0557	0.59	0.2224	1.04	0.3508	1.49	0.4319	1.94	0.4738	2.78	0.4973
0.15	0.0596	0.60	0.2257	1.05	0.3531	1.50	0.4332	1.95	0.4744	2.80	0.4974
0.16	0.0636	0.61	0.2291	1.06	0.3554	1.51	0.4345	1.96	0.4750	2.82	0.4976
0.17	0.0675	0.62	0.2324	1.07	0.3577	1.52	0.4357	1.97	0.4756	2.84	0.4977
0.18	0.0714	0.63	0.2357	1.08	0.3599	1.53	0.4370	1.98	0.4761	2.86	0.4979
0.19	0.0753	0.64	0.2389	1.09	0.3621	1.54	0.4382	1.99	0.4767	2.88	0.4980
0.20	0.0793	0.65	0.2422	1.10	0.3643	1.55	0.4394	2.00	0.4772	2.90	0.4981
0.21	0.0832	0.66	0.2454	1.11	0.3665	1.56	0.4406	2.02	0.4783	2.92	0.4982
0.22	0.0871	0.67	0.2486	1.12	0.3686	1.57	0.4418	2.04	0.4793	2.94	0.4984
0.23	0.0910	0.68	0.2517	1.13	0.3708	1.58	0.4429	2.06	0.4803	2.96	0.4985
0.24	0.0948	0.69	0.2549	1.14	0.3729	1.59	0.4441	2.08	0.4812	2.98	0.4986
0.25	0.0987	0.70	0.2580	1.15	0.3749	1.60	0.4452	2.10	0.4821	3.00	0.4987
0.26	0.1026	0.71	0.2611	1.16	0.3770	1.61	0.4463	2.12	0.4830	3.20	0.4993
0.27	0.1064	0.72	0.2642	1.17	0.3790	1.62	0.4474	2.14	0.4838	3.40	0.4997
0.28	0.1103	0.73	0.2673	1.18	0.3810	1.63	0.4484	2.16	0.4846	3.60	0.4998
0.29	0.1141	0.74	0.2703	1.19	0.3830	1.64	0.4495	2.18	0.4854	3.80	0.4999
0.30	0.1179	0.75	0.2734	1.20	0.3849	1.65	0.4515	2.20	0.4861	4.00	0.4999
0.31	0.1217	0.76	0.2764	1.21	0.3869	1.66	0.4505	2.22	0.4868	4.50	0.5000
0.32	0.1255	0.77	0.2794	1.22	0.3883	1.67	0.4525	2.24	0.4875	5.00	0.5000
0.33	0.1293	0.78	0.2823	1.23	0.3907	1.68	0.4535	2.26	0.4881		
0.34	0.1331	0.79	0.2852	1.24	0.3925	1.69	0.4545	2.28	0.4887	↓	↓
0.35	0.1368	0.80	0.2881	1.25	0.3944	1.70	0.4554	2.30	0.4893	$+\infty$	0.5
0.36	0.1406	0.81	0.2910	1.26	0.3962	1.71	0.4564	2.32	0.4898		
0.37	0.1443	0.82	0.2939	1.27	0.3980	1.72	0.4573	2.34	0.4904		
0.38	0.1480	0.83	0.2967	1.28	0.3997	1.73	0.4582	2.36	0.4909		
0.39	0.1517	0.84	0.2995	1.29	0.4015	1.74	0.4591	2.38	0.4913		
0.40	0.1554	0.85	0.3023	1.30	0.4032	1.75	0.4599	2.40	0.4918		
0.41	0.1591	0.86	0.3051	1.31	0.4049	1.76	0.4608	2.42	0.4922		
0.42	0.1628	0.87	0.3078	1.32	0.4066	1.77	0.4616	2.44	0.4927		
0.43	0.1654	0.88	0.3106	1.33	0.4082	1.78	0.4625	2.46	0.4931		
0.44	0.1700	0.89	0.3133	1.34	0.4099	1.79	0.4633	2.48	0.4934		

Table 3 The values of the function $P_k = \frac{a^k}{k!}e^{-a}$

k \ a	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1638	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0019	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4		0.0001	0.0002	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5				0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6							0.0001	0.0002	0.0003

k \ a	1	2	3	4	5	6	7	8	9	10
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.0000
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0011	0.0005
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0107	0.0050	0.0023
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150	0.0076
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339	0.0912	0.0572	0.0337	0.0189
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607	0.0378
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631
7	0.0001	0.0037	0.0216	0.0595	0.1044	0.1377	0.1490	0.1396	0.1171	0.0901
8		0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126
9		0.0002	0.0027	0.0132	0.0363	0.0688	0.1014	0.1241	0.1318	0.1251
10			0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251
11			0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137
12			0.0001	0.0006	0.0034	0.0126	0.0263	0.0481	0.0728	0.0948
13				0.0002	0.0013	0.0052	0.0142	0.0296	0.0504	0.0729
14				0.0001	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521
15					0.0002	0.0009	0.0033	0.0090	0.0194	0.0347
16						0.0003	0.0014	0.0045	0.0109	0.0217
17						0.0001	0.0006	0.0021	0.0058	0.0128
18							0.0002	0.0009	0.0029	0.0071
19							0.0001	0.0004	0.0014	0.0037
20								0.0002	0.0006	0.0019
21								0.0001	0.0003	0.0009
22									0.0001	0.0004
23										0.0002
24										0.0001

Table 4 Percentiles of the t- distribution

n \ γ	0.95	0.99	0.999	n \ γ	0.95	0.99	0.999
5	2.78	4.60	8.61	20	2.093	2.861	3.883
6	2.57	4.03	6.86	25	2.064	2.797	3.745
7	2.45	3.71	5.96	30	2.045	2.756	3.659
8	2.37	3.50	5.41	35	2.032	2.720	3.600
9	2.31	3.36	5.04	40	2.023	2.708	3.558
10	2.26	3.25	4.78	45	2.016	2.692	3.527
11	2.23	3.17	4.59	50	2.009	2.679	3.502
12	2.20	3.11	4.44	60	2.001	2.662	3.464
13	2.18	3.06	4.32	70	1.996	2.649	3.439
14	2.16	3.01	4.22	80	1.991	2.640	3.418
15	2.15	2.98	4.14	90	1.987	2.633	3.403
16	2.13	2.95	4.07	100	1.984	2.627	3.392
17	2.12	2.92	4.02	120	1.980	2.617	3.374
18	2.11	2.90	3.97	∞	1.960	2.576	3.291
19	2.10	2.88	3.92				

Table 5 The values of the function $q = q(\gamma, n)$

n \ γ	0.95	0.99	0.999
5	1.37	2.67	5.64
6	1.09	2.01	3.88
7	0.92	1.62	2.98
8	0.80	1.38	2.42
9	0.71	1.20	2.06
10	0.65	1.08	1.80
11	0.59	0.98	1.60
12	0.55	0.90	1.45
13	0.52	0.83	1.33
14	0.48	0.78	1.23
15	0.46	0.73	1.15
16	0.44	0.70	1.07
17	0.42	0.66	1.01
18	0.40	0.63	0.96
19	0.39	0.60	0.92

n \ γ	0.95	0.99	0.999
20	0.37	0.58	0.88
25	0.32	0.49	0.73
30	0.28	0.43	0.63
35	0.26	0.38	0.56
40	0.24	0.35	0.50
45	0. 22	0.32	0.46
50	0.21	0.30	0.43
60	0.188	0.269	0.38
70	0.174	0.245	0.34
80	0.161	0.226	0.31
90	0.151	0.211	0.29
100	0.143	0.198	0.27
150	0.115	0.160	0.211
200	0.099	0.136	0.185
250	0.089	0.120	0.162

Table 6 Percentiles of the χ^2 - distribution

ν \ α	0.20	0.10	0.05	0.02	0.01	0.001
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13`.815
3	4.642	6.251	7.815	9.837	11.345	16.266
4	5.989	7.779	9.488	11.668	13.237	18.467
5	7.289	9.236	11.070	13.388	15.086	20.515
6	8.558	10.645	12.592	15.033	16.812	22.457
7	9.803	12.017	14.067	16.622	18.475	24.322
8	11.030	13.362	15.507	18.168	20.090	26.125
9	12.242	14.684	16.919	19.679	21.666	27.877
10	13.442	15.987	18.307	21.161	23.209	29.588
11	14.631	17.275	19.675	22.618	24.795	31.264
12	15.812	18.549	21.026	24.054	24.217	32.909
13	16.985	19.812	22.362	25.472	27.688	34.528
14	18.151	21.064	23.685	26.783	29.141	36.123
15	19.311	22.307	24.996	28.259	30.578	37.697
16	20.465	23.542	26.296	29.633	32.000	39.252
17	21.615	24.769	27.587	30.995	32.409	40.790
18	22.760	25.989	28.869	32.346	34.805	42.312
19	23.900	27.204	30.144	33.678	36.191	43.820
20	25.038	28.412	31.410	35.020	37.566	45.315
21	26.171	29.615	32.671	36.343	38.932	46.797
22	27.301	30.813	33.924	37.659	40.289	48.268
23	28.429	32.007	35.172	38.968	41.638	49.728
24	29.553	33.196	36.415	40.270	42.980	51.179
25	30.675	34.382	37.652	41.566	42.314	52.620
26	31.795	35.563	38.885	42.856	45.642	54.052
27	32.912	36.741	40.113	44.140	46.963	55.476
28	34.027	37.916	41.337	45.419	48.278	56.893
29	35.139	39.087	42.557	46.693	49.588	58.302
30	36.250	40.256	43.773	47.962	50.892	59.703

References

1. Лебедь, С. Ф. Linear algebra and analytic geometry for foreign first-year students: учебно-методическая разработка на английском языке по дисциплине «Математика» для студентов 1-го курса / С. Ф. Лебедь, А. В. Дворниченко, И. И. Гладкий, Е. А. Крагель, Т. В. Шишко ; Брестский государственный технический университет. – Брест : БрГТУ, 2013. – 80 с.

2. Гладкий, И. И. Functions of several variables. Integrals for foreign first-year students : учебно-методическая разработка на английском языке по дисциплине «Математика» для студентов 1-го курса / И. И. Гладкий, А. В. Дворниченко, Н. А. Дерачиц, Т. И. Каримова, Т. В. Шишко ; Брестский государственный технический университет. – Брест : БрГТУ, 2014. – 60 с.

3. Гладкий, И. И. Differential equations. Multiple integrals. Infinite sequences and series : учебно-методическая разработка на английском языке по дисциплине «Математика» для студентов 1-го курса / И. И. Гладкий, А. В. Дворниченко, Н. А. Дерачиц, Т. И. Каримова, Т. В. Шишко ; Брестский государственный технический университет. – Брест : БрГТУ, 2014. – 76 с.

4. Гладкий, И. И. Elements of theory of analytic functions of one complex variable: учебно-методическая разработка на английском языке по дисциплине «Математика» / И. И. Гладкий, А. В. Дворниченко, Т. И. Каримова, С. Ф. Лебедь, Т. В. Шишко. – Брест : Брест. гос. техн. ун-т, 2015. – 48 с.

5. Бань, О. В. / Precalculus Mathematics for Calculus (part 1) : учебно-методическая разработка на английском языке по дисциплине «Математика» для слушателей факультета довузовской подготовки / О. В. Бань, А. В. Дворниченко, Е. А. Крагель, С. Ф. Лебедь. – Брест : Брест. гос. техн. ун-т, 2017. – 64 с.

6. Дворниченко, А. В. Elements of Algebra and Analytic Geometry : методические указания на английском языке по дисциплине «Математика» / А. В. Дворниченко, Е. А. Крагель, С. Ф. Лебедь, О. В. Бань, И. И. Гладкий. – Брест : Брест. гос. техн. ун-т, 2018. – 98 с.

7. Дворниченко, А. В. Integrals and Differential Equations : методические указания на английском языке по дисциплине «Математика» / А. В. Дворниченко, С. Ф. Лебедь, О. В. Бань. – Брест : Брест. гос. техн. ун-т, 2019. – 108 с.

8. Дворниченко, А. В. Functions of a Single and Several Variables: методические указания на английском языке по дисциплине «Математика» / А. В. Дворниченко, С. Ф. Лебедь, О. В. Бань. – Брест : Брест. гос. техн. ун-т, 2019. – 64 с.

9. Гладкий, И. И. Probability theory elements of mathematical statistics : учебно-методическая разработка на английском языке / И. И. Гладкий, А. В. Дворниченко, Т. И. Каримова, С. Ф. Лебедь, Т. В. Шишко ; Брестский государственный технический университет. – Брест : БрГТУ, 2014. – 60 с.

11. Гладкий, И. И. Теория вероятностей и математическая статистика. Задачи и упражнения / И. И. Гладкий, Т. И. Каримова, Л. П. Махнист, Т. А. Тузик ; Брестский государственный технический университет. – Брест : БрГТУ, 2010. – 46 с.

12. Гладкий, И. И. Теория вероятностей и математическая статистика : методические рекомендации и варианты контрольной работы по курсу «Теория вероятностей и математическая статистика» для студентов экономических специальностей заочной формы обучения (сокращенной) / И. И. Гладкий, Т. И. Каримова, Л. П. Махнист, В. С. Рубанов ; Брестский государственный технический университет. – Брест : БрГТУ, 2010. – 51 с.

13. Швычкина, Е. Н. Практикум по высшей математике для студентов технических специальностей : в 7 ч. / Е. Н. Швычкина, Л. Т. Мороз, С. Н. Наумовец, В. П. Черненко, В. Т. Джура. – Брест : Брестский государственный технический университет, 2015. – Ч. 7 : Теория вероятностей и математическая статистика. – 80 с.

14. Khoshnevisan, D. Introduction to Probability / Davar Khoshnevisan. Firas Rassoul ; ed. D. Khoshnevisan. – Utah : Cambridge University Press, 2012. – 267 p.

15. Kosolapov, J. F. Probability theory and mathematical statistics : учебное пособие по курсу «Теория вероятностей и математическая статистика» для студентов ДонГТУ (на английском языке) / Ю. Ф. Косолапов ; Донецкий гос. техн. ун-т. – Донецк : РВА ДонГТУ, 2008. – 289 с.

16. Prasanna, S. Probability and mathematical statistics / Prasanna Sahoo. – Louisville : University of Louisville, 2013. – 698 p.

17. Schay, G. Introduction to Probability with Statistical Applications / Geza Schay ; ed. G. Schay. – Birkhauser Boston : Springer Science+Business Media LLC, 2007. – 313 p.

УЧЕБНОЕ ИЗДАНИЕ

Составители:
Дворниченко Александр Валерьевич
Лебедь Светлана Федоровна
Бань Оксана Васильевна

Elements of Probability Theory And Mathematical Statistics

методические указания на английском языке
по дисциплине **«Теория вероятностей и математическая статистика»**

Текст печатается в авторской редакции, орфографии и пунктуации

Ответственный за выпуск: Дворниченко А.В.
Редактор: Боровикова Е.А.
Компьютерная вёрстка: Дворниченко А.В.

Подписано в печать 20.01.2020 г. Формат 60x84 $^{1}/_{16}$. Бумага «Performer».
Гарнитура «Times New Roman». Усл. печ. л. 6,28. Уч. изд. л. 6,75. Заказ № 45. Тираж 22 экз.
Отпечатано на ризографе учреждения образования «Брестский государственный технический университет». 224017, г. Брест, ул. Московская, 267.