Ministry of Education and Science of Ukraine
Odessa National University named after I.I. Mechnikov
Faculty of Economics and Law
Scientific Research Institute of Physics

Tyurin A.V., Akhmerov A.Yu.

THEORY OF PROBABILITY AND MATHEMATICAL STATISTICS

Odessa 2020

UDC 512.64 + 514.12
BBK 22.1

Recommended for publication by the Scientific and Methodological Council
Of Odessa National University named after I.I. Mechnikov
Proc. No. 5 of 13.12.2018

Reviewers:

A.R. Gohman – Doctor of Physical and Mathematical Sciences, Professor, Head of the Physics Department of the South Ukrainian State Pedagogical University named after K.D. Ushinsky;

A.V. Glushkov – Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Higher and Applied Mathematics of the Odessa Ecological University;

A.V. Drozd – Doctor of Technical Sciences, Professor, Department of Computer Intelligent Systems and Networks, Odessa National Polytechnic University

Tyurin A.V., Akhmerov A.Yu.

Probability theory and mathematical statistics: a textbook /
A.V.Tyurin, A.Yu. Akhmerov – Odessa: «Odessa I.I. Mechnikov National University», 2020. - 138 p.

The textbook is written in accordance with the program of the course "Probability theory and mathematical statistics", which is read to junior specialists and bachelors in the specialty "Economics". The basics of the theory, methods of obtaining, description and processing of experimental data are presented in order to study the laws of random mass phenomena.

UDC 512.64 + 514.12
BBK 22.1

CONTENT

INTRODUCTION

The surrounding nature and the society in which we live, there are various events, such as the rain, there was an accident, and others. In order for an event has occurred it must be preceded by a set of conditions. In practice, we are often faced with the events, conditions, the occurrence of which are determined by many factors. Most of these factors are not known to us, or they are random, which in principle can not be taken into account. Therefore, when the same starting conditions, these events may occur, and possibly no, i.e. their appearance is random. Probability theory is just committed to studying the properties of the mass of random events that can be repeated many times when playing a particular set of conditions. The main task of the theory of probability is a quantitative assessment of random factors on the possibility of any random event, regardless of its nature, ie, to quantify the probability of its implementation.

Despite the random nature of the occurrence of events during numerous mass reproduction of these events revealed some patterns that allow you to predict the behavior of random events under certain specified conditions. The study of these patterns on the basis of weight of random events involved in mathematical statistics, which is based on probability theory and the main task of which is limited according to the data (sample) to restore a certain degree of reliability characteristics inherent in the general population, i.e. the entire set of data imaginable, describing the phenomenon under study, random nature.

Over the last few decades of the theory of probability and mathematical statistics "spun off" branches of science such as the theory of stochastic processes, queuing theory, information theory, econometric modeling, and others. This process continues today.

One of the most important areas of application of the theory of probability and mathematical statistics is the economy. It is difficult to imagine the study and forecasting of economic phenomena without the use of econometric modeling, regression analysis, trend and smoothing models and other methods based on probability theory and mathematical statistics. Statistical patterns inherent in a centrally controlled, decentralized and even more economy. The presence of these concepts firmly established in the economy as a reserve stock, spare capacity, state reserves, financial risks, etc., testifies to this.

It should also be noted that no elements of randomness impossible development. Without accident would have been impossible the emergence of life and improvement of species, unimaginable human history, the creative activity of the people, the development of socio-economic systems.

A manifestation of chance, for example, in the socio-economic system should be regarded as a departure from the established flow channel or events in a positive way (the emergence of new scientific discoveries, technologies, and methods of organization of production, etc.), or negative (natural disasters, equipment failure, illness of employees, conflicts of people, etc.), which subsequently leads to a significant change of the flow of events.

With the development of society socio-economic system becomes more complex. Consequently, according to the laws of dynamical systems is to amplify the statistical nature of the laws describing the socio-economic phenomena. All this makes it necessary to master the methods of probability theory and mathematical statistics as a tool for statistical analysis and forecasting of various phenomena and processes.

CHAPTER 1
THEORY OF PROBABILITY

The basis of the theory of probability, as the basis of any other science, are some of the initial concepts. With these concepts is given logical definition follow more complex concepts.

As the basic initial concepts, which operates probabilities are stochastic concept experiment (experiment), a random event and the probability of a random event.

Under stochastic experiment (experience test) refers to playing a set of conditions for the emergence of a certain set of specific events, but some of these events will occur during the implementation of these conditions can not be predicted in advance. Note that this set of conditions can be created spontaneously by the action of the internal processes occurring in nature, society etc., And can be played and a person in the course of the experiment.

We begin with the concept of a random event.

1.1. Random events

Definition **1.1.** Event in probability is any fact (outcome) that may occur as a result of some experiment.

Events usually denoted by capital letters A, B, C ...

For the same experimental conditions at multiple repetition of its outcome may be different events.

Definition **1.2.** All possible outcomes of an experiment is called the space of elementary events (outcomes).

Thus, any experiment can be associated with some set Ω, the elements of which are mutually elementary outcomes ω, i.e. result of this experiment is one and only one outcome. Denote the space of elementary events $\Omega = \{\omega\}$.

Sample space $\Omega = \{\omega\}$ is one of the basic concepts of probability theory. In general, the sample space may be of any nature as limited or infinite, both discrete and continuous. Consider a few examples.

*Example*1.1. Consider shooting on target. If we are interested only in the fact of hitting the target, the elementary events are $\omega 1$ - hit the target and $\omega 0$ - misses the target. If it is important to get into some of the target area (the area vary in terms of the vulnerability of the real target), then the space of elementary events can be the following outcomes: $\omega 10$ - hit the top ten, $\omega 9$ - into the top corner, $\omega 8$ - the top eight, $\omega 7$ - in seven, $\omega 6$ - in six, $\omega 5$ - in the top five, the w4 - in four, $\omega 3$ - the top three, o2 - in twos, $\omega 1$ - per unit (corresponding to the number of points attributed to falling into a certain area) and ω0 - miss. Finally, if it is essential that the shield is in any point, which shows the target (e.g., wild boar) hit occurred, then the elementary arbitrary starting ω = (x, y) are the coordinates of the point of impact,

In the first two cases considered, the sample space $\Omega = \{\omega\}$ Course and discretely in the third - and of course continuous.

*Example*1.2. Performed an experiment: toss a coin once. The set Ω = {T, R}, where the letter T means the appearance of the emblem, the letter P - appearance tails. elementary events space $\Omega = \{\omega\}$ Is finite and discrete.

*Example*1.3. Coin thrown twice, Ω = {GG, GR, WG, PP}. elementary events space $\Omega = \{\omega\}$ Is finite and discrete.

Various events differ in the degree of possibility of their occurrence and the nature of the relationship.

***Definition* 1.3.**If the implementation of a certain set of conditions necessary event occurs, the event is called authentic. Accordingly, the impossible is an event that for a given set of conditions is never going to happen. Randomly naturally called an event that for a given set of conditions can be as occur, did not occur.

Reliable and impossible events can be seen as special cases of extreme random events. Certain event will be denoted by Ω, impossible - character ∅.

***Definition* 1.4.**Two or more random events are called equally possible, if the conditions of their appearance are the same and there is no evidence to suggest that any of them as a result of the experience is more likely to occur than the other, i.e., no outcome of the experiment is not the objective advantages over others.

For example, the store received a shipment of shoes of different colors and sizes, but they are all packed in the same box. In this case, events, select the boxes of shoes of a certain color and size are equally possible. Are also equally possible events, any number falling points when tossing a die of heads or tails coin toss et al.

Definition 1.5. Two events A and B are said to be consistent if the appearance of one of them does not preclude the emergence of another.

Consider the examples.

*Example*1.4. To toss two dice. And event - loss of 5 points on the first dice, event B - loss of 5 points in the second bone. A and B - joint events. At the same time, they are both equally possible.

*Example*1.5. At the warehouse received a batch of refrigerators of different companies and different colors. Event A - the choice of a specific color of the refrigerator, in event - the choice of a particular company's refrigerator. A and B - joint events. In this case, they are not equally possible, equally possible, these events will, if refrigerators are made in different colors of the same firm.

***Definition* 1.6.** Event Group A1, A2, ..., An are called the group of joint events if the events included in the pairs together a group.

*Example*1.6. An urn is 10 balls, six balls are red, white ball 4, and 5 are numbered balls. And - appearance of the red globe, with one removing, B - the appearance of a white ball, C - the appearance of a sphere with a number. Events A, B, C form a group of joint events.

*Example*1.7. It produced three shots at targets. Event A1 - hit the target on the first shot, A2 - hit the target at the second shot, A3 - hit the target at the third shot. Events A1, A2, A3 form a group of joint events.

***Definition* 1.7.** Two events A and B are said to be inconsistent if the appearance of one of them precludes the emergence of another.

Consider the examples.

*Example*1.8. The store received a shipment of shoes of the same size, but different colors. Event A - taken at random from the box will be a shoe in black, in the event - a box with shoes will be brown. A and B - are incompatible events.

*Example*1.9. You've come to the store to buy a refrigerator. Event A - You bought a refrigerator without defects, B - You bought a refrigerator with a defect. A and B - are incompatible events.

***Definition* 1.8.** Event Group A1, A2, ..., An are called the group of incompatible events if the events included in the mutually incompatible group.

*Example*1.10. The target produced three shots. Suppose that A represents a miss, B - one hit, C - the two ingress and D - three shots. All events A, B, C and D form a group of mutually exclusive events.

Many events can be represented as a set (group) of the simplest events. Consider the example of this elementary events space $\Omega = \{\omega\}$. Any subset Ω space, i.e. some set (group) of elementary events ω is an event. Ω The space also is an event, and it is a significant event as one of its outcomes will happen.

We illustrate the concepts introduced in a number of the simplest examples relating to a random test. Random experiment (test experience) - this experiment the result of which is a random event.

Example 1.11. Throws the dice and recorded the number of dropped points. elementary events space consists of six elements $\omega_i = i$, i = 1,…,6. Example composite event: A = {2, 4, 6} - dropped even number of points.

*Example*1.12. Throw two dice space of elementary events can be represented in a matrix $\Omega = ((i,j)_{ij})$, $i = j$ = 1,…,6. An example of a composite event: the sum of more than 10 points; And the occurrence of this event is only possible if the elementary outcome (5,6), (6,5), (6,6), i.e. A = {(5,6), (6,5), (6,6)}.

*Example*1.13. Consider the example of 1.10. If it considered sample space consisting of two outcomes ω_1- slip, ω_2 - hit the target, the events *A, B, C* and *D* are components of ω1 and ω2 elementary events ω_1 и ω_2: $A = \{ \omega_1, \omega_1, \omega_1\}$, $B = \{(\omega_1, \omega_1, \omega_2), (\omega_1, \omega_2, \omega_1), (\omega_2, \omega_1, \omega_1)\}$, $C = \{(\omega_1, \omega_2, \omega_2), (\omega_2, \omega_2, \omega_1), (\omega_2, \omega_1, \omega_2)\}$ и $D = \{\omega_2, \omega_2, \omega_2)\}$. In practice, often have to consider two elementary events forming the space w, i.e. In any test necessarily takes place at least one of these events, and the latter does not occur, therefore, they are incompatible. For example, loss of the emblem and the coin toss tails. Such events are called the opposite and are designated A and $\bar{A}$. Specifically . = ∅ and $\emptyset = \overline{\Omega}$.

***Definition* 1.9.** Two of equally incompatible events that make up the entire space of elementary events Ω called opposite.

1.2. Algebra of events

Suppose we have a sample space $\Omega = \{\omega\}$ Of any nature. We will be seen as a subset of the events *A, B, C, ...* in this space. In this case, the actions of the events are the actions of the subsets.

If the random experiment of the occurrence of the event *A* implies occurrence of the event B, then we say that A implies (contains*) B* $(A \subset AT)$. If $A \subset B$ *and* $B \subset A$, then we say that the events *A* and *B* are equivalent *(A = B).* For example, in the urn there are 20 balls 12 are black and 8 white. All the white balls are numbered, and of black balls numbered only 5. Let the one removing the event *A* - the emergence of the black ball, B - the appearance of a white ball, C - the appearance of a ball with a number. In this case, the event *B* implies (contains) S means $B \subset C$. If the numbered only white balls, the events *B* and *C* are equivalent and hence *B = C*.

***Definition* 1.10.** The sum of two events *A* and *B* is called the event $A + B$ $(A \cup B)$, Which occurs if and only if the happening or event *A* or event *B* or events *A* and *B* occur together.

Event operation amount corresponds to the union of sets. It has a similar meaning the sum of any number of events. For example, made three shots. Event

A has hit the target on the first shot, *B* - in the second and *C* - in the third, the event *D = A + B + C* is hitting the target at all, no matter at what firing sequence in which and how many times out of three.

Definition **1.11. Product (intersection)** several events called the event consisting in a joint appearance of all these events.

The product corresponds to the operation events crossing sets. Artwork S events A, B, C, ..., N is denoted $S = ABC,\ldots,N$, or $S = A \cap B \cap C \cap \ldots \cap N$. For example, made three shots. Event A has hit the target on the first shot, B - C and the second - in the third, the event D = ABC has hit the target at all three shots.

Definition **1.12. Difference** A \ B of two events A and B is the event, which includes those elementary events which belong to A and are not included in B, i.e., happens if and only when there is A, but there is no B.

The difference corresponds to the event operation sets the difference. For example, an event, opposite event A, there Ω \ A. In fact, A$\bar{A}$∪(Ω \ A) = A and Ω∩ (Ω \ A) =, i.e. really = Ω \ A.∅$\bar{A}$

The concept of the amount of the work and the difference between the two events have a clear geometric interpretation. Indeed, suppose that the event A has hit the point in region A, respectively, in the event - getting into the area B, then A + B event has hit the point in the shaded area in Figure 1.1, the AV event has hit the point in the area shaded in Fig. 1.2, the event A \ B is the point of impact to the area shaded in Figure 1.3.

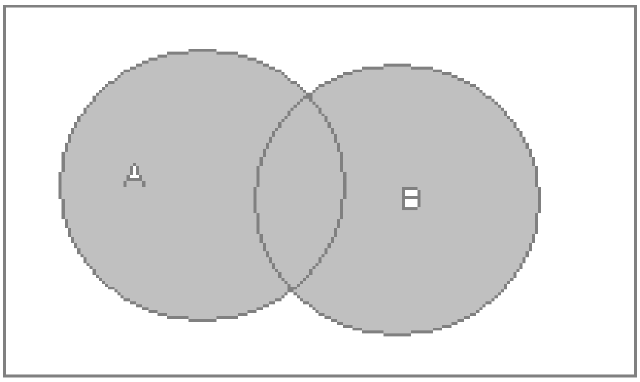

Figure 1.1

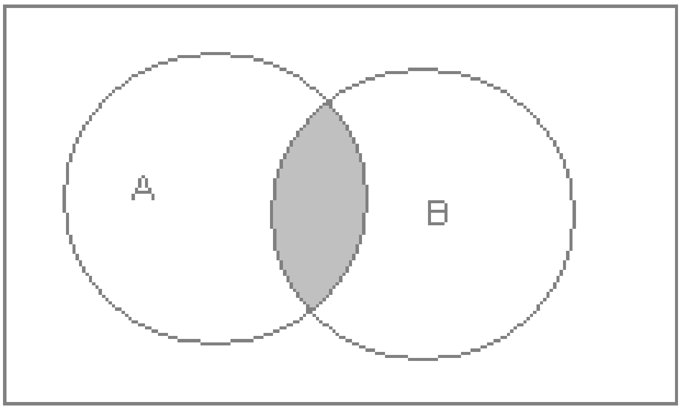

Figure 1.2

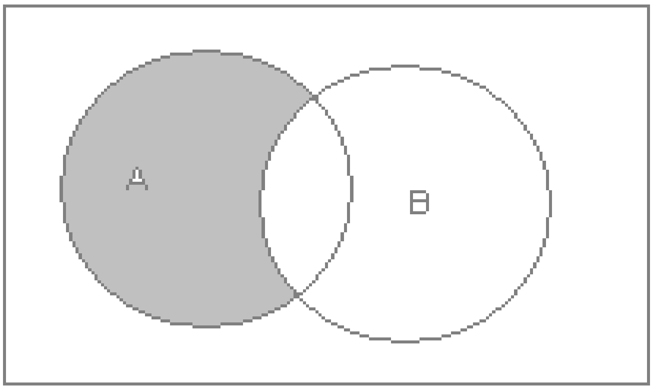

Figure 1.3

Events addition and multiplication operations as operations on sets have the following properties:

a) = A + B + A, AB = BA (commutativity);

b) (A + B) + C = A + (B + C), A (BC) = (AB) C (associativity);

c) (A + B) with respect to addition) C = AC + BC (distributivity multiplication.

We also note some obvious relations:

$AA = A$, A∅ = ∅, $A\Omega = \mathbf{A}$, $A + A = A$, $\mathbf{A} + \Omega = \Omega$, $\mathbf{A} + \varnothing = \mathbf{A}$.

For mutually exclusive events A and B, we have AB =∅. Particularly A =.$\bar{A}$∅.

If A ⊂B, the following equalities hold: A + B = A; AB = B.

By solving various problems associated with the events, very often have to be complicated in the event a combination of simpler ones (elementary) events, we have considered applying the above operations on these events. For example, suppose on target produced three shots, and covers the following elementary events:

A_1 – getting the first shot;

$\overline{A_1}$ – a miss at the first shot;

A_2 - getting the second shot;

$\overline{A_2}$ - a miss in the second shot;

A_3 - getting in the third shot;

$\overline{A_3}$ - blunder at the third shot.

Consider a complex event in which consists in the fact that as a result of three shots will be exactly one hit the target. The event can be represented by the following combination of elementary events:

$$B = A_1\overline{A_2}\,\overline{A_3} + \overline{A_1}\,A_2\,\overline{A_3} + \overline{A_1}\,\overline{A_2}\,A_3.$$

Event C, consisting in the fact that the target will be at least two results may be represented as

2. $C = A_1 A_2\overline{A_3} + A_1\overline{A_2}\,A_3 + \overline{A_1}\,A_2 A_3 + A_1 A_2 A_3.$

1.3. Probability of event

As the practice shows the possibility of occurrence of an event in a particular test can be quantified, ie number.

***Definition* 1.13.** Number expressing the quantitative assessment of the possibility of occurrence of the event, called the probability of event.

To calculate the probability are mainly used two approaches, dubbed as statistical and classical. The most general definition of probability is axiomatic that formulated Soviet mathematician Kolmogorov in 1933. However, the consideration of this definition beyond the scope of this course of lectures.

1.3.1. Statistical determination probability.

To determine the probability of an event with such a statistical approach requires that the event really occurred, and wherein multiple, i.e. this possibility can only be determined by carrying out numerous tests, or observations.

Let n made a series of experiments (tests), each of which may appear or not to appear for some event A quantitative characteristic of the possibility of occurrence of the event A in the test series is the relative frequency (often called simply the frequency) of occurrence of the event A.

***Definition* 1.14. The relative frequency** A event of the series of trials is the ratio of the number of trials in which there was an event A to the number of trials.

Designating the relative frequency of the event A through $P^*(A)$, we have, by definition: $P^*(A) = m/n$, where m –number of tests in which there was an event A, and n - the total number of tests.

The main properties of the relative frequency:

1. The relative frequency of the event is a non-negative random number between zero and one, i.e. $0 \le P^*(A) \le 1$;
2. The relative frequency of a certain event is equal to one;
3. The relative frequency is zero impossible event.

The relative frequency of the event does not remain constant with changes in the number of tests. However, experience shows that with increasing number of tests relative frequency gradually stabilizes and tends to be some specific number. Thus, with the subject event can link a certain number, which are grouped around the relative frequency and which is a quantitative characteristic of the objective connection between the complex conditions in which the tests are made and the event. This number is taken as the probability of an event in the statistical approach of its calculation.

These concepts allow starting Now give the following definition of the statistical probability of an event.

Definition **1.15. The probability of a random event** the statistical approach of its calculation is called a constant number, which are grouped around the relative frequency of the event as the number of trials.

The probability of an event A is usually denoted by $P(A)$.

It follows from the foregoing that the statistical method of determining the probability is based on a real experiment and to reliably determine the likelihood of need to do a large number of tests or observations, it is not always possible and requires large material costs. The statistical method for determining the probability is the basis of mathematical statistics, and we shall return to it in the study of this section.

However, in most cases the probability of an event can be determined theoretically never resorting to testing. The basis of this determination method is the classic definition of probability of random events.

1.3.2. The classic definition of probability.

The classic way to determine the probability based on the concept of space elementary events Ω. Moreover, this space consists of a finite number n equiprobable and independent elementary events $\omega_1, \omega_2, \ldots, \omega_n$. Suppose now that there is an event, which is kind of a subset of this space. With respect to each elementary event space Ω outcomes are divided into favorable, in which the event takes place, and unfavorable, in which it is not an event occurs. For example, when tossing dice event of occurrence of an even number of points from six elementary events space Ω three elementary favorable outcome (2, 4, 6) and is also not conducive to three (1, 3, 5).

These concepts allow starting Now give the following classical definition of probability of an event.

***Definition* 1.16. Likely occurrence** of an event is the ratio of the number of outcomes favorable to the emergence of this event to the total number of equally likely outcomes in this experiment.

Denoting the number of outcomes of the event A favorable appearance through m and the total number of equally possible outcomes in this experiment by n, this classic definition of probability can be written as formula

$$R(A) = \frac{m}{n}. \tag{1.1}$$

An important advantage, as we have noted, the classical method of determining the likelihood is that with the help of probability of an event can be determined before the experiment in advance to make conclusions for themselves. However, this method has the drawback that it is applicable only when we are dealing with equally possible outcomes of the test. In addition, the definition of equally possible outcomes in this experiment in many cases is a rather complicated task, which can not always be the solution. For example, it is not possible to establish whether the events the birthday boy or girl's equally possible, or another example of getting shots at the target, provided that the target shoot different direction.

Classical probability has the following properties:

1. The probability of a certain event Ω is 1 : P (Ω) = 1.
2. Probability impossible event is zero: P (∅) = 0.
3. The probability of a random event is contained in the interval (0,1):

$0 < P(A) < 1$.

4. The probability of any event is contained in the interval [0,1]:

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

Consider the examples in the calculation of the probability of random events, guided by the classical definition.

*Example*1.14. Once a coin toss. What is the probability of heads?

Here Ω = {Γ,P}, and equally likely outcomes of the experiment, by the condition A = {Γ} , so m = 1, n = 2, $P(A) = \frac{m}{n} = \frac{1}{2}$.

Example 1.15. Once tossed hexagonal dice. What is the probability that the number will drop points, three times?

According to the condition Ω = {1,2,3,4,5,6}, the elementary equally possible outcomes, A = {3,6}, m = 2, n = 6, $P(A) = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}$.

In more complicated problems it is not possible to clearly record all elementary outcomes of the experiment, as well as the favorable outcome of a random event. For example, there are 10 pieces, of which three defective in the drawer. From the box parts 5 are removed at random. Find the probability that among them will be two defective.

In such instances, the combinatorial methods of counting the numbers m and n.

1.3.3. Fundamentals of combinatorics.

Combinatorics – a branch of mathematics that studies questions about how many different combinations of subordinates in some conditions, can be formed from the given objects. Fundamentals of combinatorics is very important to assess the probability of random events, as they allow us to calculate the number of possible in principle favorable and all equally possible elementary events in this experiment.

Consider the basic definitions and formulas combinatorics.

We begin with the basic formula of combinatorics. Suppose that there are k groups of elements, and the i-th group consists of ni elements. We choose one element of each group. Then the total number N of the ways that you can make such a choice, is given by $N = n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k$.

Example 1.16. We explain this rule by a simple example. Suppose we have two groups of elements, and the first group is composed of n1 elements, and the second- of n_2 elements. How many different pairs of elements can be formed from these two groups, so that in the pair was one element from each group? Suppose we have the first element of the first group and, without changing it, through all possible pairs, changing only the elements of the second group. Such pairs for this element can form n_2. Then we take the second element of the first group and also makes up for it all possible pairs. Such pairs will n_2 too. Since the first group n_1 of all member of all possible options will be $n_1 \cdot n_2$.

Example 1.17. How many three-digit even numbers can be composed of the digits 0, 1, 2, 3, 4, 5, 6, may be repeated if the numbers?

Decision: The first digit can take any number of 1, 2, 3, 4, 5, 6, therefore, n_1 = 6. As the second digit can take any number of 0, 1, 2, 3, 4, 5, 6, those. n_2=7, n_3= 4 (since as the third digit can take any number of 0, 2, 4, 6). Thus, $N = n_1 \cdot n_2 \cdot n_3$= 6·7·4 = 168.

In the case where all the groups consist of the same number of elements, i.e. $n_1= n_2=...= n_k= n$ we can assume that every choice produced from the same group, and after the selection item is returned again to the group. Then the number of ways selecting the same n^k.

Such a method of choice in combinatorics is called *sampling with replacement.*

*Example*1.18. How much of all four-digit numbers can be composed of digits 1, 5, 6, 7, 8?

Decision. For each four-digit number, there are five possibilities, then N = 5·5·5·5=5^4 = 625.

Consider the set consisting of n elements. This set is called combinatorics in the *general population.*

***Definition* 1.17. Placement** of n elements in m in combinatorics is any ordered set of m different elements selected from the general population in the n elements.

EXAMPLE 1.19. Different placements of three elements {1, 2, 3} by two will sets (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2).

Accommodation differ from each other not only the elements but their order.

The number of placements in combinatorics is indicated and is calculated as follows:

$$A_n^m = n(n-1)(n-2)...(n-m+1) = \frac{n!}{(n-m)!}. \qquad (1.2)$$

Example 1.20. How many two-digit numbers in which the tens digit and the number and variety of odd items?

Solution: because odd numbers in each digit five, namely 1, 3, 5, 7, 9, this problem is reduced to the selection and placement in two different positions of two of the five different numbers, i.e. these numbers would be: $A_5^2 = 5 \cdot 4 = 20$.

Definition* 1.18. *Combination of n elements in m in combinatorics is any *unordered* set of m different elements selected from the general population in the n elements.

Example 1.21. For a set {1, 2, 3} combinations of the two elements are {1, 2}, {1, 3}, {2, 3}.

Other combinations should differ from each other by at least one element.

The number of combinations is denoted by and is calculated as follows:

$$C_n^m = \frac{n!}{m!(n-m)!}. \qquad (1.3)$$

Solution large class of problems with the use of combinations of Formula (1.3) reduces to the use of so-called *boxes* scheme: an event occurrence probability is to be calculated, it can be interpreted as the result of random selection of balloons of different colors of the boxes. Balls in the urn must be identical in size, weight and other tangible features and mixed thoroughly before taking out, then the event is the appearance of a ball (or balls group) from the urn, are equally possible and inconsistent. For example, the test coin-tossing scheme reduces to boxes comprising two balls; test with a toss of the dice comes down to the circuit boxes containing six balls, etc.

Here are a few examples of the application urn scheme when the probability of equally and mutually exclusive events.

Example 1.22. In how many ways the reader can choose two books of the six available?

Decision: The number of ways equal to the number of combinations of the six books of the two, i.e., equally:

$$C_6^2 = \frac{5 \cdot 6}{2!} = 15.$$

Example 1.23. The box contains 15 balls, including 9 red and 6 blue. Find the probability that a randomly taken out two balls will be red.

Decision: In this example, the total number n of equally possible outcomes equal to the number of all combinations of two balls, i.e. n = $C_{15}^2 = \frac{15 \cdot 14}{2!} = 105$. *A* denote the event consisting in the appearance during the test two reds; then the number of outcomes favorable to the event *A* is the number of combinations of the number of red balls on the two. Therefore, $m = C_8^2 = \frac{9 \cdot 8}{2!} = 36$.

Therefore, $P(A) = \frac{m}{n} = \frac{36}{105} = \frac{12}{35}$.

Example 1.24. The box contains N = 10 parts, of which M = 3 defective. From the box randomly extracted n = 5 parts. Find the probability that among them will be m = 2 defective.

Decision: From the conditions of the problem that $M \leq N$ and $m \leq n$. Since any combination of the *N* of *m* products has the same possibility of occurrence, then all will be equally likely outcomes C_N^n. Let *A* denote the occurrence of defective products to the m selected at random *n* products. The number of ways that you can take out the m of defective products from *M* as well C_M^m. But each of these methods may be supplemented by any group of items from among the methods which will remove the remaining *n* - *m* fit of the total number of suitable *N* - *M* products. The number of such groups is C_{N-M}^{n-m}. Consequently, all outcomes favorable to the emergence of an event *A* is equal $C_M^m \cdot C_{N-M}^{n-m}$.·. Therefore, *P*(*A*) = $\frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n} = \frac{C_3^2 \cdot C_7^3}{C_{10}^5} = \frac{5}{12}$.

Definition* 1.19. *Permutation of *n* elements is any *ordered* set of elements.

Example 1.25. All possible permutations of a set consisting of three elements {1, 2, 3} are (1, 2, 3), (1, 3, 2), (2, 3, 1), (2, 1, 3), (3, 2, 1), (3, 1, 2).

The number of different permutations of *n* elements is denoted by P_n and P_n calculated by the formula $P_n = n!$.

*Example*1.26.How many ways are seven books of different authors can be arranged on a shelf in a row?

Decision. This problem of the number of permutations of seven different books. Consequently, there is $P_7 = 7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$ ways to carry out the alignment of books.

So we see that the number of possible combinations can be calculated by different rules (permutations, combinations, placement) and the result will be different, because counting principle and the formulas themselves are different. The calculation result is dependent on several factors simultaneously. Firstly, the matter of how many elements we can combine them in sets (how great a general set of elements). Secondly, the result depends on how large sets of items we need. Finally, it is important to know whether or not to have a significant order of elements in the set. Let us explain the last factor in the following example.

Example 1.27. At the parents' meeting 20 people present. How many different options n composition of the parent committee, if it should enter 5 people?

Decision. If in this example, we are not interested in the order of names in the committee list, and what exactly will engage every member of the committee. Then, if the result of the selection in its composition will be the same people, then the meaning is the same option for us. Therefore, we can use the formula to calculate the number of combinations of 20 elements 5 and, therefore, n = C_{20}^{5} = $\frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 15504$.

Otherwise there will be things were, if every member of the committee was initially responsible for a particular area of work. Then, for the same list structure of the committee, it is possible within 5! permutations of options that matter. The number of different (and composition, and area of responsibility) variants determined in this case the number of *placement* 20 elements of 5, and therefore, n = A_{20}^{5} = $\frac{20!}{15!}$ =1860480.

We have examined the properties of the classical probability of a single event, but in practice may be cases in which a series of *n*test comes not one event, but several events that are in any way with each other. In this case, the classical probability of equally likely and incompatible events has the following properties:

5. If the event is $C = A + B$, where A and B are equally possible, mutually exclusive and are subsets of elementary events Ω, then the probability of event C is the sum of the probabilities of events A and event B: $P(C) = P(A) + P(B)$. This property is called the ***addition rule of probability***.

6. The sum of the probabilities of all the elementary event space Ω is equal to 1:

$$P(\Omega) = P(\omega_1) + P(\omega_2) + P(\omega_3) +\ldots+ P(\omega_n) = 1.$$

7. The probability of the opposite event $\overline{A}$ is $P(\overline{A}) = 1 - P(A)$.

1. If $A \subset B$, then $P(A) \le P(B)$.

2. Consider a few examples.

Example 1.28. When receiving the consignment of products 80, of which six defective is checked 40 randomly selected articles. Determine the probability that the party will be accepted if the conditions of reception of defective items allowed no more than two among tested.

Decision. Let A denote the event consisting in the fact that when testing 40 items not received any defective product, by B – event, consisting in the fact that only one received a defective product, and by C – an event consisting in the fact that two defective items received. Events A, B and C are mutually exclusive.

According to the conditions of reception, game articles will be accepted if the event would occur the $A + B + C$. Therefore, in addition rule probabilities, the required probability $P(A + B + C) = P(A) + P(B) + P(C)$.

Of the 80 items 40 items can be selected C_{80}^{40} ways. Of the 74 non-defective products 40 items, you can select C_{74}^{40} ways. Therefore, $P(A) = \frac{C_{74}^{40} \cdot C_6^0}{C_{80}^{40}}$. Similarly $P(B) = \frac{C_{74}^{39} \cdot C_6^1}{C_{80}^{40}}$, $P(C) = \frac{C_{74}^{38} \cdot C_6^2}{C_{80}^{40}}$. Therefore $P(A + B + C) = \frac{C_{74}^{40} \cdot C_6^0}{C_{80}^{40}} + \frac{C_{74}^{39} \cdot C_6^1}{C_{80}^{40}} + \frac{C_{74}^{38} \cdot C_6^2}{C_{80}^{40}} \approx$ 0,337.

Example 1.29. It produced one shot at a target consisting of two concentric rings. The probability of hitting in one shot in the target facility and rings are respectively 0.11, 0.24, 0.35. Find the probability of a miss.

Decision. Events hit a target and miss are opposite events, as make all possible outcomes when fired. Denote by *A* hit in the target, then $\overline{A}$ miss. Therefore, $P(\overline{A})= 1 - P(A)$, but $A = A_1 + A_2 + A_3 =$ 0,11 + 0,24 + 0,35 = 0,7. Therefore, $P(\overline{A}) = 0.3$.

1.3.4. Geometric probability.

In the classical definition of probability is considered a complete group of a finite number of equally likely events. In practice, very often there are such tests, the number of possible outcomes which is infinite. In such cases, the classical definition is not applicable. Sometimes, however, in such cases, you can use a different method to calculate the probability, which is still the main role is played by the concept of equal possibility of some events. Applied this technique in problems of reducing to a random point of throwing the final straight section, plane or space. Hence, there is a method name itself – the *geometric probability*.

To determine the limit ourselves to the two-dimensional case. One-dimensional and three-dimensional cases differ only in that instead of the area they need to talk about the length and volume.

So, let the plane there is some area of the *G*, with an area S_G, and it contains other region *g*, which area S_g. In the area of *G* randomly throws point. The question is, what is the probability that the point gets to *g*? It is assumed that the random thrown point can get to any point of the region *G*, and the probability of being in any part of the region *G* is proportional to the square of this part does not depend on its location and shape. In this case, hit probability *p* to *g* when throwing random point *G* is equal to $\frac{S_g}{S_G}$.

Thus, in general, if the possibility of random appearance point within a certain area on the straight line, plane or in space is not determined by the position of this region and its boundaries, but only its size, i.e. length, area or volume, the *probability of a random point within a certain area is determined as* *the ratio of the size of the area to the size of the entire region, which can appear at this point.*

Comment. In the case of the classical definition of probability impossible event probability is equal to zero. Converse is also true: if the probability of an event is equal to zero, the event is not possible. In the case of the geometrical definition of probability of the converse is not the case. For example, if the thrown point can appear anywhere in the entire plane G, the probability of getting to this point in any limited area of the plane region g is zero. However, this event may occur and, therefore, is not impossible.

Consider a few examples.

Example 1.30. A segment *OA* and length *L* real axis *Ox* random set point *B(x).* Find the probability that the smaller of the segments *OB* and *BA* is longer than $\frac{L}{5}$. It is assumed that the probability of a point on the segment is proportional to the length of the segment and does not depend on its location on the number line.

Decision. We use the concept of geometric probability - the probability *p* hitting a random point on the segment length ℓ, contained within the segment of greater length *L* equal to the ratio of their lengths: $p = \frac{\ell}{L}$. For that divide the segment *OA* points *C, D, E* and *F* into five equal segments each of length equal $\frac{L}{5}$. Task requirement is satisfied if the point falls into pieces *CD, DE* and *EF*. The probability *p* of falling for these segments is the same and equal $\frac{\frac{L}{5}}{L} = \frac{1}{5}$. These events are independent and, therefore, the probability of their joint occurrence is equal to the sum of these probabilities. Thus, P = 3p = $\frac{3}{5}$.

Example 1.31.. (The task of the meeting). Two persons have agreed to the meeting, which should take place in a specific location at any given period of time *t*. Determine the probability of the meeting, if the time of arrival of each person be independent and the waiting time of one another will be no more τ.

Decision. Denote the time of arrival of a person in terms of *x*, and the second – through *y*. A necessary and sufficient that the meeting took place is $|x - y| \leq \tau,$ or which also,

$$y < x + \tau \text{ at } y > x,$$
$$y < x - \tau \text{ at.} y < x$$

We will consider both the *x* and *y* cartesian coordinates of the points on the *xOy* coordinate plane. Then all the possible outcomes of the meeting *x* < *t*, *y* < *t* (area *G*) be depicted with a side of the square dots *t* (Fig. 1.4). Outcomes favoring the meeting $|x - y| \leq \tau$ (area g), represented by a point, which are located in the shaded area of the hexagon (Fig. 1.4).

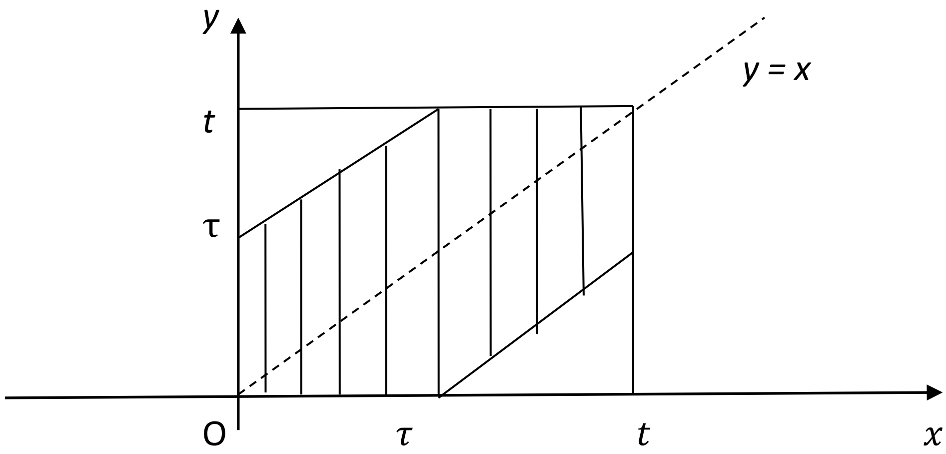

Fig. 1.4

The required probability is the ratio of the area of the shaded hexagon to square all square

$$P = \frac{t^2 - (t - \tau)^2}{t^2}.$$

1.3.5. Conditional probabilities. Independence of events.

As noted above, to speak of the probability *P(A)* as far as possible of the random event *A* has meaning only if certain set of conditions. When conditions change, change, and probability. For example, if a set of conditions under which studied the probability *P(A),* add a new condition, which consists in the occurrence of event B, we obtain a different value of the probability *P (A / B)* – the *conditional* probability of the event *A* under the condition that an event *B* has occurred. The probability *P (A),* in contrast to conventional we call *unconditional.*

Definition **1.20.** The probability of the event *A* calculated at the occurrence of another condition of the event *B*, called the ***conditional probability*** and is denoted *P (A / B).*

Conditional probabilities are determined by the nature of the relationship between events. This can be illustrated by the following examples.

*Example*1.32. There are 6 white and 4 black balls in the urn. One ball is taken from the urn at random, then the taken ball is returned to the urn and the test is repeated. Event *B* – the appearance of a white ball in the first test, event A - the appearance of a white ball in the second test. Obviously, the probability of event A does not depend on the result of the first test: *P (A)* = 3/5.

*Example*1.33. A box contains 60 surprise chocolates. In 50 of them inside as a surprise contained cheburashka and ten crocodile Gene. At random from the box first take a piece of candy, and then another. In the event - within the first selected candy would cheburashka, event *A* – within a second selected candy would also cheburashka. The probability of occurrence of an event *A* under the condition that an event *B* has occurred, is equal to *P (A)* =$\frac{49}{59}$. If in the first test event does not occur (in candy turned Crocodile Gena), the probability *P (A)* = $\frac{50}{59}$. Thus, the probability of event *A* depends on whether the event *B* occurred or not.

Thus, the nature of the event relationship can be divided into independent (first case considered by us) and dependent (second case). The concept of independent and dependent events, plays a very important role in the further research on the theory of probability. In this regard, for further discussion, we introduce the concept of dependence and independence of events.

Definitionse* 1.21.** The event *A* is ***independent in relation to an event B if the probability of an event A is independent of whether the event occurred on or not, otherwise the event *A* is ***dependent*** on event *B.*

Mathematically independence condition A of events in the event recorded in the form: *P (A / B) = P (A),* and the condition of the *P (A / B) ≠ F (A).*

For independent events the following condition: if the event does not depend on the event *B,* and event *B* is not dependent on event *A*, ie the concept of dependence and independence of events mutually. In this regard, we can give a new definition of independent events.

Definitionsi*e 1.21a.** Two events are called ***independent if the occurrence of one of them does not change the probability of occurrence of the other.

We extend the concept of independence of events in the case of an arbitrary number of events.

Definitionse* 1.22.** Several events are called mutually ***independent if each of them and any combination of other events containing or other events, or some of them, there are independent events. For example, if the event A_1, A_2 and A_3 are mutually independent, it means that there will be independent of the following events: A_1 and A_2; A_1 and A_3; A_2 and A_3; A_1A_2 and A_3; A_1A_3 and A_2; A_2 A_3 and A_1.

At the beginning of the presentation, we were told that many of the events can be presented as the result of algebraic operations on simple events. The question arises, how to determine the probability of such an event, if you know the probabilities of the elementary events. In the future, such events will be called the complex (composite) events.

1.4. **Probability calculations for complex event**

In Section 1.3.3, we have found that if a complex event is the sum of equally inconsistent and elementary events, the probability of a compound event is the sum of the probabilities of elementary events. In order to establish a rule for calculating the probability of a complex event, which is the sum of both inconsistent and joint events it is first necessary to consider the rule for calculating the probability of the event, which is a product of elementary events.

1.4.1. Theorem of multiplication of probabilities.

Before we formulate the theorem of multiplication of probabilities, we recall that the event, which is the product of two or more events, there is an event consisting in a joint appearance of all these events.

***Theorem*. 1.1.** *The probability of the product, or co-occurrence, of multiple events is the product of the probability of one of them on the conditional probabilities of other events, calculated on the assumption that all previous events have taken place:*

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)P(A_4/A_1A_2A_3)\ldots P(A_n/A_1A_2\ldots A_{n-1}).\quad(1.4)$$

For arbitrary two events A and B: $P(AB) = P(A)\,P(B/A) = P(B)P(A/B)$.

Theorem assume, without proof, and consider two consequences of the theorem of multiplication of probabilities:

Result 1.1.1. If the event A does not depend on the event B, and event B does not depend on A.

*Result*1.1.2. The probability of a multiplication of independent events together is the product of the probabilities of these events. Formula (1.4) takes the form of independent events:

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2)P(A_3)P(A_4)\ldots P(A_n)\qquad(1.5)$$

Consider a few examples:

*Example*1.32. The mechanism consists of three identical parts. The work mechanism is broken, if its assembly will be delivered to all three sizes of the items are more marked on the drawing. At the collector left 15 parts, 5 of which are larger. Find the probability of abnormal operation of the first assembled from these parts of the mechanism, if the collector takes part at random.

Decision. Let A denote the event of abnormal operation of the first mechanism assembled, and by A_1, A_2 and A_3 – events consisting in the fact that the first, second and third parts, respectively, set in the mechanism of a larger size. Then $A = A_1 A_2 A_3$, as the event A occurs when the condition of simultaneous occurrence of the events A_1, A_2 and A_3 . By the theorem of multiplication find

$P(A) = P(A_1A_2A_3) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2) = \frac{5}{15}\cdot\frac{4}{14}\cdot\frac{3}{13} \approx 0{,}011.$

*Example*1.33. At the warehouse received the consignment. The probability that each consignment to be defective products (Product) is 0.1. Determine the probability that the three parties of the goods received in a row does not contain defective products.

Decision. Let A denote the event consisting in the fact that in three games in a row received the goods do not contain defective products, and through A_1, A_2, A_3 – events consisting of the fact that each received consignment does not contain defective products.

According to the problem $A_1 = A_2 = A_3 = 1 - 0{,}1 = 0{,}9$, because developments, consisting of the fact that the incoming consignment does not contain defective products and defective products are found opposite.

Now, as the event A occurs provided the simultaneous occurrence of events A_1, A_2 and A_3, consequently A = $A_1{\cdot}A_2{\cdot}A_3$.

Events A_1, A_2 and A_3 are independent, the probability of occurrence for each of these events is not dependent on whether there have been two other events or not.

Therefore, by Theorem 1.1 multiplication of probabilities for independent events, we have:

$P(A) = P(A_1{\cdot}A_2{\cdot}A_3) = P(A_1){\cdot}P(A_2){\cdot}P(A_3) = 0{,}9^3 = 0{,}729.$

1.4.2. Probability addition theorem.

After we reviewed the multiplication theorem of probability we can now formulate the theorem of addition of probabilities is not only incompatible events but also for the joint.

***Theorem*. 1.2.** *The probability of occurrence of at least one of the two joint events is equal to the sum of the probabilities of these events without the likelihood of co-occurrence:*

$$P(A + B) = P(A) + P(B) - P(AB). \qquad (1.6)$$

Theorem accept without proof.

From (1.6) it follows as a consequence of addition and probability formula for exclusive events. Indeed, as for exclusive events $P(AB) = 0$, so from (1.6), we obtain

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

Consider a few examples of the application of Theorem 1.2.

Example 1.34. Determine the probability that a randomly chosen product is top-notch, if you know that 5% of all production is marriage, and 80% of the products without marriage meet the requirements of the 1st grade.

Decision. We denote by A – selected product is not defective, B – selected product meets the requirements of the 1st class, while AB – selected product is top-notch, and the desired probability

$P(AB) = P(A)P(B/A) = 0{,}95 \cdot 0{,}8 = 0{,}76$, where $P(A) = 1 - 0.05$, $P(B/A) = 0{,}8$.

Example 1.35. Produced two shots of the same target. The probability of hitting the first shot is 0.6 second - 0.8. Find the probability that the target will be struck at least once.

Decision. Consider the event A – hit at the first shot and the event B – in the second shot. Their probabilities $P(A) = 0.6$, $P(B) = 0.8$. Since A and B are joint and independent events, the probability that the target will be at least one hit, according to the formula (1.6), equal to

$P(A + B) = P(A) + P(B) - P(AB) = 0.6 + 0.8 - 0{,}6 \cdot 0{,}8 = 0{,}92$.

There is taken into account that the events are independent, which means $P(AB) = P(A)P(B)$.

If we use the concept of the opposite event and using the formula (1.6) we get

$$P(A + B) = 1 - P(\bar{A})\,P(\bar{B}) = 1 - 0{,}4 \cdot 0{,}2 = 0{,}92.$$

Here $P(A) + P(B) = 1$, $\bar{A}$ and $\bar{B}$ – a miss in the first and second shot, respectively, and are independent events.

1.4.3. The formula of the total probability.

Let some event of interest to us A may or may not occur with one of a number of mutually exclusive events of $H_1, H_2, \dots, H_n$, constituting the space of elementary events (also say that these events constitute a *complete group*). Events of such a series is usually called *hypotheses*. The probabilities of all hypotheses are known, i.e. are $P(H_1), P(H_2), \dots, P(H_n)$, while $P(H_1) + P(H_2) + \dots + P(H_n) = 1$. There are also conditional probability of the event A with implementation of each of these hypotheses, i.e. are $P(A/H_1), P(A/H_2), \dots, P(A/H_n)$. The probability P (A) we are interested in an event A is defined by the following theorem.

***Theorem*. 1.3.** *A probability of an event that may occur with this of the hypotheses H1, H2, ..., Hn, is the sum of pairwise products of probabilities of each of these hypotheses on their corresponding conditional probability of event A*:

P (A) = (1.7)$\sum_{i=1}^{n} P(H_i) \cdot P(A/H_i)$

Formula (1.7) is called ***total probability formula.***

Evidence. Since the hypotheses H_1, H_2,…, H_n form a complete group, the event A can be represented as the sum of the following events:

BUT AH1 AH2 = + + + ... = ANn.$\sum_{i=1}^{n} AH_i$

Since events H_i are incompatible, then the events AH_i (i = 1,2,…,n) also incompatible. This circumstance makes it possible to apply for determining the probability of an event A Theorem 1.2 summation of incompatible events probabilities

$$P(A) = \sum_{i=1}^{n} P(AH_i) \qquad (1.8)$$

The probability of event A and H_i the product is found by Theorem 1.1 of multiplication of probabilities:

$$P(AH_i) = P(H_i) \cdot P(A/H_i). \qquad (1.9)$$

Substituting (1.9) in (1.8), we obtain

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

Consider a few examples of the application of Theorem 1.3.

*Example*1.36.Let one of the three boxes located 3 white and 2 black balls, the second – 2 white and 3 black, in the third - only the white balls. From randomly selected box removed one ball. Find the probability that it is white.

Decision. Consider the events: H_1–ball is taken from the first box, H_2 – ball is taken from the second box, H3 - the ball is taken from the third box. Taking ball of any drawer occurs at random, therefore, the probability of any event box ball is taken out, the same and are as follows: $P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$. Let A denote the event consisting in the fact that the chosen from any of the three boxes will be a white ball. The probability of this event depends on from which the box selected ball and the conditional probability of the event A with the implementation of each event H_1, H_2, H_3 are: $P(A/H_1) = \frac{3}{5}$, $P(A/H_2) = \frac{2}{5}$ $P(A/H_3) = 1$.

Events H_1, H_2, H_3 are incompatible and form a complete band, so for event A may be regarded as a hypothesis, then using equation (1.7) for the total probability, we obtain

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3) = \frac{1}{3}\cdot\frac{3}{5}+\frac{1}{3}\cdot\frac{2}{5}+\frac{1}{3}\cdot 1 = \frac{10}{15} = \frac{2}{3}.$$

*Example*1.37.On two automatic machines produced identical rollers. Probability of production of premium roller on the first machine is equal to 0.95, while the second - 0.80. Made on both machines are not sorted beads in stock, including rolls made on the first machine, three times more than the second. Determine the probability that a random roll will be taken premium.

Decision. Denoted by the letter A – an event consisting in the fact that the cushion would be taken at random premium; B_1 – an event consisting in the fact that taking a random roller produced the first machine; B_2 – an event consisting in the fact that the roller is produced on the second machine. Applying obtain the total probability formula:

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) = \frac{3}{4}\cdot 0{,}95 + \frac{1}{4}\cdot 0{,}80 = 0{,}91.$$

Here, $P(B_1) = \frac{3}{4}$, $P(B_2) = \frac{1}{4}$, Since the rollers produced in the first machine, a 3-fold greater than the second. The conditional probabilities are given in the condition of the problem: $P(A/B_1) = 0{,}95$, $P(A/B_2) = 0{,}80$.

1.4.4. hypotheses Theorem (Bayes' formula).

Until now we have considered the probability of events prior to the tests, that is, in a complex environment did not appear the result of the experiment.

We now pose the following problem. There is a full group of incompatible hypotheses $H_1, H_2,\ldots, H_n$. Known each hypothesis probability $P(H_1), P(H_2),\ldots, P(H_n)$. Produced experience and its result carried out an event A is the probability for each hypothesis are known, i.e. known P (A / H1) and P (A / H-2), ..., P (A / Hn).

The question is what are the probabilities of the hypothesis H1, H2, ..., Hn in connection with the occurrences of A? In other words, we are interested in the conditional probabilities $P(A/H_1), P(A/H_2),\ldots, P(A/H_n)$.

***Theorem*. 1.4.** *The probability of the hypothesis after the test is the product of the probability of the hypothesis to test its corresponding conditional probability of an event that occurred during the test divided by the total probability of the event:*

$$P(H_i / A) = \frac{P(H_i)\cdot P(A/H_i)}{\sum_{i=1}^{n} P(H_i)\cdot P(A/H_i)}\,. \qquad (1.10)$$

Formula (1.10) is known as ***Bayes' formula***.

In a particular case, if all the hypotheses H_i ($i = 1,2,\ldots,n$) to the test have the same probability $P(H_i) = p$, equation (1.10) becomes:

$$P(H_i / A) = \frac{P(A/H_i)}{\sum_{i=1}^{n} P(A/H_i)}\,. \qquad (1.11)$$

We assume theorem hypotheses without proof and look at some examples on the use of this theorem.

Example 1.38. Let the conditions of Example 1.37, taken at random roller appeared premium. Determine the probability that it is produced on the first machine.

Decision. Using the notation of Example 1.37, Bayes' formula obtain

$$P(B_1/A) = \frac{P(B_1)\cdot P(A/B_1)}{\sum_{i=1}^{n} P(B_i)\cdot P(A/B_i)} = \frac{0{,}71}{0{,}71+0{,}20} = 0{,}78.$$

*Example*1.39. At the warehouse received three boxes of the same products from different companies: in the first 10 items including 3 non-standard, the second product 15, of which 5 and the third non-standard products 20, of which 6 nonstandard. Randomly selected one product, and it turned out to be non-standard. Determine the probability that the product is taken belongs to the company who sent the second box.

Decision. Denote by H_1, H_2, H_3 respectively the hypothesis that the product taken at random belonging to the first, second, third boxes. Then the probabilities of these hypotheses to the test are the same and equal $\frac{1}{3}$, i.e. $P(H_1) = P(H_2) = {}$ $=P(H_3)= = \frac{1}{3}$. As a result of the test is observed event A, consisting in the fact that randomly selected product is non-standard. Conditional probability of this event when the hypotheses H_1, H_2, H_3, respectively: $P(A/H_1) = \frac{3}{10}$, $P(A/H_2) = \frac{5}{15} = \frac{1}{3}$, $P(A/H_3) = \frac{3}{10}$. Now, in the Bayes formula (1.10), we find the probability of the hypothesis H_2 after test, i.e., the probability that it belongs to the non-standard products company, had sent a second box:

$$P(H_2/A) = \frac{P(A/H_2)}{\sum_{i=1}^{3} P(A/H_i)} = \frac{\frac{1}{3}}{\frac{3}{10}+\frac{1}{3}+\frac{3}{10}} = \frac{5}{14}.$$

1.4.5. The sequence of independent trials. Bernoulli formula.

In practice, faced with tasks that can be represented in the form of repetitive tests, the result of each of which may appear or not appear event A. In this case, interest is not the outcome of each test, and the total number of occurrences of A as a result of a certain number of tests . In such cases, you need to be able to determine the probability of any number m of occurrences of A as a result of n trials.

Definition* 1.23.** If the probability of an event A in each trial is independent of the outcome of other tests, such tests are called ***independent in relation to the event A.

The following problem stated.

Determine the probability that a result of n independent trials some event A occurs exactly m times, if each of these tests, this event occurs at a constant probability $P(A) = p$.

The desired probability will be denoted by $P_{m,n}(A)$. For example, $P_{4,12}(A)$ symbol denotes that the event A twelve tests appeared 4 times.

Consider a sought probability calculation method based on the application of Bernoulli's formula.

***Bernoulli formula*.** Assume that under the same conditions produced n independent trials, the result of each of which may be offensive or probability of the event A $P(A) = p$, or its opposite $\bar{A}$ probability $P(\bar{A}) = 1 - p$. Denote by A_i (i = 1,2,…,n) occurrence of an event A in the i-th trial. Due to the constancy of conditions (independent) tests

$$P(A_1) = P(A_2) = P(A_3) = \ldots = P(A_n) = p,$$
$$P(\overline{A_1}) = P(\overline{A_2}) = P(\overline{A_3}) = \ldots = P(\overline{A_n}) = 1 - p.$$

We are interested in the probability that an event A when n trials may appear exactly m times in different sequences or combinations equal to the number of combinations of n elements by m, i.e. C_n^m . An example of such a combination would be an event B in which the event A occurs m times in succession, starting with the first test:

$$B = A_1A_2\ldots A_m\,\bar{A}_{m+1}\ldots\,\bar{A}_n \qquad (1.12)$$

By hypothesis testing independent. This means that the events are independent of the combination of (1.12), therefore, using the multiplication theorem for independent events, we get:

$$P(B) = P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_m)\cdot P(\bar{A}_{m+1})\cdot\ldots\cdot P(\bar{A}_n) = p^m(1-p)^{n-m}.$$

Since all combinations of events, similar combinations are mutually exclusive events, and we do not care in what order will the event A and its opposite event $\bar{A}$appears in a sequence, using the addition theorem for the probabilities of mutually exclusive events, we get:

$$P_{m,n}(A) = C_n^m\, p^m(1-p)^{n-m} = \frac{n!}{m!(n-m)!}\, p^m\big(1-p\big)^{n-m} \qquad (1.13)$$

The resulting formula (1.13) is called the ***Bernoulli formula.***

Bernoulli formula is very important in probability theory, since it is related to a repetition test in the same conditions, i.e. with such conditions, which appear just laws of probability theory.

Consider the examples of using the Bernoulli formula.

Example 1.40.Coin tossed 5 times. Find the probability that the coat of arms appears 3 times.

Decision. Conditions problem corresponds circuit test sequences under identical conditions. Denote events: *A* - the appearance of the coat of arms in one trial, in *B* - the emblem will be 3 times in a series of five tests.

Now applying formula (1.13) with $n = 5$, $m = 3$ and $p = \frac{1}{2}$ we obtain

$P_{3,5}(A) = C_5^3 \ (\frac{1}{2})^3(\frac{1}{2})^2 = \frac{10}{32} = \frac{5}{16}$.

Example 1.41. Some products contain 5% production defective. Find the probability that among the six items taken at random:

1) are two defective;
2) It will not be rejected;
3) All six will be defective.

Decision. Conditions problem corresponds circuit test sequences under identical conditions. We introduce the notation: *A* - an event consisting in the appearance of defective products *B* - event consisting in the appearance of defective products during *m* of successive $n = 6$ tests. From task condition that $P(A) = 0{,}05$, $P(\overline{\mathbf{A}}) = 1 - 0{,}05 = 0{,}95$.

According to the Bernoulli formula

1) when $m = 2$, $P_{2,6}(A) = C_6^2 \cdot (0{,}05)^2 \cdot (0{,}95)^4 = 0{,}03$;
2) when $m = 0$, $P_{0,6}(A) = (0{,}95)^6 \approx 0{,}73$;
3) when $m = P_{6,6}(A) = (0{,}05)^6 \approx 0{,}156\ 10^{-7}$.

1.4.6. The most probable number of occurrences in successive tests.

***Definition*1.24. Most probable number m_0 occurrence of an event** *A* in *n* number of independent tests, for which the probability $P_{m,n}(A)$ is greater than or at least no less than the probability of each other possible test outcomes.

In order to determine the most probable number m_0 do not need to calculate the probability of different combinations of occurrence of the event *A*, and it is enough to know the number of trials n and probability p of occurrence of the event *A* in one trial.

It was found that the most probable number m0 satisfies the double inequality

$np - q \leq m_0 \leq np + p$, где $q = 1 - p$. (1.14)

This double inequality and serves to determine most probable number m_0.

Comment. Segment length $[np - q, np + p]$, defined by the inequality (1.14) is equal to unity: $(np + p) - (np - q) = 1 - p + p = 1$.

Therefore, if the boundaries of this segment has a fractional number, we get a single value most probable number m_0, if borders are integers, we get two values of the most probable number m_0: $np + p$ и $np - q$.

Consider the examples.

Example 1.42. Wholesale warehouse supplies 10 stores, each of which can do the application for the next day with a probability of 0.4, regardless of claims of other shops. Find the most probable number of applications per day and the probability of obtaining that number of applications.

Decision. In this task, $n = 10$, $p = 0,4$, $q = 1 - p = 1 - 0,4 = 0,6$. Substituting these data in inequality (1.14), we obtain

$$10 \cdot 0,4 - 0,6 \le m_0 \le 10 \cdot 0,4 + 0,4,$$

$$3,4 \le m_0 \le 4,4 ,$$

and therefore, $m_0 = 4$. The most probable number of applications is 4.

Let us now find the probability of getting four applications of the Bernoulli equation

$$P_{4,10}(A) = C_{10}^4 \cdot (0,4)^4 \cdot (0,6)^6 = 0,25.$$

Example 1.43. With this process 85% of all premium products. Find the most probable number of premium items in the lot of 150 items.

Decision. According to the problem $n =150$, $p =0,85$, $q =1-p =1- 0,85 = 0,15$. Substituting these data in inequality (1.14), we obtain

$150 \cdot 0,85 - 0,15 \le m_0 \le 150 \cdot 0,85 + 0.85$, so $m_0 = 128$.

Consequently, the most probable number of premium items in the lot of 150 products 128 is equal.

*Example*1.44. Determine the most probable number of the affected aircraft in a group of 13 bombers, if the planes are affected independently of each other, and the probability of hitting one aircraft is $\frac{4}{7}$.

Decision. According to the problem $n = 13$, $p = \frac{4}{7}$, $q = 1 - p = 1 - \frac{4}{7} = \frac{3}{7}$. Consequently, according to the inequality (1.14), we have

$$13 \cdot \frac{4}{7} - \frac{3}{7} \le m_0 \le 13 \cdot \frac{4}{7} + \frac{4}{7}.$$

Hence, we obtain:

$$7 \le m_0 \le 8.$$

This means that there are two values of 7 and 8, each of which is the most probable number of the affected aircraft.

CHAPTER 2
Random variables

In the first part of our presentation, we had to deal with random events. The event is by definition a qualitative characteristic of a random result of the experiment. The possibility of accidental events can be quantified, and it is the probability of occurrence of the events expressed as a number. But the result of a random experiment can also be characterized quantitatively, i.e. compare him to a certain number. For example, all the economic indicators are expressed numerically. This is the wages of workers with piece-rate wages, and production volume and profitability, and productivity, and the number of parts, leaving in their size is outside the tolerance, etc.

Further, any random experiment can be associated with the space of elementary events Ω = {ω}. If now the end of each elementary ω random experiment is also mapped and certain numerical value, the set of numeric values for all the elementary space Ω outcomes may serve as a quantitative result it characteristic random experiment, and is a random variable. The concept of a random variable is a fundamental concept in probability theory, and plays a very important role in its applications.

2.1. The concept of random variable

First, consider the concept of a random variable, which was historically.

Definition* 2.1. *Random variable is a quantity that the experiment could take one or the other (but only one) the value (expressed by integer), the advance of the experiment, it is unknown what exactly.

Random values usually denoted by capital end Latin letters *X, Y*, ..., and their possible numerical values are indicated by the respective small letters *x, y,…*

Among random variables, which are encountered in practice, we can distinguish two main types: discrete random variables and continuous random variables.

Definition* 2.2. *Discrete random variable called this quantity, the number of possible values is either finite or infinite countable set.

Here are some examples of discrete random variables: 1. The number of customers visiting the store a day; 2. The number of calls arriving at a telephone exchange during the day; 3. The number of days required to hail in June. In the last example the random variable can take on an infinite but countable set of values.

Definition* 2.3. *Continuous random variable defined to be a value, the possible numerical values of which vary in a continuous or continuous manner, i.e. represent a numerical interval.

Obviously, the number of possible continuous random variable, even in the case where it is limited and infinite. Give examples of continuous random variables: random deviations from a predetermined weight of the goods weight of the goods items in the accompanying document; deviation size produced part on the size specified in the drawing. Here, if the deviation of the weight or size occurs to the next higher value of the random variable is taken with the plus sign, if a smaller, then less.

So, with every random event can be related to two of its numerical characteristics of the probability of events, and the numerical value of a random variable, which is a kind of abstract expression of a random event. The distinctive feature of these characteristics is the fact that the probability of a random event positive value, the random value as a discrete, continuous and can take both positive and negative values. For example, suppose an event is to fill the bus for the whole route, or filling the store buyers at different times of the work shop. In both of these cases, with the event can be associated a random variable, which is expressed by the difference between the incoming and outgoing passengers on the bus, or customers into the store.

Operating with the concept of a random variable, in some cases it is more convenient than operation with random events. In what will be seen in further our discussion.

Let us now consider a modern interpretation of the concept of a random variable.

Assume a random experiment characterized space elementary events $\Omega = \{\omega\}$ and let the end of each elementary ω experiment is assigned a certain numerical value x of the plurality R of real numbers. This means that in the space of elementary events $\Omega = \{\omega\}$ is given a numerical function arguments are elementary outcomes w experiment, and the real number of their images. In other words, this function provides $\Omega = \{\omega\}$ mapping elementary events space R into a plurality of real numbers, i.e. each elementary event ω the function assigns to a real number. With this approach, and this function is considered as a random variable.

Accordingly, for the random variable is defined as follows.

Definition* 2.4. *Random variable called numerical function defined on the space of elementary events $\Omega = \{\omega\}$.

This function is denoted *X* as accepted for the random variable *X, Y*, ..., and their possible numerical values of *x, y*, ..., respectively, $X(\omega)$, $Y(\omega)$,... Sometimes for this function, we will also use the following entry:: $X \forall\omega \in \Omega \to X = X(\omega) \in R$, i.e. function itself and its value are denoted by the same letter: $X = X(\omega)$.

It should be noted that the range of values of this function is the set of real numbers R, and therefore, this feature enables the space $\Omega = \{\omega\}$ elementary events displayed as a set of points on the real axis.

Knowledge of the possible values of a random variable does not yet allow us to fully describe a random variable, so we can not say how often you should expect to see these or other possible values of the random variable in the experiment of repetition in the same conditions. For this purpose it is necessary to know the probability of the experiment each numerical value of the random variable, ie, you must know the law of probability distributions for all numeric values of a random variable. Based on this understanding, we give the following definition.

Definition 2.5. The law of random variable is any ratio, establishing the relationship between the possible values of the random variable and the corresponding probabilities.

Knowing the probability distribution between the possible values of a random variable, it is possible to judge the experiment about which values of the random variable will appear more often and which less.

It should be noted that the methods of presentation or the law of distribution of the random variable may be different. Consider some of them.

2.2. The laws of probability distribution of the random variable

We begin with distributions that are used for a discrete random variable.

2.2.1. Several discrete probability distribution of the random variable.

The simplest form of job distribution law is a discrete random variable table listing the possible values of the random variable and the corresponding probabilities. This table is called the number of the distribution of a discrete random variable.

Definition* 2.6. *Raw of the distribution or distribution law discrete random variable it called list of random variable and the corresponding values of these probabilities.

Several discrete random variable distribution can be specified in a graphic form. In the graphical representation of all possible values of the random variable are deposited on the coordinate plane on the abscissa axis and the ordinate axis - corresponding probabilities. Vertices obtained ordinates are usually joined by straight line segments. It should be noted that the compound ordinates of vertices is done only for illustrative purposes, as the intervals between vertices a random variable values can take, so the probability of its occurrence in these intervals are zero.

Such a graphical form of representation of the law of distribution of a discrete random variable is called a ***polygon distribution***.

For different conditions of experiment number distribution of a discrete random variable are distinct, but they all have one thing in common.

Property distribution series. The sum of (the ordinate or polygon) distribution representing the sum of the probabilities of all possible values of the random variable, is always equal to unity.

This is a random variable basic property of a probability distribution law and follows from the next.

Let there be given a finite discrete space of elementary events Ω and let A_1, $A_2,\dots,A_n$ are incompatible events that make up this space. Each of these events is characterized by the probability of its occurrence, respectively, $P(A_1)$, $P(A_2),\dots,P(A_n)$. The event A, representing the sum of all incompatible elementary space Ω events: $A = A_1 + A_2 +\dots+ A_n$ is a significant event, and means, respectively, that $P(A_1) + P(A_2) +\dots+ P(A_n) = 1$. If now with the space Ω can associate a random variable x is the numerical value of x which correspond to the events of this space: $x_1 = A_1$, $x_2 = A_2,\dots,x_n = A_n$, then since the probability of occurrence of events A_i and the random variable values x_i ($i = 1,2,\dots,n$) are identical, hence it follows $P(x_1) + P(x_2) +\dots+ P(x_n) = 1$. Consider an example.

*Example*2.1. Coin toss three times. Find a number of distribution and to build a polygon distribution of the number of occurrences of the emblem.

Decision. In this experiment, a random variable X is the number of occurrence of the emblem, and the numerical values which it can take is $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$.

To determine the probabilities of these values, because the trials are independent, and the emblem appearance occurs with a constant probability of $p = 1/2$, we can use the Bernoulli formula.

Pm, n=. (2.1)$C_n^m\, p^m(1-p)^{n-m}\, \frac{n!}{m!(n-m)!}\, p^m(1-p)^{n-m}$

Substituting this formula n = 3, and m values 0,1,2,3 obtain desired values of the probabilities: $P(0) = p^3 = 1/8$ (note that 0! = 1); $P(1) = 3/8$; $P(2) = 3/8$; $P(3) = 1/8$.To the resulting law of a random variable probability distribution is performed its main feature:

$$\sum_{i=1}^{4} P(x_i) = 1/8 + 3/8 + 3/8 + 1/8 = 1.$$

The appropriate number of random variable distribution for the given experiment is shown in Table 1, and polygon distribution shown in Fig. 2.1.

Considered distribution series is a convenient form of presentation of the distribution law discrete random variable with a finite number of possible values and is not suitable for a continuous random variable. Consider now the most common form of representation of the probability distribution law of a random variable, the so-called distribution function, which is applicable for all the random variables: both discrete and continuous.

Table 1

x_i	0	1	2	3
$P(x_i)$	1/8	3/8	3/8	1/8

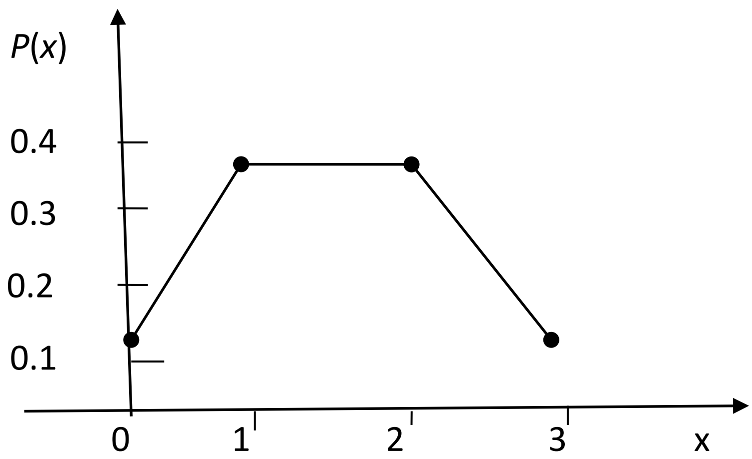

Fig. 2.1

2.2.2. The distribution function of the random probability values.

Consider a random variable $X = X(\omega)$ values which belong to the set of real numbers R and let x be a numerical value of the random variable X denote the probability that the random variable X will result in the test value is less than x $P(X < x)$. Then, if $\forall x \in X$ it is determined correspondence $x \to P(X < x)$, this correspondence is a function defined on the set values of the random variable $X \subset$R real numbers with values in the set as R, i.e. real function of one real variable. Denote this function F, and its value $F(x)$. Call this function of the random variable distribution function of X. Thus the function $F: \forall x \in \to F(x) = P(X < x)$. The domain of the distribution function of the entire set of real numbers R, even in the event that a random variable takes discrete values as those numerical values that lie below the minimum value of the received random quantity can be regarded as impossible events and the probability of their occurrence count equal to zero. Now let us define the random variable distribution function X.

Definition* 2.7. *The function of the random variable X is a function $F(x)$ defined on the whole set R of real numbers, i.e. x ∈ R, whose values have $P(X < x)$ is the probability that the random variable X will result in the test value is less than x.

The distribution function completely characterizes a random variable with a probability point of view, this means that it is a form of distribution. A significant advantage of the distribution function is that it is a real function of one real variable defined on the entire set of real numbers R, and therefore, to study it, we can use all the methods and techniques inherent in mathematical analysis for the study of such functions. One of the features of such a representation, for example, is that a function can be represented graphically in the form of a graph on a coordinate plane.

For a discrete random variable X, which can take on values $x_1, x_2, \ldots, x_n$, the distribution function will have the form:

$$F(x) = \sum_{x_i < x} P(X = x_i), \qquad (2.2)$$

where the summation is extended over all the values of the index i, for which $x_i < x$. Consider an example.

*Example*2.2. The target produced three separate shots. The probability of hitting the target with every shot is 0.4. Build results of the distribution function.

Decision. Denote the random variable number of hits through X, then the possible values of X are as follows: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$.

To determine the probabilities of these values, because the shots are independent and entering at a constant probability p = 0.4 can use the Bernoulli formula (2.1).

Substituting this formula n = 3, and m values 0,1,2,3 obtain desired values of the probabilities: $P(0) = (1 - p)^3 = 0{,}216$ (note that 0! = 1); $P(1) = 0{,}432$; $P(2) = 0{,}288$; $P(3) = 0{,}064$.

The corresponding number distribution is given in Table 2.

Table 2

x_i	0	1	2	3
R(Xi)	0,216	0.432	0,288	0.064

Now, according to c (2.2) to construct a distribution function resulting discrete random variable X:

1. When x ≤ 0 $F(x) = P(X < 0) = 0$, since the event in question for a given numerical interval is impossible event;
2. When 0 < x ≤ 1 $F(x) = P(X = 0) = 0{,}216$;
3. When 1 < x ≤ 2 $F(x) = P(X = 0) + P(X = 1) = 0{,}216 + 0{,}432 = 0{,}648$;
4. When 2 <x ≤ 3 F (x) = P (X = 0) + P (X = 1) + P (X = 2) = 0.216 + 0.432 + 0.288 = 0.936;
5. When 2 < x ≤ 3 $F(x) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}216 + 0{,}432 + 0{,}288 = 0{,}936$;

Schedule a consideration of the distribution function of the random variable X is shown in Fig. 2.2.

From the above example it follows that the distribution function of a discrete random variable X while remaining continuous from the left to the right is discontinuous and irregular increases at the transition point through its possible values x_1, x_2,…, x_n, and the magnitude of the jump is the probability corresponding value. The values of a discrete random variable divided numerical intervals, in which no other possible X values. These numerical intervals *F (x)* is the distribution function is constant, i.e. a graph of the distribution function is a stepped line.

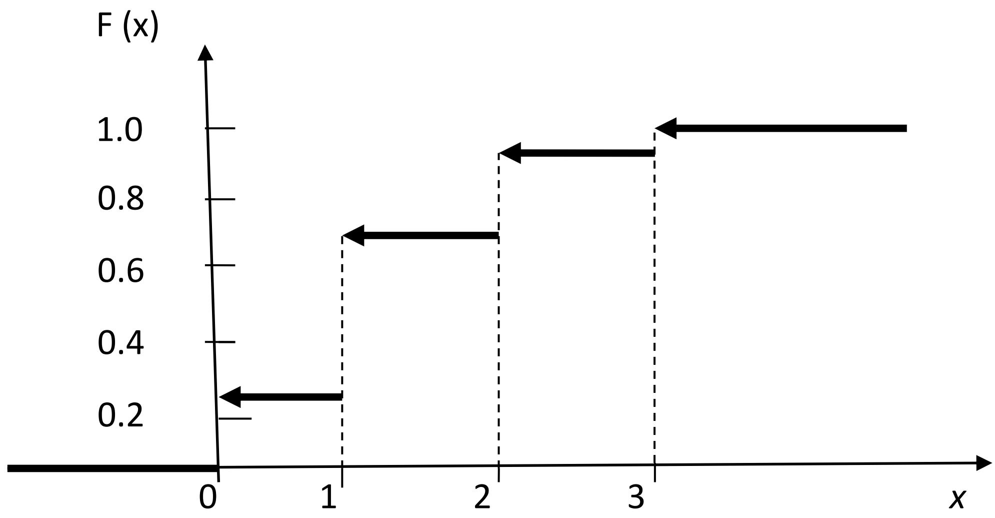

Fig. 2.2

Consider the common properties of the distribution function inherent in both discrete and continuous random variable *X.*

1. The distribution function $F(x)$ is a non-negative function, values of which lie between zero and one $0 \le F(x) \le 1$;
2. When $x \to -\infty$ $F(x) \to 0$;
3. When $x \to +\infty$ $F(x) \to 1$;
4. The probability of a random variable X in contact with an arbitrary number of the real axis interval (x_1, x_2) is defined by the right unclosed

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

Let us prove this property. To do this, consider the event $(X < x_2)$. It is evident that this event can be written as the sum of:

$(X < x_2) = (x_1 \le X < x_2) + (X < x_1)$, using the addition formula for the exclusive events, we get

$P(X < x_2) = P(x_1 \le X < x_2) + P(X < x_1)$, which implies

$F(x_2) = P(x_1 \le X < x_2) + F(x_1)$ или $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$.

5. The distribution function $F(x)$ – nondecreasing function throughout its domain R of real numbers, i.e. if $x_2 > x_1$, then $F(x_2) \ge F(x_1)$.

6. The distribution function is continuous on the left, that is,

$$\lim_{x \to x_0 - 0} F(x) = F(x_0).$$

Among the above properties, it follows that if the graph of the distribution function $F(x)$ represents a coordinate plane on the interval (a, b) monotonically increasing continuous line, it corresponds to a continuous random variable with possible values which continuously fill this interval. The converse is not true, and to him we shall return later. As an example of a continuous function graph F (x) line can serve as shown in Fig. 2.3.

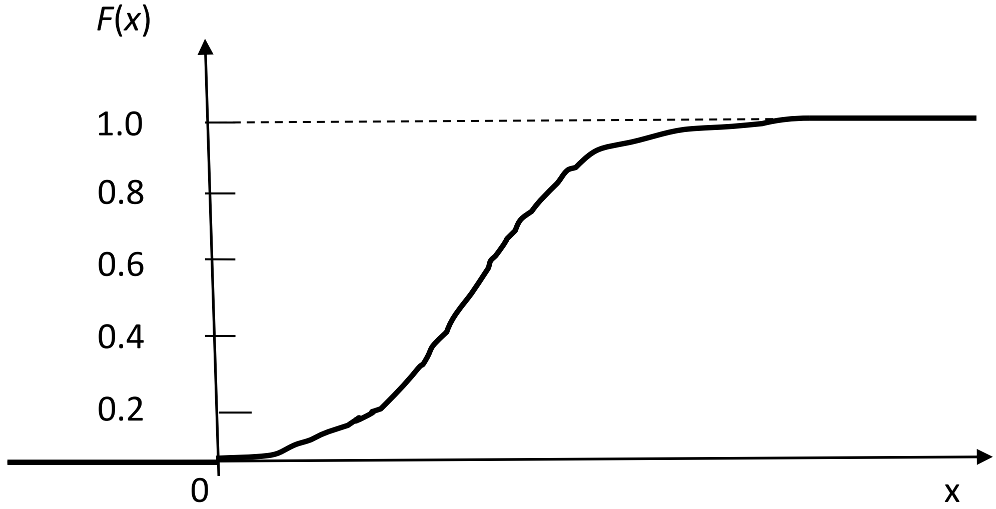

Fig. 2.3

Continuous random variable is a function of its exhaustive probability characteristic. But it has the disadvantage that, as it is difficult to judge the nature of a continuous random variable in a small neighborhood of one or the other point of the numerical axis, i.e. how often fall continuous random variable values in a neighborhood of the selected point. More visual representation of the character of a continuous random variable distribution in the vicinity of different points is given a function called probability density law or differential distribution of the random variable.

2.2.3. Continuous density probability distribution of the random variable.

Suppose there is a continuous random variable *X* with distribution function F (x). We compute the probability of getting this random variable to a numeric interval (x, $x + \Delta x$). According to property 4 of the previous section we have:

$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x)$.

We form the ratio of the length to the probability interval Δx:

$$\frac{P(x < X < x + \Delta x)}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x}. \qquad (2.3)$$

Definition 2.8. Relationship (2.3) is the average probability that per unit of length Δx interval.

Assuming that *F (x)* is a differentiable function, proceed in equation (2.3) to the limit $\Delta x \to 0$, then we get:

$$\lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x). \qquad (2.4)$$

The expression (2.4), and reflects the concept of a continuous density distribution probability of the random variable.

Definition **2.9.** Limit of the ratio of the probability that the random variable on a continuous numerical interval from *x* to *x* + Δ*x* to the length of the interval Δ*x* when Δ*x* tends to zero, is called the density distribution of the random variable probability at the point *x*, and is denoted *f (x*).

By virtue of (2.4) is the density distribution *f (x*) is the derivative of the distribution function *F (x*), i.e. $f(x) = F'(x)$.

The meaning of the density distribution *f (x*) is that it indicates how often there is a random variable X in a neighborhood of *x* by repeating tests.

Definition **2.10.** The curve showing the distribution of density *f (x*) of a random variable, the ***distribution curve*** is called.

A typical form of *f (x*) of the distribution curve is shown in Fig. 2.4 the solid curve.

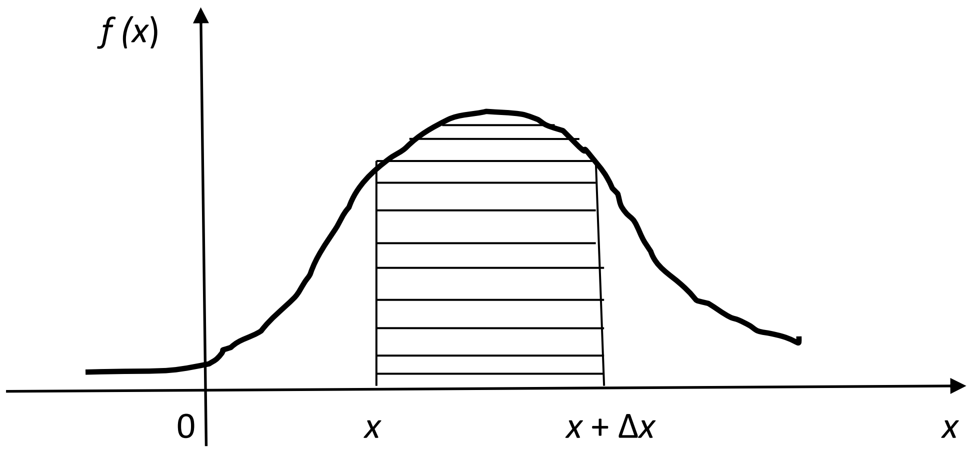

Fig. 2.4

Note that if the possible values of the random variable is filled with some numerical finite interval, the density distribution *f (x*) = 0 outside of this interval.

We have already drawn attention to the fact that not all continuous random variables, possible values are continuously filled with a numerical interval, the distribution function they will also be continuous. There are random variables, whose possible values are continuously filled with some numerical interval, but for which the distribution function is not always continuous, and in some locations is discontinuous. Such random values called mixed. For mixed random variables at the points of rupture limit of the distribution function *F*(*x*) when $\Delta x \to 0$ does not exist, therefore, does not exist at these points and characteristic density distribution. Accordingly, now give a more rigorous definition of a continuous random variable.

***Definition* 2.11.** The random variable *X* is said to be continuous if its distribution function *F*(*x*) is continuous on the entire axis *Ox*, and the density distribution *f*(*x*) exists everywhere, except perhaps a finite number of points.

Consider the properties of probability density.

1. The density distribution is non-negative, i.e., $f(x) \geq 0$.

2. The distribution function of the random variable *F*(*x*) is equal to the integral of density *f*(*x*) in the interval from - ∞ to *x*, i.e. $F(x) = \int_{-\infty}^{x} f(z)dz$. you take into account that $F(-\infty) = 0$.

3. The probability of hitting a continuous random variable *X* in the interval (*x*, *x* + Δ*x*) is equal to the integral of the density distribution taken along the interval, i.e., $P(x < X < x + \Delta x) = \int_{x}^{x+\Delta x} f(z)dz$. Geometrically, this property may be interpreted as follows: the probability that a continuous random variable will take the value belonging to the interval (*x*, *x* + Δ*x*), equal to the area of the curvilinear trapezoid, shaded in Fig. 2.4.

Note that the probability $P(x < X < x + \Delta x)$ can be obtained by summing the probabilities of elements in the areas for the entire interval (*x*, *x* + Δ*x*),

4. The integral in the infinite limits on the density distribution is equal to one: $\int_{-\infty}^{+\infty} f(x)dx = 1$.

If the range of possible values of the random variable has finite limits a and b, the density distribution *f* (*x*) = 0 outside the interval (*a, b*) and feature 4 can then be written as: $\int_{a}^{b} f(x)dx = 1$.

Geometrically, this property is the density distribution of the random variable means that the whole area under the distribution curve bounded by the *x*-axis is equal to unity.

Example. 2.3. The random variable X subject to the law with the density

$$f(x) = \begin{cases} a \cdot sinx \ \ at \ \ 0 \leq x \leq \pi, \\ 0 \ at \ x < 0 \ \ or \ \ x > \pi. \end{cases}$$

Requires:

1) Find *a* factor.

2) Draw the graph of the density distribution.

3) Find the probability of a random variable falling in the interval from 0 to $\frac{\pi}{4}$.

Decision. 1) For the determination of the coefficient *a*, we use the distribution 4 density property: $\int_{-\infty}^{+\infty} f(x)dx = \int_0^{\pi} asinxdx = 2a = 1.$

Where *a* = 1/2.

2) Graph density distribution *f (x)* shown in Fig. 2.5.

3) By property 3 we have:

$P(0 < X < \frac{\pi}{4}) = \int_0^{\frac{\pi}{4}} \frac{1}{2} sinxdx = -\frac{1}{2}(cos\frac{\pi}{4} - cos0) \approx 0{,}15.$

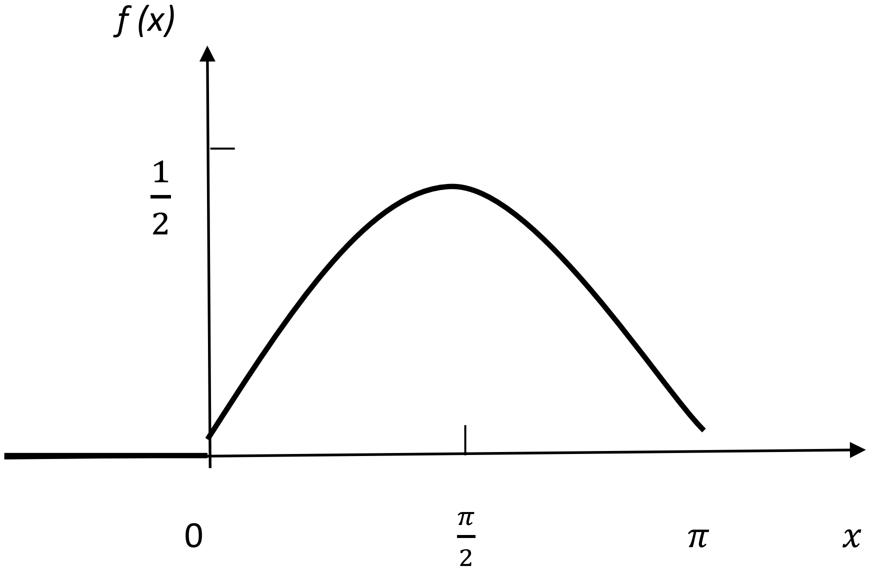

Fig. 2.5

As we have said, the law of distribution is fully characterized by a random variable with a probability point of view. Knowing the law of the random variable, you can specify where are the possible values of a random variable, and what is the likelihood of it in a given interval.

However, in solving many practical problems there is no need to characterize the random variable completely, but enough to have a random variable is only a general idea. It is often not enough to indicate the distribution of the whole law, but only some of the characteristics of the distribution law.

The probability theory for the overall characteristics of the random values are used, some values that are called random numerical characteristic values.

Their main purpose – to succinctly express the most essential features of a distribution.

2.3. Numerical characteristics of the random variable

For each random variable is necessary, first of all, to know its mean value, which are grouped about the possible values of the random variable, and any number that describes the degree of dispersion of the relative average values. In addition to these numerical specifications, for a more complete description of the random variable using a number of other numerical characteristics. they all help in one way or another as to understand the characteristics of the random variable. Consider the most common numerical characteristics.

2.3.1. Expected value.

The expectation is an essential characteristic of the position of the random variable. The mathematical expectation of a random variable is sometimes referred to simply as the average value of the random variable. Consider first a discrete random variable.

2.3.1.1. The expectation of a discrete random variable.

Let X be a discrete random variable with possible values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$.

Definition* 2.12. *Expectation discrete random variable X is the sum of the products of all possible values of the random variable on the probability of those values.

Denoting the mathematical expectation of a random variable X by $M(X)$, then by definition, the expectation of $M(X)$ of the random variable X is defined by

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i. \quad (2.5)$$

If a discrete random variable can take on an infinite countable set of values $x_1, x_2, \ldots, x_n, \ldots$ with probabilities $p_1, p_2, \ldots, p_n, \ldots$, then its expectation is defined by

$$M(X) = \sum_{i=1}^{+\infty} x_i p_i. \quad (2.6)$$

In the statistical approach of determining the expectation of a random variable, i.e., empirically from tests expectation is approximately equal to the arithmetic mean of the observed values of the random variable, and the more precise the greater the number of trials.

If produced several series of tests, the expectation there is a fixed number, which will fluctuate around the arithmetic mean value of a random variable, calculated for each test series.

Consider now a continuous mathematical expectation of a random variable.

2.3.1.2. Continuous mathematical expectation of a random variable.

Let a continuous random variable X, all possible values which belong to the interval [a, b], and f (x) the density of its distribution.

Definition **2.13.** ***Continuous random variable X is the mathematical expectation,*** all possible values which belong to the interval [*a, b*] is called the definite integral

$$M(X) = \int_a^b xf(x)dx. \qquad (2.7)$$

If the possible values of continuous random variable *X* belong throughout Ox, the expectation is determined by the integral

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx. \qquad (2.8)$$

It should be noted that there are also such random variables for which there is no expectation, since the corresponding amount (2.6) or the corresponding integral (2.8) diverge. However, these random variables are rare, and we are with them, we will not be faced.

We note the simplest properties of the expectation.

2.3.1.3. Properties of the expectation of a random variable.

Property 1. The expectation constant value equal to the most constant, i.e. $M(C) = C$.

Indeed, a constant value can be regarded as a special case of a certain event, then (2.8), $M(C) = C\cdot 1 = C$.

Property 2. If the random value is multiplied by some number k, then, and the expectation is multiplied by the same number or a constant factor can be taken as a sign of the expectation, that is, $M(kx) = kM(X)$.

Property 3. The expectation of the sum of random variables is equal to the sum of their mathematical expectations

$$M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n).$$

Property 4. $M(X_1 - X_2) = M(X_1) - M(X_2)$.

Property 5. For independent random variables $X_1, X_2, \ldots X_n$ mathematical expectation of the product is equal to the product of their mathematical expectations:

$$M(X_1, X_2, \ldots X_n) = M(X_1)\, M(X_2) \ldots M(X_n).$$

Property 6. The expectation deviations $X - M(X)$ values of the random variable *X* from its mean value $M(X)$ is zero:

$M(X - M(X)) = M(X) - M(M(X)) = M(X) - M(X) = 0.$

Consider the examples.

*Example*2.4. The company holds cash prizes for loyal customers. For this purpose, placed in the box 20, identical balls, inside which is a coupon indicating the monetary prize amount which is 1000 UAH, 2000 UAH, ..., 20000 UAH. Determine the average of the winning amount.

Decision. Prize value is the random variable X, which takes values x_1 = 1000 UAH, x_2 = 2000 UAH,…, x_{20} = 20000 UAH. The probability of winning any amount of the same and is p = 1/20. Then, in accordance with the formula (2.5) we have

$$M(X) = p(x_1 + x_2 +\ldots + x_{20}) = p\sum_{i=1}^{20} x_i = 10500 \text{ UAH}.$$

Example 2.5. Let the random variables X_1, X_2 respectively, given the laws of distribution Table 3 and Table 4:

X_1 TABLE 3

x_{1i}	- 0.1	- 0.01	0	0.01	0.1
R_{1i}	0.1	0.2	0.4	0.2	0.1

X_2 TABLE 4

x_{2i}	-20	- 10	0	10	20
R_{1i}	0.3	0.1	0.2	0.1	0.3

Compute $M(X_1)$ and $M(X_2)$

$M(X_1) = (-\,0{,}1)\,0{,}1 + (-\,0{,}01)\,0{,}2 + 0 \cdot 0{,}4 + 0{,}01 \cdot 0{,}2 + 0{,}1 \cdot 0{,}1 = 0.$

$M(X_2) = (-\,20)\,0{,}3 + (-\,10)\,0{,}1 + 0 \cdot 0{,}2 + 10 \cdot 0{,}1 + 20 \cdot 0{,}3 = 0.$

The expectations, i.e. the average values of the two random variables are equal to zero. However, their distribution is different. If the values X_1 differs little from its mathematical expectation, the value X_2 to a large extent from its mathematical expectation, and the likelihood of such deviations are not small. These examples show that the average value of the random variable cannot be determined, which deviations from it have taken place value of the random variable when tested as a smaller and a larger side. So at the same average size drop-down in the two areas for the year precipitation is not to say that these areas are equally favorable for agricultural works. Similarly, in terms of average wages cannot be judged on the proportion of high- and low-paid workers.

2.3.2. Variance and standard deviation.

The values observed in the practice of random variables is always more or less fluctuate around an average value. This phenomenon is called *dispersion values around its mean value.*

Numerical characteristics showing how closely grouped possible values of the random variable around the scattering center (expectation), called scattering characteristics.

As a random variable scattering characteristic *X* from its mean value can not be used deviation *X* - *M (X*) of its value from the average, since according property 6 expectation expectation of such deviation is zero: *M (X - M (X))* = 0. it only indicates that the value of deviation - the number of different signs. Therefore, as the random variable scattering measures take the expectation of a square deviation of the random variable from its expectation, which is called the variance of the random variable *X* and is denoted *D (X)* or D_X.

Definition* 2.14. *Variance of the random variable X is a mathematical expectation of the square of the random variable *X* values of deviations from its expectation, i.e.

$$D(X) = M(X - M(X))^2. \tag{2.9}$$

For a discrete random variable *X* takes values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ dispersion expressed as the sum

$$D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 \cdot p_i\,, \tag{2.10}$$

and continuous - integral

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 \cdot f(x)dx. \tag{2.11}$$

If the range of possible values of the random variable has finite limits *a* and *b*, then *x* - *M (X)* = 0 outside the interval (*a, b*) and then (2.11) can be written as: $D(X) = \int_a^b (x - M(X))^2 \cdot f(x)dx$.

Formulas (2.10) and (2.11) follow directly from the definition of the expectation (2.5) and (2.8).

The variance of the random variable is a convenient characteristic dispersion of possible values of the random variable. However, it is devoid of clarity, as has the dimensions of a square random variable.

For convenience it is desirable to have a characteristic dimension of the random variable which coincides with the dimension as well as the mathematical expectation. Such a characteristic is the standard deviation of the random variable, which represents the positive square root of the variance.

Definition* 2.15. *Standard deviation of the random variable X is the value equal to the square root of the variance of a random variable.

Represent the standard deviation of the random variable X symbol $\sigma_X$and, consequently,

$$\sigma_X = \sqrt{D(X)}. \qquad (2.12)$$

Consider the simple properties of the dispersion.

Property 1. Dispersion constant value is zero: $D(C) = 0$.

Property 2. The dispersion product of a constant value for the random variable is equal to the product of the square of the constant value to the variance of a random variable: $D(CX) = C^2 D(X)$.

Property 3. The dispersion of the random variable X is the mathematical expectation of a random variable a square minus the square of its expectation: $D(X) = M(X^2) - M^2(X)$.

Property 4. pairwise independent random variables X_1, X_2, …, X_n variance of the sum is equal to the sum of the variances.

$D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n)$.

Consider the examples.

Example 2.6. Calculate variance for the random variables X_1, X_2 from Example 2.5. for which the mathematical expectation of random variables both equal to zero: $M(X_1) = M(X_2) = 0$.

Decision. Using formula (2.10), and tables 3 and 4.

$D(X_1) = 0{,}01 \cdot 0{,}1 + 0{,}0001 \cdot 0{,}2 + 0{,}0001 \cdot 0{,}2 + 0{,}01 \cdot 0{,}1 = 0{,}001 + 0{,}00002 + 0{,}00002 + 0{,}001 = 0{,}00204$.

$D(X_2) = (-20)^2 \cdot 0{,}3 + (-10)^2 \cdot 0{,}1 + 10^2 \cdot 0{,}1 + 20^2 \cdot 0{,}3 = 240 + 20 = 260$.

The closer the value of the dispersion to zero, the less random variable scatter about the mean.

*Example*2.7. A random variable is defined density distribution

$$f(x) = \begin{cases} \dfrac{cosx}{2} & at \quad -\dfrac{\pi}{2} \le x \ll \dfrac{\pi}{2}, \\ 0 & at \quad |x| > \dfrac{\pi}{2}. \end{cases}$$

The variance and standard deviation of the random variable X.

Decision. Using formula (2.7), we find the expectation

$$M(X) = \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x\cos x dx = \frac{1}{2} x\sin x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} - \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin x dx = 0\,.$$

Since *M (X)* = 0, then applying (2.11), we find:

$$D(X) = \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\cos x dx = \frac{1}{2}\left(2x\cos x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} + (x^2-2)\sin x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\right) = \frac{\pi^2-8}{4},$$

hence

$$\sigma_X = \sqrt{\frac{\pi^2-8}{4}} = \frac{1}{2}\sqrt{\pi^2-8.}$$

2.3.3. Mode and median of the random variable.

Furthermore expectation, which is the main characteristic of the numerical position of the random variable, in practice, apply position and other characteristics, in particular the mode and median of the random variable.

Definition* 2.16. *Mode M_d discrete random variable X we mean a value of the random variable, which corresponds to the highest probability.

Definition* 2.17. *Mode M_d of a random variable X is the continuous type we mean a value of the random variable, which corresponds to the maximum point of the probability density *f (x).*

If discrete random variable distribution of polygon (or a continuous curve of the random variable) has two or more maxima, then the distribution is called ***bimodal*** or ***multimodal.***

Sometimes there are distributions that have a minimum but do not have a maximum. Such distributions are called antimodal.

Definition* 2.18. *M_D median of the random variable X is the continuous type we mean a value of a random variable with respect to which equiprobably obtain larger or smaller value of the random variable, i.e., $P(X < M_D) = P(X > M_D)$.

Geometrically median – it abscissa of the point where the area bounded by the distribution curve is divided in half. Each of these areas is equal to 0.5, as the entire area enclosed by the distribution curve is equal to one. Therefore, the distribution function at M_D

$$F(M_D) = P(X < M_D) = 0{,}5.$$

Note that if the distribution of the unimodal and symmetric, then all three numerical characteristics of a random variable position – mean, mode and median – are the same.

Let us now consider the specific laws of distribution of the random variable is most often encountered in practice.

2.4. Examples of random variable distribution laws

We start with discrete random variables.

2.4.1. Examples of a discrete random variable distribution laws.

Among the laws of distribution for discrete random variables are the most common binomial distribution and Poisson distribution.

2.4.1.1. Binomial distribution.

The binomial distribution takes place under the following conditions.

Let the random variable *X* represents the number of occurrences of an event *A* in *n* independent trials conducted in the same conditions. The probability of an event *A* is constant and equal to *p*. Consequently, the probability of occurrence of an event *A* is not equal to *q* = 1 - *p*.

The possible values of the random variable *X* are $x_0 = 0$, $x_1 = 1$, $x_2 = 2$,…, $x_n = n$. The probabilities $P(X = x_i)$, $i = 0, 1, 2,…, n$ of the possible values are determined by the Bernoulli formula (2.1) with the proviso that $m = x_i$:

$$P(X = x_i) = C_n^{x_i}\, p^{x_i}(1-p)^{n-x_i} = \frac{n!}{x_i!(n-x_i)!}\, p^{x_i}(1-p)^{n-x_i}\,. \qquad (2.13)$$

Since the numerical values $x_0 = 0$, $x_1 = 1$, $x_2 = 2$,…, $x_n = n$, the received random variable *X*, consisting in a different number of occurrence of an event *A* in a series of *n* trials are incompatible and form a space elementary events for the random variable *X* therefore, in order to formula (2.13) was close to the distribution of the random variable *X*, it is necessary that the sum of the probabilities defined by (2.13) for all elementary events of a random variable *X* is equal to unity, i.e.,

$$\sum_{i=0}^{n} P(X = x_i) = C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \cdots + C_n^n p^n q^0 = 1.$$

Here, *q* = 1 - *p*.

Indeed, we consider expression $(p + q)^n = 1$ is decomposable binomial $(p + q)^n$ by the binomial theorem. We obtain

$$C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \cdots + C_n^n p^n q^0 = 1,$$

those. the sum of the probabilities of all values of the random variable is equal to one, therefore (2.13) is a distribution law.$\sum_{i=0}^{n} P(X = x_i)$

Definition* 2.19.** The distribution of the discrete random variable X for which the number distribution is given by (2.13) is a ***binomial distribution.

We find the expectation and the variance of the random variable X having a binomial distribution.

By definition, the expectation of a discrete random variable, we have:

$$M(X) = \sum_{i=0}^{n} x_i p_i = \sum_{i=0}^{n} x_i C_n^{x_i} \, p^{x_i} (1-p)^{n-x_i}. \qquad (2.14)$$

Consider the random variables $X_1, X_2, \ldots, X_n$, with the same distribution law:

$$X_k = \begin{cases} 1, if \ in \ the \ k-th \ test \ \ event \ A \ occured, \\ 0, if \ in \ the \ k-th \ test \ \ event \ \overline{A} \ occured, \end{cases}$$

where $k = 1,2,\ldots,n$. Then $X = X_1 + X_2 + \ldots + X_n$.

Using the properties of the expectation, we get:

$$M(X) = M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n) .$$

We find expectation X_k , $M(X_k) = 0 \cdot (1 - p) + 1 \cdot p = p$, hence, $M(X) = n \cdot p$.

Thus, the expected number of occurrences of an event in a series of independent and identical trials is the product of the number of tests on the probability of occurrence of an event in one trial.

Similarly find variance for the binomial distribution:

$D(X) = D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n)$.

$D(X_k) = (0 - p)^2 (1 - p) + (1 - p)^2 \ p = \ p^2 (1 - p) + (1 - p)^2 p =$

$= p (1 - p) (p + 1 - p) = p (1 - p) = p \, q$.

Therefore, $D(X) = n \cdot p \cdot q$, thus $\sigma_X = \sqrt{npq}$.

Example 2.8. Delivered to the warehouse a lot of products. For the adoption of the party is necessary that the probability of the presence in the party of the defective product is not more than 0.6. For this is done on a sample of 50 items from each batch. Find what should be the expectation, variance and standard deviation of the number of defective parts in the sample to the party was adopted.

Decision. Let the random variable X is a number x_i, $i = 0,1,2,\ldots,50$ of defective products in a sample of $n = 50$ articles. Probability of defectiveness one-piece $p = 0.06$. The random variable X is the binomial distribution defined by (2.13). Therefore, the expectation, ie, the average value of the number of defective components in the sample is calculated according to the formula $M(X) = n \cdot p = 50 \cdot 0{,}6 = 3$ (product). Accordingly dispersion $D(X) = n \cdot p \cdot q = 50 \cdot 0{,}06 \cdot 0{,}94 = 2{,}82$, and standard deviation. $\sigma_X = \sqrt{npq} = \sqrt{2{,}82} \approx 1{,}68 x_i \sigma_X = \sqrt{npq} = \sqrt{2{,}82} \approx 1{,}68.$

Thus, in order to product batch was adopted in the sample of 50 products should not contain more than $M(X) + \sigma_X \approx 5$ five defective products.

2.4.1.2. Poisson distribution.

Typical examples of the random variable having a Poisson distribution, are the number of calls to the telephone station for some time, the number of failures complicated apparatus for a certain period of operation time, if it is known that failures are independent from each other, and failures have an λ average per unit time, and so on.

Consider the general problem of the theory of probability, leading to a Poisson distribution.

Let a real axis *Ox* certain points distributed random variable *X*, the host infinite set of numerical values in such a way that the probability of any given number of points on any segment ℓ from *Ox* does not depend on the number of points falling to other nonoverlapping segments axis *Ox*, and the distribution of these segments, and depends only on the size of the segment ℓ**.**

Required to find the probability $P(X = m)$ that the segment length ℓ of real axis *Ox* fall exactly *m* points, assuming that the points are distributed over the entire axis with the same average density. We denote this density, i.e., the average number of points (expectation) per unit length, through ρ. Then the required probability is given by

$$P(X = m) = \frac{(\rho\,\ell\,)^m}{m!} e^{-\rho\,\ell}\,.$$

The value $\rho\,\ell$ of this expression is nothing else than the expected number of points falling on the segment length ℓ . Denoting $\rho\,\ell = \lambda$, we get

$$P(X = m) = \frac{(\lambda)^m}{m!} e^{-\lambda}\,. \qquad (2.15)$$

Definition* 2.20.** The distribution of the discrete random variable X, described by formula (2.15) is called ***Poisson distribution.

The Poisson distribution is dependent on a single parameter λ, which is the mathematical expectation of a random variable *X* ($M(X) = \lambda$).

The sum of probabilities for all numerical values of the random variable *X*, despite their infinite number ($m \to \infty$), as for any of the distribution law is equal to one:

$$\sum_{m=0}^{\infty} P(X = m) = \sum_{m=0}^{\infty} \frac{(\lambda)^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(\lambda)^m}{m!} = e^{-\lambda} e^{\lambda} = 1.$$

We define the variance of a random variable *X*, which has a Poisson distribution (2.15). According to a third dispersion property have: $D(X) = M(X^2) - M^2(X)$.

Found: $M(X^2)$: $M(X^2) = \sum_{m=0}^{\infty} m^2 \frac{(\lambda)^m}{m!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{m=1}^{\infty} m \frac{(\lambda)^{m-1}}{(m-1)!} =$
$\lambda e^{-\lambda} \sum_{m=1}^{\infty} [(m-1)+1] \frac{(\lambda)^{m-1}}{(m-1)!} =$
$\lambda e^{-\lambda} \sum_{m=1}^{\infty} (m-1) \frac{(\lambda)^{m-1}}{(m-1)!} + \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{(\lambda)^{m-1}}{(m-1)!} = \lambda^2 e^{-\lambda} \sum_{m=2}^{\infty} \frac{(\lambda)^{m-2}}{(m-2)!} +$
$\lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{(\lambda)^{m-1}}{(m-1)!}$.

Since each of these sums is e^{λ} then
$M(X^2) = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.$
Therefore, the variance value *X* is equal to
$D(X) = M(X^2) - M^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$
Thus, the variance of a random variable having a Poisson distribution, is equal to its expectation.

Consider the examples.

Example 2.9. In an ambulance for a specific hour of the day receives an average of 90 calls. Find the probability that in a minute goes up to three calls.

Decision. The expected number of calls per minute is equal to 90/60 = 3/2. The probability that for a given moment there is not more than three calls, equals the sum of the probabilities that during this minute is either 0 or 1 or 2 or 3 call. Therefore, the probability sought

$$\sum_{m=0}^{3} P(X=m) = \sum_{\mathrm{m}=0}^{3} \frac{\left(\frac{3}{2}\right)^m}{m!} e^{-\frac{3}{2}} = e^{-\frac{3}{2}} \left(1 + \frac{3}{2} + \frac{9}{8} + \frac{27}{48}\right) \approx 0{,}81.$$

Poisson distribution can be used as an approximation in cases where the exact distribution of the random variable is a binomial distribution and when the expectation differs little from the dispersion, i.e. when $n \cdot p \approx n \cdot p \cdot q$.

Example 2.10. The plant sent to the base 500 of benign products. The likelihood that the product will be damaged in a way, is 0,002. Find the probability that the base will arrive three worthless products.

Decision. The problem is solved by using the Poisson approximation formula. We have: p =0,002, $q = 1 - p = 0{,}998$ и $n = 500$.

According to the formulas $M(X) = n \cdot p$ и $D(X) = n \cdot p \cdot q$ define respectively the expectation and variance of the number of unacceptable products: $M(X) = 500 \cdot 0{,}002 = 1$, $D(X) = 500 \cdot 0{,}002 \cdot 0{,}998 = 0{,}998$.

Since $M(X) \approx D(X)$, the setting $\lambda = M(X) = 1$, we obtain approximately the desired probability of the Poisson formula:

$$P(X=3) = \frac{(\lambda)^m}{m!} e^{-\lambda} = \frac{1}{3!} e^{-1} \approx 0{,}06.$$

2.4.2. Examples of continuous random variable distribution laws.

2.4.2.1. Uniform distribution.

***Definition*. 2.21.** The random variable X is called a continuous type distributed uniformly on the interval $[a, b]$, if its distribution density is constant on this segment, and the segment is equal to zero, i.e.

$$f(x) = \begin{cases} 0, \; if \;\; x \notin [a,b], \\ C, \;\; if \;\; x \in [a,b]. \end{cases}$$

The graph of $f(x)$ the probability density for uniform distribution shown in Fig. 2.6.

Since the area bounded by the distribution curve is equal to one, and this area (Fig.2.6) has a height of the rectangle area C and a base $b - a$, hence, a uniform distribution takes the form $C = \frac{1}{b-a}$

$$f(X) = \begin{cases} 0, \; if \;\; x \notin [a,b], \\ \dfrac{1}{b-a}, \;\; if \;\; x \in [a,b]. \end{cases} \quad (2.16)$$

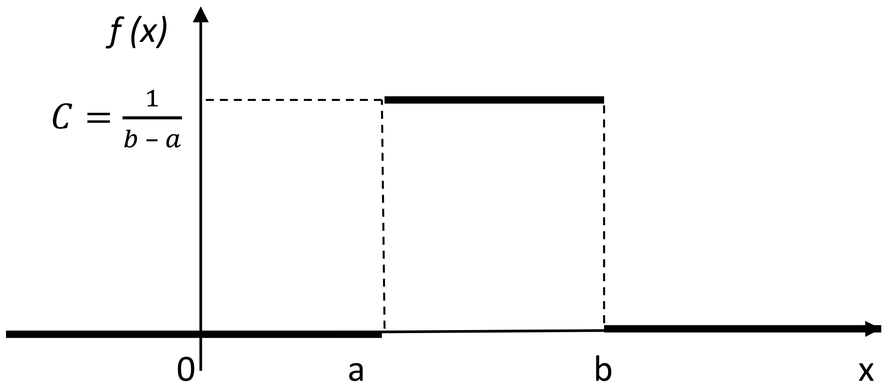

Fig. 2.6

The distribution function $F(x)$ for the uniform distribution. According to the property of density distribution, defined by the formula $F(x) = \int_{-\infty}^{x} f(z)dz$, we have

$$F(x) = \int_{-\infty}^{x} f(z)dz = \int_{a}^{x} \frac{1}{b-a} dx = \frac{x-a}{b-a} \;\; if \; x \in [a,b].$$

If $x < a$ $F(x) = 0$, a $x > b$ $F(x) = 1$.

In this way,

$$F(x) = \begin{cases} 0 & at \quad x < a, \\ \dfrac{x-a}{b-a} & at \quad a \le x \le b, \\ 1 & at \quad x > b. \end{cases}$$

Graf function *F (x)* shown in Fig. 2.7.

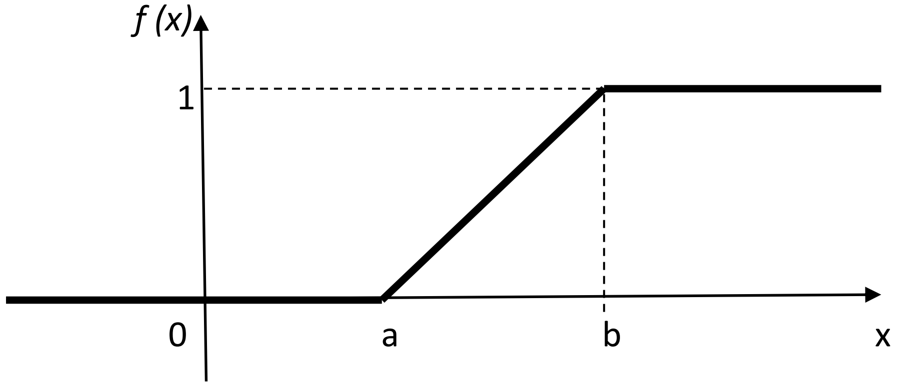

Fig. 2.7

With the random variable having a uniform distribution, is frequently encountered in practice when rounding measuring instrumentation readings divisions to integers scales. Error by rounding to the nearest whole frame dividing *X* is a random variable which may take a constant probability density any value between two adjacent integer divisions.

We compute the expectation and the variance of the random variable *X* having a uniform distribution at the site of *a* to *b.*

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

Those the expectation of uniform distribution is in the middle range of its distribution.

Variance of the random variable *X* from the formula (2.11):

$$D(X) = \int_{-\infty}^{\infty} \left(x - M(X)\right)^2 \cdot f(x)dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a}dx = \frac{(b-a)^2}{12}.$$

From standard deviation $\sigma_X = \sqrt{D(X)} = \frac{|b-a|}{2\sqrt{3}}$.

2.4.2.2. Exponential distribution.

In practical applications of probability theory, especially in queuing theory, operations research, issues of security and other applications often deal with random variables having the so-called exponential or exponential distribution.

Definition*. 2.22.** Continuous random variable *X* is distributed according to an ***exponential distribution with parameter $\lambda > 0$, if the density distribution is given by

$$f(x) = \begin{cases} 0, & if\ x < 0, \\ \lambda e^{-\lambda x}, & if\ x \geq 0. \end{cases} \qquad (2.17)$$

The graph of *f (x)* is shown in Figure 2.8.

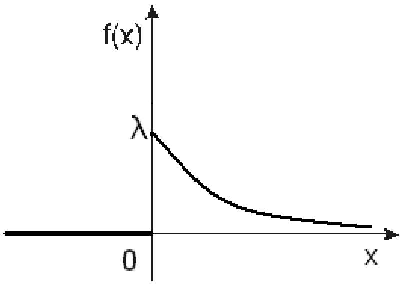

Fig. 2.8

At the exponential distribution expectation M (X) and standard deviation σ_X the same (give proof do yourself) and make the inverse of the parameter λ: $M(X) = \sigma_X = \frac{1}{\lambda}$.

2.4.2.3. Normal distribution.

Among the distributions of continuous random variables the central place occupies the normal law, it is also called Gaussian distribution. It manifests itself in all cases where the random variable X is the result of many different factors. Each factor alone affects the value of X is insignificant and can not specify which to a greater extent than others. Examples of random variables with a normal distribution, are: arrival time deviation of the vehicle from the arrival time specified in the schedule; deviation of the actual dimensions of articles of the nominal size of the data sheet and other products.

***Definition* 2.23.** The random variable *X* is distributed according to a normal distribution (Gaussian law) with parameters a and $\sigma > 0$, if the probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (2.18)$$

To plot the function, carry out its study. To do this, we calculate the derivative:

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

If $x < a$ $f'(x) > 0$, and therefore, the interval $(-\infty, a)$ function increases, and when $x > a$ $f'(x) < 0$, – function decreases. At the point $x = a$ – function has a maximum equal $\frac{1}{\sigma\sqrt{2\pi}}$.

Graph of the function shown in Fig. 2.9 and is called the ***Gaussian curve***.

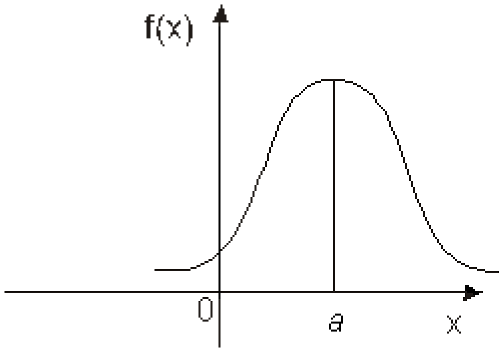

Fig. 2.9

Note some properties Gaussian curve:

1) The distribution curve is symmetrical about the ordinate passing through the point a;

2) When $|x| \to \infty$the branch of the curve asymptotically approach to the axis Ox;

3) Changing a parameter σ when a = const leads to a narrowing of the distribution curve, if $0 < \sigma < 1$ and if broadening $\sigma > 1$.

We calculate the mean and variance of the normal distribution:

$$M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Making the change of variables $\frac{x-a}{\sigma\sqrt{2}} = t$, we have:

$$M(X) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} (a + \sigma\sqrt{2}\, t) \cdot e^{-t^2} dt = \frac{a}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt + \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2} dt.$$

The second integral is equal to zero, the integral of an odd function symmetric in the range, the first integral is an integral of Poisson known:

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = 2\int_{0}^{+\infty} e^{-t^2} dt = \sqrt{\pi}.$$

therefore

$$M(X) = \frac{a}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{a}{\sqrt{\pi}} \sqrt{\pi} = a.$$

Thus, the parameter *a* is the mathematical expectation of a random variable X, having a normal distribution.

Similarly it can be shown that *D* (*X*) = σ^2 Consequently, parameter σ for the normal distribution of the random variable *X* is the standard deviation: σ = $\sigma_X \;\; = \sqrt{D(X).}$

Importance in applications is a special case of the normal distribution density parameters $a=0$ and $\sigma=1$ called ***standard (normalized) normal distribution:***

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{2.19}$$

The function (2.19) - even, i.e., $f_0(-x) = f_0(x).$ The curve of the standard normal distribution is symmetric with respect to the y-coordinate axis and has a maximum of 1 / $\sqrt{2\pi}$.

2.4.3. The probability of hitting a random value having a normal distribution on the predetermined portion.

We have found that if the random variable *X* is defined by the probability density *f (x*), the probability *P*(x_1< *X* < x_2) falling values of the random variable *X* in the interval (x_1,x_2) is calculated by the formula

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

Let the random variable *X* has a normal distribution (2.18). Then the probability that *X* takes the value belonging to the interval (x_1,x_2),

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Making the change of variable $\frac{x-a}{\sigma\sqrt{2}} = t$, we get:

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{\pi}} \int_{\frac{x_1-a}{\sigma\sqrt{2}}}^{\frac{x_2-a}{\sigma\sqrt{2}}} e^{-t^2} dt. \tag{2.20}$$

Since the integral $\int e^{-t^2} dt$ can not be expressed by elementary functions, then for calculating the integral (2.20) are tables of values of the special function, which is called the ***Laplace function*** or ***probability integral***, to which we now turn consideration

2.4.3.1. Laplace function.

Laplace function, or probability integral is of the form:

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

Laplace function has the following properties:

1) Laplace function defined on the whole set of real numbers, and $-\infty < x < +\infty$ и $\overline{\Phi}$ (0) = 0;
2) Laplace function $\overline{\Phi}$ (x) is an odd function, i.e., $\overline{\Phi}$ (− *x*) = − $\overline{\Phi}$ (*x*);
3) $\overline{\Phi}$ (+ ∞) = 1. In fact, the $\overline{\Phi}$ (+ ∞) = $\frac{2}{\sqrt{\pi}}\int_0^{+\infty} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2}$ = 1,

respectively $\overline{\Phi}$ (− ∞) = − 1.

Laplace function is shown in Fig. 2.10.

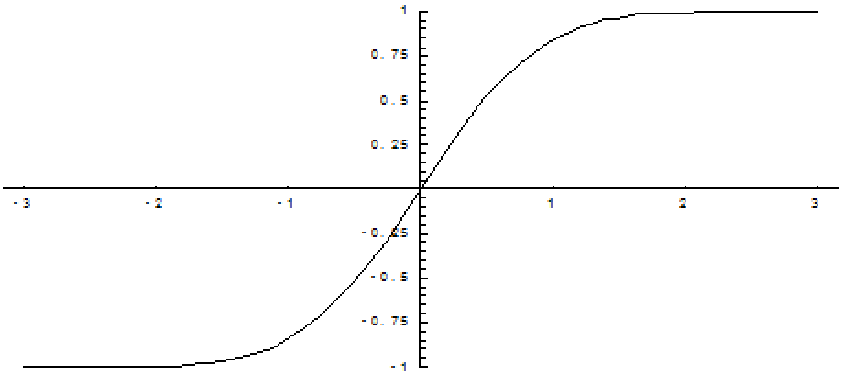

Fig. 2.10

Table Laplace function $\overline{\Phi}(x)$ values shown in Appendix 1.

Now using the Laplace function of the probability (2.20) of the random variable *X* entering a predetermined interval (x_1,x_2),we obtain

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{\pi}} \int_{\frac{x_1-a}{\sigma\sqrt{2}}}^{\frac{x_2-a}{\sigma\sqrt{2}}} e^{-t^2} dt = \frac{1}{2}\left[\frac{2}{\sqrt{\pi}} \int_0^{\frac{x_2-a}{\sigma\sqrt{2}}} e^{-t^2} dt - \frac{2}{\sqrt{\pi}} \int_0^{\frac{x_1-a}{\sigma\sqrt{2}}} e^{-t^2} dt\right] =$$

$$= \frac{1}{2}\left[\overline{\Phi}\left(\frac{x_2-a}{\sigma\sqrt{2}}\right) - \overline{\Phi}\left(\frac{x_1-a}{\sigma\sqrt{2}}\right)\right]. \qquad (2.21)$$

In addition to the Laplace function has used the *normalized Laplace function* (often it is this function called Laplace function), which is associated with the Laplace function of the ratio

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

Fig. 2.11 shows a graph of normalized Laplace function.

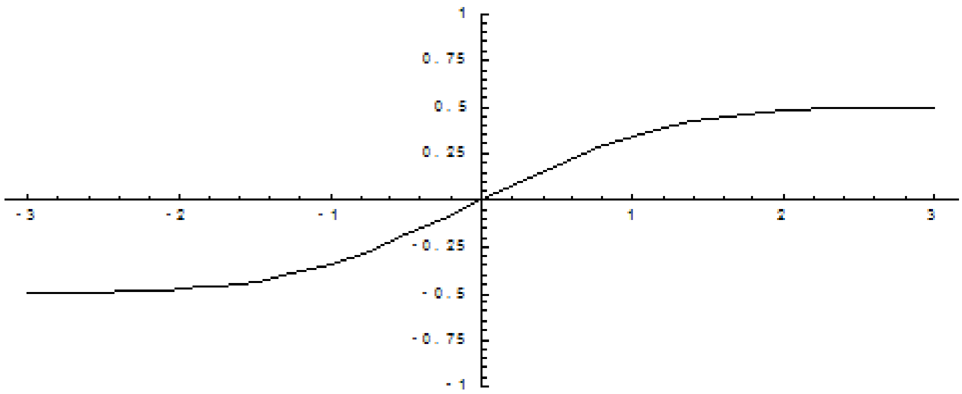

Fig. 2.11

A distinctive feature of the normalized Laplace function is as follows. If the $\overline{\Phi}(+\infty) = 1$, то $\Phi(+\infty) = 0{,}5$, respectively $\overline{\Phi}(-\infty) = -1$, a $\Phi(-\infty) = -0{,}5$.

Table values normalized Laplace function $\Phi(x)$ is shown in Appendix 2.

When using a normalized function of the Laplace equation (2.21) for the probability of a random variable *X* entering a predetermined interval (x_1,x_2), takes the form

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}}\int_{\frac{x_1-a}{\sigma}}^{\frac{x_2-a}{\sigma}} e^{-\frac{t^2}{2}}dt = \left[\frac{1}{\sqrt{2\pi}}\int_0^{\frac{x_2-a}{\sigma}} e^{-\frac{t^2}{2}}dt - \frac{1}{\sqrt{2\pi}}\int_0^{\frac{x_1-a}{\sigma}} e^{-\frac{t^2}{2}}dt\right] =$$

$$= \left[\Phi\left(\frac{x_2-a}{\sigma}\right) - \Phi\left(\frac{x_1-a}{\sigma}\right)\right]. \qquad (2.22)$$

Here we use the change of variable $\frac{x-a}{\sigma} = t$.

Note that for the unlimited (infinite) interval $(-\infty, +\infty)$ expressions (2.21) and (2.22) give the probability equal to unity:

$$P(-\infty < X < +\infty) = \frac{1}{2}[\overline{\Phi}(+\infty) - \overline{\Phi}(-\infty)] = [\Phi(+\infty) - \Phi(-\infty)] = 1.$$

For the standard normal distribution $f_0(x)$ (2.19) formula (2.21) and (2.22) respectively take the form:

$$P(x_1 < X < x_2) = \frac{1}{2}\left[\overline{\Phi}\left(\frac{x_2}{\sqrt{2}}\right) - \overline{\Phi}\left(\frac{x_2}{\sqrt{2}}\right)\right];$$

$$P(x_1 < X < x_2) = [\Phi(x_2) - \Phi(x_1)].$$

2.4.3.2. Three sigma rule.

When considering the normal distribution stands out an important special case when it is necessary to determine the probability $P(M(X) - \varepsilon < X < M(X) + \varepsilon$ enter the random variable X having a normal distribution in the neighborhood of its mean value (expectation) $M(X)$ with standard deviation σ_X. Applying this case formula (2.21) with the proviso that $M(X) = a$, and $\sigma_X = \sigma$, get

$$P(M(X) - \varepsilon < X < M(X) + \varepsilon) = \frac{1}{2}\left[\Phi\left(\frac{a+\varepsilon-a}{\sigma\sqrt{2}}\right) - \Phi\left(\frac{a-\varepsilon-a}{\sigma\sqrt{2}}\right)\right] = \frac{1}{2}\left[\Phi\left(\frac{\varepsilon}{\sigma\sqrt{2}}\right) - \right.$$
$$\left.-\Phi\left(\frac{-\varepsilon}{\sigma\sqrt{2}}\right)\right] = \Phi\left(\frac{\varepsilon}{\sigma\sqrt{2}}\right). \qquad (2.23)$$

We now determine the probability $P(|X - M(X)| < 3\sigma_X)$ deviation values of the random variable X is a normal distribution from its expectation $M(X) = a$ to a value less than ε three times the standard deviation $\varepsilon < 3\sigma = 3\sigma_X$. For calculating the table using the Laplace function, we obtain:

$$P(|X - M(X)| < 3\sigma_X) = \Phi\left(\frac{3}{\sqrt{2}}\right) \approx 0{,}9973.$$

Thus, the probability that the random variable deviates from its expectation by an amount greater than three times the standard deviation is practically zero. This is the so-called *rule of three sigma*.

In practice, it is believed that if a random variable is performed usually three sigma, then this random variable has a normal distribution.

Consider the examples.

*Example*2.11. Error locator in measuring the distance to the target subject to the normal law. The expectation of the errors is equal to 5 *m*, and the standard deviation is 10 *m*. Find the probability that the measured distance will deviate from the true not more than 20 *m*.

Decision. Solution of the problem reduces to determining the probability of a random variable X contact (error locator) in the interval (- 20.20) with the proviso that $a = 5$, $\sigma = 10$. According to the formula (2.21), we have:

$$P(-20 < X < 20) = \frac{1}{2}\left[\Phi\left(\frac{20-5}{10\sqrt{2}}\right) - \Phi\left(\frac{-20-5}{10\sqrt{2}}\right)\right] = \frac{1}{2}[\Phi(1{,}06) + \Phi(1{,}77)] \approx 0{,}875.$$

Example 2.12 Diameter size bushings, manufactured by the company may assume a normally distributed random variable with expectation a = 2.5 *cm* and the dispersion $D[X] = 10^{-4}\ cm^2$. In what limits can be guaranteed sleeve diameter size, if the probability of practical reliability adopted three sigma rule.

Decision. Denote the amount by which canεSleeve diameter size deviate from the mathematical expectations with probability three sigma 0.997. Then, according to equation (2.23), we have

$$P(|X - M(X)| < \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma\sqrt{2}}\right) \approx 0{,}997.$$

Now, using the table of values of the Laplace function, we obtain

$\frac{\varepsilon}{\sigma\sqrt{2}} = 2{,}1$, where (since $\sigma = 0{,}01$) $\varepsilon = 0{,}01 \cdot 2{,}1 \cdot \sqrt{2} \approx 0{,}03$.

Thus, the diameter of the sleeve size with probability 0.997 belongs to the interval (2.47, 2.53).

2.5. The moments of the random variable

A generalization of the basic numerical random variable characteristics is the concept of the moments of the random variable. In probability theory distinguish the moments of two types: primary and central.

2.5.1. The initial moments of the random variable.

***Definition*. 2.24. The starting point μ_k degree *k* (*k*-th order)** the random variable *X* is called the expectation *k*-th degree of the random variable *X*, i.e.

$$m_k = M(X^k),\ k = 1,2,\ldots,n.$$

Consequently, for a discrete random variable initial time is expressed by the sum of

$$m_k = \sum_{i=1}^{n} x_i^k p_i,$$

and continuous – integral $m_k = \int_{-\infty}^{+\infty} x^k f(x)dx$.

It is the first order moment of the initial moments of the random variable of particular importance, which is nothing other than the mathematical expectation of a random variable.

Initial moments of higher orders are mainly used to calculate central moments.

2.5.2. The central moments of the random variable.

Definition. 2.25. The central point μ_k of degree *k* (*k*-th order) is the mathematical expectation of the *k*-th degree of the variance of the random variable *X* from the mean value (*X* - *M* (*X*)), i.e.

$$\mu_k = M((X - M(X))^k), k = 1,2,\ldots,n.$$

For a discrete random variable central point is expressed by the sum of

$$\mu_k = \sum_{i=1}^{n} (x_i - M(X))^k p_i,$$

and continuous – integral

$$\mu_k = \int_{-\infty}^{+\infty} (x - M(X))^k f(x)dx.$$

Central moments can be expressed through the initial moments always. For example:

$\mu_1 = M\ (X - M(X)) = M\ (X - m_1) = 0$. The central point of the first order is always zero.

$\mu_2 = M(X - M(X))^2 = M(X^2 - 2XM(X) + M^2(X) = M(X^2) - M(2XM(X)) + M(M^2(X)) = M(X^2) - 2M(X)M(X) + M^2(X) = \mathrm{M}(X^2) - M^2(X) = m_2 - (m_1)^2 = D(X)$. The central point of the second order is not simply the dispersion random variable.

The central point of degree k can be converted to an expression through the initial moments, using the binomial theorem.

We write the formula for the 3rd and 4th central moments which are widely used for the description of a random variable:

$\mu_3 = m_3 - 3m_1m_2 + 2(m_1)^2$;

$\mu_4 = m_4 - 4m_1m_3 + 6m_1(m_2)^2 - 3(m_1)^4$.

The third central moment μ_3 is characteristic allocation asymmetry. If the random variable X is distributed symmetrically with respect to its mathematical expectation, the third central moment $\mu_3 = 0$.

Since the third central moment is the random variable dimension of the cube, it is usually considered dimensionless quantity a_X – ratio μ_3 to the standard deviation in the third degree, called the asymmetry coefficient.

***Definition*. 2.26. Asymmetry coefficient** curve distribution of the random variable X with respect to their expectation is a quantity defined by the formula

$$a_X = \frac{\mu_3}{\sigma_X^3}.$$

If $a_X < 0$ the curve of distribution with respect to its mathematical expectation is left-sided asymmetry, if $a_X > 0$ – right asymmetry.

Fourth central moment μ_4 is used to characterize the peaked or flat-topped random variable distribution. These distribution characteristics are described by the so-called excess.

Definition*. 2.27. Kurtosis of the random variable *X is the value

$$E_X = \frac{\mu_4}{\sigma_X^4} - 3.$$

Number 3 is subtracted from the ratio $\frac{\mu_4}{\sigma_X^4}$ because for the most common normal distribution $\frac{\mu_4}{\sigma_X^4} = 3$, and therefore, the normal distribution law $E_X = 0$.

Thus, the normal distribution curve for which the kurtosis is equal to zero, as if taken as the standard against which to compare the other distribution.

Curves more peaked than normal have a positive kurtosis; curves are more flat – topped - a negative kurtosis (Figure 2.12.).

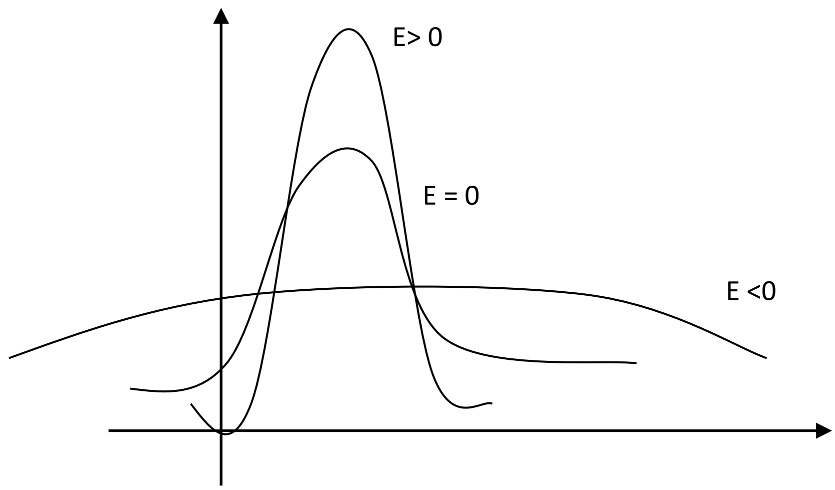

Fig. 2.12

CHAPTER 3

Systems of random variables

In the previous section we considered random variables and acquainted with different characteristics of the random variable.

In the study of random phenomena, depending on their complexity, we have to use two, three or a greater number of random variables. For example, the point of contact of the projectile is determined not one, but two random variables: the abscissa and ordinate. Random satellite deviation to bring the orbit is determined by a complex of three random variables - three satellite coordinates.

With different dimensions, we often have to deal with two or more random variables.

Simultaneous consideration of two or more random variables leads to random values system.

3.1. The concept of random variables

We agree to a system of several random variables $X_1, X_2,…,X_n$ is denoted by $(X_1, X_2,…,X_n)$. In the study of random variables is not enough to study in the random variables individually, make up the system, you must take into consideration the connection or relationship between these values. Here there are new and different from those considered earlier task.

When considering random variables is convenient to use a geometric interpretation of the system. For example, a system of two random variables *(X,* *Y)* can be regarded as a random point on the *xOy* plane with coordinates *X* and *Y* or as a random vector in a plane with random components *X* and *Y*. The system of three random variables *(X, Y, Z)* It can be viewed as a random point in three-dimensional space, or as a random vector in space. Similarly, system *n* random variables (X_1, X_2,…,X_n) can be considered as a random point in n-dimensional space, or as a random vector in n-dimensional vector space R^n .

Since the system of n random variables can be interpreted as a system of *n* random vectors in a vector space R^n, the theory of random variables systems can be seen as a theory of random vectors in linear space.

Depending on the type of random values constituting the system can be a system of discrete and continuous random variables, as well as hybrid systems, which include any type of random variables.

In the study of random variables systems restrict ourselves to a detailed study of the system of two random variables, because all the provisions relating to the system of two random variables, can be easily extended to a system of three, four or more random variables.

3.2. Law of distribution system of two random variables

In the study of a random variable, we became acquainted with the law of its distribution and considered its various forms. A similar role is played by the law of distribution of random variables.

***Definition* 3.1. The law of random variables distribution** is a relation, establishing the relationship between the range of possible values of the system of random variables and the probabilities of occurrence of system in these areas.

Just as for the one of the random variable, the distribution law of random variables can be set in various forms. Consider the main ones for the system of two random variables.

3.2.1. The distribution function of a system of two random variables.

***Definition*3.2. The distribution function of a system of two random variables**is a function of two variables *F (x, y),* which is equal to the probability of joint perform two inequalities *X <x* and *Y <y*,

$$F(x,y) = P(X < x, Y < y). \tag{3.1}$$

The geometric system function of two random values represents the probability distribution of the random contact point *(X, Y)* in the left (shaded) endless lower quadrant of the plane (Fig. 3.1) with vertex at point (*x, y).*

This geometric interpretation of a system of two random variables distribution function allows visually illustrate the following properties.

Property 1. If one of the arguments tends to plus infinity, then the distribution function of the system is committed to the distribution function of a random variable, according to another argument, ie,

$$\lim_{x\to+\infty} F(x,y) = F_1(y); \lim_{y\to+\infty} F(x,y) = F_2(x).$$

Symbolically, it is written as:

$$F(+\infty, y) = F_1(y); F(x, +\infty) = F_2(x).$$

This property of the distribution function is easy to see clearly, pushing one of the quadrants boundaries $+\infty$ (Figure 3.1.); wherein the quadrant is converted into half-plane. The probability of getting a random point in a half-plane is a function of the distribution of one of the variables in the system.

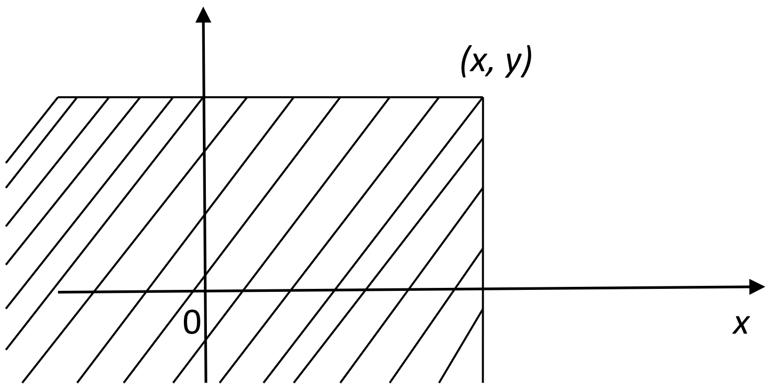

Fig. 3.1

Property 2. If both arguments seek to plus infinity, the system of distribution function tends to unity;

$$\lim_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1, \qquad \text{or} \;\; F(+\infty, +\infty) = 1.$$

Indeed, when $x \to +\infty$ and $y \to +\infty$ the quadrant and with the vertex (*x,y)* becomes full coordinate plane *xOy*, accidental ingress point into which a certain event. $x \to +\infty y \to +\infty$

Property 3. When one or both aspiration arguments to minus infinity distribution function tends toward zero, i.e.

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

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

Indeed, removing a particular quadrant boundary (or both bounds) to minus infinity, we see that the probability of a random point in the quadrant in the limit of zero.

Property 4. The distribution function is a decreasing function in each variable, ie,

$$F(x_2, y) \geq F(x_1, y), \qquad if \;\; x_2 > x_1;$$
$$F(x, y_2) \geq F(x, y_1), \qquad if \;\; y_2 > y_1.$$

Indeed, increasing the x (moving the border right quadrant) or increasing in (shifting the border up), we obviously can not reduce the likelihood of getting a random point in a quadrant.

Property 5. The probability of hitting a random point *(X, Y)* in an arbitrary rectangle (Fig. 3.2) with sides parallel to the coordinate axes calculated by the formula

$$P(a \leq X < b, c \leq Y < d) = F(b, d) - F(a, d) - F(b, c) + F(a, c). \quad (3.2)$$

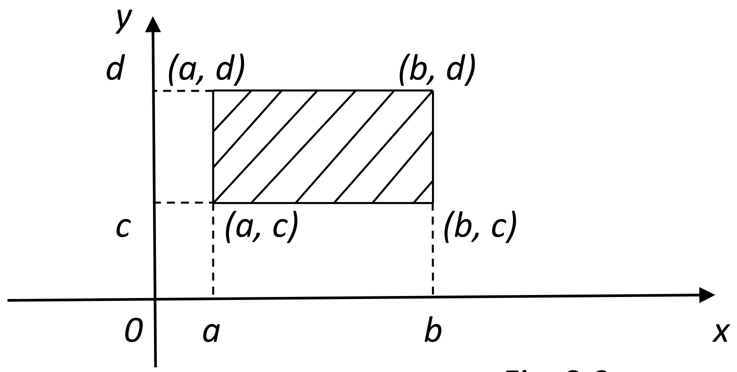

Fig. 3.2

The distribution function is a universal characteristic of random variables. It can be used to describe the systems of both discrete and continuous random variables. However, the basic system of practical significance are continuous random variables whose distribution is not characterized by the distribution function and the distribution density.

3.2.2. The density of the distribution system of two random variables.

The density distribution characteristic of the system is to be exhaustive continuous random variables with which the calculation of the probabilities in the various areas is made easier, and the description of the distribution system becomes more visible.

We define the density of the system of two random variables is similar to the way we have defined the density distribution for a random variable.

Suppose that there is a system of two continuous random variables *(X, Y).* Consider the probability of hitting a random point *(X, Y)* in the elementary rectangle with sides $\Delta x \;\; and \;\; \Delta y$ of and adjacent to the point with coordinates *(x, y*) (Fig. 3.3).

Using formula (3.2), we obtain:

$$P(x \leq X < x + \Delta x, y \leq Y < y + \Delta y) =$$
$$= F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y).$$

Obtaining the probability of the rectangle on the area and take the limit as $\Delta x \to 0$ and $\Delta y \to 0$:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x \le X < x + \Delta x, y \le Y < y + \Delta y)}{\Delta x \cdot \Delta y} =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x \cdot \Delta y}. \qquad (3.3)$$

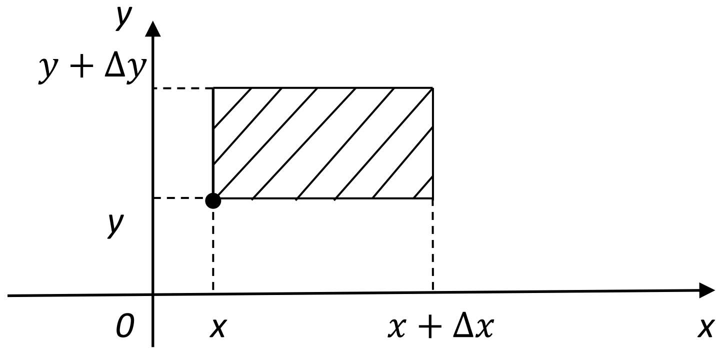

Fig. 3.3

Assume that the distribution function $F(x, y)$is not only continuous, but also twice differentiable, then the right side of the formula (3-3) represents a second mixed partial derivative functions $F(x, y)$. We denote this derivative via $f(x, y)$:

$$f(x, y) = F''_{xy}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \qquad (3.4)$$

Definition **3.3.** The function $f(x, y)$given by expression (3.4) is the **density distribution of continuous system of two random variables *(X, Y).***

Thus, the density distribution of a system of two random variables is the limit of the ratio of the probability that a random point *(X, Y)* in the elementary rectangle (Figure 3.3.) To the area of the rectangle when both its size tends to zero; it can be calculated as the second partial derivative of the mixed system of the distribution function.

Geometrically function $f(x, y)$ can be represented by a surface, which is called the distribution surface.

Considering the density distribution for one of the random variable X, we have introduced the concept of "probability element", which expresses the probability of hitting a random variable X elemental portion. Similarly introduced the concept of "probability element" and a system of two random variables. Element system probability of two random variables gives the probability of hitting a random point in the elementary rectangle with sides adjacent to the point (x, y). Geometrically probability element is parallelepiped elementary volume, based on the elementary rectangle with sides and height.$f(x) f(x) dx dx f(x,y) dx dy dx$ и $dy dx$ и $dy f(x,y)$

Thus, knowing the density distribution can determine the probability of hitting a random point (X, Y) to an arbitrary region D. This probability can be obtained by summing the probabilities of elements across the area D and the limit process when the largest rectangle with sides and shrinks to a point: $f(x,y)\Delta x = dx \Delta y = dy$

$$P\big((X,Y) \subset D\big) = \iint f(x,y) dx dy. \qquad (3,5)$$

Here, the double integral is taken over the domain D.

Geometrically, the probability of hitting the region D of the cylindrical body is represented by the volume bounded by the surface distribution and resting on this area.

Using formula (3.5), we can express the distribution system through a distribution density function. The distribution function is the probability of falling into the quadrant bounded by the abscissa and ordinate$F(x,y) f(x,y) F(x,y) - \infty, x - \infty, y$, so

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y) dx dy. \qquad (3.6)$$

Consider the properties of the density distribution of a system of two random variables.

Property 1. The distribution density is a function of a non-negative: $f(x,y) \geq 0$.

Indeed, the density distribution is the limit ratio of the probability that a random point in a rectangle with the parties Δx and Δy to the area of the rectangle (if $\Delta x \to 0$ and $\Delta y \to 0$) both of these values − non-negative. Consequently, the limit of their relationship can not be negative, i.e. $f(x,y) \geq 0$.

Property 2. Double improper integral with infinite limits on the density distribution system is equal to one:

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

Evidence. On the basis of formula (3.6) and the properties of the distribution function $F(+\infty, +\infty) = 1$ we have

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

Geometrically, this property means that the volume of the body bounded by the distribution surface and the *xOy* plane, is equal to unity.

Example 3.1. System distribution density of two random variables (X, Y) is given by the expression

$$f(x,y)=\frac{a}{1+x^2+x^2y^2+y^2}.$$

Find *a*. To determine the distribution function $F(x,y)$, and to find the probability of hitting a random point in the rectangle (Fig. 3,4) with vertices *O*(0,0), *A*(0,1), $B(\sqrt{3},1)$, $C(\sqrt{3},0)$.

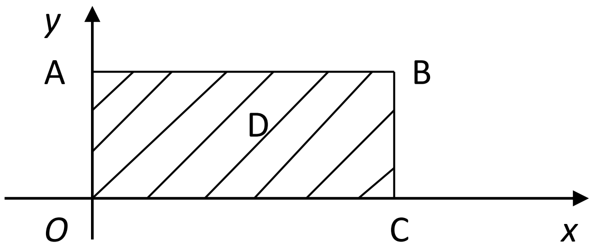

Fig. 3.4

Decision. Using the property 2 density distribution, we obtain the value of *a* constant:

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

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{a}{1+x^2+x^2y^2+y^2}dxdy=$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{a}{(1+x^2)(1+y^2)}dxdy=$$

$$=a\int_{-\infty}^{+\infty}\frac{dx}{(1+x^2)}\int_{-\infty}^{+\infty}\frac{dy}{(1+y^2)}=a\cdot arctgx|_{-\infty}^{+\infty}\cdot arctgy|_{-\infty}^{+\infty}=a\pi^2$$

$$=1\,.$$

Hence, $a=\frac{1}{\pi^2}$.

Distribution function $F(x,y)$ is defined by the formula (3.6):

$$F(x,y)=\frac{1}{\pi^2}\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{dxdy}{(1+x^2)(1+y^2)}=\left(\frac{1}{\pi}arctgx+\frac{1}{2}\right)\cdot\left(\frac{1}{\pi}arctgy+\frac{1}{2}\right).$$

The probability of hitting a random point (*X, Y*) in a given rectangle (Fig. 3.4) according to formula (3.5) is

$$P\big((X,Y)\subset D\big)=\iint\frac{dxdy}{(1+x^2)(1+y^2)}dxdy=$$

$$=\frac{1}{\pi^2}\int_0^{\sqrt{3}}\frac{dx}{(1+x^2)}\int_0^1\frac{dy}{(1+y^2)}=\frac{1}{\pi^2}\cdot arctgx|_0^{\sqrt{3}}\cdot arctgy|_0^1=$$

$$=\frac{1}{\pi^2}\cdot\frac{\pi}{3}\cdot\frac{\pi}{4}=\frac{1}{12}.$$

Example 3.2. System distribution density of two random variables (*X, Y*) has a uniform distribution in the region *D*, i.e. it is given by the expression

$$f(x,y)=\begin{cases} C \text{ inside } D, \\ 0 \text{ outside } D. \end{cases}$$

Find the probability of hitting a random point (*X, Y*) in a certain part ω of the region *D*.

Decision. By property 2 density distribution, we have:

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

Where S_D – field area *D*. Hence $C=\frac{1}{S_D}$.

Now according to (3.5) we have

$$P\big((X,Y)\subset\omega\big)=\iint f(x,y)dxdy=\frac{1}{S_D}\iint dxdy=\frac{S_\omega}{S_D}.$$

Here S_ω − the field area ω, and a double integral is taken over the area ω contained in *D*.

Thus, if the system density distribution of two random variables (*X, Y*) is uniformly distributed in the area *D*, the probability accidental contact point (*X, Y*) in some part ω of region *D* equal to the ratio of their areas.

In concluding this section we note that if we know the density of the system of two random variables, it is possible to determine the distribution and density of each of the random variables in the system.

Thus, according to property 1 of section 3.2.1, we have:

$$F_1(x)=F(x,+\infty);\ F_2(y)=F(+\infty,y).$$

Therefore, using the formula (3.6), the relationship between $f(x,y)$and $F(x,y)$ can be represented $F_1(x)$ and $F_2(y)$ in the form of:

$$\left.\begin{aligned} F_1(x) = F(x,+\infty) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(x,y)dxdy, \\ F_2(y) = F(+\infty,y) = \int_{-\infty}^{y}\int_{-\infty}^{+\infty} f(x,y)dxdy. \end{aligned}\right\}$$

Differentiating the first equality with respect to *x*, and the second - in *y*, we obtain the expression for the density distribution of individual random variables in the system:

$$\left.\begin{aligned} f_1(x) = F_1'(x) = \int_{-\infty}^{+\infty} f(x,y)dy, \\ f_2(y) = F_2'(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx. \end{aligned}\right\} \qquad (3.7)$$

Thus, in order to obtain a density distribution of one of the random variables in the system, it is necessary to integrate the density distribution in the infinite within the field of the argument corresponding to each random variable.

3.3. Dependent and independent random variables

The notion of dependence or independence of random variables is one of the most important concepts of probability theory.

***Definition* 3.4.**The random variables *X* and *Y* are called independent if the law of distribution of each of them depends on what value to accept each value. Otherwise, the random variables *X* and *Y* are called the dependent.

Simply enter a sign of independence of random variables, which we state in the following theorem, which we assume without proof.

***Theorem* 3.1.** *In order to continuous random variables X and Y are independent, it is necessary and sufficient that the density of the distribution system (X, Y) is equal to the product of the density distribution of the individual quantities appearing in*: $f(x,y) = f_1(x)f_2(y)$.

Result. If the distribution density $f(x,y)$ expressed as a product of two factors, the first of which contains only *x*, and the second − only *y*, then the random variables *X* and *Y* are independent.

*Example*3.3. system random variables (*X, Y*) is the probability density:

$$f(x,y) = \frac{1}{\pi^2(1+x^2+y^2+x^2y^2)}.$$

Required to determine dependent or independent random variables *X* and *Y*.

Decision. The answer follows from the possibility of expanding the probability density $f(x,y)$ distribution of factoring:

$$f(x,y)=\frac{1}{\pi^2(1+x^2)(1+y^2)}.$$

Hence it follows that the random variables *X* and *Y* are independent, since $f(x,y)=f_1(x)f_2(y)$ each of them is subject to a so-called Cauchy's law:

$$f_1(x)=\frac{1}{\pi(1+x^2)},\qquad f_2(y)=\frac{1}{\pi(1+y^2)}.$$

Example 3.4. System random variables (*X, Y*) is uniformly distributed inside a circle of radius r:

$$f(x,y)=\begin{cases}\dfrac{1}{\pi r^2} & \text{at } x^2+y^2\le r^2,\\ 0 & \text{at } x^2+y^2>r^2.\end{cases}\qquad(3.8)$$

Find density distributions of the random variables *X* and *Y*, and set, dependent or independent of these random variables.

Decision. Substituting (3.8) into the formula (3.7), we find the density distribution of the random variables *X* and *Y*:

$$f_1(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\displaystyle\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{dy}{\pi r^2}=\frac{2}{\pi r^2}\sqrt{r^2-x^2} & \text{при } |x|\le r,\\ 0 & \text{at } |x|>r.\end{cases}$$

Similarly, computing, we get:

$$f_2(y)=\begin{cases}\displaystyle\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}}\frac{dx}{\pi r^2}=\frac{2}{\pi r^2}\sqrt{r^2-y^2} & \text{at } |y|\le r,\\ 0 & \text{at } |y|>r.\end{cases}$$

From these expressions it follows that the product of random variables distribution densities *X* and *Y* is not equal to their joint distribution density $f(x,y)\neq f_1(x)f_2(y)$, which means that the random variables *X* and *Y* are dependent.

3.4. Numerical characteristics of a system of two random variables. Correlation time. Correlation coefficient

Random variables distribution laws are exhaustive probability characteristics of it. However, very often such an exhaustive description can not be applied. Sometimes the limitations of the experimental material does not allow the law to build distribution system.

In these cases, very great application found numerical characteristics of the distribution law of random variables, which to some extent may also give an idea about the nature of law and the system of distribution.

The basis of the receipt of the numerical characteristics of the system of random variables on the concept of moments. As for one of the random variable, here distinguish the initial and central moments.

Definition* 3.5. *The starting point a_{ks} of order k + s system (X, Y) it is a mathematical expectation of the product *k*-th power of the random variable *X* in *s*-th power of the random variable *Y*:

$$a_{ks} = M[X^k Y^s]. \tag{3.9}$$

Formulas for calculating the initial moments a_{ks} are written as follows:

for a system of discrete random variables

$$a_{ks} = \sum_i \sum_j x_i^k y_j^s P_{ij}, \tag{3.10}$$

where $P_{ij} = P\left(X = x_i,\ Y = y_j\right)$ – the probability that the system (*X, Y*) takes the values $\left(x_i, y_j\right)$, and the summation is over all possible values of random variables XY;

for a system of continuous random variables

$$a_{ks} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^k y^s f(x,y) dxdy, \tag{3.11}$$

where $f(x,y)$ – density of the distribution system.

In practice, the most used are the initial moments of the first order:

$$\left.\begin{aligned} a_{10} &= M[X^1 Y^0] = M[X] = m_x, \\ a_{01} &= M[X^0 Y^1] = M[Y] = m_{y,} \end{aligned}\right\}$$

which are the mathematical expectations of random variables *X* and *Y*, are entering the system. These expectations determine coordinates of a point, called the dispersion system centered on the plane.

We turn now to the central points.

Definition* 3.6. *The central point μ_{ks} of order k + s system (X, Y) is a mathematical expectation of the product *k*-th and *s*-th degree corresponding centered values (deviations of random variables X and Y from their mean values):

$$\mu_{ks} = M[(X - M[X])^k (Y - M[Y])^s]. \tag{3.12}$$

Formulas to calculate central moments μ_{ks} are written as follows:

for a system of discrete random variables

$$\mu_{ks} = \sum_i \sum_j (x_i - M[X])^k \left(y_j - M[Y]\right)^s P_{ij}, \qquad (3.13)$$

for a system of continuous random variables

$$\mu_{ks} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - M[X])^k (y - M[Y])^s f(x,y) dxdy. \qquad (3.14)$$

In practice, the most widely used are of second order central moments. Two of them are already known to us the variance of the random variables *X* and *Y*:

$$D_X = \mu_{20} = M[(X - M[X])^2 (Y - M[Y])^0] = M[(X - M[X])^2],$$
$$D_Y = \mu_{02} = M[(X - M[X])^0 (Y - M[Y])^2] = M[(Y - M[Y])^2],$$

which characterize the dispersion random point in the direction of the axes *Ox* and *Oy*.

Definition* 3.7.** A special role in the system study of two random variables plays second mixed central point μ_{11}, which is called ***the correlation point or connection point. It is usually indicated k_{xy} .

$$k_{xy} = \mu_{11} = M[(X - M[X])(Y - M[Y])]. \qquad (3.15)$$

Moment connection k_{xy} defined as the expectation of product deviations of two random variables from their mathematical expectations, addition of dispersion values *X* and *Y*, can characterize the mutual influence of these random variables. To assess the degree of this influence is not generally used very moment of connection k_{xy}, and the dimensionless ratio

$$r_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y}, \qquad (3.16)$$

where σ_x and σ_y − the standard deviations of random variables, respectively X, and Y.

Definition* 3.8.** The value r_{xy} given by relation (3.16) is called the ***coefficient of correlation of random variables X and Y.

The correlation moment and correlation coefficient have the following property.

If the random variables X and Y independent, then the time correlation and the correlation coefficient are zero.

Proof for continuous random variables. Let *X* and *Y* − are independent random variables with the distribution density (x,y) . Then, according to Theorem 3.1, we have: $f(x,y) = f_1(x) f_2(y)$ where $f_1(x)$ и $f_2(y)$ − density distribution, respectively, the values of *X* and *Y*.

Hence,

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

$$= \int_{-\infty}^{+\infty}(x - M[X])f_1(x)dx \int_{-\infty}^{+\infty}(y - M[Y])f_2(y)dy\,,$$

i. e. double integral is converted into a product of two integrals each of which is zero, because they represent the mathematical expectation of random variables of deviations from their mathematical expectations. In other words the first order central moments.

So, for independent random variables *X* and *Y* correlation moment $k_{xy} = 0$.

The vanishing points and the correlation of Formula (3.16) correlation coefficient equal to zero.

Similarly we can prove this property for discrete random variables.

The vanishing of correlation coefficient is only necessary but not sufficient condition for independence of the random variables. This means that there may be system dependent random variables, the correlation coefficient is equal to zero. An example of such a system is a system of random variables (*X, Y*), distributed uniformly inside a circle of radius r centered at the origin. In Example 3.4, we have shown that the random variables X and Y system having such distribution are dependent. We now calculate the correlation moment.

Since for a system of random variables (*X, Y*), evenly distributed inside a circle centered at the origin, then $M[X] = 0$, $M[Y] = 0$

$$k_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy.$$

Then we have:

$$k_{xy} = \frac{1}{\pi r^2}\int_{-r}^{r} x\left(\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} ydy\right)dx,$$

Where $f(x,y) = \begin{cases} \frac{1}{\pi r^2} & \text{at } x^2 + y^2 \le r^2, \\ 0 & \text{at } x^2 + y^2 > r^2. \end{cases}$

Here, the inner integral is zero (integrand is odd, the integration limits are different only in sign), so that $k_{xy} = 0$ or, equivalently, the correlation coefficient $r_{xy} = 0$.

Definition **3.9.** Two random variables *X* and *Y* are called ***uncorrelated*** if the correlation coefficient is equal to zero; *X* and *Y* are called ***correlated***, if the correlation coefficient is nonzero.

Thus, if the random variables *X* and *Y* are independent, then they are not correlated, but are uncorrelated random variables cannot generally conclude their independence.

Furthermore moment correlation coefficient and correlation, mutual coupling of two random variables can be described by the *regression lines*. Indeed, while for each value of *X* = *x* value *Y* is a random variable, allowing its dispersion values, but the dependence on Y X often affects a change in its average value (mathematical expectations $M[Y] = m_y$) in the transition from one value *x* to another's. This last relationship and describes the regression curve $y = m_y(x)$.

Similarly, the dependence of *X* from *Y*, which affects the change in the average values (expected m_x) *X* at the transition from one value to another have described the regression curve $x = m_x(y)$.

3.5. Function and distribution density of the system a random number of random variables and their numerical characteristics

In practice, very often it is necessary to consider the system for more than two random variables. These systems are, as we mentioned earlier in this chapter, it is interpreted as random points or random vectors in the space of the corresponding dimension.

Complete characterization system of arbitrary number of random variables is the law of distribution systems that can be expressed by a distribution function or a distribution density.

Definition **3.10.** ***The distribution function of a system of n random variables*** **(*X_1, X_2,…, X_n*)** is the function *n* arguments x_1, x_2,…, x_n, equal sharing of *n* probability inequalities $X_i < x_i$ (*i* = 1, 2,…, *n*), i.e.

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n). \qquad (3.17)$$

This function is a decreasing function of each variable with fixed values of the other variables. If at least one of the variables x_1, x_2,…, x_n tends to minus infinity ($-\infty$), the distribution function $F(x_1, x_2, \dots, x_n)$ tends to zero.

Distinguish from the values of (X_1, X_2,…, X_n) private system (X_1, X_2,…, X_m), while the distribution function of the system is determined by the formula $F_{1,2,\dots,m}(x_1, x_2, \dots, x_m) = F(x_1, x_2, \dots, x_m, +\infty, \dots, +\infty)$. In particular, the distribution function of each of the variables in the system are obtained by putting all the other arguments are in the system distribution function equal to $+\infty$: $F_1(x_1) = F(x_1, +\infty, \dots, +\infty)$.

If all of the variables x_1, x_2,…, x_n aspire to $+\infty$, then $F(x_1, x_2, \dots, x_n)$ tends to unity: $F(+\infty, \dots, +\infty) = 1$.

The distribution function is quite a common characteristic of random variables. Any system of random variables has a distribution function. For a description of the law of distribution system of continuous random variables are commonly used density distribution systems.

Definition* 3.11. *Density distribution $f(x_1, x_2, \dots, x_n)$ for system n random variables (X_1, X_2,…, X_n) is defined as the limit of the ratio of probability of occurrence of the system (X_1, X_2,…, X_n) in a small neighborhood of points $(x_1, x_2, \dots, x_n)$ to the size of this neighborhood its $\Delta x_1 \Delta x_2 \dots \Delta x_n$ unconfined decrease, i. e.

$$f(x_1, x_2, \dots, x_n) = \lim_{\substack{\Delta x_1 \to 0 \\ \Delta x_2 \to 0 \\ \vdots \\ \Delta x_n \to 0}} \frac{P(x < X_1 < x_1 + \Delta x_1, \dots, x_n < X_n < x_n + \Delta x_n)}{\Delta x_1 \Delta x_2 \dots \Delta x_n}.$$

The main properties of the density distribution of random variables (X_1, X_2,…, X_n):

1) The density of the distribution system cannot be negative: $f(x_1, x_2, \dots, x_n) \geq 0$.

2) The probability of hitting a random point with coordinates (X_1, X_2,…, X_n) in n-dimensional domain D is expressed by n-multiple integral of the area

$$P\big((X_1, X_2, \dots, X_n) \subset D\big) = \iint \dots \int f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n .$$

3) Probability Density system (X_1, X_2,…, X_n) is expressed by the function formula of the distribution system

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}.$$

4) The normalization condition system density distribution is as follows:

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1.$$

5) The distribution density of each of the variables in the system are obtained if the density of integrated distribution system in the infinite limits for all other arguments:

$$f_1(x_1) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_2 \dots dx_n.$$

6) For the system ($X_1, X_2, \dots, X_n$) of independent random variables is the product of the density distribution of the density distribution of the individual values included in the system:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n).$$

The converse also holds. If the distribution system density ($X_1, X_2, \dots, X_n$) of independent random variables is the product of the density distribution of selected variables in the system, then the random variables ($X_1, X_2, \dots, X_n$) are independent.

The main numerical characteristics with which the system may be characterized by n random variables ($X_1, X_2, \dots, X_n$) are the following:

The expectations random variables in the system $m_{x_1}, m_{x_2}, \dots, m_{x_n}$, which together define the center of the system of dispersion or the expectation of an n-dimensional random vector.

Dispersions $D_{x_1}, D_{x_2}, \dots, D_{x_n}$ characterizing dissipation random point in the direction of the axes.

Correlation times each pair of *n* random variables $k_{x_i x_j} = M\left[\left(X_i - m_{x_i}\right)\left(X_j - m_{x_j}\right)\right]$ $(i \neq j)$ describing the pairwise correlation of random variables in the system.

Knowing correlation times can find correlation coefficients $r_{x_i x_j} = \frac{k_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$ that characterize the degree of connection between each pair of random variables.

CHAPTER 4 Limit theorems of probability theory

4.1. Preliminaries

We already know that the laws of probability theory studies the inherent mass random phenomena. Like any other science, the theory of probability is used to accurately predict the possible outcome of a particular phenomenon or experiment. However, if the phenomenon is isolated, not widespread, the probability theory can predict the likely outcome is usually only a very wide range. Quite another matter when the phenomenon - mass. Regularities is manifested in a large number of random events occurring in similar conditions. With a sufficiently large number of test characteristics of random events and random variables observed during the test, it is almost non-random. For example, the frequency of the event with a large number of tests becomes stable, the same applies to the average values of the random variables. This fact allows the use of the results of observations of random events to predict the results of future tests.

Group theory that the agreement between the theoretical and experimental characteristics of random variables and random events with a large number of tests on them, as well as relating to limit distribution laws, are collectively called the limit theorems of probability theory.

In this chapter, we look at two types of limit theorems: the law of large numbers and the central limit theorem.

4.2. The law of large numbers: Chebyshev inequality and the theorem. Bernoulli's theorem and Poisson

The law of large numbers, which occupies an important place in the theory of probability, is the link between the theory of probability as a mathematical science and laws of random phenomena in mass observations on them. The law of large numbers plays a very important role in the practical applications of probability theory to natural phenomena and technical processes associated with mass production.

In proving theorems, belonging to the group of the "law of large numbers" are the Chebyshev inequality.

***Chebyshev's inequality*.** The probability that the deviation of the random variable X from its expectation m_x of an absolute value of not less than any positive number ε, is bounded above by the value $\frac{D_X}{\varepsilon^2}$, wherein D_X − the dispersion of the random variable X.

$$P(|X - m_x| \geq \varepsilon) \leq \frac{D_X}{\varepsilon^2}. \qquad (4.1)$$

Chebyshev's inequality may be written in another form, as applied to the opposite event - random deviation from the expectation value is less than ε:

$$P(|X - m_x| < \varepsilon) \geq 1 - \frac{D_X}{\varepsilon^2}. \qquad (4.2)$$

One of the most important forms of the law of large numbers is a theorem of Chebyshev. It establishes a relationship between the average of the observed values of the random variable and its expectation.

***Theorem Chebyshev*.** *With an unlimited increase in the number of independent trials arithmetic average of the observed values of the random variable having a finite variance converges in probability to its expectation.*

Let us explain the meaning of the term "converges in probability." The sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ converges in probability to the value and if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X - a| < \varepsilon) = 1,$$

or in more detail: a sequence of random variables X_1, $X_2, \ldots$ converges in probability to the value a, if for any $\varepsilon > 0$ and $\delta > 0$ there exists $n(\varepsilon, \delta)$, beyond which the inequality

$$P(|X - a| < \varepsilon) > 1 - \delta.$$

Thus Theorem Chebyshev means that if X_1, $X_2, \ldots$ independent and identically distributed random variables with mean m_x and finite variance D_X, for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i-1}^{n} X_i}{n} - m_x\right| < \varepsilon\right) = 1.$$

The proof is omitted.

Chebyshev's theorem may be extended to the more general case, when the characteristics of the observed random quantities vary from experiment to experiment. It turns out that in this case, under certain conditions the arithmetic mean is stable and converges to a certain probability of not a random variable. More precisely, the following theorem of generalized Chebyshev.

Generalized Chebyshev theorem. *With an unlimited increase in the number of independent tests on random variables with limited dispersion, the arithmetic average of the observed values converge in proximity to the arithmetic mean of the expectations of these quantities, i.e.,*

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i-1}^{n} X_i}{n} - \frac{\sum_{i-1}^{n} m_{x_i}}{n}\right| < \varepsilon\right) = 1.$$

The next most important and historically the first form of the law of large numbers is Bernoulli's theorem. It establishes a relationship between the frequency of the event and its probability. The proof given by Bernoulli, was very difficult. A simple proof is given by P.L. Chebyshev – as a simple consequence of his theorem.

Bernoulli's theorem. *With an unlimited increase in the number of independent tests at constant conditions relative frequency of the event in question A converges in probability to its probability p in a separate test.*

If we denote the relative frequency of the event A in n trials through p^*, Bernoulli's theorem can be written as:

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

For the case when the tests take place at the same situation, we have the Poisson theorem.

Poisson theorem*. *With an unlimited increase in the number of independent variables in the test conditions, the relative frequency of the event in question converges in probability A to the arithmetic mean of the probabilities* p_i *of its data at tests.

Proof of the Poisson theorem of generalized Chebyshev's theorem.

4.3. The central limit theorem. De Moivre – Laplace theorem

Discussed in the previous section of the theorem are different forms of the law of large numbers. The law of large numbers of sets of convergence in probability fact some random variables to their permanent characteristics. In this case, none of the forms of the law of large numbers, we are not dealing with the laws of distribution of random variables.

In this section we consider the problem associated with finding the limit of the law sum distribution

$$Y_n = \sum_{i=1}^{n} X_i, \tag{4.3}$$

when the number of terms increases indefinitely. The central limit theorem of probability theory (Lyapunov's theorem) establishes the conditions under which this limit law is normal. Various forms of the central limit theorem differ in terms that overlap the distribution of random variables X_i forming amount of (4.3). We formulate (proof omitted) the simplest form of the central limit theorem, when the random variables $X_1, X_2, \dots, X_n$ are mutually independent and identically distributed.

***Central Limit Theorem*.** *If the random variables $X_1, X_2, \dots, X_n$ are mutually independent and have the same distribution law with mean m and dispersion σ^2, wherein there is a third absolute moment μ_3, when n unlimited increase amount distribution law (4.3) gets arbitrarily close to normal.*

The random variables $X_1, X_2, \dots, X_n$ appearing in the central limit theorem may have arbitrary probability distributions. If it is assumed that the random variables X_i are identically distributed, discrete and take only two possible values 0 or 1, then we arrive at a theorem DeMoivre - Laplace represents simply a special case of the central limit theorem.

***De Moivre* – Laplace *Theorem*.** *If performed n independent trials, in each of which an event A occurs with probability p, then for every interval* (α, β) *of the relation*

$$P\left(\alpha < \frac{Y - np}{\sqrt{npg}} < \beta\right) = \frac{1}{2}\left[\Phi\left(\frac{\beta}{\sqrt{2}}\right) - \Phi\left(\frac{\alpha}{\sqrt{2}}\right)\right],$$

where Y*- the number of occurrences of A in the* n *tests,* $g = 1 - p\Phi(x)$ *– the Laplace function.*

De Moivre - Laplace theorem describes the behavior of the binomial distribution for large values n. This circumstance makes it possible to simplify the calculations associated with the binomial distribution at large n. In fact, the count probability of the random variable Y to reach interval (α, β) the exact formula

$$P(\alpha < Y < \beta) = \sum_{\alpha<k<\beta} C_n^k p^k g^{n-k}$$

associated with large n cumbersome calculations. It is much easier to use the approximate formula

$$P(\alpha < Y < \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - np}{\sqrt{2npg}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{2npg}}\right)\right] \qquad (4.4)$$

in the assumption that the random variable Y is normally distributed with mean $m_Y = np$ and variance $\sigma_Y^2 = npg$.

Example 4.1. Find the probability that the resulting 1,000 tosses a coin number of heads to be concluded in the interval (475, 525).

Decision. In this task, p = 0,5, $n = 1000$. Hence, , $np = 500npg = 250$.

Putting in equation (4.4) $\alpha = 475, \beta = 525$ we obtain:

$$P(475 < Y < 525) \approx \frac{1}{2}\left[\Phi\left(\frac{525 - 500}{\sqrt{500}}\right) - \Phi\left(\frac{475 - 500}{\sqrt{500}}\right)\right] = \Phi\left(\frac{25}{10\sqrt{5}}\right) \approx 0{,}8854.$$

Example 4.2. The plant produces 90% of the products of the first variety, and 10% of the products of the second variety. Randomly selected 1000 products. Find the probability that the number of first-class products will be in the range from 900 to 940.

Decision. Probability of selecting items of Class 0.9, the number of trials $n = 1000$. Hence, $np = 900, npg = 90$.

Using formula (4.4), we obtain:

$$P(900 < Y < 940) \approx \frac{1}{2}\left[\Phi\left(\frac{940 - 900}{\sqrt{180}}\right) - \Phi\left(\frac{900 - 900}{\sqrt{180}}\right)\right] = 0{,}5.$$

CHAPTER 5
MATH STATISTICS
5.1. Methods of statistical data collection

The methods of collecting statistical data play a decisive role in establishing laws for statistical objects and form the basis in planning for data collection experiment.

5.1.1. General and sample.

Suppose you want to examine the totality of similar objects with respect to a qualitative or quantitative features that characterize these objects. For example, if there is a batch of goods, the quality sign may be the name of the product, and quantitative - controlled by the weight of the goods.

Sometimes a census conducted, i.e. each of the examined objects with respect to the aggregate feature that interest. In practice, however, a continuous survey is used relatively rarely. For example, if the set contains a very large number of objects that carry out a census is physically impossible. If the examination of the object associated with its destruction, or require large material costs, then conduct a census almost does not make sense. In such cases, use a so-called sampling method. Of the entire collection of objects accidentally selected a limited number of objects and subjected to their study. In this connection, distinguished general, and the total sample:

Definition*. 5.1. *Parent population is the set of all possible objects mentally given type, on which observations are made in order to obtain specific values of the random variable, or a combination of the results of all conceivable observations taken at constant conditions over one of random variables associated with this type of objects.

Comment: Part of the universe has a finite number of objects. However, if the number is high enough, sometimes in order to simplify calculations assume that the general population is made up of countless objects. Such an assumption is justified by the fact that the increase in the general population (of sufficiently large size) has practically no effect on the sampling data results.

Definition* 5.2. *The total sample or just ***part of the sample*** referred to randomly selected objects from the general population.

Definition* 5.3. *The volume of the sample (selective or general) is the number of the plurality of objects.

Thus, the sampling method is that the population of volume N is taken sample volume n (where $n << N$) and determined characteristics of the sample that is taken as the approximate values of the relevant characteristics of the entire population. In this case, the greater n, the more informed judgment can be made based on a sample of the population properties. Also note that the sample gives the most information about the general population only in the case when the results of surveys that make up the sample are independent.

5.1.2. Sampling methods.

In drawing up the sample can proceed in two ways: after the selected object and on it made the observation, it can be returned or not returned to the general population. In accordance with the above sample is divided into repeated sampling without replacement.

Definition **5.4.** ***Re called*** sample in which (before the next selection) the selected object is returned in the general population.

Definition **5.5.** ***Without repetitions*** called sampling, in which the selected object in the general population is not refundable.

In practice, usually are repetition-free random selection.

To according to the sample can be judged fairly confident about the traits of interest the general population, it is necessary to sample the objects it properly represented. In other words, the sample must be properly proportion the total population. This requirement is briefly stated as follows: the sample should be representative (representative).

The law of large numbers can be argued that the sample is ***representative***, if it is carried out by chance: every object selected sample randomly from the general population, if all objects have the same probability of being sampled.

In practice, there are different methods of selection. In principle, these methods can be divided into two types:

1. The selection does not require dissection of the total population in the part. These include: a) simple random selection of repetition-free; b) simple random repeated sampling.

2. Selection, wherein the general population is divided into parts. These include: a) a typical selection; b) mechanical selection; c) Serial selection.

Definition **5.6.** ***Just random*** called this selection, in which the objects are removed one by one from the entire population and do not return after inspection (nonrepetitive selection) or return (re-selection) in the general population.

Definition **5.7.** ***Typical*** called selection, in which objects are selected from the entire population, and from each of its "typical" part. For example, if the items are made on several machines, the selection is made not from the entire population of components produced by all machines and production of each machine individually.

Typical selection used when the survey characteristics significantly vary in different parts of the typical general population. For example, if products are produced on several machines, some of which are more or less worn, there is a typical selection is appropriate.

Definition* 5.8. *Mechanical called screening, in which the general population "mechanically" is divided into as many groups as objects should be included in the selection, and each group selected a single object. For example, if you want to take away 20% of the manufactured machine parts that are taken every fifth part; if 5% is required to select the items, taken every twentieth item, and so on. d. It should be noted that sometimes a mechanical selection may not provide representative sample.

Definition* 5.9. *Serial called selection, in which objects are selected from the general population is not one, and "series" which be screened. For example, if products are produced by a large group of automatic machines, then be screened products only a few machines. serial
selection used when the survey characteristics vary slightly in different series.

We emphasize that in practice often combined selection used in which the above methods are combined.

5.2. Statistical sampling distribution

Let us assume that we study some general set of objects, and we are interested in a sign of these objects that make up the random variable X, the law of which the distribution is unknown. To this end, a number of independent measurements (sampling) is performed on a random variable X. The primary form of sample processing results, representing a set of values of the random variable *X*, is a statistical (variation) and a number of statistical sampling the distribution.

5.2.1. Statistical (variational) number. Statistical sample allocation.

Results for samples x_i of independent tests obtained in a series of *n* over the random variable *X* represent a sequence recorded in the ascending order of $x_1 < x_2 < \ldots < x_k$ record them in a table:

i	1	2	...	k
x_i	x_1	x_2	...	x_k

Here, $k \leq n$, because the results of the measurements in the sample can be repeated.

Definition* 5.10.** The observed values x_i of the random variable *X* in the sample are considered ***variants and variant sequence recorded in ascending order, – the ***variation (statistical) row***.

Statistical row can be processed in various ways. One way this processing is to construct a statistical distribution of the sample.

Suppose that in a number x_1 of variations observed n_1 value $x_2 - n_2$ of time $x_k - n_k$ - time and $\sum n_i = n$ sample size.

***Definition* 5.11.** Numbers observation n_i values x_i of the random variable *X* in the sample are called *frequencies*, and their ratio to volume of the sample $n_i/n = W_i -$ *relative frequencies*.

Definitionse* 5.12. *Statistical sampling distribution called list option recorded in ascending order and their respective frequencies or relative frequencies.

Usually such conformity issued in the form of a table:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k
W_i	n_1/n	n_2/n	...	n_k/n

Note that, in theory, under the probability distribution understand the correspondence between the possible values of the random variable *X,* quantitative trait mapping some general population and their probabilities and mathematical statistics – correspondence between the observed variations (observed values of the random variable) and their relative frequencies. Obviously, when $n \to \infty$ statistical sampling the distribution approaches to the distribution of the random variable *X*.

Consider now, as in the case of the probability distribution of the random variable *X*, the most general form of presentation of the statistical distribution of the sample, so-called statistical (empirical) sample distribution function.

5.2.2. Statistical (empirical) distribution function.

Suppose it is known statistical sampling the distribution of the random variable *X* for displaying a quantitative indication of the total population. We introduce the notation: n_x – the number of cases in which the observed value of the random variable *X* in the sample is less than *x; n* – number of observations (sample size). Clearly, the relative frequency that the random variable *X* takes in the sample values *x* is less than n_x/n. If *x* varies, that is, generally speaking, and the relative frequency change, i.e., relative frequency n_x/n is a function of *x.*

Definition* 5.13.** Function F^*determining for each $x \in R$ relative frequency that the random variable *X* takes in the sample values are less than *x* is called ***statistical sampling the distribution function.

Thus, by definition $F^*(x) = \frac{n_x}{n}$, where n_x – version number smaller than *x*, *n* - sample size.

Since this function is an empirical (experimental) way, then it is called ***empirical***.

Unlike the empirical distribution function of the sample distribution function *F (x)* of the total population is called the ***theoretical distribution function.*** The difference between the empirical and theoretical functions is that the theoretical function *F (x)* determines the probability of the event *X <x,* and the empirical function $F^*(x)$determines the relative frequency of the event. This implies that the statistical distribution function for any random sampling values (discrete or continuous) is always discontinuous step function jumps which correspond to the observed values of the random variable and magnitude equal to the relative frequencies of these values.

According to Bernoulli's theorem with an indefinitely large number of experiments *n* relative frequency of the event *X <x* converges in probability to the probability of this event. This means that the statistical distribution function $F^*(x)$ with increasing n converges to the true probability of the random variable X is the theoretical F (x) of the distribution function, i.e., $\lim_{n\to+\infty} F^*(x) = F(x)$.

This conclusion is confirmed by the fact that $F^*(x)$ has all the properties of *F (x).* Really:

1. The values of the empirical distribution function belongs to the interval [0,1];
2. $F^*(x)$ – nondecreasing function;
3. If x_1 – the smallest version, the $F^*(x) = 0$ when $x \le x_1$; if x_k is the greatest option, then $F^*(x) = 1$ when $x > x_k$

Thus, the empirical sample distribution function can be used to estimate the theoretical distribution function of the population. Consider an example.

Example 5.1. Construct statistical error distribution function 10 measurement range to the target using the rangefinder. The measurement results are given in the table:

i	1	2	3	4	5	6	7	8	9	10
x_i,м	5	– 8	10	15	5	– 10	– 6	20	10	15

Decision. Taking the measurement results, first construct a variation number, then the statistical distribution of the sampling frequency.

Variation number and the statistical distribution of the sampling frequency are respectively:

i	1	2	3	4	5	6	7
x_i	- 10	- 8	- 6	5	10	15	20

x_i	- 10	- 8	- 6	5	10	15	20
n_i	1	1	1	2	2	2	1

Since the sample size *n* = 10, therefore, the statistical distribution of relative sampling rate, have

x_i	- 10	- 8	- 6	5	10	15	20
W_i	0.1	0.1	0.1	0.2	0.2	0.2	0.1

Now we proceed to the construction of the statistical distribution function $F^*(x)$. Since the smallest embodiment is - 10, therefore $F^*(x) = 0$, when $x \leq -10$.

Meaning of *X* <- 8, wherein *X* – random value range measurement error of the target observed in the sample with a relative frequency of 0.1, namely $x_1 = -10$ therefore $F^*(x) = 0$, when $1 - 10 < x \leq -8$.

Values *X* <- 6, namely $x_1 = -10$, and $x_2 = -8$ observed with a relative frequency of 0.1 and 0.1, respectively, hence at $F^*(x) = 0{,}2 - 8 < x \leq -6$.

The values of *X* <5, namely $x_1 = -10$, $x_2 = -8$ and $x_3 = -6$, observed with a relative frequency of respectively 0.1, 0.1, 0.1, therefore $F^*(x) = 0{,}1 + 0{,}1 + 0{,}1 = 0{,}3$ with $-6 < x \leq 5$.

Values *X* <10, namely $x_1 = -10, x_2 = -8, x_3 = -6,\ x_4 = 5,$ were observed relative frequency of respectively 0.1, 0.1, 0.1, 0.2, therefore $F^*(x) = 0{,}1 + 0{,}1 + 0{,}1 + 0{,}2 = 0{,}5$ with $5 < x \leq 10$.

Values *X* <15, namely $x_1 = -10,\ x_2 = -8,\ x_3 = -6, x_4 = 5,\ x_5 = 10$ observed relative frequency of respectively 0.1, 0.1, 0.1, 0.2, 0.2, therefore $F^*(x) = 0{,}7$ with $10 < x \leq 15$.

Values *X* <20, namely $x_1 = -10,\ x_2 = -8,\ x_3 = -6, x_4 = 5,\ x_5 = 10,$ $x_6 = 15$ were observed relative frequency of respectively 0.1, 0.1, 0.1, 0.2, 0.2, 0.2, therefore $F^*(x) = 0{,}91$ with $5 < x \leq 20$.

Since *x* =20 highest embodiment, therefore $F^*(x) = 0{,}1 + 0{,}1 + 0{,}1 + 0{,}2 + 0{,}2 + 0{,}2 + 0{,}1 = 1$, when *x*> 20.

Thus, the required statistical function $F^*(x)$ at 10 measurement error distribution target range has the form

$$F^* = \begin{cases} 0 & \text{at} \quad x \le -10, \\ 0{,}1 & \text{at} \quad -10 < x \le -8, \\ 0{,}2 & \text{at} \quad -8 < x \le -6, \\ 0{,}3 & \text{at} \quad -6 < x \le 5, \\ 0{,}5 & \text{at} \quad 5 < x \le 10, \\ 0{,}7 & \text{at} \quad 10 < x \le 15, \\ 0{,}9 & \text{at} \quad 15 < x \le 20, \\ 1 & \text{at} \quad x > 20. \end{cases}$$

The graph of this function is shown in Fig. 5.1.

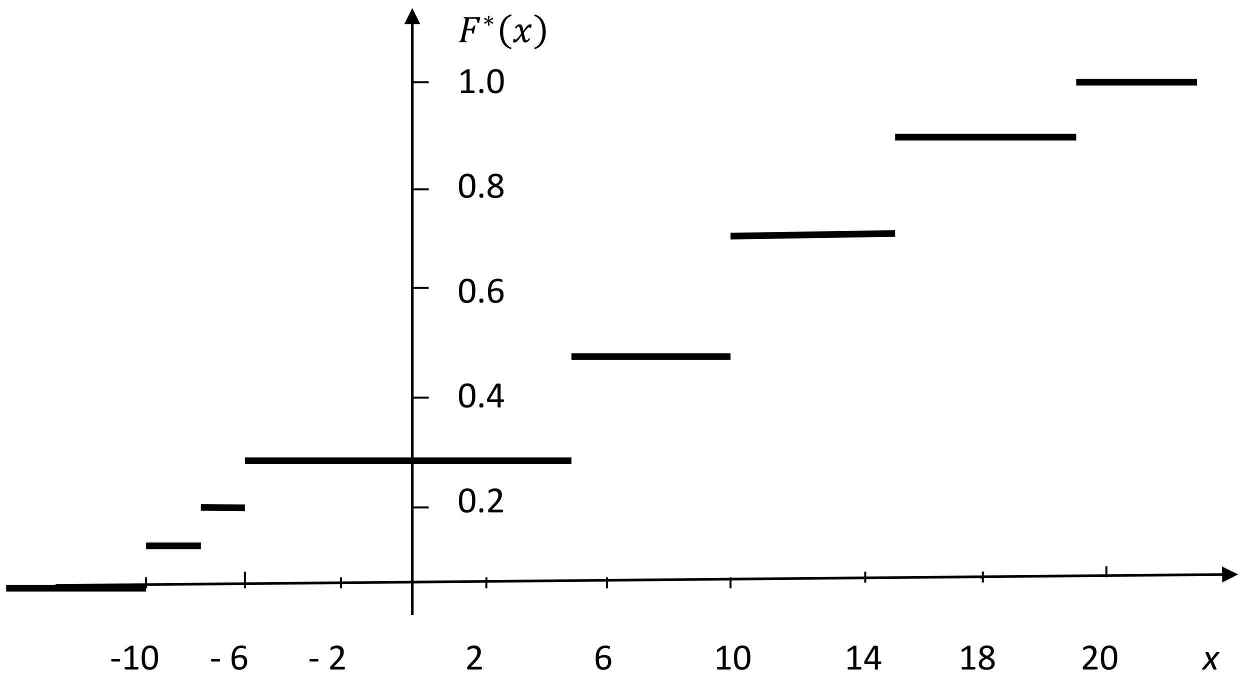

Fig. 5.1

It is noted that a large number of experiments n build statistical distribution function $F^*(x)$ – a very difficult operation, so it is often convenient to use other characteristics of statistical distributions, are not the same distribution function F (x), a probability density $f(x)$. Such visual graphical representations of the statistical distribution of the sample are polygon and the histogram of the relative frequencies.

5.2.3. Polygon and histogram.

If the sign of the total population under study displayed a random variable X is discrete in nature and the sample size is small, in this case to represent the statistical distribution of the sample is more convenient to use the polygon relative frequencies.

To construct the polygon relative frequency on the abscissa axis represents variants x_i, and the ordinate axis – corresponding relative frequencies W_i. Points (W_i, x_i) and connect the straight segments obtained polygon relative sampling frequencies.

Definition* 5.14. *Polygon relative sampling frequency is called a broken line, which segments connect points (x_i, W_i) wherein x_i – embodiments, and W_i – the corresponding relative frequency.

Fig. 5.2 depicts the polygon relative frequencies of Example 5.1.

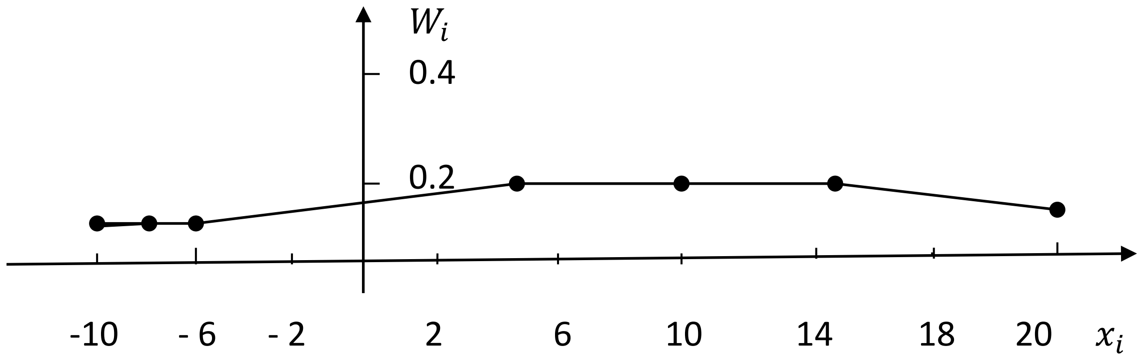

Fig. 5.2

If the sample size is large enough, then the representation of the statistical distribution of the samples in the form of a polygon is difficult, and in solving many problems and inappropriate, for example, when we study the random variable is continuous. In such cases, to represent the statistical distribution of the sample is more convenient to use the histogram of relative frequencies.

To construct a histogram of the relative frequency interval in which are enclosed all the observed value of the random variable, divided into partial intervals and the length h for each partial interval are - the sum of the relative frequencies embodiment caught in the i-th interval. ih_iW_i

Definition* 5.15. *Relative frequency histogram samples called a stepped shape consisting of rectangles whose bases are partial length *h* intervals, and the height equal to the ratio W_i/ *h* (density relative frequencies).

As for the number of partitions, their number is chosen so that the measurement results are best visible and contained quite a lot of information.i

Relative frequency histogram is constructed as follows: the x-axis represents the partial intervals, and each of them is constructed rectangle, area of which equals W_i – relative frequency variant, caught in this partial interval respectively equal to the height of this rectangle W_i/ h.

From the method of constructing a histogram of the relative frequencies, it follows that the total area of its equal to the sum of all relative sampling frequency variant, i.e. unit.

As an example, we construct a histogram of the relative frequencies of statistical population range measurement errors using EDM.

*Example*5.2. Build a histogram of the relative frequencies for the next statistical distribution of measurement error frequency:

h_i	[- 20, -15]	(- 15, -10]	(-10, - 5]	(- 50]	(0.5]	(5,10]	(10,15]	(15,20]
n_i	2	8	17	24	26	13	6	4

Decision. For each partial interval h_i define W_i – relative frequency measurement errors. Since the sample volume $n = \sum n_i = 100$, thus statistical distribution of relative frequencies W_i have

h_i	[- 20, -15]	(- 15, -10]	(-10, - 5]	(- 50]	(0.5]	(5,10]	(10,15]	(15,20]
W_i	0.02	0.08	0.17	0.24	0.26	0.13	0.06	0.04

Now, defining for each partial rectangle its height W_i/ *h*, where *h* = 5, we construct a histogram of the relative measurement error rates (Fig. 5.3).

Obviously, if the histogram points connect smooth line, this line as a first approximation to represent a graph of the probability density of the random variable *X*. In this case, if the number of dimensions increase, and select smaller partitions of partial intervals, the histogram will be increasingly closer to the probability density of the random value *X*.

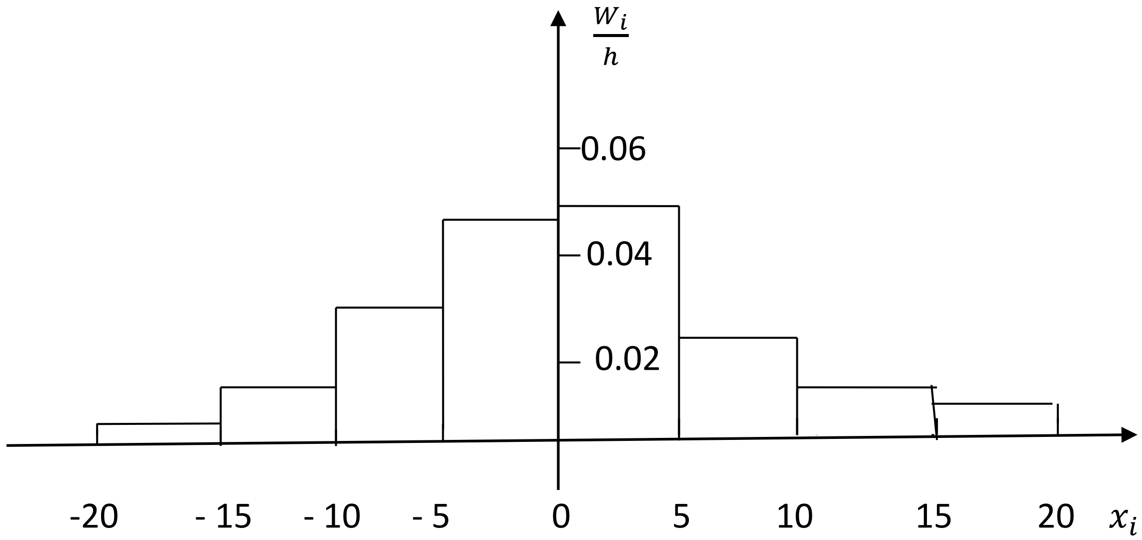

Fig. 5.3

5.2.4. Numerical characteristics of the statistical distribution of the sample.

The distribution of the random variable is a function, an indication of this function completely describes a random variable with a probability standpoint. However, in solving many practical problems there is no need to characterize the random variable exhaustively, and it is sufficient for some numerical characteristics that characterize the essential features of the random variable. The main characteristics of the random numerical values are the mean and variance. The expectation characterizes the mean value, which are grouped about the possible values of the random variable, and characterizes the degree of dispersion of the dispersion values of the relative average.

Similar numerical characteristics exist for sampling statistical distributions. An analogy mathematical expectation of a random variable *X* is the arithmetic average of the observed values of the random variable in the sample:

$$M^*[X] = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{j=1}^{k} x_j^* \cdot n_j}{n} = \sum_{j=1}^{k} x_j^* \cdot W_j\,, \qquad (5.1)$$

wherein x_i – the value of the random variable, the observed in *i*-th trial, *n* - number of tests in a sample, *k* - number of variant in the sample, x_j^* – sample embodiment, *j* - sequence number of options in the variational several sample, n_j – frequency variations, W_j – relative frequency variations.

This characteristic is referred to as ***the statistical average of the sample*** of the random variable *X*.

With a large number of tests n arithmetic average of the observed values of the random variable in the sample approaches (converges in probability) to its expectation and can be made approximately equal to the expectation.

The analogy of the random variable *X* is the ***statistical dispersion*** of the dispersion, which is defined as follows:

$$D^*[X] = \frac{\sum_{i=1}^{n}(x_i - M^*[X])^2}{n}, \qquad (5.2)$$

Wherein x_i – the random variable value, observed in the *i*-th trial, *n* – number of tests in a sample, $M^*[X]$ – average statistical sampling.

Similarly determined statistical initial and central moments of any order for the sample:

$$M^*[X^k] = \frac{\sum_{i=1}^{n} x_i^k}{n},$$

$$M^*[(X - M^*[X^k])^k] = \frac{\sum_{i=1}^{n}(x_i - M^*[X^k])^k}{n}.$$

Note that as the number of observation all the statistical characteristics of the probabilities converge to respective numerical values of the random characteristics and a sufficiently large number n of observations can be made approximately equal to them.

*Example*5.3. To determine a statistical average and variance of the statistical distribution of the errors of 10 measurements target range by using a range finder. The measurement results are given in the table:

i	1	2	3	4	5	6	7	8	9	10
$x_i.m$	5	- 8	10	15	5	- 10	- 6	20	10	15

Decision. By applying the formula (5.1) and (5.2), we obtain

$$M^*[X] = \frac{\sum_{i=1}^{n} x_i}{n} = 5{,}6\ m,$$

$$D^*[X] = \frac{\sum_{i=1}^{n}(x_i - M^*[X])^2}{n} \approx 98{,}64\ m^2.$$

Statistical mean square deviation **sampling**, will be $\sigma_X^* = \sqrt{D^*[X]} \approx 9{,}93\ m$.

5.3. Statistical evaluation of random variable distribution parameters

Suppose it is required to study the random variable *X*, displaying a quantitative indication of the total population. Assume that from theoretical considerations it was found on what is legally distributed random variable *X*. Of course there is a problem of parameter estimation, which is determined by this distribution. For example, if it is known beforehand that the random variable studied normal distribution, it is necessary to evaluate (found approximately) the mean and standard deviation, as these two parameters completely determine the normal distribution. If there is reason to believe that a random variable is, for example, the Poisson distribution, it is necessary to estimate the parameter λ, which is determined by the distribution.

Usually there is only available to the researcher data sample, x_1, x_2 ,… x_n, obtained by n independent observations of the random variable *X*. Through these data, and express the estimated parameter. For example, for the evaluation of the expectation of a normal distribution can use the formula (3.1) $M^*[X] = \frac{\sum_{i=1}^{n} X_i}{n}$ for the arithmetic mean of the observed values x_1, x_2,…, x_n in a sample of a random variable. Observed value x_1, x_2,…, x_n the random variable *X* in the sample, with its repeated repetition, in turn, should be considered as independent random variables X_1, X_2,…, X_n, each of which is allocated to the same law as the random variable X for each Expressing sampling estimated parameter, we come to the fact that the estimated parameter will be a function of the observed random variables X_1, X_2,…, X_n. So, find a statistical estimate of the unknown parameter of the theoretical distribution $M^*[X] = \frac{\sum_{i=1}^{n} x_i}{n}$ – it means finding a function of the observed random variables, which gives the approximate value of the estimated parameter.

Thus, a statistical estimate of the unknown parameter of the theoretical distribution of the random variable X is called a function of the observed random variables $X_1, X_2,…, X_n$.

To statistical estimates were given "good" approximation of the estimated parameters, they must meet certain requirements. Consider these requirements.

5.3.1. Unbiased, efficient and consistent estimates.

Assume that the sample volume *n* found statistical evaluation parameter Q_1^* of the theoretical distribution of the random variable X. Repeat experience, i.e. extract from a population another sample of the same volume and its data will find statistical evaluation Q_2^*. Repeating the experience many times, we get the numbers Q_1^*, Q_2^*,…, Q_k^*, which, generally speaking, different from each other. Thus, a plurality of numbers Q_1^*, Q_2^*,…, Q_k^*, can be considered as a random variable Q^* , and figures Q_1^*, Q_2^*,…, Q_k^*, representing a statistical parameter estimation Q – both its possible values. It is obvious that in order for a random variable Q^* whose values have statistical estimates of the unknown parameter Q of practical value, it must have the following property. It is necessary that the expectation $M[Q^*]$ random value Q^* is equal to the estimated parameter: Q: $M[Q^*] = Q$. Otherwise, a random value Q^* will produce a systematic error in the estimate of the unknown parameter Q or overcharge ($M[Q^*] > Q$), or underestimated ($M[Q^*] < Q$). In other words, the compliance $M[Q^*] = Q$ guarantee of receiving systematic errors, i.e., errors that distort the measurement results in a certain direction. While this requirement also does not resolve the error (one value Q^* may be larger and the other smaller Q), but in this case, the error of opposite sign will occur equally often.

Definition **5.15.** The random variable Q^* value, which is the statistical estimate the unknown parameter Q of the theoretical distribution of the random variable X is called ***unbiased*** if the expectation $M[Q^*]$ random value is equal to the estimated parameter Q, i.e.

$$M[Q^*] = Q. \qquad 5.3.$$

***Definition* 5.16. Offset** called the random value Q^* , the expectation $M[Q^*]$ which is not equal to the estimated parameter Q , i.e. $M[Q^*] \neq Q$.

However, it would be wrong to believe that an unbiased estimator always gives a good approximation of the estimated parameter. Valid possible values of the random variable Q^* may be highly dispersed around its mean value $M[Q^*]$, i.e. dispersion $D[Q^*]$ can be significant. In this case, according to the found small amount of random samples expectation $M[Q^*]$ of the random variable Q^* may be quite remote from the estimated parameter Q . Therefore, taking the $M[Q^*]$ as an approximate value of the estimated parameter Q , we would have made a big mistake. If the request to the dispersion $D[Q^*]$ random value Q^* was small, the possibility to avoid a large error will be eliminated. For this reason, to the random variable Q^* statistical parameter Q estimation are required efficiency.

Definition* 5.17.** Random statistical evaluation value parameter Q^* (for a given volume of sample *n*) is called ***effective if the dispersion $D[Q^*]$ is minimal, i.e. $D[Q^*] = D_{min}$.

When considering a large amount of samples (*n* is large!) to a random variable statistical parameter Q^* estimation Q are required consistency.

Definition* 5.18.** Random statistical evaluation Q^* value parameter Q called ***consistent if it converges to the estimated probability parameter Q with unbounded increase sample size *n*, i.e.

$$\lim_{n\to\infty} P[|Q^* - Q| < \varepsilon] = 1, \tag{5.4}$$

where ε – an arbitrarily small positive number. To meet the requirements (5.3) is sufficient to dispersion $D[Q^*]$ of the random variable Q^* converges to zero at $n \to \infty$, i.e. to satisfy the condition

$$\lim_{n\to\infty} D[Q^*] = 0. \tag{5.5}$$

From the definition follows that the viability assessment means that a sufficiently large sample size n with arbitrarily high reliability evaluation deviation from the true value of the parameter is less than a prescribed value no matter how small it is.

Thus, for development of practical methods for processing experimental data to obtain estimates taken as approximate values of the unknown parameters of the random quantity, should be guided formulated above properties changes (5.3), (5.4) and (5.5).

5.3.2. Determining the approximate values of the expectation and the variance of a random variable.

Suppose we have a random variable *X* with mean M [X] = and the dispersion *M*[*X*] = m_X both parameters are unknown. Required on the basis of experimental data to find a wealthy and unbiased estimates of these parameters.

Denote x_1, x_2,…, x_n the values of the random variable *X* obtained as a result of *n* independent equally accurate measurements, i.e. measurements, which were conducted under identical conditions. Usually consider these conditions to be met if the measurements were performed by one device. At the same time, as we have said, the sample values x_1, x_2,…, x_n the random variable *X* obtained in the end *n* independent equally accurate measurements, may be considered as random variables, ..., (e.g., at multiple repetition of n independent equally accurate measurements or theoretical considerations). Each of the observed random variables X_1, X_2,…, X_n distributed according to the same law as the random variable *X* and therefore has the same numerical characteristics, which is the random variable *X*.

As a statistical evaluation for the expectation m_X of a random variable *X* take the arithmetic average of the observed values X_1, X_2,…, X_n and is denoted $\overline{X}$ therefore

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}. \tag{5.6}$$

We show that this estimate of the mathematical expectation m_X of a random variable *X* is unbiased and consistent. Indeed, according to the law of large numbers,

$$\lim_{n\to\infty} P\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_X\right| < \varepsilon\right] = 1.$$

This means that $\bar{X}$ a consistent estimate. Evaluation $\overline{X}$ is also unbiased as

$$M[\bar{X}] = M\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{\sum_{i=1}^{n} M[X_i]}{n} = \frac{\sum_{i=1}^{n} m_X}{n} = m_X.$$

We now turn to the estimation variance D_X. If the evaluation D_X take statistical sampling variance (5.2)

$$D^*[X] = \frac{\sum_{i=1}^{n}(x_i - M^*[X])^2}{n} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n},$$

it can be shown (proof omitted), that this estimate is consistent, but is not an unbiased estimator, i. e. $M[D^*[X]] \neq D_X$.

This evaluation leads to systematic errors, giving lower values of the population variance. The reason is that the arithmetic average $\overline{X}$ of the observed values X_1, X_2,…, X_n and expectation M [X] = random variable X are different. If the expectation *M*[*X*] = m_X random variable *X* is known, then it is natural that as an estimate for the variance D_X of the random variable *X* with a known mathematical expectation *M*[*X*] = m_X the statistical variance should be taken

$$D^*[X] = \frac{\sum_{i=1}^{n}(X_i - m_X)^2}{n}.$$

In order to correct this systematic error in the case when the evaluation D_X is not used expectation *M*[*X*] = m_X random variable *X*, and the arithmetic average $\bar{X}$ of the observed values X_1, X_2,…, X_n and obtain an estimate for the population variance D_X, which is a consistent and unshifted sufficient for this statistical variance multiplied sample $D^*[X]$by a fraction $\frac{n}{n-1}$. By doing this, we get

$$M\left[\frac{n}{n-1}D^*[X]\right] = \frac{n}{n-1}\, M[D^*[X]] = D_X.$$

Thus, if as a result of *n* equally accurate measurements independent of the random variable *X* with unknown expectation *M*[*X*] = m_X and the dispersion *D*[*X*] = D_X the value obtained X_1, X_2,…, X_n, to determine these parameters should be used, respectively, the following approximate estimates:

$$\overline{m_X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$\overline{D_X} = \frac{n}{n-1}D^*[X] = \frac{\sum_{i=1}^{n}(X_i - \overline{m_X})^2}{n-1}. \qquad (5.7)$$

Both of these estimates are approximate, and wealthy and unbiased.

For example, using the estimates (5.7) in practice. When processing measuring results, such as a physical quantity necessary to estimate the true value of the measured quantity and accuracy of its measurement. To estimate the measured value and the accuracy of its measurement is used of the formula (5.7). Since usually measurements are mutually independent, are the same expectation (actual measured value) and the same dispersion (in the case of equally accurate measurements), the formula (5.7) are applicable. The true measured value estimated by the arithmetic mean of random measurements $\bar{X}$, and the measurement accuracy (measurement error) in the error theory is commonly characterized using the standard deviation $\sqrt{\bar{D}}$ of random measurement results. However, the definition $\bar{X}$ and $\bar{D}$ according to the formulas (5.7) this leads to cumbersome calculations, so in practice it is advisable to use the formulas

$$\overline{m_X} = \frac{\sum_{i=1}^{n}(X_i - a)}{n} + a,$$

$$\overline{D_X} = \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n}(X_i - a)^2}{n} - \left(\overline{X} - a\right)^2\right], \qquad (5.8)$$

that the skillful selection of *a* (usually it expresses the expected measured value) greatly facilitate processing of statistical material. Note that the formula (5.8) by simple transformations leads to formula (5.7).

Example. 5.4. After each hour measured voltage in the mains. The measurement results are presented in volts as a random number:

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i, B	222	219	224	220	218	217	221	220	215	218	223	225

i	13	14	15	16	17	18	19	20	21	22	23	24
x_i, B	220	226	221	216	211	219	220	221	222	218	221	219

Find bounds for the values of the measured value and the accuracy of its measurements according to the measurement results.

Decision. Guest $\bar{x}$ and $\overline{d}$ accordingly to the measured value and variance for a particular measurement results found from the formula (5.8) by setting *a* = 220 V (the measured value of certificate data) and replacing them random values X_i and $\overline{X}$ specific values x_i and $\bar{x}$ these random values obtained from measurements . Estimate $\overline{\sigma}$ for the standard deviation of specific dimensions, i.e. measuring accuracy, we find from Eq. $\overline{\sigma} = \sqrt{\overline{d}}$. All the necessary calculations are given in the following table

i	$x_i - a$	$(x_i - a)^2$	*i*	$x_i - a$	$(x_i - a)^2$	*i*	$x_i - a$	$(x_i - a)^2$
1	2	4	9	-5	25	17	1	1
2	-1	1	10	-2	4	18	-1	1
3	4	16	11	3	9	19	0	0
4	0	0	12	5	25	20	1	1
5	-2	4	13	0	0	21	2	4
6	-3	9	14	6	36	22	-2	4
7	1	1	15	1	1	23	1	1
8	0	0	16	-4	16	24	-1	1
Σ	1	35		4	116		1	13

Hence,

$$\bar{x} = \frac{\sum_{i=1}^{24}(x_i - 220)}{24} + 220 = \frac{6}{24} + 220 = 220.25 \text{ B},$$

$$\overline{d} = \frac{24}{24-1}\left[\frac{\sum_{i=1}^{24}(x_i - 220)^2}{24} - (220{,}25 - 220)^2\right] \approx 7{,}06\ \text{B}^2,$$

$$\overline{\sigma} = \sqrt{7{,}06} \approx 2{,}66\ \text{B}.$$

So the measured values in the data measurements was 220.25 B, and measurement accuracy of 2.66 V.

The above evaluation (5.7) is called point, since they are defined by a single number. It is clear that with a small volume of the sample point *n* of $\bar{X}$ (3.6) and $\overline{D} = \frac{n}{n-1}D^*[X]$(5.7) may differ significantly from the estimated parameters, respectively, *M [X]* and D_X , i.e. lead to gross errors. For this reason, a small volume of sample should be used Interval estimation.

Definition* 5.19. *Interval called the assessment, which is defined by two numbers – the ends of the interval.

Interval estimates allow us to establish the accuracy and reliability of the assessment. Reveal the essence of these concepts.

5.3.3. Accuracy of the estimation. Confidence probability (reliability). Confidence interval.

Let found according to statistical characteristics Q^* of the sample is unknown parameter Q estimation, the random variable *X*. We assume Q a constant number (Q can also be a random value). Clearly, the more accurately Q^*determines the parameter Q is less than the absolute value of the difference $|Q - Q^*|$. Consequently, the estimation accuracy can be characterized by a positive number $\delta > 0$, which establishes that the evaluation Q^* carried out with this accuracy, if it satisfies the inequality $|Q - Q^*| < \delta$. However, since the evaluation Q^* is random, therefore, affirm that the evaluation satisfies this inequality is possible only with a certain probability β.

Definition* 5.20. *Reliability (confidence probability) parameter estimation Q at Q^* is called the probability β with which the inequality $|Q - Q^*| < \delta$.

Let the probability that $|Q - Q^*| < \delta$ is β (usually β to take a number close to unity, for example, 0.95, 0.99): $P[|Q - Q^*| < \delta] = \beta$. Replacing the inequality $|Q - Q^*| < \delta$ is equivalent to the double inequality $-\delta < Q - Q^* < \delta$ or $Q^* - \delta < Q < Q^* + \delta$, we have

$$P[Q^* - \delta < Q < Q^* + \delta] = \beta. \tag{5.9}$$

Note that in (5.9) the unknown parameter value Q is non-random value, and the interval $(Q^* - \delta, Q^* + \delta)$ is a random variable, since the position of the interval on the real axis depends on the random variable Q^* (center interval), length of interval 2δ is also in the general case is a random variable. Therefore, in this case, the probability β is better interpreted as the probability of not getting the unknown parameter Q in the interval $(Q^* - \delta, Q^* + \delta)$, and the probability that a random interval $(Q^* - \delta, Q^* + \delta)$ will cover the unknown parameter Q. In other words, the probability of a trust should not be linked with the estimated parameters; it is connected only with the borders of the confidence interval which vary from sample to sample.

Definition* 5.21.** Interval $(Q^* - \delta, Q^* + \delta)$, which covers the unknown parameter Q with a predetermined reliability (confidence probability) β are called ***confidence.

As an example, consider the problem of the confidence interval for the expectation.

5.3.4. Construction of a confidence interval for estimating the expectation of a random variable normally distributed with known variance.

Let X quantitative indication of the total population are normally distributed $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ with standard deviation σ = $\sqrt{D[X]}$ (or dispersion $D[X]$) of this distribution is known. It is required to estimate the unknown expectation *a* and the general population on selective medium of the random variable $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, where $X_1,\ X_2,\dots, X_n$ – the random variables independent equally accurate measurements set themselves the task of finding confidence intervals covering parameter *a* and reliability (confidence probability) β and what is the accuracy δ of estimation of the parameter a depends on the amount of n samples.

Since the value of *X* has a normal distribution, then the sample mean of the random variable $\bar{X}$, found by independent observations, as is normally distributed (accept this assertion without proof). The distribution parameters $\bar{X}$ are as follows: $M[\bar{X}] = a, D[\bar{X}] = \frac{D[X]}{n},\ \sigma[\bar{X}] = \frac{\sigma}{\sqrt{n}}$. We now require that the relation

$$P\left[\left|\bar{X} - a\right| < \delta\right] = \beta. \qquad (5.10)$$

Given that the random variable $\bar{X}$ is normally distributed, we express confidence probability β in the left side of (5.10) over the normalized Laplace function:

$$P\left(a-\delta<\overline{X}<a+\delta\right)=\frac{1}{\sqrt{2\pi}}\int\limits_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}e^{-\frac{t^2}{2}}dt=\left[\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)-\Phi\left(\frac{-\delta\sqrt{n}}{\sigma}\right)\right].\quad(5.11)$$

Recall that the normalized Laplace function $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ allows you to determine the probability $P(x_1<X<x_2)$ of getting a random variable X is distributed normally for a predetermined interval (x_1,x_2):

$$P(x_1<X<x_2)=\frac{1}{\sqrt{2\pi}}\int\limits_{\frac{x_1-a}{\sigma}}^{\frac{x_2-a}{\sigma}}e^{-\frac{t^2}{2}}dt=\left[\frac{1}{\sqrt{2\pi}}\int\limits_{0}^{\frac{x_2-a}{\sigma}}e^{-\frac{t^2}{2}}dt-\frac{1}{\sqrt{2\pi}}\int\limits_{0}^{\frac{x_1-a}{\sigma}}e^{-\frac{t^2}{2}}dt\right]=$$

$$=\left[\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right)\right].$$

In this case. $x_1=a-\delta x_2=a+\delta$.

Since the Laplace function is odd, then the equation (5.11) takes the form:

$$P\left(a-\delta<\overline{X}<a+\delta\right)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right).$$

Taking into account that the probability $P\left(a-\delta<\overline{X}<a+\delta\right)$ is given and equal to β then find the equation $2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=\beta$ of accuracy $\delta=\delta_\beta$ of the estimate of the parameter *a*: $\delta_\beta=\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\beta}{2}\right)$ where $\Phi^{-1}\left(\frac{\beta}{2}\right)$ – the inverse of the normalized Laplace function. Denote the value $\Phi^{-1}\left(\frac{\beta}{2}\right)$ through t_β : $\Phi^{-1}\left(\frac{\beta}{2}\right)=t_\beta$. The number t_β is determined by the Laplace function table (Appendix 2). Find the argument t_β, which corresponds to the value of the Laplace function equal to $\frac{\beta}{2}$. Now for the accuracy δ of the estimate of the parameter *a*, we obtain $\delta=\delta_\beta=\frac{\sigma t_\beta}{\sqrt{n}}$.

Thus, the aforesaid problems completely solved, and its final solution has the form

$$P\left(\overline{X}-\frac{\sigma t}{\sqrt{n}}<a<\overline{X}+\frac{\sigma t}{\sqrt{n}}\right)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=\beta.$$

The meaning of this relation is that the reliability β can be argued that the confidence intervals $\left(\overline{X}-\frac{\sigma t}{\sqrt{n}};\overline{X}+\frac{\sigma t}{\sqrt{n}}\right)$ coated with an unknown parameter *a*; estimation accuracy $\delta=\frac{\sigma t}{\sqrt{n}}\overline{X}$ – half of the length of the confidence interval, symmetrical relative.

Let us explain the meaning of which is specified reliability. Reliability, for example $\beta = 0{,}95$, indicates that if made sufficiently large number of samples, the 95% of them defines such confidence intervals in which the parameter is actually concluded; only 5% of cases it can go beyond the boundaries of the confidence interval.

Comment 1. If you want to evaluate the expectation of a given accuracy δ and reliability β, the minimum sample size that will ensure the accuracy of finding the formula $n = \frac{\sigma^2 t^2}{\delta^2}\left(\text{follows from } \delta = \frac{\sigma t}{\sqrt{n}}\right)$.

The magnitude of the confidence intervals is dependent on the amount of n sample and as a consequence of the selective medium of the random variable $\bar{X}$. If a confidence interval calculated for the particular sample size *n* in the expression $\left(\overline{X} - \frac{\sigma t}{\sqrt{n}}; \overline{X} + \frac{\sigma t}{\sqrt{n}}\right)$ sample average random quantity $\bar{X}$ must be replaced by its specific value $\bar{x}$ in the sample: $\left(\bar{x} - \frac{\sigma t}{\sqrt{n}}; \bar{x} + \frac{\sigma t}{\sqrt{n}}\right)$

Comment 2. The assessment $\left|\overline{X} - a\right| < \frac{\sigma t}{\sqrt{n}} \delta = \frac{\sigma t}{\sqrt{n}}$ referred to classical. The formula that determines the accuracy of the classic evaluation, the following conclusions:

1) with increasing volume *n* of sample number δ decreases, and hence the accuracy of evaluation increases;
2) increase reliability evaluation $\beta = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)$ leads to an increase $\frac{\delta\sqrt{n}}{\sigma}$ (*Φ* (*x*) – increasing function), consequently, to an increase δ; in other words, increasing the reliability of the classical evaluation entails a decrease in its accuracy

Example. 5.5. The random variable *X* is a normal distribution with a known mean square deviation σ = 3. Find the confidence intervals for the estimates of the unknown and the mathematical expectation *a* of a random sample $\bar{x}$ average value $\overline{X}$ when the sample size is *n* = 36 and the set reliability evaluation $\beta = 0{,}95$.

Decision. Found $t_\beta = \Phi^{-1}\left(\frac{\beta}{2}\right)$ 2F From correlation 2Φ(t_β) = 0.95 obtain Φ(t_β) = 0.475. According to Laplace function table values (Appendix 2) find $t_\beta = 1.96$.

An estimate of accuracy: $\delta_\beta = \frac{\sigma t_\beta}{\sqrt{n}} = \frac{1{,}96 \cdot 3}{\sqrt{36}} = 0{,}98$.

Now we can write the confidence interval:$(\bar{x} - 0{,}98; \bar{x} + 0{,}98)$. For example, if $\bar{x} = 4{,}1$ the confidence interval has the following confidence limits: $\bar{x} - 0{,}98 = 3{,}12$: $\bar{x} + 0{,}98 = 5{,}08$.

Thus, the values of the unknown parameter *a* as consistent with the sample data satisfy the inequality 3,12 < *a* < 5,08. We emphasize that it would be wrong to write *P* (3,12 <*a* <5,08) = 0,95. Indeed, since *a* – a constant, then either it is enclosed within results interval (time event 3.12 < *a* <5.08 significantly and its probability equal to unity), or it is not enclosed (in this case the event is 3.12 < *a* <5,08 is not possible and its probability is zero). Therefore, as we have said, the confidence probability should not be linked with the estimated parameters; it is connected only with the borders of the confidence interval which vary from sample to sample.

5.3.5. Construction of a confidence interval for estimating the expectation of a random variable normally distributed with unknown dispersion.

Let X quantitative indications of the total population are normally distributed $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, with an average deviation σ =$\sqrt{D[X]}$ (or the dispersion $D[X]$) of this distribution is not known. Is required to estimate the unknown expectation and the general population using confidence intervals. Of course, you cannot use the results of the previous section, in which σ it assumed known.

For an accurate construction of the confidence interval must know the law of distribution of a random variable average $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, where X_1, X_2,..., X_n – independent random variables equally accurate measurements, which generally depends on the unknown parameters themselves value X.

It turns out that in some cases, a random variable $\overline{X}$ can change to another random variable , which is a function of the observed values X_1, X_2,..., X_n, which the law of distribution does not depend on the unknown parameters of the variable X, and depends only on the sample size n and the type of distribution law of the random variable *X* .

For example, it is proved that the values for a normal distribution random variable *X* (it’s possible values will be denoted by *t*)

$$T = \sqrt{n}\frac{\bar{X} - a}{\sqrt{\bar{D}}}\,, \qquad (5.12)$$

Where $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, $\overline{D} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ – estimating unknown parameters *a* and $D[X]$ of a random variable *X*, and is subject to the Student distribution with *k = n* - 1 degrees of freedom. Student distribution probability density is given by:

$$S_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}, \qquad (5.13)$$

Where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du$- gamma function.

Formula (5.13) shows that the Student distribution does not depend on estimates of $\overline{X}$ and $\overline{D}$ and depends on the sample size *n* (or, equivalently, the number of degrees of freedom *k = n*- 1); this feature is its great advantage. At the same time $S_{n-1}(t)$ is an even function of *t*.

Consider the application of the Student distribution in the construction of the confidence interval for the expectation.

Let n manufactured sample volume for the random variable *X* is distributed according to a normal distribution with unknown expectation *a* and variance D. On the basis of experimental data for these parameter estimates $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and $\overline{D} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ are constructed and which are considered as random variables.

Required to construct confidence interval, confidence level is appropriate for the mathematical expectation β of a random variable *X*.

Denote by δ_β half the length of the confidence interval, symmetrical with respect to $\overline{X}$, then we have: $P\left(\left|\overline{X} - a\right| < \delta_\beta\right) = \beta$. Let us turn to the left side of this equation by a random value $\overline{X}$ to the random variable *T*, distributed by Student's law. For this multiply both side of $\left|\overline{X} - a\right| < \delta_\beta$ a positive value $\frac{\sqrt{n}}{\sqrt{\overline{D}}}$, we obtain

$$P\left(\sqrt{n}\frac{\left|\overline{X} - a\right|}{\sqrt{\overline{D}}} < \frac{\delta_\beta}{\sqrt{\frac{\overline{D}}{n}}}\right) = P\left(|T| < \frac{\delta_\beta}{\sqrt{\frac{\overline{D}}{n}}}\right) = \beta.$$

Given the parity function $S_{n-1}(t)$, we find that the probability β of the inequality $|T| < t_\beta = \frac{\delta_\beta}{\sqrt{\frac{\overline{D}}{n}}}$ is equal to:

$$P\left(\,|T| < t_\beta\right) = 2\int_0^{t_\beta} S_{n-1}(t)dt = \beta. \qquad (5.14)$$

Equation (5.14) determines the magnitude t_β depending on the confidence coefficient β. For this it is necessary to use a special table (Appendix 3), using which at confidence probability β, and sample size *n* found value t_β .

Having determined the value t_β of the formula $\delta_\beta = t_\beta \sqrt{\frac{\overline{D}}{n}}$, we find the half width of the confidence interval. Then, replacing the equality $P\left(\left|\overline{X} - a\right| < t_\beta \sqrt{\frac{\overline{D}}{n}}\right) = \beta$ inequality in parentheses to constitute him a double inequality, we obtain

$$P\left(\overline{X} - t_\beta \sqrt{\frac{\overline{D}}{n}} < a < \overline{X} + t_\beta \sqrt{\frac{\overline{D}}{n}}\right) = \beta.$$

So, using the Student's t distribution, we found confidence interval $\left(\overline{x} - t_\beta \sqrt{\frac{\overline{d}}{n}} < a < \overline{x} + t_\beta \sqrt{\frac{\overline{d}}{n}}\right)$ covering the unknown parameter *a* with reliability β. Here, the random variables $\overline{X}$ and $\overline{D}$ replaced with specific values of these quantities, respectively, $\overline{x}$ and $\overline{d}$, found in the sample.

Example 5.6. Let quantitative trait *X* general population are normally distributed. The sample $x_1, x_2, \dots, x_{16}$ volume $n = 16$ of the sample mean $\overline{x} = 20{,}2$, and found a "corrected" sample variance $\overline{d} = \frac{\sum_{i=1}^{16}(x_i - \bar{x})^2}{n-1} = 0{,}64$. Rate unknown expectation using a confidence interval with reliability $\beta = 0{,}95$.

Decision. Found t_β . Using the attached Table 2, for $\beta = 0{,}95$ and $n = 16$ find $t_\beta = 2{,}13$.

We find the confidence limits:

$$\overline{x} - t_\beta \sqrt{\frac{\overline{d}}{n}} = 20{,}2 - 2{,}13\sqrt{\frac{0{,}64}{16}} = 19{,}774,$$

$$\overline{x} + t_\beta \sqrt{\frac{\overline{d}}{n}} = 20{,}2 + 2{,}13\sqrt{\frac{0{,}64}{16}} = 20{,}626.$$

So, with the reliability of 0.95 and the unknown parameter *a* lies in the confidence interval of $19{,}774 < a < 20{,}626$.

Example 3.4. Produced 10 independent experiments over a random variable *X*, normally distributed with unknown parameters *a* and σ. The results are shown in the form of random number

i	1	2	3	4	5	6	7	8	9	10
x_i	2.5	2	-2.3	1.9	-2.1	2.4	2.3	-2.5	1.5	-1.7

Find an estimate $\overline{x}$ for the mean and to estimate the unknown expectation and with the help of the confidence interval with reliability $\beta = 0{,}95$.

Decision. We have; .$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 0{,}4\overline{d} = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{10-1} \approx 4{,}933$

Refer to the table of Annex 3, on $\beta = 0{,}95$ and $n = 10$ find $t_\beta = 2{,}26$ and record confidence limits:

$$\overline{x} - t_\beta\sqrt{\frac{\overline{d}}{n}} = 0{,}4 - 2{,}26\sqrt{\frac{4{,}933}{10}} = -1{,}18,$$

$$\overline{x} + t_\beta\sqrt{\frac{\overline{d}}{n}} = 0{,}4 + 2{,}26\sqrt{\frac{4{,}933}{10}} = 1{,}98.$$

So, with the reliability of 0.95 and the unknown parameter lies in the confidence interval $-1{,}18 < a < 1{,}98$.

*Example*5.7. Above the random variable *X* is made 20 experiments. The results of the experiments are given in the following table:

i	x_i	i	x_i
1	10.9	8	10.8
2	10.7	12	10.3
3	11.0	13	10.5
4	10.5	14	10.8
5	10.6	15	10.9
6	10.4	16	10.6
7	11.3	17	11.3
8	10.8	18	10.8
9	11.2	19	10.9
10	10.9	20	10.7

It is required to estimate the unknown and the expectation of a random variable *X* using a confidence interval with reliability $\beta = 0{,}95$. In this case, the confidence intervals constructed using the Laplace function and Student distribution and compare these estimates.

Decision. 1) Construction of using Laplace function. We have $\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 10{,}78$. By Laplace function table (Appendix 2) we find the argument $t_{0{,}95} = \Phi^{-1}\left(\frac{0{,}95}{2}\right) \approx 1{,}96$ that corresponds to the value of the Laplace function equal to 0.475. Then, by the formula $\delta_\beta = \frac{\sigma t_\beta}{\sqrt{n}}$ we find half the length of the confidence interval. The final formula for the standard deviation σ of the random variable *X*, which we have not specified, use his assessment $\overline{\sigma} = \sqrt{\frac{\sum_{i=1}^{20}(x_i - \bar{x})^2}{20-1}} \approx 0{,}253.$ Then

$$\delta_\beta = 0{,}253 \frac{1}{\sqrt{20}} 1{,}96 \approx 0{,}111.$$

Now we can write the confidence limits:

$$\overline{x} - \delta_\beta = 10{,}78 - 0{,}111 \approx 10{,}669,$$
$$\overline{x} + \delta_\beta = 10{,}78 + 0{,}111 \approx 10{,}891.$$

So, with the reliability of 0.95 and the unknown parameter lies in the confidence interval of 10.669.$< a < 10{,}891$

2) Construction using the Student's t distribution. We have .$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 10{,}78;$ $\overline{d} = \frac{\sum_{i=1}^{20}(x_i - \bar{x})^2}{20-1} \approx 0{,}064.$

Refer to the table of Annex 3, on $\beta = 0{,}95$ and $n = 20$ find $t_\beta = 2{,}093$ and record confidence limits:

$$\overline{x} - t_\beta \sqrt{\frac{\overline{d}}{n}} = 10{,}78 - 2{,}093 \sqrt{\frac{0{,}064}{20}} \approx 10{,}78 - 0{,}118 \approx 10{,}662,$$

$$\overline{x} + t_\beta \sqrt{\frac{\overline{d}}{n}} = 10{,}78 + 2{,}093 \sqrt{\frac{0{,}064}{20}} \approx 10{,}78 + 0{,}118 \approx 10{,}898.$$

So, with the reliability of 0.95 and the unknown parameter *a* lies in the confidence interval $10{,}662 < a < 10{,}898$.

Comment. The last example, the results indicate that the use for estimating the expectation, a normally distributed random variable *X* instead of the Student distribution (5.13) The Laplace function, assuming that the sample average random value $\bar{X}$ is also allocated under the normal law that the random variable *X* leads to unnecessarily narrowing the confidence interval, i.e. to improve the accuracy of the estimate. This discrepancy occurs for small sample volumes ($n < 30$), in particular for small n. With an unlimited increase in the sample size n, the Student distribution tends to a normal distribution. Really,

$$\lim_{n\to\infty}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}\right) = \frac{1}{\sqrt{2\pi}};\ \lim_{n\to\infty}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} = e^{-\frac{t^2}{2}}.$$

Therefore, practically at the average $n > 30$ for the sample of the random variable $\bar{X}$ may be instead of the Student distribution in determining t_β the expression (5.14), use of normal distribution, i.e. Laplace function.

Consider an example application of the confidence interval for the evaluation of the mathematical expectation of a random variable X in practice. Suppose made independent equally accurate measurements of some physical quantity and the true value is unknown. We consider the results of individual measurements as random values $X_1, X_2,..., X_n$. These values are independent of (independent of measurement) have the same expectation of *a* (true measured value), the same dispersion σ^2 (equal observations) and normally distributed (this assumption is confirmed by experiment). Thus, all the assumptions that have been made in the derivation of confidence intervals in the previous two sections, are performed, and therefore, we have the right to use the formula in which, taking into account the comments made in the previous section.

Example. 5.8. According to nine independent equally accurate measurements of the physical quantity found arithmetic mean of the individual measurements $\bar{x} = 42{,}319$ and the "corrected" standard deviation $\overline{\sigma} = 5{,}0$. It is required to estimate the true value of the measured value with reliability $\beta = 0{,}95$.

Decision. The true measured value is equal to the expectation of a random variable *X*, whose values are the results of measurements. Therefore, the problem reduces to evaluating the expectation *a* of a random variable *X* and the unknown sample σ and a small volume $n = 9$ by using a confidence interval $\left(\overline{x} - t_\beta\sqrt{\frac{\overline{d}}{n}} < a < \overline{x} + t_\beta\sqrt{\frac{\overline{d}}{n}}\right)$ covering unknown parameter and with a given reliability $\beta = 0{,}95$.

Using Table from the application 3 (the sample size is small), on $\beta = 0{,}95$ and $n = 9$ find $t_\beta = 2{,}31$. Now we find the accuracy of the estimate $\delta_\beta = t_\beta\sqrt{\frac{\overline{d}}{n}} = t_\beta\frac{\overline{\sigma}}{\sqrt{n}} = 2{,}31\frac{5}{\sqrt{9}} = 3{,}85$ and record confidence limits:

$$\overline{x} - t_\beta\sqrt{\frac{\overline{d}}{n}} = 42{,}319 - 3{,}85 = 38{,}469,$$

$$\overline{x} + t_\beta \sqrt{\frac{\overline{d}}{n}} = 42{,}319 + 3{,}85 = 46{,}169.$$

Thus, the reliability 0.95 true measured value lies in the confidence interval $38{,}469 < a < 46{,}169$.

5.3.6. Construction of a confidence interval for estimating the standard deviation of the random variable distributed normally.

Let quantitative trait X general population is normally distributed. It is required to estimate the unknown general standard deviation σ of "correction" of the sample mean square deviation $\overline{\sigma}$. We set ourselves the task of finding confidence intervals covering the parameter σ with a predetermined reliability β.

Require that the ratio

$$P(|\sigma - \overline{\sigma}| < \delta) = \beta, \text{ or } P(\overline{\sigma} - \delta < \sigma < \overline{\sigma} + \delta) = \beta.$$

In order to be able to use the finished table, we transform the double inequality is equivalent to the inequality. Putting we obtain$\overline{\sigma} - \delta < \sigma < \overline{\sigma} + \delta\overline{\sigma}\left(1 - \frac{\delta}{\overline{\sigma}}\right) < \sigma < \overline{\sigma}\left(1 + \frac{\delta}{\overline{\sigma}}\right)\frac{\delta}{\overline{\sigma}} = \mu$

$$\overline{\sigma}(1 - \mu) < \sigma < \overline{\sigma}(1 + \mu). \qquad (5.15)$$

It remains μ to be found. For this purpose, we introduce a random value

$$\chi = \frac{\overline{\Omega}}{\sigma}\sqrt{n-1},$$

wherein $\overline{\Omega}$ – the random variable taking values which are "corrected" sample standard deviations $\overline{\sigma}, n$ – the sample volume.

The density distribution χ is given by

$$R(\chi, n) = \frac{\chi^{n-2} e^{-\frac{\chi^3}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right)}, \qquad (5.16)$$

where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du$ – the gamma function. This allocation is independent of the estimated parameter σ and depends only on the sample size n.

We transform the inequality (5.15) so that it takes the form $\chi_1 < \chi < \chi_2$. The probability of this inequality is given probability β i.e. $P(\chi_1 < \chi < \chi_2) = \int_{\chi_1}^{\chi_2} R(\chi, n) d\chi$.

Assuming that $\mu < 1$ rewrite (5.15) as follows:

$$\frac{1}{\overline{\Omega}(1 + \mu)} < \frac{1}{\sigma} < \frac{1}{\overline{\Omega}(1 - \mu)}.$$

Multiplying all the members of the inequality by $\overline{\Omega}\sqrt{n-1}$, we get $\frac{\sqrt{n-1}}{(1+\mu)} < \frac{\overline{\Omega}\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{(1-\mu)}$ *или* $\frac{\sqrt{n-1}}{(1+\mu)} < \chi < \frac{\sqrt{n-1}}{(1-\mu)}$. The probability that this inequality and, consequently, it is equivalent to the inequality (5.15) will be implemented, is

$$\int_{\frac{\sqrt{n-1}}{(1+\mu)}}^{\frac{\sqrt{n-1}}{(1-\mu)}} R(\chi, n) d\chi = \beta.$$

From this equation it can be on the set n and β find . Almost for finding μ used applications table 4.

Calculating sample $\overline{\sigma}$ and finding in the table , we obtain the desired confidence interval (5.15), coating σ with a predetermined reliability β, i.e. interval $\overline{\sigma}(1-\mu) < \sigma < \overline{\sigma}(1+\mu)$.

Example 5.9. Quantitative *X* sign of the population is normally distributed. Sample volume $n = 25$ found "correction" of the standard deviation $\overline{\sigma} = 0{,}8$. Find confidence interval covering general standard deviation σ with reliability $\beta = 0{,}95$.

Decision. According to Table application 3 of data $\beta = 0{,}95$ and $n = 25$ find $\mu = 0{,}32$. Therefore, the desired range (5.15) is:

$$0{,}8(1-0{,}32) < \sigma < 0{,}8(1+0{,}32), \qquad \text{or} \quad 0{,}544 < \sigma < 1{,}056.$$

Comment. It was assumed above that $\mu < 1$. If $\mu > 1$ then (5.15) takes the form (give that $\sigma > 0$), $0 < \sigma < \overline{\sigma}(1+\mu)$ or (after transformation, analogous to the case $\mu < 1 \frac{\sqrt{n-1}}{(1+\mu)} < \chi < \infty$). Consequently, the values $\mu > 1$ can be found from the equation

$$\int_{\frac{\sqrt{n-1}}{(1+\mu)}}^{\infty} R(\chi, n) d\chi = \beta.$$

To find the values $\mu > 1$ use Table from the application 3.

Example 5.10. Quantitative sign *X* of the population is normally distributed. Sample volume $n = 10$ found "correction" of the standard deviation $\overline{\sigma} = 0{,}16$. Find confidence interval covering general standard deviation σ with reliability $\beta = 0{,}999$.

Decision. According to Table application 3 of data $\beta = 0{,}999$ and $n = 10$ find $\mu = 1{,}80$. Therefore, the desired range (3.14) is:

$$0 < \sigma < 0{,}16(1+1{,}80), \textit{или}\ 0 < \sigma < 0{,}448.$$

Example 5.11. Using 15 equally accurate measurements found "correction" of the standard deviation $\overline{\sigma} = 0{,}12$. Find the accuracy of measurements with the reliability of 0.99.

Decision. The measurement accuracy characterized by standard deviation σ of random measurement errors, and therefore the problem is reduced to finding a confidence interval (3.14), coating σ with a predetermined reliability 0.99.

According to Table from application 3 of data $\beta = 0{,}99$ and $n = 15$ find $\mu = 0{,}73$. Therefore, the desired range (5.15) is:

$$0{,}12(1 - 0{,}73) < \sigma < 0{,}12(1 + 0{,}73), \qquad \text{or } 0{,}03 < \sigma < 0{,}21.$$

5.3.7. Estimation of probability (binomial distribution) from the relative frequency.

Let performed independent tests with the unknown probability *p* of occurrence of the event *A* in each trial. Is required to estimate the unknown probability *p* of the relative frequency, ie, it is necessary to find its point and interval estimation.

point estimate. As a point estimate unknown probability *p* are random relative frequency $W = \frac{X}{n}$, wherein *X* − random variable with a binomial distribution values which *m* is the number of occurrences, A, n − the number of trials.

This estimate is unbiased, i.e. its expectation is equal to the estimated probability. Indeed, given that for binomial distribution *M*(*X*) = np, obtain

$$M(W) = M\left[\frac{X}{n}\right] = \frac{M(X)}{n} = \frac{np}{n} = p.$$

We find the variance of the estimate, taking into account that for binomial distribution *D (X) = npq,* where *q* = 1 - *p*:

$$D(W) = D\left[\frac{X}{n}\right] = \frac{D(X)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

Hence, the standard deviation $\sigma_W = \sqrt{D(W)} = \sqrt{\frac{pq}{n}}$.

Interval estimation. We find a confidence interval for estimating the probability of the event A p by the random relative frequency *W*. For this purpose we define the probability $P(|W - p| < \delta)$ that the absolute value of the relative frequency deviation of the random probability p from *W* does not exceed a positive number δ. Use formula obtained in probability theory enables us to find the probability of hitting a random variable *X* values distributed normally with mean and a standard deviation σ in the interval (x_1,x_2):

$$P(x_1 < X < x_2) = \left[\Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right)\right],$$

where *F (x)* – Laplace function.

Putting in this formula $x_1 = a - \delta, x_2 = a + \delta$, and considering that the Laplace function is even arrive at the formula

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right). \quad (5.17)$$

If n sufficiently large, and the probability *p* is very close to zero to unity, then we can assume that the random relative frequency *W* approximately normally distributed, and, as shown above $M(W) = p$.

Thus, substituting in relation (5.17) is the random variable *X* and its expectation *a* of a random variable W, respectively, and her expectation *p*, we obtain an approximation (as the relative frequency *W* approximately normally distributed) Equality

$$P(|W \ - p| < \delta) = 2\Phi\left(\frac{\delta}{\sigma_W}\right). \quad (5.18)$$

We proceed to the construction of the confidence interval (p_1,p_2), which covers the reliability β of the estimated parameter *p*. What we require that the ratio (5.18) was performed with reliability β:

$$P(|W \ - p| < \delta) = 2\Phi\left(\frac{\delta}{\sigma_W}\right) = \beta.$$

Replacing σ_W through, we get $\sqrt{\frac{pq}{n}} P(|W \ - p| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{pq}}\right) =$ $2\Phi(t) =$ where $\beta t = \frac{\delta\sqrt{n}}{\sqrt{pq}}$. Hence $\delta = t\frac{\sqrt{pq}}{\sqrt{n}}$ therefore

$$P\left(|W \ - p| < t\frac{\sqrt{pq}}{\sqrt{n}}\right) = 2\Phi(t) = \beta.$$

Thus, the reliability β of the inequality

$$|W \ - p| < t\frac{\sqrt{pq}}{\sqrt{n}}.$$

In order to obtain a working formula random variable W replace nonrandom observed relative frequency $\omega = \frac{m}{n}$, where m – the number of $n$occurrences of an event *A* under testing and substituting 1 - *p* instead of *q*:

$$|\omega \ - p| < t\frac{\sqrt{p(1 - p\ )}}{\sqrt{n}}.$$

Given that the probability *p* is unknown, we solve this inequality with respect to *p*. Assume that $\omega > p$. Then $\omega \ - p < t\frac{\sqrt{p(1 - p\ )}}{\sqrt{n}}$. Both parts are positive inequality; bring them to the square, we get the equivalent quadratic inequality with respect to *p*:

$$\left(\frac{t^2}{n}+1\right)p^2-2\left(\omega+\frac{t^2}{n}\right)p+\omega^2<0.$$

The discriminant of the trinomial is positive, so its roots are real and distinct:

smaller root

$$p_1=\frac{n}{t^2+n}\left[\omega+\frac{t^2}{2n}-t\sqrt{\frac{\omega(1-\omega)}{n}+\left(\frac{t}{2n}\right)^2}\right], \tag{5.19}$$

larger root

$$p_2=\frac{n}{t^2+n}\left[\omega+\frac{t^2}{2n}+t\sqrt{\frac{\omega(1-\omega)}{n}+\left(\frac{t}{2n}\right)^2}\right]. \tag{5.20}$$

Thus, the desired confidence interval $p_1<p<p_2$, where p_1 and p_2 is given by (5.19) and (5.20).

In the derivation we have assumed that $\omega>p$; the same results were obtained with $\omega<p$.

Example. 5.12. Produce independent tests with the same, but unknown probability *p* of occurrence of an event *A* in each test. Find a confidence interval for estimating the probability *p* with reliability 0.95 if in 80 tests event *A* appeared 16 times.

Decision. By condition, $n=80m=16\beta=95$. We find the relative frequency of occurrence of the event *A*: $\omega=\frac{m}{n}=\frac{16}{80}=0{,}2$.

Found t from the relation $2\Phi(t)=0{,}95$; on Laplace function table (see para. Annex 1) find $t=1{,}96$.

Substituting $n=80\omega=0{,}2t=1{,}96$, in Formula (5.19) and (5.20), we obtain respectively $p_1=0{,}128$, $p_2=0{,}299$.

Thus, the desired confidence interval $0{,}128<p<0{,}299$.

5.3.8. Determining the approximate values of numerical characteristics of a system of two random variables.

Let the random variables (X,Y) of the system is produced under the same conditions *n* independent trials. The test results (X_1,Y_1), (X_2,Y_2),…, (X_n,Y_n) … are independent random variables systems $X_1,X_2,\dots,X_n$ and Y_1, $Y_2,\dots,Y_n$, expectations, variance and correlation moments are the same, i.e. . $M[X_i]=m_X$, $M[Y_i]=m_Y$, $D_{X_i}=D_X$, $D_{Y_i}=D_Y$, $k_{X_iY_i}=k_{XY}$. Required by treatment of the experimental data to find the approximate values of said numerical characteristics.

This problem is solved in the same way as we solved it for a random variable.

Since the unknown expectations m_X and m_Y , and the dispersion characteristics D_X and D_Y are also individual random variables within the system, to determine their approximate values, applying the formula (5.7), we obtain:

$$\overline{m_X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad \overline{D_X} = \frac{\sum_{i=1}^{n}(X_i - \overline{m_X})^2}{n-1};$$
$$\overline{m_Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \qquad \overline{D_Y} = \frac{\sum_{i=1}^{n}(Y_i - \overline{m_Y})^2}{n-1}.$$

Since the correlation time is the expectation of the product of deviations of random variables X and Y their mathematical expectations, the approximate value of the correlation moment $\overline{k_{XY}}$ looking as a linear combination of the form

$$\overline{k_{XY}} = \sum_{i=1}^{n} C_i (X_i - \overline{m_X})\,(Y_i - \overline{m_Y}), \qquad (5.21)$$

where C_i – constant coefficients, and by virtue of measurement equal accuracy, $C_i = C$.

Unknown factor C determined from the condition that the value $\overline{k_{XY}}$ was unbiased estimate the time for the correlation moment k_{XY} i.e. that

$$M\left[\overline{k_{XY}}\right] = M\left[\sum_{i=1}^{n} C_i (X_i - \overline{m_X})\,(Y_i - \overline{m_Y})\right] = C\sum_{i=1}^{n} M[(X_i - \overline{m_X})(Y_i - \overline{m_Y})] =$$
$$= k_{XY}.$$

It can be shown that this condition is met and the estimate is unbiased if $C = \frac{1}{n-1}$.

Finding this dispersion evaluation: $D\left[\overline{k_{XY}}\right] = D\left[\frac{\sum_{i=1}^{n}(X_i - \overline{m_X})(Y_i - \overline{m_Y})}{n-1}\right]$, it can be shown that if $n \to \infty$ the variance of the estimates $\left[\overline{k_{XY}}\right] \to 0$, which means that the estimate of the correlation points (5.21) and is consistent with $C = \frac{1}{n-1}$.

In this way,

$$\overline{k_{XY}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{m_X})\,(Y_i - \overline{m_Y})$$

it is an unbiased and consistent estimator for the correlation moment k_{XY} of random variables (X, Y).

Experienced correlation coefficient $\overline{r_{XY}}$ is determined by the formula

$$\overline{r_{XY}} = \frac{\overline{k_{XY}}}{\overline{\sigma_X} \cdot \overline{\sigma_Y}}.$$

In this experimental standard deviation of the correlation coefficient calculated by the formula

$$\sigma_{\overline{r_{XY}}} = \frac{1 - r_{XY}^2}{\sqrt{n-1}} \approx \frac{1 - \overline{r}_{XY}^2}{\sqrt{n-1}}.$$

5.3.9. Maximum likelihood method for finding the distribution of the parameter estimates.

In addition to the methods of point estimates of the distribution parameters, we have discussed in section 5.2, there are other methods of distribution of point estimates of unknown parameters. These include one of the most important methods for finding estimates of parameters according to experience, which is called the maximum likelihood method.

***The discrete random variables*.** Let *X* – a discrete random variable, which is a result of the n tests adopted values x_1, x_2, …,x_n. Assume that the form of the distribution law of the value of *X* is set but unknown parameter , which is determined by the law. It is required to find its point estimate.

Denote the probability that the result of the test variable *X* takes the value x_i $(i = 1,2, \ldots, n)$ through $p(x_i; \theta)$.

Definition* 5.22. *Likelihood function of a discrete random variable X call function of the argument θ:

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

where $x_1, x_2, \ldots, x_n$ – fixed numbers.

As a point estimate of the parameter θ taking a value $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$ at which the likelihood function is maximized. Assessment θ^* called the *maximum likelihood estimate.*

Functions L and $\ln L$ reach a maximum at the same value θ, so instead of finding the maximum of the function L looking for $\ln L$ (which is more convenient), the maximum of which is called the *log-likelihood function.*

As you know, the argument θ of the maximum point $\ln L$ can be searched, for example, as follows:

1) find the derivative $\frac{d \ln L}{d\theta}$;

2) equating the derivative to zero, and to find the fixed point – the root of the equation obtained (called Likelihood equation);

3) find the second derivative $\frac{d^2 \ln L}{d\theta^2}$; if the second derivative $\theta = \theta^*$ is negative then $\theta^* -$ the maximum point.

Found maximum point θ^* taken as the maximum likelihood estimation parameter θ.

Maximum likelihood method has several advantages: it always leads to an affluent (though sometimes biased) estimates having the smallest possible dispersion compared to other well and uses the sample data on the estimated parameter.

The disadvantage of this method is that it often requires complex calculations.

Comment. The likelihood function − a function of the argument θ; maximum likelihood estimation − a function of the independent variables $x_1, x_2, \dots, x_n$.

Example 5.13. Find the method of maximum likelihood estimate of the parameter λ of the Poisson distribution

$$P_m(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda},$$

where x_i − the number of occurrences in the *i*-th (*i* = 1,2, ..., n) experiment (experiment comprises m test), m − the number of trials made.

Decision. Construct the likelihood function, given that: $\theta = \lambda$.

$$L(x_1, x_2, \dots, x_n; \lambda) = p(x_1; \lambda) \cdot p(x_2; \lambda) \cdot \dots \cdot p(x_n; \lambda) =$$
$$= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \cdot \dots \cdot \frac{\lambda^{x_n}}{x_n!} e^{-\lambda}.$$

We find the log-likelihood function:

$$\ln L = \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - n\lambda - \ln(x_1! \cdot x_2! \cdot \dots \cdot x_n!).$$

We find the first derivative λ of the equation and write the likelihood, which equate to zero the first derivative:

$$\frac{d \ln L}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0.$$

We find a stationary point, which will solve the resulting equation for λ:

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x_в}.$$

We find the second derivative with respect to $: \lambda \frac{d^2 \ln L}{d\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}.$

If $\lambda = \overline{x_в}$ the second derivative is negative; consequently, $\lambda = \overline{x_в}$ − the maximum point, and therefore, it is necessary to take the sample mean as an estimate of the maximum likelihood parameter λ of the Poisson distribution $\overline{x_в} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Example 5.14. Find the method of maximum likelihood estimate of a parameter *p* of the binomial distribution

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

if n_1 independent trials the event *A* appeared $x_1 = m_1$ again in n_2 independent trials the event *A* appeared again $x_2 = m_2$.

Decision. Construct the likelihood function, given that $\theta = p$:

$$L = P_{n_1}(m_1) \cdot P_{n_2}(m_2) = C_{n_1}^{m_1} C_{n_2}^{m_2} p^{m_1+m_2} (1-p)^{[(n_1+n_2)-(m_1+m_2)]}.$$

We find the log-likelihood function:

$$\ln L = \ln\left(C_{n_1}^{m_1} C_{n_2}^{m_2}\right) + (m_1+m_2)\ln p + [(n_1+n_2)-(m_1+m_2)]\ln(1-p).$$

We find the first derivative with respect to p

$$\frac{d\ln L}{dp} = \frac{m_1+m_2}{p} - \frac{[(n_1+n_2)-(m_1+m_2)]}{1-p}.$$

We write the likelihood equation, which equate to zero the first derivative:

$$\frac{m_1+m_2}{p} - \frac{[(n_1+n_2)-(m_1+m_2)]}{1-p} = 0.$$

We find a stationary point, which will solve the resulting equation for p: $p = \frac{m_1+m_2}{n_1+n_2}$

We find the second derivative with respect to p

$$\frac{d^2\ln L}{dp^2} = -\frac{m_1+m_2}{p^2} - \frac{[(n_1+n_2)-(m_1+m_2)]}{(1-p)^2}.$$

It is easy to see that for $p = \frac{m_1+m_2}{n_1+n_2} \le 1$ the second derivative is negative; consequently, $p = \frac{m_1+m_2}{n_1+n_2}$ – the maximum point, and therefore it should be taken as the maximum likelihood estimation of unknown probability of binomial distribution: $pp^* = \frac{m_1+m_2}{n_1+n_2}$.

Thus, the probability estimate *p* by the method of maximum likelihood coincides with the relative frequency $\frac{m_1+m_2}{n_1+n_2}$ of occurrence of an event *A* under testing.

***Continuous random variables*.** Let *X* – continuous random variables, which as a result of the n tests adopted values x_1, x_2, …,x_n. Suppose that the form of the distribution density $f(x)$ of *X* is given, but unknown parameter θ that determines the function. It is required to find its point estimate.

Definition **5.23.** ***The function of the continuous random variable X Likelihood*** call function of the argument θ

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta),$$

Where $x_1, x_2, \dots, x_n -$ fixed numbers.

Maximum likelihood estimate of the unknown parameter continuous random variable distribution looking like in the case of discrete values.

Example 5.15. Find the method of maximum likelihood estimate of the parameter λ of the exponential distribution

$$f(x) = \lambda\, e^{-\lambda x} \qquad (0 < x < +\infty)$$

if the result of the n test the random variable *X*, distributed according to the exponential law, adopted values $x_1, x_2, \dots, x_n$.

Decision. Construct the likelihood function, given that $\theta = \lambda$

$$L(x_1, x_2, \dots, x_n; \lambda) = \left(\lambda\, e^{-\lambda x_1}\right) \cdot \left(\lambda\, e^{-\lambda x_2}\right) \cdot \dots \cdot \left(\lambda\, e^{-\lambda x_n}\right).$$

We find the log-likelihood function:

$$\ln L = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i.$$

We find the first derivative with respect to λ

$$\frac{d \ln L}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

We write the likelihood equation, which equate to zero the first derivative: $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i. = 0$:

We find a stationary point, which will solve the resulting equation for λ:

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i.} = \frac{1}{\overline{x_в}}.$$

We find the second derivative with respect to λ: $\frac{d^2 \ln L}{dp^2} = -\frac{n}{\lambda^2}$.

If $\lambda = \frac{1}{\overline{x_В}}$ the second derivative is negative: consequently $\lambda = \frac{1}{\overline{x_В}}$ – the maximum point, and hence, as the maximum likelihood estimation of the parameter λ of the exponential distribution is necessary to take the reciprocal of selective medium: $\lambda^* = \frac{1}{\overline{x_В}}$.

Comment. If the density distribution $f(x)$ of continuous random variable *X* is defined by two unknown parameters θ_1 and θ_2 the likelihood function is a function of two independent variables θ_1 and θ_2 :

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2) = f(x_1; \theta_1, \theta_2) \cdot f(x_2; \theta_1, \theta_2) \cdot \dots \cdot f(x_n; \theta_1, \theta_2),$$

where $x_1, x_2, \dots, x_n$ – the observed values of *X*. Next, find the log-likelihood function for finding the maximum of its make up and solve the system

$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1} = 0, \\ \dfrac{\partial \ln L}{\partial \theta_2} = 0. \end{cases}$$

Example 5.16. Find the method of maximum likelihood estimation of the parameters *a* and *σ* normal distribution

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

if the result of the n test value *X* accepted values $x_1, x_2, \dots, x_n$.

Decision. Construct the likelihood function, given that $\theta_1 = a$ and $\theta_2 = \sigma$:

$$L = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-a)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-a)^2}{2\sigma^2}} \cdot \dots \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-a)^2}{2\sigma^2}}.$$

Here

$$L = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} e^{-\sum_{i=1}^{n} \frac{(x_i-a)^2}{2\sigma^2}}.$$

We find the log-likelihood function:

$$\ln L = -n \ln \sigma + \ln\left(\frac{1}{\left(\sqrt{2\pi}\right)^n}\right) - \frac{\sum_{i=1}^{n}(x_i - a)^2}{2\sigma^2}.$$

We find the partial derivatives of *a* and *σ*:

$$\frac{\partial \ln L}{\partial a} = \frac{\sum_{i=1}^{n} x_i - na}{\sigma^2}; \qquad \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i - a)^2}{\sigma^2}.$$

By equating the partial derivatives to zero and solving the resulting system of two equations for *a* and *σ* obtain:

$$a = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x_в}; \qquad \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x_в})^2}{n} = D_B.$$

Thus, the maximum likelihood estimation of the unknown: $a^* = \overline{x_в}$; $\sigma^* = \sqrt{D_B}$. Note that the first estimate of unbiased, and the second offset.

5.4. **Statistical hypothesis testing**

In practice often falls on the basis of sample surveys to test various assumptions (hypotheses) on the parameters of the law, or the distribution of the population.

Definition* 5.24.** The hypothesis of an unknown distribution or the parameters of the known distribution of the random variable, verifiable by sampling ***statistical call. Examples of statistical hypotheses:

1) general population Poisson distribution;
2) the variance of two normal populations are equal.

The first hypothesis was assumed as an unknown distribution, in the second - on the parameters of the two known distributions.

Along with the hypothesis put forward by considering and contrary to her hypothesis. If put forward the hypothesis is rejected, then there is a contrary hypothesis. For this reason, it is advisable to distinguish between these hypotheses.

Definition* 5.25.** Extended (verifiable) statistical hypothesis called ***zero (base) and H0 denote the contrary it is called a ***competing hypothesis (alternative)*** and designated H_1.

For example, if the null statistical hypothesis is to assume that the expectation a of normal distribution is equal to 5, then the competing hypothesis, in particular, may consist in the assumption that $a \neq 5$. Briefly is written as follows: H_0: $a = 5$; H_1: $a \neq 5$.

Statistical hypotheses are distinguished also by the number of the assumptions contained in the hypothesis.

Definition* 5.26.** Statistical hypothesis is called ***simple if it contains only one assumption about the parameter, or the distribution of the random variable. Otherwise, the statistical hypothesis is called ***complex***.

For example, the hypothesis H_0: $a = 5$ – simple hypothesis H_0: $a > 5$ – complex, because it involves a number of possible values of a.

Matching procedure nominated statistical hypotheses with the sample and the decision on the admissibility of these hypotheses is called statistical hypothesis testing.

Before formulating the problem of testing of statistical hypotheses in general, consider two examples.

*Example*5.17. There is a warehouse of finished products. It is known that products arrive at the warehouse parties with two companies that produce products of different quality, and the same batch are dispensed to the consumer. The products' quality is characterized by the probability p that the product is defective randomly selected from the lot. For one firm $p = p_0$, the other for $p = p_1$ ($p_0 > p_1$). Random user selects one batch of products. It should be based on test results to decide which company manufactured the selected batch of products.

Decision. H_0 – hypothesis consists in the fact that the chosen batch of poor quality products, i.e. the probability of defect is p_0; H_1 – the opposite hypothesis, the probability of defect is p_1. We will call H_0 – null and H_1 – competing hypotheses.

Shall select from the lot at random *n* units sold. *Y* denote the number of defective products from selected. It is clear that all possible values of *y*: 0, 1, 2, ..., *n* define the random variable which is denoted Y. A solution of the task is meant generation of the decision rule (criterion), which assigns to each possible value of the random variable Y from one or hypotheses H_0 or H_1.

We denote the set of possible values of the random variable *Y* through Δ, whereas according to the above, the desired decision rule consists in a plurality of partitioning Δ into pieces Δ_0 and Δ_1 . After contact with the possible values of the random variable *Y* into a plurality of Δ_0 hypothesis H_0 is adopted and, conversely, penetration of possible values at a plurality of Δ_1 hypothesis H_1 is adopted.

The question is what criterion for partitioning Δ sets into parts Δ_0 and Δ_0 to choose.

Example 5.18. At some point in time input receiving unit receives the random variable *Y*, which is the sum of a known signal *X* and the random noise *Z,* or audio interference. Produced measurement value *Y*. According to the numerical value obtained from the need to decide whether to present the input signal *X,* that is, to select one of two options: $y = x + z$ or $y = z$.

Decision. As the null hypothesis H_0 take the absence of a signal, and as a competing hypothesis H_1 – presence of signal. The task is to test the hypotheses H_0 relatively hypothesis H_1.

The set Δ of possible values of the random variable *Y* is the set of all real numbers *R* and displayed by all the points of the coordinate axis *Oy*. Seeking a decision rule is to break the *y*-axis coordinate into two parts: Δ_0 and Δ_1. You must select one of these partitions, resulting in the smallest possible risk in this problem in the decision.

***General statement of the problem*.** There are two opposing statistical hypothesis H_0 and H_1. Required to produce test the null hypothesis H_0 with respect to the competing hypothesis H_1 on the basis of the test results.

To test the null hypothesis using specially chosen random variable, exact or approximate distribution is known.

***Definition* 5.27.** Random variable, which is used to test the null statistical hypothesis, called a statistical criterion (criterion of consent or a decision rule).

After selecting a specific criterion, for example, the random variable *Y* set Δ of all possible values *y* at random variable *Y* is partitioned into two disjoint subsets Δ_0, and Δ_1 with the condition accepting the hypothesis H_0 when hit received value y of the random variable Y in the result of the experience Δ_0 and H_1- hypothesis – in contact y at Δ_1.

***Definition* 5.28.** The set of test values for which the hypothesis is accepted, called the domain hypothesis decision (tolerance range) at which reject the null hypothesis, called the critical region.

Since the criterion Y – dimensional random variable, all its possible values belong in a certain interval, so the choice of decision rule, i.e. rule set Δ partition into two parts Δ_0 and Δ_1 in any possible hypothesis testing problem as follows. Thus, in example 5.18. you can specify any number l_1 and put $\Delta_0 = (-\infty, l_1)\Delta_1 = (l_1, +\infty)$ (Fig. 5.4).

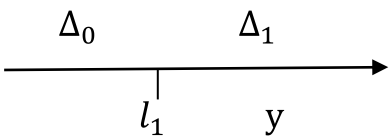

Fig. 5.4.

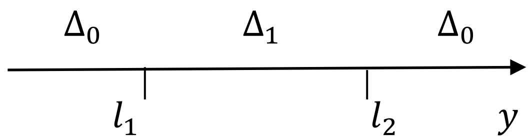

Fig.5.5

Another option is a decision rule is shown in Fig. 5.5.

The question is: which of these partitions should be preferred, or distracted from this example, which of all possible partitions in each task is considered the best, i.e., resulting in the lowest possible risk in this problem in the decision?

As a result, the statistical hypothesis testing in the two cases may be taken the wrong decision, i.e. may be mistakes of two kinds.

Error of the first kind It is that correct hypothesis is rejected.

Error of the second kind It is to be taken the wrong hypothesis.

Comment 1. The right decision can be taken in two ways:

1) the hypothesis is accepted, and in reality it is correct;
2) the hypothesis is rejected, and in fact it is false.

Comment 2. The probability of making a mistake of the first kind can be designated: it is called the *level of significance*. Most often taken to be the level of significance of 0.05 or 0.01. If, for example, the level of significance taken as equal to 0.05, this means that in five cases out of a hundred there is a risk to make a mistake of the first kind (to reject the correct hypothesis).α

Thus, the main principle of verification of statistical hypotheses can be formulated as follows: if the criterion value belongs to the observed critical area -

the hypothesis is rejected if the observed value of the criterion belongs to the area of acceptance of the hypothesis – the hypothesis is accepted.

5.5. The concept of the criteria for approval

In many cases, the practice on the basis of certain data to make assumptions about the form of the legitimate interests of our random variable X. However, the final decision on the form of the distribution law in such cases it is advisable to check whether this assumption is consistent with the experience. At the same time due to the limited number of observations experienced distribution law will usually be in some way different from the expected, even if the assumption that the distribution is done correctly the law. In this connection there is a need to solve the following problem: if the discrepancy between the experimental and anticipated distribution law distribution law consequence of the limited number of observations, or it is a significant and due to the fact that the actual distribution of the random variable is different from the intended. To solve this problem are so-called "approval criteria".

The idea of applying the criteria of fit is as follows.

Suppose, for example, on the basis of the statistical data we will test the hypothesis H, consists in the fact that the random variable *X* is the distribution function $F(x)$.

In order to accept or reject the hypothesis *H*, we consider the random variable Y, the degree of divergence characterizing the statistical and theoretical distributions. The value *Y* can be selected in various ways. For example, as *Y* can take the maximum deviation of the statistical distribution function $F^*(x)$ of the theoretical $F(x)$. Obviously, the law of distribution of the random variable Y depends on the law of distribution of the random variable *x*, over which the experiments were carried out, and the number of experiments n.

Assume that the distribution law of a random variable *X* is known to us.

Suppose that as a result of n experiments on the value of a random X variable Y is taken at a certain value *y*. The question is, is it possible to explain the assumed value of Y = *y* from accidental causes or whether it is too high and indicates a significant difference between the theoretical and statistical distributions, i.e. the unsuitability of the hypothesis *H*? To answer this question, let us assume that the true hypothesis *H*, and calculate the probability that a random variable *Y* due to accidental causes, associated with a limited amount of experimental material, will be set to no less than the observed value of *y*, i.e. calculate the probability $P(Y \geq y)$. If this probability is small, the hypothesis *H* should deny how improbable, and if this probability is significant, the experimental data do not contradict the hypothesis *H*.

To calculate the probability $P(Y \geq y)$ it is necessary to know the distribution law of a random variable Y, which, as we have said, depends on the distribution law of the random variable X (distribution function $F(x)$) and the number of experiments n. It turns out that in some ways to the selection of a random variable Y its distribution law is sufficiently large n is virtually independent of the random variable X. That law measures such differences and use in mathematical statistics as a criterion for approval.

The simplest criterion for testing the hypothesis that the distribution law is a Kolmogorov criterion represents a maximum value of the absolute value of the difference between the statistical distribution function $F^*(x)$ and the corresponding theoretical distribution function $F(x)$, i.e. $D = max|F^*(x) - F(x)|$.

A.N. Kolmogorov proved that no matter what kind of had a continuous distribution function $F(x)$ with an unlimited increase in the number of independent observations n, the probability of the inequality $D\sqrt{n} \geq \lambda$ tends to limit

$$P(\lambda) = 1 - \sum_{k=-\infty}^{+\infty} (-1)^k \, e^{-2k^2\lambda^2}. \qquad (5.22)$$

Probability $P(\lambda)$, a brief excerpt of which is given below:

λ	0.828	1,224	1,358	1,627	1,950
$P(\lambda)$	0.5	0.1	0.05	0.01	0,001

The scheme of the Kolmogorov test is as follows.

1. According to the n results of the measurements is based statistical distribution function $F^*(x)$.
2. On the same graph is constructed supposed theoretical distribution function $F(x)$.
3. It is determined by the maximum value of the module of the difference of their ordinates D (Fig. 5.6).
4. Calculated value $\lambda = D\sqrt{n}$.
5. According to the above table it is the probability $P(\lambda)$ corresponding to the fact that due to accidental causes maximum difference between $F^*(x)$ and $F(x)$ will not be less than is actually observed.

If the probability is very small, the hypothesis is rejected: the relatively high probability of the hypothesis is considered compatible with the experimental results. $P(\lambda)P(\lambda)$

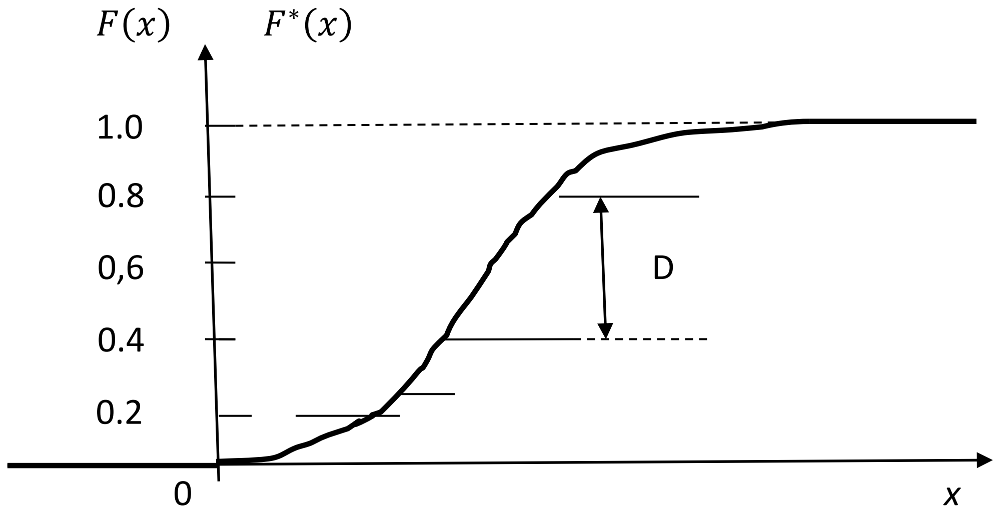

Fig. 5.6

Note that the Kolmogorov criterion can be applied only in the case where the hypothetical distribution $F(x)$ is completely known, i.e. not only known form of the distribution function $F(x)$, but also all the parameters appearing in it. If the parameters of the alleged law of distribution are determined on the basis of experimental data, the consent of the other criteria, such as test χ^2 (chi-square) are used.

A list of recommended basic and additional literature

Main literature.

1. E.S. Wentzel. Probability. // Textbook. Publishing house "Nauka", Moscow, 1969, 576 p.
2. E.I. Gursky. Probability theory with elements of mathematical statistics. // Tutorial. M: Publishing house of the "High School", 1971, 328 p.
3. V.E. Gmurman. Theory of Probability and Mathematical Statistics. // Tutorial. M: Publishing house of the "High School", 2003, 479 p.
4. L.Z .Rumshiskii. Elements of probability theory. // Tutorial. M: Publishing house "Science", 1966, 155 p.

Additional literature.

1. I.N. Volodin. Lectures on the theory of probability and mathematical statistics. // Textbook. Kazan, 2006, 271 p.
2. V.A. Coleman, V.N. Kalinin. Theory of Probability and Mathematical Statistics. // Textbook. Publ KNORUS 2009, 384 p.
3. A.I. Kibzun, E.R. Goryainov, A. Naumov, A.N. Sirotin. Theory of Probability and Mathematical Statistics. // Tutorial. M: Publishing house FIZMATLIT, 2002, 224 p.
4. A.N. Borodin Elementary course in probability and mathematical statistics. // Tutorial. SPb .: Lan, 1999.- 224 p.
5. V.A. Vatutin, G.I. Ivchenko and others. The theory of probability and mathematical statistics in problems. //2nd ed., Rev. - M .: Bustard, 2003.- 328 p.

APPENDIX 1. TABLE function of the Laplace $\overline{\Phi}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$

x	Φ (x)	x	Φ (x)	x	Φ (x)
0.00	0.0000	1.00	.8427	2.00	.9953
0.05	.0564	1.05	.8624	2.05	.9963
0.10	0.1125	1.10	.8802	2.10	.9970
0.15	.1680	1.15	.8961	2.15	0.9976
0.2	.2227	1.20	.9103	2.20	.9981
0.25	.2763	1.25	0.9229	2.25	.9985
0.3	.3286	1.30	.9340	2.30	.9988
0.35	.3794	1.35	.9438	2.35	0.9991
0.40	.4284	1.40	.9523	2.40	.9993
0.45	.4755	1.45	.9597	2.45	0.9995
0.50	.5205	1.50	.9661	2.50	0.9996
0.55	.5633	1.55	.9716	2.55	0.9997
0.60	.6039	1.60	.9736	2.60	0.9998
0.65	.6420	1.65	0.9804	2.65	0.9998
0.70	.6778	1.70	.9838	2.70	0.9999
0.75	.7112	1.75	.9867	2.75	0.9999
0.80	.7421	1.80	.9891	2.80	0.9999
0.85	.7707	1.85	.9911	3.00	1.0000
0.90	.7969	1.90	.9928		
0.95	.8209	1.95	.9942		

APPENDIX 2. Table of the Laplace function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	Φ (x)	x	Φ (x)	x	Φ (x)	x	Φ (x)	x	Φ (x)	x	Φ (x)
0.00	0,00000	0.50	0.19146	1.00	0.34134	1.50	0.43319	2.00	0.47725	3.00	0.49865
0.01	0.00399	0.51	0.19497	1.01	0.34375	1.51	0.43448	2.02	0.47831	3.05	0.49886
0.02	0.00798	0.52	0.19847	1.02	0.34614	1.52	0.43574	2.04	0.47932	3.10	0.49903
0.03	0.01197	0.53	0.20194	1.03	0.34849	1.53	0.43699	2.06	0.48030	3.15	0.49918
0.04	0.01595	0.54	0.20540	1.04	0.35083	1.54	0.43822	2.08	0.48124	3.20	0.49931
0.05	0.01994	0.55	0.20884	1.05	0.35314	1.55	0.43943	2.10	0.48214	3.25	0.49942
0.06	0.02392	0.56	0.21226	1.06	0.35543	1.56	0.44062	2.12	0.48300	3.30	0.49952
0.07	0.02790	0.57	0.21566	1.07	0.35769	1.57	0.44179	2.14	0.48382	3.35	0.49960
0.08	0.03188	0.58	0.21904	1.08	0.35993	1.58	0.44295	2.16	0.48461	3.40	0.49966
0.09	0.03586	0.59	0.22240	1.09	0.36214	1.59	0.44408	2.18	0.48537	3.45	0.49972
0.10	0.03983	0.60	0.22575	1.10	0.36433	1.60	0.44520	2.20	0.48610	3.50	0.49977
0.11	0.04380	0.61	0.22907	1.11	0.36650	1.61	0.44630	2.22	0.48679	3.55	0.49981
0.12	0.04776	0.62	0.23237	1.12	0.36864	1.62	0.44738	2.24	0.48745	3.60	0.49984
0.13	0.05172	0.63	0.23565	1.13	0.37076	1.63	0.44845	2.26	0.48809	3.65	0.49987
0.14	0.05567	0.64	0.23891	1.14	0.37286	1.64	0.44950	2.28	0.48870	3.70	0.49989
0.15	0.05962	0.65	0.24215	1.15	0.37493	1.65	0.45053	2.30	0.48928	3.75	0.49991
0.16	0.06356	0.66	0.24537	1.16	0.37698	1.66	0.45154	2.32	0.48983	3.80	0.49993
0.17	0.06749	0.67	0.24857	1.17	0.37900	1.67	0.45254	2.34	0.49036	3.85	0.49994
0.18	0.07142	0.68	0.25175	1.18	0.38100	1.68	0.45352	2.36	0.49086	3.90	0.49995
0.19	0.07535	0.69	0.25490	1.19	0.38298	1.69	0.45449	2.38	0.49134	3.95	0.49996

APPENDIX 3. The table of values t_β, which are the solution of the equation $2\int_0^{t_\beta} S_{n-1}(t)dt = \beta$ with given β and n, where $S_{n-1}(t)$ – Student's distribution

n	β			*n*	β		
	0.95	0.99	0,999		0.95	0.99	0,999
5	2.78	4.6	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				

APPENDIX 4. Table values μ are solutions of equation $\int_{\frac{\sqrt{n-1}}{(1+\mu)}}^{\frac{\sqrt{n-1}}{(1-\mu)}} R(\chi, n)d\chi = \beta$ given β and n, where $R(\chi, n)$ – distribution of "chi"

n	β			*n*	β		
	0.95	0.99	0,999		0.95	0.99	0,999
5	1.37	2.67	5.64	20	0.37	0.58	0.88
6	1.09	2.01	3.88	25	0.32	0.49	0.73
7	0.92	1.62	2.98	30	0.28	0.43	0.63
8	0.80	1.38	2.42	35	0.26	0.38	0.56
9	0.71	1.20	2.06	40	0.24	0.35	0.50
10	0.65	1.08	1.80	45	0.22	0.32	0.46
11	0.59	0.98	1.60	50	0.21	0.30	0.43
12	0.55	0.90	1.45	60	0.168	0.269	0.38
13	0.52	0.83	1.33	70	0.174	0,245	0.34
14	0.48	0.78	1.23	80	0.161	0.226	0.31
15	0.46	0.73	1.15	90	0,151	0,211	0.29
16	0.44	0.70	1.07	100	0,143	0.198	0.27
17	0.42	0.66	1.01	150	0,115	0,160	0.21
18	0.40	0.63	0.96	200	0,099	0.136	0,185
19	0.39	0.60	0.92	250	0.089	0.120	0.162

Questions for self-control Chapter 1

1. What does the theory of probability? Its basic concepts and tasks.
2. What events are called random? Give examples of random events.
3. What events are called equally possible? Give examples.
4. What events are called joint and uncooperative? Give examples.
5. What events form the sample space? Give examples.
6. What events are called the opposite? Give examples.
7. What events are called the amount, or the association of several events? Give examples.
8. What events are called the product, or the intersection of the (combined), multiple events? Give examples.
9. What events are called the difference, more events? Give examples.
10. Formulate statistical and classical definition of probability events. What is the difference between them and the similarity?
11. What is the sum of the probabilities of incompatible events, forming a space of elementary outcomes?
12. Formulate basic definitions and record the basic formula of combinatorics.
13. What is called the conditional probability? What events are called independent? Give examples.
14. Formulate the theorem of multiplication of probabilities and consequences of it.
15. Formulate the theorem of addition of probabilities and consequences of it.
16. Prove total probability formula.
17. Formulate theorem hypotheses and record Bayes formula.
18. When deciding which tasks apply the formula of total probability?
19. When deciding which tasks use the formula hypothesis probability (Bayes)?
20. Prove Bernoulli formula.
21. When deciding which tasks Bernoulli formula used?
22. Define the the most probable number of occurrences of events at successive independent tests, and provide its computing rule.

Questions for self-control Chapter 2

23. What does the theory of probability? Its basic concepts and tasks.
24. What events are called random? Give examples of random events.
25. What events are called equally possible? Give examples.
26. What events are called joint and uncooperative? Give examples.
27. What events form the sample space? Give examples.
28. What events are called the opposite? Give examples.
29. What events are called the amount, or the association of several events? Give examples.
30. What events are called the product, or the intersection of the (combined), multiple events? Give examples.
31. What events are called the difference, more events? Give examples.
32. Formulate statistical and classical definition of probability events. What is the difference between them and the similarity?
33. What is the sum of the probabilities of incompatible events, forming a space of elementary outcomes?
34. Formulate basic definitions and record the basic formula of combinatorics.
35. What is called the conditional probability? What events are called independent? Give examples.
36. Formulate the theorem of multiplication of probabilities and consequences of it.
37. Formulate the theorem of addition of probabilities and consequences of it.
38. Prove total probability formula.
39. Formulate theorem hypotheses and record Bayes formula.
40. When deciding which tasks apply the formula of total probability?
41. When deciding which tasks use the formula hypothesis probability (Bayes)?
42. Prove Bernoulli formula.
43. When deciding which tasks Bernoulli formula used?
44. define the the most probable number of occurrences of events at successive independent tests, and provide its computing rule.

Questions for self-control to Chapter 3

45. What does the theory of probability? Its basic concepts and tasks.
46. What events are called random? Give examples of random events.
47. What events are called equally possible? Give examples.
48. What events are called joint and uncooperative? Give examples.
49. What events form the sample space? Give examples.
50. What events are called the opposite? Give examples.
51. What events are called the amount, or the association of several events? Give examples.
52. What events are called the product, or the intersection of the (combined), multiple events? Give examples.
53. What events are called the difference, more events? Give examples.
54. Formulate statistical and classical definition of probability events. What is the difference between them and the similarity?
55. What is the sum of the probabilities of incompatible events, forming a space of elementary outcomes?
56. Formulate basic definitions and record the basic formula of combinatorics.
57. What is called the conditional probability? What events are called independent? Give examples.
58. Formulate the theorem of multiplication of probabilities and consequences of it.
59. Formulate the theorem of addition of probabilities and consequences of it.
60. Prove total probability formula.
61. Formulate theorem hypotheses and record Bayes formula.
62. When deciding which tasks apply the formula of total probability?
63. When deciding which tasks use the formula hypothesis probability (Bayes)?
64. Prove Bernoulli formula.
65. When deciding which tasks Bernoulli formula used?
66. define the the most probable number of occurrences of events at successive independent tests, and provide its computing rule.

Questions for self-control to Chapter 4

67. Define a random variable. What are the random variables are called discrete and continuous? Give examples.
68. What it is called the law of the random variable? Several distribution and polygon distribution of a discrete random variable, and their properties.
69. Define the probability distribution function. List and prove properties of the distribution function.
70. How knowing the distribution function to find the probability of a random variable falling in the specified range?
71. What is the difference between the graphs of the distribution of discrete and continuous random variables?
72. Define the probability density. List and prove properties of probability density. Does the concept of suitable probability density function for a discrete random variable?
73. How knowing the density of distribution, find the probability of a random variable falling in the specified range?
74. Give the definition of the expectation of a discrete random variable. As it can be interpreted statistically?
75. Give the definition of the expectation of a continuous random variable. As it can be interpreted statistically?
76. List the properties of the expectation of a random variable.
77. What is called fashion and the median of a random variable?
78. Define a random variable dispersion and list its properties.
79. Define the standard deviation of the random variable and clarify its meaning as the numerical characteristic of a random variable.
80. What is called the initial point of k-th order of a random variable?
81. What is called the focal point of the k-th order of a random variable?
82. What is called the binomial probability distribution? What is the expectation and variance of the random variable having a binomial distribution?
83. What is the probability distribution is called a Poisson distribution? What is the expectation and variance of the random variable distributed by the Poisson law?

84. What is the distribution of the random variable is called a uniform distribution? What is the expectation and variance of the random variable having a uniform distribution?
85. What is the distribution of the random variable is called the exponential distribution? What is the expectation and variance of the random variable having an exponential distribution?
86. What is the distribution of the random variable is called the normal distribution? What is the expectation and variance of the random variable distributed by the normal law?
87. How it is called the graph of the probability density of the normal distribution, and what are its properties?
88. What is the probability of a random variable contact having a normal distribution on the predetermined portion. Three sigma rule.
89. What is called the Laplace function and what are its properties?
90. What are the properties of random variable described by the third and fourth central moments?
91. What are the properties of random variable described by a coefficient of skewness and kurtosis?

Questions for self-control to Chapter 5

92. What is the essence of the law of large numbers?
93. Specify generalized Chebyshev theorem. What is the practical significance of Chebyshev's theorem?
94. Explain using Bernoulli's theorem, stability property relative frequencies.
95. What is the essence of the central limit theorem?