MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
"IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE"

M. Onai, V. Sushchuk-Sliusarenko, A. Dychka

FOUNDATIONS OF PROBABILITY THEORY

Study aid

Recommended by the Methodical Council of Igor Sikorsky Kyiv Polytechnic Institute
as a study aid for bachelor's applicants n the educational program
"Bachelor" specialty 121 (F2) "Software engineering»

Electronic network educational edition

Kyiv
Igor Sikorsky Kyiv Polytechnic Institute
2025

UDC

Authors: M. Onai, associate professor, PhD

V. Sushchuk-Sliusarenko, senior lecturer

A. Dychka, assistant professor, PhD

Reviewer: Artem VOLOKYTA, PhD, Associate Professor of the Department of Computer Engineering Faculty of Informatics and Computer Engineering Igor Sikorsky Kyiv Polytechnic Institute

Responsible editor: Oksana TARASENKO-KLIATCHENKO,
Associate Prfessor of SPSCS, Ph.D., Assc. Prof.

The approval was granted by the
Methodical Council of the Igor Sikorsky Kyiv Polytechnic Institute
(minutes of meeting No.5 dated March 6, 2025)
upon the submission of the Academic Council of the FAM
(minutes of meeting No.7 dated 27.01.2025)

Onai M.

Foundations of Probability Theory [Electronic resource]: study aid bachelor's applicants on the educational program "Bachelor" of the specialty 121 (F2) "Software engineering"/ M. Onai, V. Sushchuk-Sliusarenko, A. Dychka: Igor Sikorsky Kyiv Polytechnic Institute.-Electronic text data (1 file). – Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2025. – 205p.

A tutorial is designed for more detailed familiarization of students with theoretical information and practical techniques of the probability theory, as well as as a guide for use in practical classes and for independent work of students in the process of preparing for independent and control works, etc. The tutorial contains examples of the application of the probability theory methods in various fields of knowledge, brief theoretical materials related to the topic are presented, and examples of solving typical problems are given. A large number of tasks allows you to create a sufficient number of different options during independent work and as homework. The collection of problems is intended for students studying in specialty 121 (F2) "Software Engineering" of the Faculty of Applied Mathematics Igor Sikorsky Kyiv Polytechnic Institute.

UDC

CONTENT

Introduction

The subject of probability theory. All events (phenomena) we observe can be divided into reliable, impossible, and accidental. The subject of probability theory is the study of probabilistic laws of homogeneous random events. Methods of probability theory are widely used in various fields of science and technology. Probability theory is the basis for mathematical and applied statistics. Theoretical and practical training that the students receive while studying probability theory, contributes to the expansion of the scientific horizons of future specialists and is the theoretical basis for the study of special technical, economic, and social disciplines.

The purpose of studying the discipline is to form a mathematical foundation sufficient for the analysis of specific technical, economic, and physical systems. The study of the basics of probability theory makes it possible to apply the acquired knowledge to solve applied problems.

Upon completion of the discipline, the students must acquire the following skills and abilities:

- mastery of the basic terminology of the discipline, the ability to explain the content of basic concepts and sections;
- ability to classify standard problems by features, ability to solve them;
- mastery of skills in working with thematic literature, knowledge of basic textbooks, reference books, and tables. Ability to find the necessary information in the literature (ways to solve problems);
- knowledge of basic practical techniques for solving standard problems of probability theory (combinatorial methods, methods related to basic theorems and distributions.

Chapter 1

The concept of probability theory. Subject and tasks of the discipline. Basic definitions. Events and operations on events. Direct calculation of event probabilities

1.1. Classification of events. Basic definitions

One of the basic concepts of probability theory is the concept of an **event**.
A **random event** (or simply an event) is any fact that may or may not occur as a result of a trial.
A trial (research, experiment) will be considered the fulfillment of a certain set of conditions under which a particular result is recorded. The trial can be performed by a person, but it can also take place independently of the person acting as an observer.
The event is not an accident, but only a possible trial result. Events are indicated by capital letters of the Latin alphabet: $A, B, C...$
If in each trial, when an event A occurs, an event B occurs, it is said that the event B includes the event A and denotes $A \subset B$.
For example, if the event A is a 1st grade product, B - a 2nd-grade product, C - a standard product, then $A \subset C$ and $B \subset C$.
If at the same time $A \subset B$ and $B \subset A$, then in this case the events are called **equivalent** and denote $A = B$.
Events are said to be **incompatible** if one of the trials that occurs excludes the possibility of another. Otherwise, the events are called **compatible**.
For example, receiving two winnings on one lottery ticket is incompatible, and receiving the same winnings on two tickets is compatible. Receiving a student's exam in one discipline as "excellent", "good" and "satisfactory" - the events are

incompatible, and obtaining the same grades in exams in three disciplines - the events are compatible.

An event is said **to be valid** (denoted by the letter Ω or U) if it inevitably occurs as a result of the trial.

An event is said **to be impossible** (denoted by a symbol $\varnothing$ or V) if it cannot occur as a result of the trial.

For example, if all integrated circuits in a batch are standard, then the choice from a given batch of a standard circuit is a valid event, and the choice of a defective circuit under these conditions is an impossible event.

Events are said **to be equally possible** if, as a result of the trial, none of these events is objectively more possible.

For example, drawing an ace, a jack, a king, or a queen from a deck of cards or the appearance of a coat of arms or numbers when tossing a coin - events are equally possible. Thus, if the coin is made symmetrically, then there is no reason to consider the "appearance of the coat of arms" when tossing a coin as an objectively more possible event than the "emergence of numbers".

Equivalent events occur only in trials that have symmetry of possible outcomes, and our ignorance of which of the events is objectively more possible in the absence of symmetry of results cannot be a reason to consider events equally possible.

Several events are said **to be the only possible** ones if at least one of them is required as a result of the trial.

For example, events that may occur if there are two children in the family: A - "two boys", B - "one boy, one girl", C - "two girls" are the only possible.

Another example. Events that can occur in the case of 10 shots and get a certain number m: D - m <2, E - m <8, F - m> 5 are also the only possible because any result of the shooting will be at least one of these events (for example, at m = 9 – event F, at m= 1 – event D or E, and other events).

Several events form a **complete group** (complete system) if they are the only possible and incompatible trial results. This means that one and only one of these events must occur as a result of the trial. Thus, in the last two examples, events A, B, C form a complete group because they are the only possible and incompatible, and events D, E, F do not form a complete group because they are the only possible but compatible.

A special case of events that form a complete group are opposite events. Two incompatible events that form a complete group are called **opposites**. An event opposite to the event A is denoted $\overline{A}$.

For example, "appearance of the coat of arms" and "appearance of the number" when tossing a coin, "absence of defective processors" and "presence of at least one defective processor" in the batch of manufactured processors are the opposite events.

1.2. Operations on events

The sum of several events is an event that consists in the occurrence of one of these events. If A and B are compatible events, then their sum $A+B$ means that at least one of the events will occur. If these A and B are incompatible events, then their sum $A+B$ means that only one of the events will occur (A or B).

The product of several events is an event that consists in the simultaneous occurrence of all these events. If A, B, C are compatible events, then their product means that an event A, an event B, and an event C all will occur.

The difference between the two events A and B is the event, which means that event A occurs and event B does not occur.

◀**Example 1.1.** The winner of the competition receives an award: prize (event A), money reward (event B), or medal (event C). Name the events:

a) $A+B$;

b) ABC;

c) $AB-C$.

Solutions.

a) The event $A+B$ is that the winner receives either a prize, a money reward, or both awards.

b) The event ABC consists in awarding the winner with a prize, a money reward, and a medal.

c) The event $AB-C$ is that the winner receives a prize and a money reward, but is not awarded with a medal. ▶

1.2.1. Geometric interpretation of basic operations on events using Venn diagrams

Suppose, for example, that a point (a valid event Ω) is chosen at random inside a rectangle (Fig. 1.1), and the event A consists in hitting this point in a smaller circle (Fig. 1.1a) and the event B consists in hitting this point in a larger circle (Fig. 1.1b). Then the sum of events $A+B$ means that the point hits the entire shaded area of both circles (Fig. 1.1c), and the product AB means that the point hits the general part of the circles (Fig. 1.1d). In Fig. 1.1d and Fig. 1.1e,f shaded areas show that the events $\overline{A}$ and $\overline{B}$ are opposite to the events A and B, and Fig. 1.1g and 1.1h represent the differences of events $A-B$ and $B-A$.

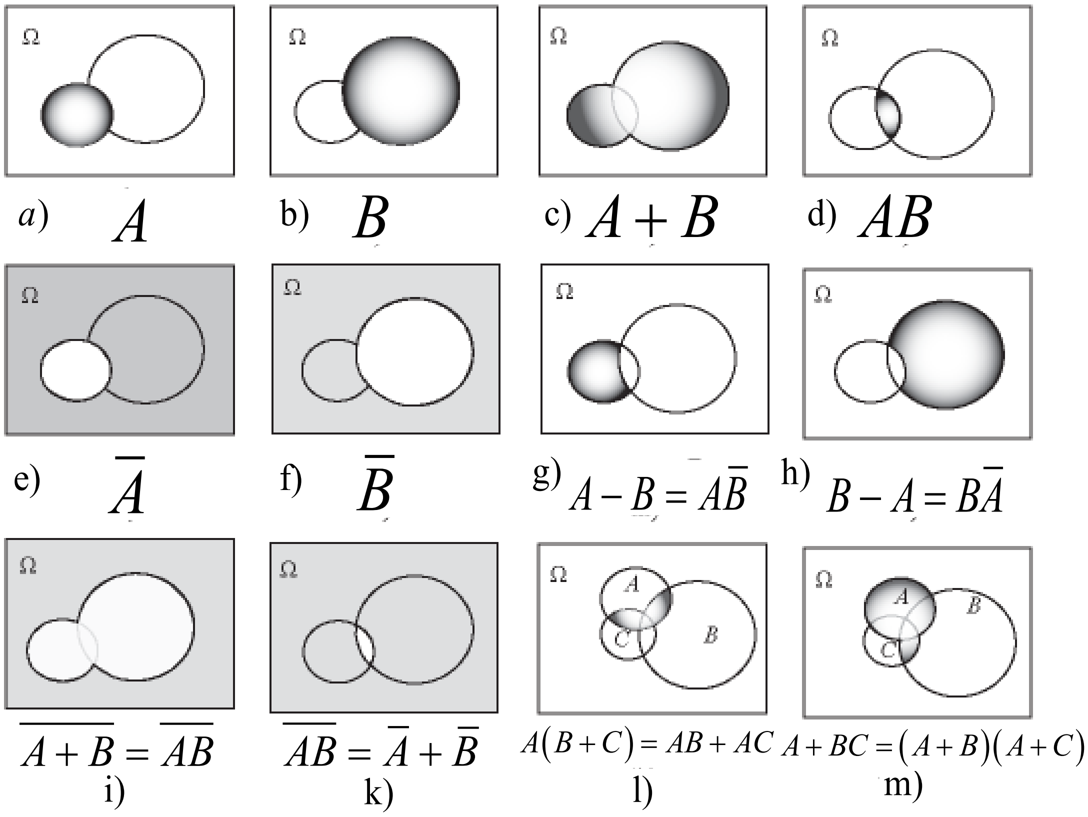


Fig. 1.1. Venn charts

◄ **Example 1.2.** Make sure the equations are true:

a) $\overline{A+B+...+K}=\overline{A}\overline{B}...\overline{K}$; *b*) $\overline{AB...K}=\overline{A}+\overline{B}+...+\overline{K}$.

The solution.

a) If the event $A+B+...+K$ is the occurrence of at least one of these events $A,B,...,K$ then the opposite event $\overline{A+B+...+K}$ means that none of these events will occur, i.e. the product of events $\overline{A}\overline{B}...\overline{K}$.

b) If the event $AB...K$ is the joint occurrence of all events $A,B,...,K$, then the opposite event $\overline{AB...K}$ means that none of these events will occur, i.e. we have a sum $\overline{A}+\overline{B}+...+\overline{K}$. In Fig. 1.1i and 1.1k the example of two events graphically represents similar relationships between events. ►

If the event A is the sum of incompatible events $A_1, A_2, ..., A_n$ $\left(A = A_1 + A_2 + ... + A_n\right)$, it is said that the occurrence of the event A is facilitated by individual n cases (options).

1.2.2. Properties of operations of addition and multiplication of events

1. $A + B = B + A$ - commutative addition.
2. $A + \left(B + C\right) = \left(A + B\right) + C$ - associative addition.
3. $AB = BA$ - commutativity of the product.
4. $A\left(BC\right) = \left(AB\right)C$ - associativity of the product.
5. $A\left(B + C\right) = AB + AC$; $A + BC = (A + B)(A + C)$ – distributivity laws.

The last two properties are graphically represented in Fig. 1.1l and 1.1m.
From the definition of operations on events, the obvious equations follow:
$A + A = A$, $AA = A$, $A + \Omega = \Omega$, $A\Omega = A$, $A + \varnothing = A$, $A\varnothing = \varnothing$.

1.3 Elements of Combinatorics

To successfully solve problems using the classical definition of probability, you need to know the basic rules and formulas of **combinatorics** - a branch of mathematics that studies, in particular, methods of solving combinatorial problems i.e. problems involving counting different combinations.
Let A_i, $\left(i = 1, 2, ..., n\right)$ be the elements of a finite large set. Here are two important rules that are often used to solve combinatorcial problems.

The amount rule

If the element A_1 can be selected in n_1 ways, the element A_2 - in other n_2 ways, the element A_3 - different from the first two n_3 ways, and so on, the element A_k - in n_k ways other than the first $(k-1)$, the choice of one of the elements: either A_1, or A_2, ..., or A_k can be done in $n_1 + n_2 + ... + n_k$ ways.

◄**Example 1.3.** From a box containing 150 integrated circuits of the 1st grade, 120 integrated circuits of the 2nd grade, and 30 integrated circuits of the 3rd grade, one integrated circuit is selected. In how many ways can you choose one integrated circuit of the first or second grade?

The solution.

The integrated circuit of the 1st grade can be selected in $n_1 = 150$ ways, of the 2nd grade - in $n_2 = 120$ ways. As a rule of sum, we have $n_1 + n_2 = 150 + 120 = 270$ ways to choose one integrated circuit of the 1st or 2nd grade. ►

Product rule

If the element A_1 can be selected by n_1 methods, after each such selection the element A_2 can be selected by n_2 methods and so on, after each $(k-1)$ selection the element A_k can be selected by n_k methods, then the selection of all elements A_1, A_2,..., A_k in the specified order can be done by $n_1 n_2 ... n_k$ methods.

◄**Example 1.4.** There are 30 people in the group. It is necessary to elect the mayor, his deputy, and trade unionist. In how many ways can this be done?

The solution.

Any of the 30 students may be elected as a mayor, any of the remaining 29 students may be elected as his deputy, and any of the remaining 28 students may be elected as a trade unionist, i.e. $n_1 = 30, n_2 = 29, n_3 = 28$. According to the rule of the product, the total number of ways to choose a mayor, his deputy, and trade unionist is equal to $n_1 n_2 n_3 = 24360$ ways. ►

Let us have a set of n different elements. From this set subsets of m elements ($0 \le m \le n$) can be formed. For example, from 5 elements *a, b, c, d, e* can be selected combinations of 2 elements - *ab, cd, eb, ba, ce,* and so on, combinations of 3 elements - *abc, cbd, cba, ead* and so on.

If subsets of n elements by m differ either in the composition of elements or in the order of their arrangement (or both), then such ordered m — element subsets are called **placements of n elements by m**. The number of placements of n elements by m is equal to

$$A_n^m = \underbrace{n(n-1)(n-2)...(n-m+1)}_{m \text{ multipliers}} \qquad (1.1)$$

or

$$A_n^m = \frac{n!}{(n-m)!}, \qquad (1.1a)$$

where $n!$ is the product of the first numbers of the natural series, or $n! = 1 \cdot 2 \cdot ... \cdot n$.

◄**Example 1.5.** The schedule of one day consists of 5 classes. Determine the number of possible schedule options when choosing from 11 disciplines.

The solution. Each version of the schedule is a set of 5 disciplines out of 11, which differs from other options in the composition of disciplines and their sequence (or both), that is, it is an arrangement of 11 elements to 5. The number of possible variants of the schedule, i.e. the number of placements from 11 to 5, is calculated by the formula (1.1 or 1.1a):

$$A_{11}^{5} = \underbrace{11\cdot 10\cdot 9\cdot 8\cdot 7}_{\text{5 multipliers}} = 55\ 440 \ ►$$

If subsets of elements differ only in the composition of elements, they are called combinations of n elements to m. The number of combinations of n elements to m is equal to

$$C_n^m = \frac{n(n-1)(n-2)...(n-m+1)}{1\cdot 2...m}, \tag{1.2}$$

or

$$C_n^m = \frac{n!}{m!(n-m)!}. \tag{1.2a}$$

Because by definition $0! = 1$ then $C_n^0 = 1$.

Properties of the number of combinations

$$C_n^m = C_n^{n-m} \tag{1.3}$$

$$C_n^m + C_n^{m+1} = C_{n+1}^{m+1} \tag{1.4}$$

◄**Example 1.6.** 16 participants take part in the chess tournament. How many games will be played in the tournament if every two of the 16 participants meet only once?

The solution. Each game, which is played between two participants out of 16, differs from the others only in the composition of the participants, i.e. it is a combination of 16 elements of 2. According to the formula (1.2 or 1.2a):

$C_{16}^2 = \dfrac{16 \cdot 15}{1 \cdot 2} = 120$. ►

Ordered sets that differ only in the order of the elements (i.e., can be formed from the same set) are called permutations of this set. The number of permutations of the set of n elements is equal to:

$$P_n = n! \qquad (1.5)$$

◄**Example 1.7.** The order of performance of 7 participants in the competition is determined by drawing lots. In how many ways can you make a list of contrialants according to the draw?

The solution. Each draw variant differs only in the order of the contrial participants, i.e. it is a permutation of 7 elements. Their number by formula (1.5): $P_7 = 7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$. ►

If placements from n elements on some of the m elements (or all) are the same, then such placements are called **placements with repetitions of n elements on m**.

For example, from 5 elements *a, b, c, d, e* on 3 placements with repetitions will be *abc, cba, bed, cdb, bbe, ebb, beb, ddd* and so on, combinations with repetitions will be *abc, bed, bbe, ddd* and so on.

The number of placements with repetitions of n elements on t is equal to

$$\tilde{A}_n^m = n^m, \qquad (1.6)$$

and the number of combinations with repetitions of *n* elements by *t* is equal to

$$\tilde{C}_n^m = C_{n+m-1}^m , \qquad (1.7)$$

where C_{n+m-1}^m is determined by the formula (1.2) or the formula (1.2a).

◄**Example 1.8.** 10 films take part in the competition with 5 nominations. In how many ways can the prizes be distributed, if for each nomination there are: *a*) different prizes; *b*) the same prizes?

The solution.

a) Each of the prize distribution options is a subset of 5 films out of 10, which differ from others both in the composition of films and their order of nominations (or both), and the same films can be repeated several times, i.e. we have a placement with repetitions of 10 elements of 5. Their number by formula (1.6) is equal to

$$\tilde{A}_{10}^5 = 10^5 = 100\ 000 .$$

b) If the same prizes are set for each nomination, the order of the films in the subset of 5 winners does not matter, and the number of prize distribution options is the number of combinations with repetitions of 10 elements of 5, which is determined by the formula (1.7):

$$\tilde{C}_{10}^5 = C_{10+5-1}^5 = C_{14}^5 = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 2002 \ . \ ►$$

If there are different k elements in the permutations of the total number of n elements, and the 1st element is repeated n_1 times, the 2nd element - n_2 times, k -th element - n_k times, then such permutations are $n_1 + n_2 + ... + n_k = n$, are called **permutations with repetitions of *n* elements**. The number of permutations with repetitions of the elements is equal to

$$P_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1! \cdot n_2! \cdot ... \cdot n_k!} \quad . \qquad (1.8)$$

◄**Example 1.9.** How many seven-digit numbers exist, consisting of the digits 4, 5 and 6, in which the digit 4 is repeated 3 times, and the digits 5 and 6 are repeated 2 times?

Solutions.

Each seven-digit number differs from the other in the order of digits ($n_1 = 3$, $n_2 = 2$, and $n_3 = 2$, and their sum is 7), i.e. it is a permutation with repetitions, which includes 7 elements. Their number by formula (1.8) is equal to:

$$P_7(3;2;2) = \frac{7!}{3! \cdot 2! \cdot 2!} = 210 \ . \ ►$$

1.4. The classic definition of probability

In practice, it is important to be able to compare events according to the degree of possibility of their occurrence. The events: "precipitation in the form of rain" and "precipitation in the form of snow" on the first day of summer, as well as "winning on one ticket" and "winning on each of the n purchased tickets" in the lottery have different degrees of their onset. The numerical measure of the degree of objective possibility of an event is called the **probability of the event**. This definition, although it reflects the concept of probability of an event, is not mathematical. So, it is necessary to quantify it. Let the results of some trial form a complete group of events, i.e. these events are the only possible, incompatible, and equally possible. Such results are called **elementary results**, cases or chances. Under such conditions, the trial is reduced to a case diagram or "urn scheme" (because any probabilistic task for this trial can be replaced by an equivalent task with urns and balls of different colors).

A case is said to be **favorable for event *A*** if its occurrence causes event *A*. According to the **classical definition**, the probability of event *A* is equal to the ratio of the number of cases that contribute to event *A* to the total number of cases, i.e.

$$P(A) = \frac{m}{n}, \tag{1.9}$$

where $P(A)$ - the probability of the event A; m - the number of cases that contribute to the occurrence of the event; n - total number of cases.

◄**Example 1.10.** The dice are thrown once. The number of possible outcomes is the number of points that can be received (1, 2, 3, 4, 5, 6). What is the probability of an even number of points?
Solutions. All results $n = 6$ form a complete group of events and are equally possible. Event A - "the appearance of an even number of points" contributes to 3 results (cases) - 2, 4, and 6 points. According to the formula (1.9):

$$P(A) = 3/6 = 1/2.$$ ►

1.4.1. Event probability properties

1. For any event A, the inequality holds:

$$0 \le P(A) \le 1. \tag{1.10}$$

2. The probability of a reliable event is equal to one, i.e.

$$P(\Omega) = 1. \tag{1.11}$$

3. The probability of an impossible event is zero, i.e.

$$P(\varnothing) = 0. \tag{1.12}$$

The properties are obvious because $P(A)=\frac{m}{n}$, while the number of m contributing cases for any event satisfies the inequality $0 \le m \le n$, for a reliable event it is equal $m=n$ and for an impossible event $m=0$.
Events whose probabilities are close to zero or close to one are therefore called **practically impossible** or **practically reliable** events.

1.5. Statistical definition of probability

The classical definition of probability is used only for those events that may occur as a result of trials with symmetry of results, i.e. those that can be reduced to the urn scheme. However, there is a large class of events whose probabilities cannot be calculated using the classical formula. These are primarily events that are not equivalent to trial results. For example, if a coin is deformed, then obviously the events "appearance of the coat of arms" and "appearance of a number" when tossing a coin cannot be considered equally possible, and formula (1.9) to calculate the probability of any of them will be unsuitable.
However, there is another approach to estimating the probability of events, which is based on how often this event will appear in the trials. In this case, the **statistical definition of probability** is used.
The **statistical probability of event A** is the relative frequency of occurrence of this event in the trials, i.e.

$$\tilde{P}(A)=w(A)=\frac{m}{n}, \tag{1.13}$$

where $\tilde{P}(A)$ - the statistical probability of the event A;

$w(A)$ - relative frequency of the event A;

m - the number of trials in which the event A occurred;

n - total number of trials.

In contrast to the "mathematical" probability $P(A)$, which is considered in classical definition, statistical probability $\tilde{P}(A)$ is an experimental characteristic.

The statistical definition of probability does not apply to any events with an uncertain outcome, which in practice are considered random, but only to those that have certain properties, namely:

1. These events must be the result of only those trials that can be reproduced an unlimited number of times without any change in the trial conditions. For example, it is impractical to ask the question of determining the likelihood of hostilities, the emergence of ingenious works of art, and other similar events, given that it is impossible to fully reproduce the conditions for the necessary trials.

2. Events must have so-called statistical stability or relative frequency stability. This means that in different series of trials, the relative frequency of the event varies insignificantly (the smaller, the greater the number of trials), fluctuating within certain limits relative to a constant number. Such a constant number is the probability of the event.

3. The number of trials that result in an event A should be large enough, because only in this case can the probability $P(A)$ of the event be considered approximately equal to its relative frequency.

Thus, it can be argued that probability theory studies only those events for which it is valid not only to assert their randomness but also to estimate the relative frequency of their occurrence.

1.6. Geometric definition of probability

To calculate the probability of an event A according to the classical formula, the number of possible trial results must be finite. This shortcoming of the classical definition can be eliminated by using the **geometric definition of probability**, i.e. by finding the probability of a point hitting a certain area (segment, part of a plane, etc.). Suppose, for example, that the flat figure *g* is part of the flat figure *G*. A point is randomly thrown at the figure *G*. This means that all points in the area *G* are "equal" to the randomly thrown point. Assuming that the probability of event A is the hit of a thrown point on the figure *g* is proportional to the area of this figure and does not depend neither on its location about *G*, nor on the form of *g*, we find

$$P(A) = \frac{S_g}{S_G}, \tag{1.14}$$

where S_g and S_G are the areas of *g* and *G*, respectively (Fig. 1.2).

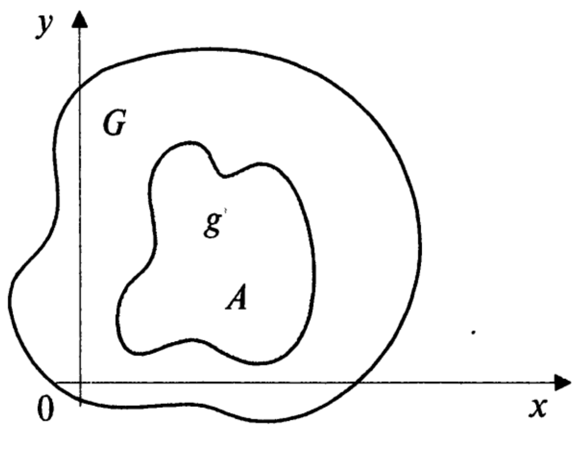
The figure *g* is called the one that contributes to the event. The area covered by the concept of geometric probability can be one-dimensional (straight line, segment) and three-dimensional (some object in space). Denoting the measure (length, area, volume) of the area through *mes*, we come to the following definition.

The **geometric probability** of an event is the ratio of the measure of the region that contributes to the occurrence of event A to the measure of the whole region, i.e.

$$P(A) = \frac{mes\ g}{mes\ G}. \tag{1.15}$$

◄ **Example 1.11.** Two people (A and B) agreed to meet in a certain place, agreeing only that everyone appears there at any time between 11 and 12 o'clock and waits for 30 minutes. If the partner has not yet arrived or has already left the agreed place, the meeting will not take place. Find the probability that the meeting will take place.

The solution. Let us denote the moments of arrival at the meeting place of persons A and B, respectively, by x and y. In the rectangular coordinate system of Oxy, take 11 hours as the starting point and 1 hour as the unit of measurement. According to the task $0 \le x \le 1,\ 0 \le y \le 1$. These inequalities are satisfied by the coordinates of any point belonging to the square *OKLM* with a side equal to 1 (Fig. 1.3). Event C - a meeting of two people - will take place if the difference between x and y does not exceed 0,5 hours. (in absolute value), i.e. $|y - x| \le 0{,}5$. The solution to the last inequality is the band $x - 0{,}5 \le y \le x + 0{,}5$ inside the square of Fig. 1.3 is represented by the shaded area g. By formula (1.14): $P(C) = \dfrac{S_g}{S_G} = \dfrac{1 - 2(1/2) \cdot 0{,}5^2}{1^2} = 0{,}75$, since the area of the region g is equal to the area of the square G without the sum of the areas of the two angular (unshaded) triangles. ►



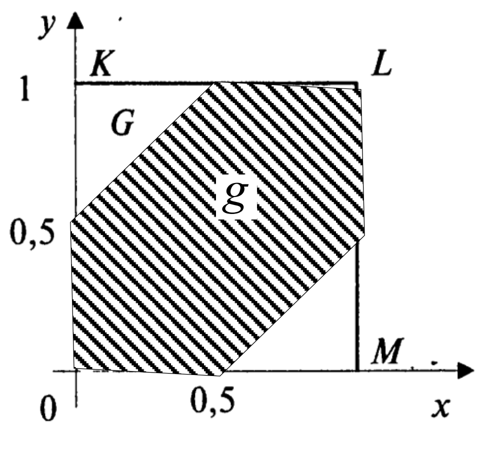


Fig. 1.2. Geometric probability   Fig. 1.3 Illustration to example 1.11

1.7. Direct calculation of probabilities

Its classical definition is used to directly calculate the probability.

◄**Example 1.12.** The letters T, E, H, Y, R, O are written on separate cards. The child takes cards at random and lays out: a) 3 cards; b) all 6 cards. What is the probability that the word will come out: *a*) "TOR";
b) "THEORY"?

The solution. *a*) The event A - the word "TOR" is composed. The different combinations of the three letters out of the available six are the placement, as they may differ both in the composition of the letters included in the combination and their order in the resulting sequence (or both), i.e. the total number of cases is represented by the formula $n = A_6^3$, and favorable for A is only 1 case (let it be *m*). According to the formula (1.9):

$$P(A) = \frac{m}{n} = \frac{1}{A_6^3} = \frac{1}{6 \cdot 5 \cdot 4} = \frac{1}{120}.$$

.

b) the event B - the word "THEORY" is composed. Different combinations of six letters are permutations, as they differ only in order in possible sequences; that is, the total number of cases $n = P_6 = 6!$, and there are only one case favorable to the event. So $P(B) = \frac{m}{n} = \frac{1}{P_6} = \frac{1}{6!} = \frac{1}{720}$. ►

◄**Example 1.13.** Using the condition of example 1.3, find the probability that the word "PINEAPPLE" is composed, if on separate cards are written three letters P, two letters E, one letter N, one letter I, one letter A and one letter L.

The solution. Let the event B be the receipt of the word "PINEAPPLE". The total number of cases $n = P_9 = 9!$, but now the number of cases m that contribute to the event B is much higher, because the permutation of the three letters P can be carried out $P_3 = 3!$ ways, and the permutation of two letters E - $P_2 = 2!$ ways without changing the word "PINEAPPLE" made up of cards. As a rule, the product $m = P_3 \cdot P_2$. So, $P(B) = \frac{m}{n} = \frac{P_3 \cdot P_2}{P_9} = \frac{3! \cdot 2!}{9!} = \frac{1}{30240}$.

The problem can be solved in another way, considering the combinations of letters as permutations with repetitions, of which the event B is facilitated by 1 combination:

$$P(B) = \frac{1}{P_9(3;2;1)} = 1 \Big/ \left(\frac{9!}{3! \cdot 2! \cdot 1!} \right) = \frac{1}{30240} \text{ .} \blacktriangleright$$

◄**Example 1.14.** Out of 30 students, 10 have athletic titles. What is the probability that 3 randomly selected students have athletic titles?

Solutions. The event A - 3 students, who were randomly selected have athletic titles. The total number of cases of choosing 3 students out of 30 is equal $n = C_{30}^3$. Similarly, the number of cases favoring event A is equal $m = C_{10}^3$.

So, $P(A) = \frac{m}{n} = \frac{C_{10}^3}{C_{30}^3} = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} : \frac{30 \cdot 29 \cdot 28}{1 \cdot 2 \cdot 3} = \frac{61}{203} \approx 0{,}030$. ►

◄**Example 1.15.** 4 passengers entered the elevator on the 1st floor of a nine-story building, each of whom can get out independently of each other on any floor from the 2nd to the 9th. What is the probability that all passengers will leave: *a*) on the 6th floor; *b*) on the same floor?

The solution. *a*) Let's consider the event A - all passengers will leave on the 6th floor. Each passenger can get from the 2nd to the 9th floor in 8 ways. According to the product rule, the total number of ways for four passengers to get out of the elevator is equal to $n = 8 \cdot 8 \cdot 8 \cdot 8 = 8^4$. The number of cases favoring the event A is equal $m = 1$. So,

$$P(A) = \frac{m}{n} = \frac{1}{8^4} = 0{,}00024.$$

b) Let's consider the event B - all passengers will leave on the same floor. Now $m = 8$ cases will favor the event B (all passengers will leave either on the 2nd floor, or on the 3rd, ..., or on the 9th floor). So,

$$P(B) = \frac{m}{n} = \frac{8}{8^4} = \frac{1}{8^3} = 0{,}00195.$$

(The total number of ways to get passengers out of the elevator can be found differently if we take into account that the combinations of floor numbers on which each of the four passengers can get off the elevator, for example, 3456, 4356, 4433, 5666, 5555, 9785 and so on, are arrangements of 8 elements (floors) of 4 allowing the use of place more than once. Their number is equal to $n = \tilde{A}_8^4 = 8^4$). ►

◄**Example 1.16.** Under the terms of the "Sportloto 6 out of 45" lottery, a lottery participant who guessed 4, 5, 6 sports receives a cash prize. Find the probability that: a) all 6 digits; b) 4 digits will be guessed.

The solution. a) Let's consider the event A - the participant of the lottery guessed all 6 sports out of 45. The total number of all cases (all options for filling in sports lotto cards) equals $n = C_{45}^{6}$, because each option differs only in the composition of sports. Favorable for the event A is only one case. So,

$$P(A) = \frac{1}{C_{45}^{6}} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40} \approx 10^{-8} .$$

b) Let's consider the event B - the lottery participant guessed 4 sports out of 6 winners. First, find the number of ways you can choose 4 sports out of 6 winnings, i.e. C_{6}^{4}. But to each combination of 4 winning sports, you need to add a combination of 2 sports that didn't win. The number of such combinations equals C_{39}^{2}. As a rule, the total number of cases that are favorable for the event B is equal to $m = C_{6}^{4} \cdot C_{39}^{2}$. So,

$$P(B) = \frac{m}{n} = \frac{C_{6}^{4} \cdot C_{39}^{2}}{C_{45}^{6}} = 0{,}00136 \ . \blacktriangleright$$

◄**Example 1.17.** There are 100 integrated circuits (ICs) in the batch, 4 of which are defective. The batch is arbitrarily divided into two equal parts, which are sent to two consumers. What is the probability that all defective ICs : *a*) will reach one consumer; *b*) will be distributed equally among consumers?

The solution. *a*) Let's consider the event A - all defective ICs will come to one consumer. The total number of ways to select 50 ICs out of 100 is $n = C_{100}^{50}$.

Cases,in which out of the 50 ICs sent to one consumer, there will be either 46 standard out of 96 (and all 4 defective) products, or 50 standard out of 96 (and 0 defective) are favorable for the event A; their number equals

$m = C_{96}^{46} \cdot C_{4}^{4} + C_{96}^{50} C_{4}^{0}$. So,

$$P(A) = \frac{m}{n} = \frac{C_{96}^{46} \cdot C_{4}^{4} + C_{96}^{50} \cdot C_{4}^{0}}{C_{100}^{50}} = \frac{C_{96}^{46} \cdot 1 + C_{96}^{46} \cdot 1}{C_{100}^{50}} = \frac{2C_{96}^{46}}{C_{100}^{50}} = \frac{2 \cdot 96! \cdot 50! \cdot 50!}{46! \cdot 50! \cdot 100!} =$$

$$= \frac{2 \cdot 96! \cdot 46! \cdot 47 \cdot 48 \cdot 49 \cdot 50}{46! \cdot 96! \cdot 97 \cdot 98 \cdot 99 \cdot 100} = 0{,}117,$$

(100! = 96! · 97 · 98 · 99 · 100, 50! = 46! · 47 · 48 · 49 · 50).

b) Let's consider the event B - there are 2 defective ICs in each delivery. Cases in which out of 50 ICs sent to one consumer, there will be 48 standard ones out of 96 and 2 defective ones out of 4 are favorable for the event B, their number equals $m = C_{96}^{48} \cdot C_4^2$.

So, $$P(B) = \frac{m}{n} = \frac{C_{96}^{48} \cdot C_4^2}{C_{100}^{50}} = \frac{96! \cdot 4! \cdot 50! \cdot 50!}{48! \cdot 48! \cdot 2! \cdot 2! \cdot 100!} =$$

$$= \frac{96!(2! \cdot 3 \cdot 4)(48! \cdot 49 \cdot 50)^2}{(48!)^2 2! \cdot 2(96! \cdot 97 \cdot 98 \cdot 99 \cdot 100)} = \frac{3 \cdot 4(49 \cdot 50)^2}{2 \cdot 97 \cdot 98 \cdot 99 \cdot 100} = 0{,}383 \text{ . ►}$$

◄ **Example 1.18.** The store sold 21 of the 25 computers of three brands (5, 7 and 13 computers, respectively). Assuming that the probability of being sold for a computer of each brand is the same, find the probability that computers of : a) one brand; b) three different brands were not sold.

The solution. *a*) Let's consider the event A - there are computers of only one brand left not sold. The total number of ways to get 4 (unsold) computers out of 25 is $n = C_{25}^4$. The number of ways to get 4 computers of the first brand out of 5 is equal to $m_1 = C_5^4$; of the second brand out of 7 - $m_2 = C_7^4$ and of the third brand out of 13 - $m_3 = C_{13}^4$. By rule of sum,

$m = m_1 + m_2 + m_3 = C_5^4 + C_7^4 + C_{13}^4$ cases are favorable to the event A. So,

$$P(A) = \frac{m}{n} = \frac{C_5^4 + C_7^4 + C_{13}^4}{C_{25}^4} = \frac{5 + 35 + 715}{12650} = \frac{755}{12650} = 0{,}060.$$

b) Let's consider the event B - there are computers of three different brands left not sold. The event B can take place in one of three ways. According to the first option, the event B will take place if there are 1, 1, 2 computers of the 1st, 2nd, and 3rd brands left, respectively; the second option - 1, 2, 1 and the third option will be 2, 1, 1 computers, respectively, of the 1st, 2nd and 3rd brands. Since there were 5 computers of the 1st brand, 7 of the 2nd brand, and 13 computers of the 3rd brand before the sale, the number of cases favorable for the first option is equal to $m_1 = C_5^1 C_7^1 C_{13}^2$ according to the product rule; for the second option - $m_2 = C_5^1 C_7^2 C_{13}^1$; for third option - $m_3 = C_5^2 C_7^1 C_{13}^1$. The total number of favorable cases is equal $m = m_1 + m_2 + m_3$. Now

$$P(B) = \frac{m}{n} = \frac{m_1 + m_2 + m_3}{n} = \frac{C_5^1 C_7^1 C_{13}^2 + C_5^1 C_7^2 C_{13}^1 + C_5^2 C_7^1 C_{13}^1}{C_{25}^4} =$$

$$= \frac{5 \cdot 7 \cdot 78 + 5 \cdot 21 \cdot 13 + 10 \cdot 7 \cdot 13}{12650} = \frac{5005}{12650} = 0{,}396 \ . \ ►$$

◄ **Example 1.19.** There are $m = 25$ students in the classroom. Find the probability that at least two students have birthdays on the same date. At what number of students the probability of this event is not less than 0.95%? (We consider that the possibility of having a birthday on any day of the year is equal.)

The solution. Let's consider the event A - the birthdays of at least two students out of m students present in the classroom coincide. Find the probability of the opposite event— the birthdays of all students are different. The number of events that are favorable for the event $\overline{A}$ is the number of placements from $n = 365$ the elements (days of the year) by m, i.e. A_n^m. The total number of cases is also determined by the number of placements of n elements on m, but placements with repetitions, i.e. $\tilde{A}_n^m = n^m$. According to the classical definition of probability:

$$P(\overline{A}) = \frac{A_n^m}{n^m} = \frac{n(n-1)...(n-m+1)}{n^m}$$

and for $n = 365$: $P(A) = 1 - P(\overline{A}) = 1 - \dfrac{365 \cdot 364...(365-m+1)}{365^m}$.

Provided that the $m = 25$ probability of the event A will be $P(A) = 0{,}569$.

When calculating the probability $P(A)$ for different m, it is easy to see that the inequality $P(A) > 0{,}95$ will be met with $m \geq 47$, i.e. only 47 students in the classroom, with a probability of at least 0.95, to say that at least two of them have the same birthdays. ►

Control questions

1. What events are called compatible (incompatible)? Give examples.
2. Using Venn diagrams, illustrate and describe the events:

   a) $\overline{A+B}$; b) $\overline{A}+\overline{B}$.
3. Calculate: P_5; A_7^3; C_{10}^4 .
4. Definition of probability: classical, statistical, geometric. Their differences and areas of application.
5. The task. The store sold 25 of the 30 computers from three manufacturers, 7, 8 and 15 computers were sold, respectively. Assuming that the probability of a computer being sold for each firm is the same, find the probability that the remaining computers are: *a*) from one manufacturer; *b*) three different manufacturers.

Chapter 2

Basic theorems and formulas of probability theory

2.1. Probability addition theorem

Theorem 2.1. The probability of the sum of a finite number of incompatible events is equal to the sum of the probabilities of these events:

$$P\left(A+B+...+K\right)=P\left(A\right)+P\left(B\right)+...+P\left(K\right) . \quad (2.1)$$

Proof. We prove the theorem for two events. Suppose that as a result of trialing from n the total number of equally possible and incompatible (elementary) trial results, m_1 cases are favorable for the event A, and m_2 cases are favorable for the event *B* (Fig. 2.1). According to the classical definition $P(A)=\frac{m_1}{n}$, $P(B)=\frac{m_2}{n}$.

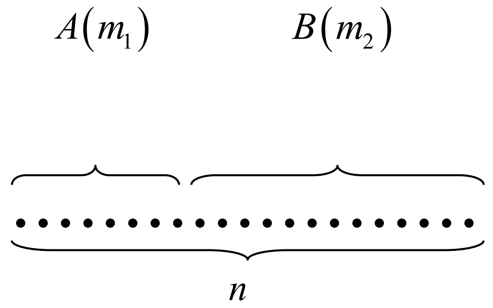


Fig. 2.1. Probability of the sum of two incompatible events

Because the events A and B are incompatible, none of the cases are favorable for one of these events are favorable for the other. Therefore, cases m_1+m_2 are favorable for the events *A*+*B*. So,

$$P(A+B)=\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}=P(A)+P(B).$$

Similar considerations can be applied to any finite number of events. ●

Corollary 1. The sum of the probabilities of events that form a complete group is equal to one:

$$P(A)+P(B)+...+P(K)=1. \quad (2.2)$$

Proof. If events $A,B,...,K$ form a complete group, they are the only possible and incompatible. Then the event $A+B+...+K$ is the occurrence as a result of trialing at least one of these events and is reliable, i.e. its probability is equal to one: $P(A+B+...+K)=1$ ●

Corollary 2. The sum of the probabilities of opposite events is equal to one:

$$P(A)+P(\overline{A})=1 \quad (2.3)$$

Proposition (2.3) follows from the fact that opposite events form a complete group.

◄**Example 2.1.** The probability of failure of the product during operation up to one year is 0.13, and during operation up to 3 years - 0.36. Find the probability of product failure during operation from 1 year to 3 years.

The solution. Let's consider the events A,B,C - the failure of products during operation for a period of 1 year, from 1 year to 3 years, more than 3 years, and according to the task $P(A)=0{,}13$; $P(C)=0{,}36$. Obviously, $C=A+B$ where A and B are incompatible events. According to the addition theorem, $P(C)=P(A)+P(B)$, whence

$$P(B)=P(C)-P(A)=0{,}36-0{,}13=0{,}23. \text{ ►}$$

Remark. The considered addition theorem is applied only to incompatible events and the attempt to use it for compatible events leads to incorrect and even absurd results. For example, let the *A* be the probability of winning in any lottery, i.e. *P* (*A*) = 0.05.
100 tickets were purchased
(i = 1, 2, ..., 100). Then, applying the addition theorem, we obtain that the probability of winning for at least one of the 100 tickets will be

$$P(A_1 + A_2 + \ldots + A_i + \ldots + A_{100}) = P(A_1) + P(A_2) + \ldots + P(A_{100}) =$$
$$= \underbrace{0,05 + 0,05 + \ldots + 0,05}_{\text{100 times}} = 5 \cdot$$

The absurdity of the answer (the probability of any event can not be more than 1, is due to the unsuitability in this case of the addition theorem, because the winnings for each ticket, i.e. events $A_1, A_2, \ldots, A_{100}$ are compatible events.

2.2. Conditional probability of the event. Probability multiplication theorem. Independent events

Probability $P(B)$ as a measure of the degree of objective possibility of the event B makes sense under a certain set of conditions. When conditions change, the probability of the event B may change. Thus, if a new condition *A* is added to the set of conditions under which the probability $P(B)$ is determined, then the obtained probability of the event B, found under the condition that the event occurred, is called the **conditional probability of event *B*** and is denoted $P_A(B)$.

◄**Example 2.2.** There are 5 details in the box, including 3 standard and 2 defective. Two details are removed from it in turn (with return and without return). Find the conditional probability that the second detail is standard, provided that the first detail is: a) standard; b) non-standard.

The solution. Let the events A and B be the removal of the standard detail, respectively, on the 1st and 2nd turn. Then $P(A)=\frac{3}{5}$. If the first removed detail is returned to the box, the probability of removing the standard detail a second time $P(B)=\frac{3}{5}$. If the detail taken for the first time is not returned to the box, the probability $P(B)$ of taking the standard detail for the second time depends on what was the first detail taken - standard (event A) or defective (event $\overline{A}$). In the first case $P_A(B)=\frac{2}{4}$, in the second case $P_{\overline{A}}(B)=\frac{3}{4}$, because among the four remaining details, the standard will be respectively 2 or 3. ►

From the total number n of equally possible and incompatible (elementary) trial results, m cases are favorable fort he event A, k cases are favorable fort he event B, and l cases ($l \le m$, $l \le k$) are favorable for the joint occurrence of events A and B i.e. AB event. (Fig. 2.2).

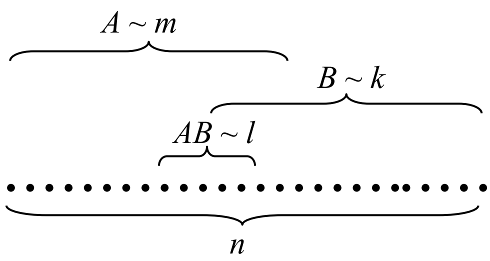


Fig. 2.2. Conditional probability

Then, according to the classical definition of probability,

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

After the event A occurred, the number of all equally possible outcomes decreased from n to m, and the number of cases favorable for the event B from k to l. Therefore, the conditional probability

$$P_A(B) = \frac{l}{m} = \frac{l/n}{m/n} = \frac{P(AB)}{P(A)} \ . \qquad (2.4)$$

Similarly

$$P_B(A) = \frac{P(AB)}{P(B)}. \qquad (2.5)$$

Multiplying the right and left parts of equations (2.4) and (2.5) by $P(A)$ and $P(B)$, respectively, we obtain:

$$P(AB) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A). \qquad (2.6)$$

Theorem 2.2. Probability multiplication rule. The probability of the product of two events is equal to the product of the probability of one of them and the conditional probability of the other, found under the condition that the first event occurred.

Proof. (singularly)●

The probability multiplication theorem is easy to generalize for the case of an arbitrary number of events:

$$P(ABC...KL) = P(A) \cdot P_A(B) \cdot P_{AB}(C)...P_{ABC...K}(L), \qquad (2.7)$$

that is, the probability of the product of several events is equal to the product of the probability of one of these events and the conditional probabilities of the others; the conditional probability of each subsequent event is calculated provided that all previous events have occurred.

◄**Example 2.3.** The operation of the electronic device stopped due to the failure of one of the five unified units. Each unit is successively replaced with a new one until the device starts working. What is the probability that you will have to replace: a) 2 blocks, b) 4 blocks?

The solution. a) Denote the events: A_i - the *i*-th block is working, *i* = 1,2, ..., 5; B - replacement of two blocks. You will have to replace 2 blocks if the 1st block is working (4 chances out of 5), and the 2nd – is not (1 chance out of the remaining 4), i.e. $B = A_1\overline{A}_2$. Now by the multiplication theorem (2.6)

$$P(B) = P(A_1\overline{A}_2) = P(A_1)\cdot P_{A_1}(\overline{A}_2) = \frac{4}{5}\cdot\frac{1}{4} = \frac{1}{5}.$$

b) Let the event C be the replacement of 4 blocks. Obviously, $C = A_1A_2A_3\overline{A}_4$, by the multiplication theorem

$$P(C) = P(A_1A_2A_3\overline{A}_4) = P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)P_{A_1A_2A_3}(\overline{A}_4) =$$

$$= \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = \frac{1}{5}\ . \ ►$$

◄**Example 2.4.** The child takes the cards in random order and lays them out. Find the probability that the word "PINEAPPLE" will appear if three letters P, two letters E and one letter I, N, L are written on separate cards.

The solution. Let's consider the event B be when the word "pineapple" appears. Event B will occur if the first is a card with the letter P (3 chances out of 9), the second - with the letter I (1 chances out of the remaining 8), the third - with the letter N (1 chances out of the remaining 7) and so on. By the multiplication theorem

$$P(B) = \frac{3}{9}\cdot\frac{1}{8}\cdot\frac{1}{7}\cdot\frac{2}{6}\cdot\frac{1}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{1} = \frac{1}{30240}\ . \ ►$$

The probability multiplication theorem takes the simplest form when the events that form the product are **independent**.

An event B is called **independent** of event A if its probability does not change depending on whether the event occurred or not, i.e. $P_A(B) = P(B)$ or $P_B(A) = P(A)$. Otherwise, if $P_A(B) \neq P(B)$ or $P_B(A) \neq P(A)$ the event is called **dependent on**. Dependence and independence of events are always mutual. Therefore, we can give the following definition of the independence of events: two events are called **independent** if the appearance of one of them does not change the probability of the other.

◄**Example 2.5.** Determine whether or not the events A and B depend on the condition of example 2.2.

The solution. In case of return of the taken detail $P_A(B) = P_{\bar{A}}(B) = P(B) = \frac{3}{5}$, i.e. events A and B are independent. If the detail taken from a box is not returned, then $P_A(B) \neq P_{\bar{A}}(B) \left(\frac{2}{4} \neq \frac{3}{4} \right)$, $P_A(B) \neq P(B)$, so both events A and B are dependent. ►

Several events A, B,..., L are called **independent in aggregate** (or simply independent) if any two of them are independent and any of these events and any combination (product) of other events are independent. Otherwise, the events
A, B,..., L are called **dependent.**

For example, three events *A*, *B*, *C* are independent (independent in total), if the ivents *A* and *B*, *A* and *C*, *B* and *C*, *A* and *BC*, *B* and *AC*, C and *AB are independent*.

For independent events, the theorem (rule) of multiplication of probabilities for two and more events takes the form:

$$P(AB) = P(A)P(B) \tag{2.8}$$

$$P(ABC...KL) = P(A)P(B)...P(L), \tag{2.9}$$

that is, the probability of the product of two or more independent events is equal to the product of the probabilities of these events.

◄**Example 2.6.** The probability of hitting the target for the first shooter is 0.8, for the second - 0.7, for the third - 0.9. Each shooter fires one shot. What is the probability that the target has 3 holes?

The solution. Let's denote the events: A_i - hitting the target of the *i*-th shooter (i = 1, 2, 3);

B - three hits on the target. It is obvious that $B = A_1A_2A_3$, and events A_1, A_2, A_3 are independent. By the multiplication theorem for independent events

$P(B) = P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3) = 0{,}8 \cdot 0{,}7 \cdot 0{,}9 = 0{,}504$ ►

Remark. 1**.** The independence of events is based on their physical independence. This means that the sets of random factors that lead to one or another trial result do not overlap (or almost do not overlap). For example, if there are two installations in the workshop, which are not related to each other by production conditions, then the failure of each installation is an independent event. If these installations are connected by a single technological cycle, the failure of one of the installations depends on the state of operation of the others.

However, if the sets of random factors overlap, then the events that appear as a result of the trial are not necessarily dependent.

Let, for example, consider the events:

A - draw a card of the spades at random from the deck;
B - draw an ace at random from the deck.

It is necessary to find out whether events A and B are dependent. At first glance, we can assume the dependence of events A and B due to the intersection of cases, they contribute: among the cards of spades there is an ace, and among the aces - a card of spades. We make sure, however, that events A and B are independent.

$P(B) = \frac{4}{36} = \frac{1}{9}$ (there are 4 aces out of 36 cards in the deck),

$P_A(B) = \frac{1}{9}$ (1 ace out of 9 cards of the spades in the deck).

Therefore, $P_A(B) = P(B)$, that is, events A and B are independent.

2. Pairwise independence of several events (i.e., independence taken from them of any two events) does not yet mean their independence in aggregate.

Let's make sure of this on the example of S. N. Bernshtein.

Suppose that the faces of a regular tetrahedron (a triangular pyramid with identical edges) are painted: 1st - in red (event A), 2nd - in green (B), 3rd - in blue (C) and 4th - in all three colors (ABC event). When tossing a tetrahedron, the probability of any face it lands on having the same color in its coloring is 1/2 (since there are 4 faces in total, and 2 with the corresponding color, that is, two chances out of four). Thus, $P(A) = 1/2$, $P(B) = 1/2$, $P(C) = 1/2$.

You can also calculate that

$$P_B(A) = P_C(A) = P_A(B) = P_B(C) = \frac{1}{2}$$

(one chance out of two), that is, events A, B, C are pairwise independent. If two events occurred at the same time, for example, A and B, i.e. AB, then the third event C will necessarily occur, i.e. $P_{AB}(C) = 1$ and similarly $P_{AC}(B) = 1$, $P_{BC}(A) = 1$; therefore, the probability of each of the events A, B, or C has changed, and the events A, B, and C are collectively dependent.

When solving a number of problems, you need to find the probability of the sum of two or more compatible events, that is, the probability of the appearance of at least one of these events. In this case, the theorem of addition of probabilities cannot be applied.

Theorem 2.3. The probability of the sum of two compatible events is equal to the sum of the probabilities of these events without the probability of their product, i.e.

$$P(A+B)=P(A)+P(B)-P(AB) \tag{2.10}$$

Proof.

Let us present the event $A+B$, which consists in the appearance of at least one of the two events A or B, in the form of the sum of three incompatible options: $A+B=A\overline{B}+\overline{A}B+AB$. Then by the addition theorem

$$P(A+B)=P(A\overline{B})+P(\overline{A}B)+P(AB) \tag{2.11}$$

Taking into account that $A=A\overline{B}+AB$, from where $P(A)=P(A\overline{B})+P(AB)$ and $P(A\overline{B})=P(A)-P(AB)$, and similarly, we obtain, substituting the found expressions:

$$P(A+B)=\left(P(A)-P(AB)\right)+\left(P(B)-P(AB)\right)+P(AB)=$$
$$=P(A)+P(B)-P(AB)$$

The correctness of formula (2.10) can be visually verified by Fig. 2.3.

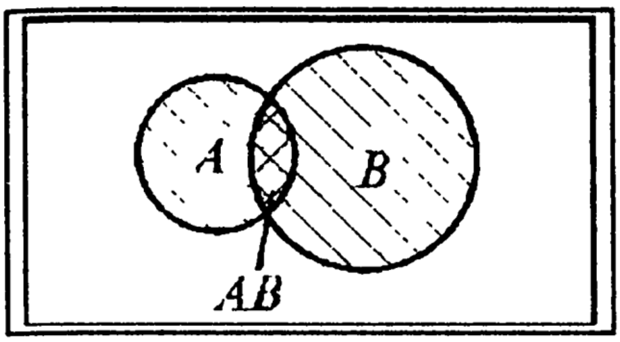


Fig.2.3

In the case of three or more common events, the corresponding formula for the probability of the sum *P(A + B +...+ K)* is very cumbersome and it is easier to go to the opposite event L:

$$L = \overline{A + B + ... + K} = \overline{A}\overline{B}...\overline{K}\ .$$

Then *P(A + B +...+ K)* = 1 - *P (L)*, or

$$P(A + B + ... + K) = 1 - P(\overline{A}\overline{B}...\overline{K}), \qquad (2.12)$$

that is, the probability of the sum of several joint events *A, B,..., K* is equal to the difference between unity and the probability of the product of opposite events $\overline{A}, \overline{B}, ..., \overline{K}$.. If the events *A, B, ..., K* are independent, then

$$P(A + B + ... + K) = 1 - P(\overline{A})P(\overline{B})...P(\overline{K})\ . \qquad (2.13)$$

In a separate case, if the probabilities of independent events are equal, i.e. $P(A) = P(B) = ... = P(K) = p$, then the probability of their sum equals

$$P(A + B + ... + K) = 1 - (1 - p)^n, \qquad (2.14)$$

because in this case $P(\overline{A})P(\overline{B})...P(\overline{K}) = \underbrace{(1 - p)...(1 - p)}_{n \text{ times}} = (1 - p)^n$).

◄ **Example 2.7.** There are 5 winning tickets for every 100 lottery tickets. What is the probability of winning at least one ticket if: a) 2 tickets are purchased; b) 4 tickets?

The solution. Let the event A_i be the winnings for the *i*-th ticket (*i* = 1,2,3,4).

a) The probability that at least one of the two tickets will be winning

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2) =$$

$$= P(A_1) + P(A_2) - P(A_1)P(A_2) = \frac{5}{100} + \frac{5}{100} - \frac{5}{100} \cdot \frac{4}{99} = 0{,}098\ .$$

b) The probability that at least one of the four tickets will be winning

$$P(A_1 + A_2 + A_3 + A_4) = 1 - P(\overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4) =$$

$$=1-\frac{95}{100}\cdot\frac{94}{99}\cdot\frac{93}{98}\cdot\frac{92}{97}=0,188. \blacktriangleright$$

2.3. Formula of total probability. Bayes formula

The corollary of the two main theorems of probability theory - the addition theorem and the multiplication theorem is the full probability formula and the Bayes formula.

Theorem 2.4. If the event F can occur only if one of the events (hypotheses) $A_1, A_2, ..., A_n$, which form a complete group, occurs, then the probability of the event F is equal to the sum of the products of the probabilities of each of these events (hypotheses) by the corresponding conditional probabilities of the event F:

$$P(F)=\sum_{i=1}^{n}P(A_i)P_{A_i}(F). \tag{2.15}$$

Proof. According to the condition of the event (hypotheses) $A_1, A_2, ..., A_n$ form a complete group, therefore, they are the only possible and incompatible ones. Since the hypotheses $A_1, A_2, ..., A_n$ — are only possible, and the event F under the condition of the theorem can occur only together with one of the hypotheses, then

$$F = A_1F + A_2F + ... + A_nF.$$

Due to the fact that the hypotheses $A_1, A_2, ..., A_n$ are incompatible, the theorem of addition of probabilities can be applied:

$$P(F)=P(A_1F)+P(A_2F)+...+P(A_nF)=\sum_{i=1}^{n}P(A_iF).$$

According to the multiplication theorem $P(A_iF)=P(A_i)P(F)$.

It has been proven that **Bayes' formula** is a **consequence** of the multiplication theorem and the formula of total probability. It is used when the event F (which can appear only with one of the hypotheses $A_1, A_2, ..., A_n$, which form a complete group of events) has occurred and when, if necessary, it is to be made a quantitative reassessment of the apriori probabilities of the hypotheses $P(A_1), P(A_2), ..., P(A_n)$, which are known before the trial, that is, it is necessary to find the posterior (obtained after the trial) conditional probabilities of the hypotheses $P_F(A_1), P_F(A_2), ..., P_F(A_n)$.

To obtain the desired formula, we write the theorem of multiplication of the probabilities of events F and A_i in two forms:

$$P(FA_i) = P(F)P_F(A_i) = P(A_i) \cdot P_{A_i}(F),$$

where

$$P_F(A_i) = \frac{P(A_i)P_{A_i}(F)}{P(F)}, \tag{2.16}$$

or with consideration (2.15)

$$P_F(A_i) = \frac{P(A_i)P_{A_i}(F)}{\sum_{i=1}^{n} P(A_i)P_{A_i}(F)}. \tag{2.17}$$

Formula (2.17) is called the **Bayes formula.**

The value of the Bayesian formula is that we can check and correct the hypotheses put forward for trialing. This approach is called **Bayesian** and makes it possible to change and clarify managerial decisions in the economy, when evaluating unknown parameters of the distribution of the investigated characteristics in statistical analysis, etc.

◄ **Example 2.8.** The trading firm received computers from three suppliers in a ratio of 1:4:5. Practice has shown that computers that come from the 1st, 2nd and 3rd suppliers will not require repair during the warranty period in 98, 88 and 92% of cases, respectively.

1) Find the probability that a computer delivered to a trading firm will not require repair during the warranty period.

2) The sold computer needs repair during the warranty period. From which supplier did this computer most likely come?

The solution. 1) Let's mark the events:

A_i — the computer arrived at the trading firm from the *i*-th supplier ($i = 1, 2, 3$);

F — the computer will not require repair during the warranty period.

According to the task

$$P(A_1)=\frac{1}{1+4+5}=0{,}1; \quad P_{A_1}(F)=0{,}98;$$

$$P(A_2)=\frac{4}{1+4+5}=0{,}4; \quad P_{A_2}(F)=0{,}88;$$

$$P(A_3)=\frac{5}{1+4+5}=0{,}5; \quad P_{A_3}(F)=0{,}92;$$

According to the formula of total probability (2.15)

$$P(F)=0{,}1\cdot 0{,}98+0{,}4\cdot 0{,}88+0{,}5\cdot 0{,}92=0{,}91.$$

2) Event $\overline{F}$ — computers will require repair during the warranty period;

$$P\left(\overline{F}\right)=1-P\left(F\right)=1-0{,}91=0{,}09.$$

According to the task

$$P_{A_1}(\overline{F})=1-0{,}98=0{,}02;$$

$$P_{A_2}(\overline{F})=1-0{,}88=0{,}12;$$

$$P_{A_3}(\overline{F}) = 1 - 0{,}92 = 0{,}08.$$

According to the Bayes formula (2.17)

$$P_{\overline{F}}(A_1) = \frac{0{,}1 \cdot 0{,}02}{0{,}09} = 0{,}022; \quad P_{\overline{F}}(A_2) = \frac{0{,}4 \cdot 0{,}12}{0{,}09} = 0{,}533;$$

$$P_{\overline{F}}(A_3) = \frac{0{,}5 \cdot 0{,}08}{0{,}09} = 0{,}444.$$

Thus, after the occurrence of the event $\overline{F}$, the probability of the hypothesis A_2 increased from $P(A_2) = 0{,}4$ to the maximum $P(A_2) = 0{,}533$, and the probability of the hypothesis A_3 — decreased from the maximum $P(A_3) = 0{,}5$ to $P(A_3) = 0{,}444$; if earlier (before the occurrence of the event) hypothesis A_3 was the most likely, now, in the light of new information (the occurrence of the event), the most probable hypothesis is A_2 — the receipt of this computers from the 2nd supplier. ►

◄ **Example 2.9.** It is known that on average 95% of manufactured products meet the standard. The simplified control scheme recognizes products as suitable with a probability of 0.98 if they are standard and with a probability of 0.06 if they are non-standard. Determine the probability that: 1) a product taken at random will pass the simplified control, 2) the product is standard, if it: a) passed the simplified control, b) passed the simplified control twice.

The solution. 1) Let's mark the hypothesis:

A_1, A_2 — the product taken at random is standard or non-standard, respectively;

F - the product passed simplified control.

According to the task

$$P(A_1) = 0{,}95, \ \ P(A_2) = 0{,}05, \ \ P_{A_1}(F) = 0{,}98; \ \ P_{A_2}(F) = 0{,}06.$$

The probability that a product taken at random will pass simplified control, according to the formula of full probability (2.15):

$$P(F) = 0{,}95 \cdot 0{,}98 + 0{,}05 \cdot 0{,}06 = 0{,}934.$$

2) *a*) Bayesian probability that a product that passed simplified control is standard (2.17):

$$P_F(A_1) = \frac{0{,}95 \cdot 0{,}98}{0{,}934} = 0{,}997.$$

b) Let the event F^* be the case when the product passed the simplified control twice. Then by the theorem of multiplication of

$$P_{A_1}(F^*) = 0{,}98 \cdot 0{,}98 = 0{,}9604 \text{ and } P_{A_2}(F^*) = 0{,}06 \cdot 0{,}06 = 0{,}0036.$$

According to the Bayes formula:

$$P_{F^*}(A_1) = \frac{0{,}95 \cdot 0{,}9604}{0{,}95 \cdot 0{,}9604 + 0{,}05 \cdot 0{,}0036} = 0{,}9998.$$

Since $P_{F^*}(A_2) = 1 - P_{F^*}(A_1) = 1 - 0{,}9998 = 0{,}0002$ is very small, then hypothesis A_2 that a product that passed simplified control twice is non-standard should be rejected as a practically impossible event. ►

◄ **Example 2.10.** Two shooters independently of each other shoot at the target, making one shot each. The probability of hitting the target for the first shooter is 0.8; for the second - 0.4. After the shots, one hole was found in the target. What is the probability that it belongs to: a) the 1st shooter, b) the 2nd shooter?

The solution. Let us denote the hypotheses:

A_1 — both shooters missed the target;

A_2 — both shooters hit the target;

A_3 — the 1st shooter hit the target, the 2nd did not;

A_4 — the 1st shooter missed the target, the 2nd hit the target;

F - one hole in the target (one hit).

Let's find the probabilities of the hypotheses and the conditional probabilities of the event F for these hypotheses:

$$P(A_1) = 0,2 \cdot 0,6 = 0,12, \quad P_{A_1}(F) = 0;$$

$$P(A_2) = 0,8 \cdot 0,4 = 0,32, \quad P_{A_2}(F) = 0;$$

$$P(A_3) = 0,8 \cdot 0,6 = 0,48, \quad P_{A_3}(F) = 1;$$

$$P(A_4) = 0,2 \cdot 0,4 = 0,08, \quad P_{A_4}(F) = 1.$$

So, according to the Bayes formula (2.17)

$$P_F(A_3) = \frac{0,48 \cdot 1}{0,12 \cdot 0 + 0,32 \cdot 0 + 0,48 \cdot 1 + 0,08 \cdot 1} = \frac{6}{7} = 0,857,$$

$$P_F(A_4) = \frac{0,08 \cdot 1}{0,12 \cdot 0 + 0,32 \cdot 0 + 0,48 \cdot 1 + 0,08 \cdot 1} = \frac{1}{7} = 0,143,$$

that is, the probability that the 1st shooter hit the target in the presence of one hole is 6 times higher than this for the second shooter. ►

2.4. Theoretical-multiple interpretation of the main concepts and axiomatic construction of the theory of probabilities

Let Ω be the set of all possible results of some trial (experiment). Each element ω of the set Ω, ($\omega \in \Omega$), is called an **elementary event** or an elementary result, and the set Ω itself is the **space of elementary events**. Any event A is considered as some subset of the set Ω, that is, $A \subset \Omega$.

For example, when throwing a dice, 6 elementary results (events) are possible: ω1 - loss of 1 point, ω2 - loss of 2 points, ... , ω6 - loss of 6 points, that is, the space of elementary events Ω = {ω1, ω2, ω3, ω4 , ω5, ω6}. Event A, which occurs when an even number of points appears, is A = {ω2, ω4, ω6}.

In the meeting problem, an infinite and uncountable set of elementary outcomes (events) is possible — points (*x, y*) of the OKLM square, the *x* and *y* coordinates of which are equal to the moments of arrival at the meeting place of two people, that is, the space of elementary events Ω is the OKLM square. The event A, which consists in the fact that the meeting of two people took place, is the shaded region g — a part of the square, that is, a subset of the space Ω: $A \subset \Omega$.

The space of elementary events Ω itself is an event that always occurs (for any elementary outcome ω), and is called a reliable event. Thus, Ω appears in two qualities: the set of all elementary results and a reliable event.

An empty set is added to the entire space Ω of elementary events, which is considered as an event $\varnothing$ and is called an **impossible event**.

The sum of several events $A_1, A_2, ..., A_n$ is called the union of sets $A_1 \cup A_2 \cup ... \cup A_n$.

The product of several events $A_1, A_2, ..., A_n$ is called the intersection of sets $A_1 \cap A_2 \cap ... \cap A_n$.

The event opposite to the event A is called the **complement of the set** A to Ω, i.e. Ω \ A.

The Euler-Venn diagrams show the sum $A + B$, the product AB of two events, and events opposite to events A, B.

Several events $A_1, A_2, ..., A_n$ form a **complete group** (a complete system) if their sum represents the entire space of elementary events, and the events themselves incompatible, that is $\sum_{i=1}^{n} A_i = \Omega$ and $A_i A_j = \varnothing (i \neq j)$.

Thus, operations on events will be understood as operations on corresponding sets. The table 2.1 shows the correspondence between the terms set theory and probability theory.

2.4.1. Axiomatic construction of probability theory

The need for a formal logical justification of probability theory, and its axiomatic construction arose in connection with the development of probability theory itself as a mathematical science and its applications in various fields. Such established sciences as geometry, theoretical mechanics, and set theory are built axiomatically. The foundation of each is several axioms, which are a generalization of many centuries of human experience, and science itself is built based on strict logical reasoning without recourse to visual representations. The axiomatics of probability theory derive from the basic properties of the probabilities of events to which the classical or statistical definition of probability is applied. Axiomatic definition of probability as individual cases includes both classical and statistical definitions and overcomes the shortcomings of each of them. For the first time, the idea of axiomatic construction of probabilities was expressed by the Russian academician S.M. Bernstein, who proceeded from a qualitative comparison of events according to their greater or lesser probability. In the early 1930s, Academician O.M. Kolmogorov developed another approach that connects probability theory with modern metric function theory and set theory, which is now generally accepted.

Table 2.1

Correspondence of the terms probability theory and set theory

Marking	Termini	
	Theory of sets	Probability theory
Ω	Set, space	The space of elementary events, a reliable event
Ω	Element of the set	Elementary event (elementary result)
A, B	Subset *A, B*	Event *A, B*
$A+B=A \cup B$	Union (sum) of sets *A* and *B*	The sum of events *A* and *B*
$AB=A \cap B$	Intersection of sets *A* and *B*	The product of events *A* and *B*
$\varnothing$	The empty set	An impossible event
$\overline{A}$	Complementary of the set *A*	The opposite event for *A*
$AB=A \cap B=\varnothing$	The sets *A* and *B* do not intersect	Events *A* and *B* are incompatible
$A=B$	Sets *A* and *B* are equal	Events *A* and *B* are equivalent
$A \subset B$	*A* is a subset of *B*	Event *A* entails event *B*

Axioms of probability theory

Each event A will be assigned a certain number, which will be called **the probability of event** A, and will be denoted by $P(A)$. Since any event is a set, the probability of an event is a function of the set.

The probability of an event must satisfy the following axioms.

1. The probability of any event is non-negative:

$$P(A) \geq 0.$$

2. The probability of a credible event equals 1:

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

3. The probability of the sum of incompatible events is equal to the sum of the probabilities of these events, i.e. if $A_i A_j = \varnothing \ (i \neq j)$, then

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

The main properties of probabilities can be deduced from the axiom:

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

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

3. $0 \leq P(A) \leq 1$.

4. $P(A) \leq P(B)$ if $A \subset B$

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

6. $P(A + B) \leq P(A) + P(B)$.

In the case of an arbitrary (not necessarily finite) space of elementary events Ω, axiom 3 must be replaced by a stronger, extended addition axiom 3' (which cannot be deduced from axiom 3).

3'. If there is a countable set of incompatible events $A_1, A_2, \ldots, A_n, \ldots$ $A_i A_j = \varnothing (i \neq j)$, then

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The axioms of probability theory allow to calculate the probabilities of any events (subsets of the space Ω) through the probabilities of elementary events (if their number is finite or countable). The question of where the probabilities of elementary events come from is not considered in the axiomatic construction of probability theory. In practice, they are determined using a classical definition (if the trial is reduced to a scheme of cases) or a statistical definition.

Formulated axioms do not determine the conditional probability of one event relative to another, which is introduced by definition.

The conditional probability of event B relative to event A is the ratio of the probability of the product of these events to the probability of event A, i.e.

$$P_A(B) = \frac{P(AB)}{P(A)}, \tag{2.18}$$

if $P(A) \neq 0$.

The theorem (rule) of multiplication of probabilities for any events automatically follows from this definition.

In more complete courses of probability theory, the concept of probability space is considered, which is defined by a trio of components (symbols) (Ω, S, P), where Ω is the space of elementary events, S is σ (sigma) - the algebra of events, and P is probability. We have already considered the first (Ω) and third (P) components of the probability space. The second component (S) of the probability space - the σ-algebra of events - represents some system of subsets of the space of elementary results (events) Ω. If Ω is finite or countable, then any subset of elementary outcomes is an event, and the σ-algebra is the system of all these subsets. If Ω is more than countable, then it turns out that not every arbitrary subset of Ω can be called an event. The reason for this is the existence of so-called non-dimensional subsets. Therefore, in this case, an event does not mean any subset of the space Ω, but only a subset from a selected class S, and the σ-algebra is a system of such subsets.

Control questions

1. Theorems for adding probabilities of compatible and incompatible events.
2. Conditional probability, theorem of multiplication of probabilities.
3. The formula of total probability and its consequence.
4. What is the value of the Bayes formula?
5. The theoretical essence of the multiple interpretations of probability theory.

Chapter 3

Repeated independent trials

3.1. Bernoulli's formula

Let's consider several problems.

1. Let there be white and black balls in the urn. Event A - a white ball is taken out of the urn. We denote the probability $P(A)$ by p and consider it independent of the number of times we remove a white ball from the urn. To ensure the independence of the process of removing the white ball, we will return the ball to the urn each time and mix it thoroughly. The opposite event – to remove the black ball (event $\overline{A}$) – will have a probability of $1-p=q$. Determine the probability $P_n(k)$ that in n trials the white ball will appear exactly k times.

2. What is the probability $P_n(k)$ that, in n flips of a coin, a head appears k times, given that the probability of a head appearing in one trial is p?

3. What is the probability $P_n(k)$ that, with n independent shots, the number of hits on the target will be k, provided that the probability of one hit is p and does not depend on the shot number?

4. What is the probability that out of n newborns, k are boys if the probability of a boy being born is p?

5. The operator is on duty during time T. It is established that for some small interval $\tau \ll T$ the probability of a call appearing is equal to p. What is the probability of k calls during the entire waiting time? Denoting $n=T/\tau$, we note that the task is to find $P_n(k)$.

All these tasks are reduced to a scheme of so-called independent **Bernoulli trials.** Let's match each trial with a random event, which we will denote by A_i. Trials are said to be independent if the outcome of each trial does not depend on the outcome of the remaining trials. It follows from the definition that the events A_i, that correspond to the trials, are independent in the aggregate, i.e.

$$P\left(A_1 \cdot A_2 \cdot \ldots \cdot A_n\right) = P\left(A_1\right) \cdot P\left(A_2\right) \cdot \ldots \cdot P\left(A_n\right).$$

Independent trials, each with only two outcomes (event A occurred or not occurred), are called Bernoulli trials.

Events $A_1, A_2, \ldots, A_n$ make up a group of elementary events. The above problems can be formulated in a general form as follows: n independent Bernoulli trials are conducted, in which the occurrence of event A is observed.

What is the probability of occurrence of event A in n trials k times, if the probability of occurrence of A in one trial is p, and the probability of non-occurrence is q?

Let B_k denote the event, which consists in the fact that in n trials the event A will appear k times. Event B_k will occur if, in n trials, event A occurs k times and fails to occur $n - k$ times. We denote by A_i, $\left(i = \overline{1,k}\right)$ the event that consists in the appearance of event A in the i-th trial, $\overline{A}_i$ – the event that consists in the non-appearance of event A in the i-th trial.

Then

$$B_k = A_1 A_2 \ldots A_k \overline{A_{k+1}}\,\overline{A_{k+2}} \ldots \overline{A_n} + A_1 \overline{A_2} A_3 \ldots \overline{A_{n-1}} A_n + \ldots + \overline{A_1}\,\overline{A_2} \ldots \overline{A_{n-k}}\,\overline{A_{n-(k-1)}} \ldots A_n \quad (3.1)$$

Each summand consists of k factors of the A type and of $(n - k)$ factors of the $\overline{A}$ type. The probability for each summand has the form $p^k q^{n-k}$. There are as many summands in (3.1) as can be composed of n elements $A_1, A_2, ..., A_n$ combinations of k elements, i.e. C_n^k. Given that the summands in (3.1) are incompatible events, we have

$$P(B_k) = C_n^k p^k q^{n-k}.$$

However $P(B_k) = P_n(k)$, therefore,

$$P_n(k) = C_n^k p^k q^{n-k} = C_n^k p^k (1-p)^{n-k}. \quad (3.2)$$

Formula (3.2) is called **Bernoulli's formula** and answers the questions above.

◄ **Example 3.1.** The probability of manufacturing a standard detail on an automatic machine is 0.8. Find the probability of the possible number of defective details among the 5 selected.

The solution. The probability of manufacturing a defective detail $p = 1 - 0,8 = 0,2$. The sought probabilities are found using Bernoulli's formula (3.2):

$P_5(0) = C_5^0 \cdot (0,2)^0 \cdot (0,8)^5 = 0,32768$; $P_5(1) = C_5^1 \cdot (0,2)^1 \cdot (0,8)^4 = 0,4096$;

$P_5(2) = C_5^2 \cdot (0,2)^2 \cdot (0,8)^3 = 0,2048$; $P_5(3) = C_5^3 \cdot (0,2)^3 \cdot (0,8)^2 = 0,0512$;

$P_5(4) = C_5^4 \cdot (0,2)^4 \cdot (0,8)^1 = 0,0064$; $P_5(5) = C_5^5 \cdot (0,2)^5 \cdot (0,8)^0 = 0,00032$.

The obtained results graphically correspond to points with coordinates $(m; P_n(m))$. By sequentially connecting these points with straight line segments, we get a broken line called a probability distribution polygon (Fig. 3.1). According to the graph, we can see that there are such values (m) that have the greatrial probability $(m_0 = 1)$. ►

The number (m_0) is called the **most probable**. To find the most probable value m_0, we write down the system of inequalities:

$$\begin{cases} P_n(m_0) \geq P_n(m_0+1) \\ P_n(m_0) \geq P_n(m_0-1) \end{cases}. \qquad (3.3)$$

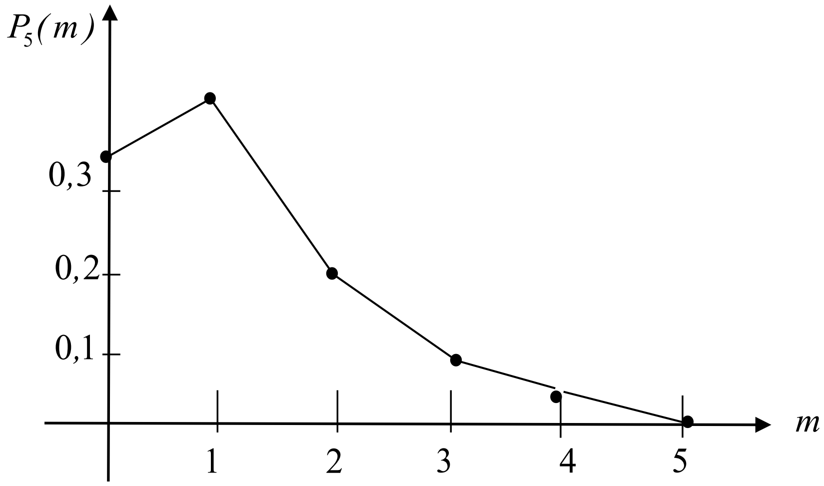


Fig. 3.1. Probability distribution polygon

Let's solve the first inequality using Bernoulli's formula:

$$\frac{n!}{m_0!(n-m_0)!} p^{m_0} q^{n-m_0} \geq \frac{n!}{(m_0+1)!(n-m_0-1)!} p^{m_0+1} q^{n-m_0-1}.$$

After simplifications, we get: $\frac{1}{(n-m_0)} q \geq \frac{1}{(m_0+1)} p$,

or $(m_0+1)q \geq (n-m_0)p$.

Let's rewrite the inequality in the form $m_0 \geq np - q,\ (p+q=1)$.

After solving the second inequality, we will have: $m_0 \leq np + p$.

Combining these two inequalities into a double one, we will have:

$$np - q \leq m_0 \leq np + p.$$

◀ **Example 3.2.** According to example 3.1, find the most likely number of defective details out of 5 selected and the probability of this event.

The solution. Let's use the double inequality $np - q \le m_0 \le np + p$:

$5 \cdot 0{,}2 - 0{,}8 \le m_0 \le 5 \cdot 0{,}2 + 0{,}2$ or $0{,}2 \le m_0 \le 1{,}2$. The only integer that satisfies this inequality, $m_0 = 1$, and its probability $P_5(1)=0{,}4096$

(see example 3.1) ►

◀ **Example 3.3.** How many times must the dice be rolled so that the most likely number of threes is 10?

The solution. In our case $p = \frac{1}{6}$, $m_0 = 10$. Let's use the inequality $np - q \le m_0 \le np + p$: $n\frac{1}{6} - \frac{5}{6} \le 10 \le n\frac{1}{6} + \frac{1}{6}$ or $n - 5 \le 10 \le n + 1$.

Therefore, $59 \le n \le 65$, that is, the die must be rolled between 59 and 65 times. ►

3.2. Limit behavior of probabilities in Bernoulli trials. Poisson's formula

Despite the compactness of formula (3.2), for large *n*, the direct calculation of $P_n(k)$ using this formula involves complex calculations. For simplification, we will find a number of approximate formulas for calculating $P_n(k)$.

To derive these formulas, we turn to the formula C_n^k:

$$C_n^k = \frac{n(n-1)...(n-k+1)}{k!} = \frac{n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right);$$

$$\frac{C_n^k \cdot k!}{n^k} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right). \tag{3.3}$$

If k is bounded, and $n \to \infty$, then $\frac{C_n^k \cdot k!}{n^k} \to 1$.

So, in formula (3.3) instead of C_n^k can be used $\frac{n^k}{k!}$.

It can be shown that

$$1-\frac{k(k-1)}{2n} \le \frac{C_n^k \cdot k!}{n^k} \le 1,\ 0 \le k \le n,\ n > 1.$$

Therefore, if $k \le 0{,}14\sqrt{n}$, then an approximate formula can be used for calculation

$$P_n(k) = C_n^k p^k q^{n-k} \approx \frac{1}{k!}\left(\frac{np}{q}\right)^k q^n \qquad (3.4)$$

Poisson's theorem. If Bernoulli trials are carried out with $n \to \infty$ and $p \to 0$, but $np = \lambda$, then the probability of occurrence of event A k times, and it is known that A appears in a single trial with probability p, leads to $e^{-\lambda}\frac{\lambda^k}{k!}$:

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (3.5)$$

Expression (3.5) is called the asymptotic **Poisson's formula**.

Proof. Let $\lambda = n \cdot p$, $p = \frac{\lambda}{n}$. Then according to (3.2)

$$P_n(k) = C_n^k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}.$$

When fixed k and $n \to \infty$ $C_n^k \to \frac{n^k}{k!}$, ago

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!} \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!}e^{-\lambda};$$

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

$$P_n(k) \approx \frac{\lambda^k}{k!}e^{-\lambda}. \tag{3.6}$$

It has been proven ●

There are special tables of Poisson function's values $P_k(\lambda)$. It is enough to find any printed textbook or download a table from the Internet.

◀**Example 3.4.** The telephone exchange serves 1000 subscribers. In a given time interval, a subscriber can make a call independently of other subscribers with a probability of 0.005. Find the probability that there were no more than seven calls in the given interval.

The solution. Let us denote by A the event, which consists of the fact that no more than seven calls took place. This is a complex event, it includes eight elementary events: ω_0 – no call; ω_1 – one call; ω_2 – two calls, etc.; ω_7 – seven calls. The probability of event A is equal to the sum of the probabilities of elementary events:

$$P(A) = P_{1000}(\omega_0) + P_{1000}(\omega_1) + \ldots + P_{1000}(\omega_7)$$

To calculate the probability of each elementary event, we use Poisson's theorem. $p = 0{,}005;\ n = 1000;\ \lambda = np = 5$.

$$P(A) = \sum_{i=0}^{7}\omega_i = \sum_{i=0}^{7} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-5}\left(\frac{5^0}{0!} + \ldots + \frac{5^7}{7!}\right) \approx 0{,}875.$$ ▶

When using Poisson's theorem it should be taken into account that it is used to calculate the probabilities of mass unlikely events (usually $p < 0{,}1;\ npq < 9$). If these conditions are not met, the so-called Moivre-Laplace limit theorem is used to calculate the probabilities $P_n(k)$.

3.3. Local and integral theorems of Moivre - Laplace

Local theorems of Moivre – Laplace.

If the probability p of occurrence of an event A in each trial is constant and different from 0 and 1, then the probability $P_n(k)$ that the event A will occur k times in n independent trials at a sufficiently large value n of the number, is approximately equal to:

$$P_n(k) \approx \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{npq}} e^{-\frac{x^2}{2}}, \tag{3.7}$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the Gaussian function; $x = \frac{k - np}{\sqrt{npq}}$.

The larger is n, the more accurate the formula. In practice, approximate values are used as exact, provided that $npq \geq 20$. To simplify calculations, a table of values of the Gaussian function was compiled (Appendix 2). To use the table, it is necessary to take into account the properties of the function $\varphi(x)$:

1) $\varphi(x)$ function is even: $\varphi(x) = \varphi(-x)$;

2) $\varphi(x)$ function is multitonically decreasing for positive values of x, if $x \to \infty$, then $\varphi(x) \to 0$. (Practically, it can be assumed $\varphi(x) \approx 0$ that already at $x > 4$).

◄**Example 3.5.** Find the probability of a coat of arms appearing 55 times in 100 independent tosses of a coin. The probability of the appearance of the coat of arms in one trial $p = 0.5$.

The solution. According to the task $k = 55;\ n = 100;\ q = 0,5;$

$$x = \frac{k - np}{\sqrt{npq}} = \frac{55 - 100 \cdot 0,5}{\sqrt{100 \cdot 0,5 \cdot 0,5}} = 1;\ \varphi(x) = 0,2420;$$

$$P_{100}(55) \approx \frac{\varphi(1)}{\sqrt{100 \cdot 0,25}} = \frac{0,2420}{5} = 0,0484 .$$►

Let's note that the local Moivre–Laplace theorem makes it possible to estimate individual probabilities and their behavior as a function of k, for large n.

In some cases, it is necessary to calculate the probability that event A will occur in the n trials no less than k_1 and no more k_2 times: $P_n(k_1, k_2)$.

Laplace's integral theorem is used for the calculation $P_n(k_1, k_2)$.

Integral theorem of Moivre – Laplace

If the probability p of the occurrence of an event A in each trial is constant, and $0 < p < 1$, then the probability $P_n(k_1, k_2)$ that the event A will occur in trials from k_1 to k_2 times (for a sufficiently large value of n) is approximately equal to

$$P_n(k_1, k_2) = \frac{1}{2}\left(\Phi(x_2) - \Phi(x_1)\right), \qquad (3.8)$$

where $\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$ - **the Laplace function** (or the integral of probabilities);

$$x_1=\frac{k_1-np}{\sqrt{npq}};\ x_2=\frac{k_2-np}{\sqrt{npq}}. \qquad (3.9)$$

The function $\Phi(x)$ is tabulated (see Appendix 3). To use the table, you need to know the properties of the Laplace function:

1) the function $\Phi(x)$ is odd, i.e. $\Phi(-x)=-\Phi(x)$;

2) the function $\Phi(x)$ is monotonically increasing, and at $x\to\infty$ $\Phi(x)\to 1$.

(Practically, it can be assumed $\Phi(x)\approx 1$ that already at $x>4$).

◄**Example 3.6.** According to statistical observations in some cities, 80 out of every 100 families have computers at home. Find the probability that 300-360 families out of 400, have computers at home.

The solution. Let's use the Moivre-Laplace integral theorem ($npq=64>20$). The probability that a family has a computer is $p=\frac{80}{100}=0,8$. $x_1=\frac{300-400\cdot 0,8}{\sqrt{400\cdot 0,8\cdot 0,2}}=-2,5;$ $x_2=\frac{360-400\cdot 0,8}{\sqrt{400\cdot 0,8\cdot 0,2}}=5,0$.

Taking into account the properties of the Laplace function, according to the table in Appendix 2, we find: $\Phi(5,0)\approx 1;$ $\Phi(-2,5)=-\Phi(2,5)=-0,9876$.

Then,

$$P_{400}(300,360)\approx\frac{1}{2}\left(\Phi(5,0)-\Phi(-2,5)\right)=\frac{1}{2}(1+0,9876)=0,9938.$$ ►

Consequence of the Moivre-Laplace integral theorem

If the probability of occurrence of an event A in each trial is constant and different from 0 and 1, then with a sufficiently large value of n of the number of independent trials, the probability that:

1) the number k of the occurrence of the event A differs from the product np by no more than a value $\varepsilon > 0$ (in absolute terms) equals to

$$P\left(\left|k - np\right| \le \varepsilon\right) \approx \Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right); \tag{3.12}$$

2) the frequency $\frac{k}{n}$ of the event A is in the range from α to β will be equal to

$$P_n\left(\alpha \le \frac{k}{n} \le \beta\right) \approx \frac{1}{2}\left(\Phi\left(z_2\right) - \Phi\left(z_1\right)\right), \tag{3.13}$$

where $z_1 = \frac{\alpha - p}{\sqrt{pq / n}}$; $z_2 = \frac{\beta - p}{\sqrt{pq / n}}$;

3) the frequency $\frac{k}{n}$ of the event A differs from its probability p by no more than value $\Delta > 0$ (by absolute value) will be equal

$$P_n\left(\left|\frac{k}{n} - p\right| \le \Delta\right) \approx \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{pq}}\right). \tag{3.14}$$

Proof

1) Inequality $\left|k - np\right| \le \varepsilon$ equals inequality $np - \varepsilon \le k \le np + \varepsilon$.

Therefore, according to the Moivre–Laplace integral formula

$$P\left(\left|k-np\right|\le \varepsilon\right)=P_n\left(np-\varepsilon\le k\le np+\varepsilon\right)\approx$$

$$\approx\frac{1}{2}\left(\Phi\left(\frac{np+\varepsilon-np}{\sqrt{npq}}\right)-\Phi\left(\frac{np-\varepsilon-np}{\sqrt{npq}}\right)\right)=\frac{1}{2}\left(\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right)-\Phi\left(\frac{-\varepsilon}{\sqrt{npq}}\right)\right)=$$

$$=\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$ Proved●

2) The inequality $\alpha\le\frac{k}{n}\le\beta$ is equivalent to the inequality $k_1\le k\le k_2$, where $k_1=n\alpha$ and $k_2=n\beta$. Let's replace in the formula $P_n(k_1,k_2)=\frac{1}{2}\left(\Phi\left(x_2\right)-\Phi\left(x_1\right)\right)$ $k_1,\ k_2$ with corresponding expressions and obtain the formula (3.13).

3) Inequality $\left|\frac{k}{n}-p\right|\le\Delta$ is equal to inequality $\left|k-np\right|\le\Delta n$. Let's replace in the formula $P\left(\left|k-np\right|\le\varepsilon\right)\approx\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right)$ $\varepsilon=\Delta n$ and get proving●

◄ **Example 3.7.** According to statistics, on average, 87% of newborns live to the age of 50 years. 1) Find the probability that out of 1000 newborns, the frequency of those who lived to the age of 50 will: a) be in the range from 0.9 to 0.95; b) will differ from the probability of this event by no more than 0.04 (by absolute value). 2) At what number of newborns with a reliability of 0.95 will the frequency of those who lived to the age of 50 be in the range from 0.86 to 0.88?

The solution. 1) a) The probability p that the newborn will live to 50 years is equal to 0,87. $n=1000$ is a large enough number, $npq=1000\cdot 0{,}87\cdot 0{,}13=113{,}1>20$; we can apply a corollary from the Moivre-Laplace integral theorem:

$$z_1=\frac{0,9-0,87}{\sqrt{0,87\cdot 0,13/1000}}=2,82;\ z_2=\frac{0,95-0,87}{\sqrt{0,87\cdot 0,13/1000}}=7,52.$$

According to the formula $\frac{1}{2}\left(\Phi(7,52)-\Phi(2,82)\right)$ we have

$$P_{1000}\left(0,9\le\frac{k}{n}\le 0,95\right)\approx\frac{1}{2}\left(\Phi(7,52)-\Phi(2,82)\right)=\frac{1}{2}(1-0,9952)=0,0024.$$

б) According to the formula $P_n\left(\left|\frac{k}{n}-p\right|\le\Delta\right)\approx\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{pq}}\right)$ we have

$$P_{1000}\left(\left|\frac{k}{n}-0,87\right|\le 0,04\right)\approx\Phi\left(\frac{0,04\sqrt{1000}}{\sqrt{0,87\cdot 0,13}}\right)=\Phi(3,76)=0,9998.$$

The obtained result means that it is almost certain that from 0.83 to 0.91 of the number of newborns will live to age 50.

2) According tot he task $P_n\left(0,86\le\frac{k}{n}\le 0,88\right)=0,95$ or

$$P_n\left(-0,01\le\frac{k}{n}-0,87\le 0,01\right)=0,95;\ P_n\left(\left|\frac{k}{n}-0,87\right|\le 0,01\right)=0,95.$$

According to the formula $P_n\left(\left|\frac{k}{n}-p\right|\le\Delta\right)\approx\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{pq}}\right)$ if $\Delta=0,01$ we will get that $\Phi\left(\frac{0.01\sqrt{n}}{\sqrt{0,87\cdot 0,13}}\right)=0,95$. According to the table in Appendix 2, we find that this value corresponds to the value $\frac{0.01\sqrt{n}}{\sqrt{0,87\cdot 0,13}}=1,96$. Therefore, $n=4345$, i.e., the condition of the problem can be guaranteed only with a significant increase in the number of newborns. ►

◄**Example 3.8.** The insurance company has 10,000 clients. Insurance contribution for each client is UAH 500. In the case of an insured event, the probability of which based on the data and estimates of experts equals to p = 0.005, the insurance company is obliged to pay the client insurance amount of UAH 50,000. What profit can an insurance company with a reliability of 0.95 expect?

The solution. The size of the company's profit is the difference between the total contribution of all clients and the total insurance amount paid to n_0 clients subject to the occurrence of an insured event, i.e.

$$\Pi = 500 \cdot 10 - 50 n_0 = 50(100 - n_0) \text{ thousand hryvnias.}$$

For the definition n_0, we will apply the Moivre - Laplace integral formula (the condition $npq = 10000 \cdot 0,005 \cdot 0,995 = 49,75 \geq 20$ is fulfilled).

According to the task

$$P_{10000}(0 \leq m_0 \leq n_0) = \frac{1}{2}\left(\Phi(x_2) - \Phi(x_1)\right) = 0,95,$$

where m_0 the number of clients to whom the insurance amount will be paid;

$$x_1 = \frac{0 - np}{\sqrt{npq}} = -\sqrt{\frac{np}{q}} = -\sqrt{\frac{10000 \cdot 0,005}{0,995}} = -7,09; \; x_2 = \frac{n_0 - np}{\sqrt{npq}},$$

therefore,

$$n_0 = np + x_2\sqrt{npq} = 10000 \cdot 0,005 + x_2\sqrt{49,75} = 50 + x_2\sqrt{49,75}.$$

From the ratio

$$P_{10000}(0 \leq m_0 \leq n_0) = \frac{1}{2}\left(\Phi(x_2) - \Phi(x_1)\right) = 0,95,$$

$$\Phi\left(x_2\right) = 1,9 + \Phi(x_1) = 1,9 + \Phi(-7,09) \approx 1,9 + (-1) = 0,9$$

According to table 3 (see appendixes) $\Phi(x_2) = 0,9$ under condition that $x_2 = 1,645$.

Now $n_0 = 50 + 1,645\sqrt{49,75} = 61,6$ and $\Pi = 50(100 - 61,6) = 1920$, that is, with a reliability of 0.95, the expected profit will be 1.92 million hryvnias.▶

3.4. Polynomial scheme

In the polynomial scheme, a transition is made from a sequence of independent trials with two outcomes (A and $\overline{A}$) to a sequence of independent trials with k mutually exclusive outcomes $\left(A_1, A_2, ..., A_k\right)$. At the same time, in each trial, events $A_1, A_2, ..., A_k$ occur according to probabilities $p_1, p_2, ..., p_k$. Then the probability $P_n(m_1, m_2, ..., m_k)$ that in n independent trials the event A_1 will occur m_1 times, $A_2 - m_2$ times, etc., $A_k - m_k$ times $(m_1 + m_2 + ... + m_k = n)$, is determined by the formula

$$P_n(m_1, m_2, ..., m_k) = \frac{n!}{m_1! m_2! ... m_k!} p_1^{m_1} p_2^{m_2} ... p_k^{m_k}. \tag{3.15}$$

Let the event A be that in n independent trials the event A_1 will occur m_1 times, $A_2 - m_2$ times, etc., $A_k - m_k$ times $(m_1 + m_2 + ... + m_k = n)$. Then the resulting formula reflects the fact that the event A can be represented as a sum of incompatible options, the probability of each of which is according to the probability multiplication theorem for independent events is equal to $p_1^{m_1} p_2^{m_2} ... p_k^{m_k}$, and the number of options is determined by the number of permutations with repetitions of the n elements.

◀**Example 3.9.** There's a 0.2 probability that a person belonging to a certain population group has dark hair, probability 0.3 that they have brown hair, probability 0.4 that they have blond hair, and probability 0.1 that they have red hair. Find the probability that the composition of a randomly selected group of 8 people will meet the following conditions: a) the group has an equal number of dark-haired, fair-haired, blond and red-haired people; b) the number of blondes in the group is three times more than the number of redheads.

The solution. *a*) According to the formula (3.15), the probability of the desired event A is equal to

$$P(A) = P_8(2,2,2,2) = \frac{8!}{2!2!2!2!} 0{,}2^2 \cdot 0{,}3^2 \cdot 0{,}4^2 \cdot 0{,}1^2 = 0{,}0145.$$

b) The probability of the desired event B is equal to the sum of the probabilities of two incompatible events (options): - B_1 in the group there are 3 blond-haired people, 1 person has red hair, and the rest have any hair color; B_2 - there are 6 blondes and 2 redheads in the group.

$$P(B_1) = P_8(3;1;4) = \frac{8!}{3!1!4!} 0{,}4^3 \cdot 0{,}1^1 \cdot 0{,}5^4 = 0{,}112;$$

$$P(B_2) = P_8(6;2) = \frac{8!}{6!2!} 0{,}4^6 \cdot 0{,}1^2 = 0{,}0011;$$

$$P(B) = P(B_1) + P(B_2) = 0{,}112 + 0{,}0011 = 0{,}1131. \blacktriangleright$$

If the probabilities $p_1, p_2, \ldots, p_k$ of occurrence of events $A_1, A_2, \ldots, A_k$ in each trial change depending on the trial result, then we get a scheme of dependent trials. Such schemes are considered in the theory of random processes, in particular in Markov chains.

Control questions

1) What are the main prerequisites for using the Bernoulli scheme?

2) What is Poisson's formula used for?

3) When solving what types of problems, the local theorem of Moivare – Laplace is used?

4) To solve which problems is used the consequence of the integral theorem of Moivre-Laplace?

Chapter 4

Random variables

4.1. The concept of a random variable

By a random variable, we consider a variable that, as a result of the trial, takes one of its possible values (which one is unknown). A random variable is called **discrete** if it has a set of values finite, or infinite, but countable. A **continuous** random variable is a variable whose infinite and uncountable set of values is some interval (finite or infinite) of the numerical axis.

Examples of random variables:

1) the number of children born during the day in the city of Kyiv is discrete random variable;

2) the number of defective integrated circuits is discrete and random size;

3) the number of shots that were fired before the first hit - discrete random variable;

4) the flight range of an artillery projectile is continuous and random size;

5) monthly electricity consumption at the enterprise is continuous random variable.

The set-theoretic interpretation of the basic concepts allows us to formulate the following definition of a random variable: **a random variable X is a function defined on a set of elementary events (or in the space of elementary events)**, i.e. $X = f(\omega)$, where ω is an elementary event belonging to the space Ω.

Random variables will be denoted by uppercase Latin letters, and their values by corresponding lowercase letters.

4.2. The distribution law of a random variable

A random variable in its entirety can be described using the distribution law.

The law of the distribution of a random variable is any relationship that establishes a connection between the possible values of a random variable and the corresponding probabilities. There is a saying that a random variable is distributed according to a given law.

For a discrete random variable, the distribution law can be given by a table, analytically or graphically.

An example of the tabular form of the law of the distribution of a random variable X.

The values of the random variable are in ascending order of values	x_1	x_2	x_3	...	x_n	...
The probability of the corresponding value appearing	p_1	p_2	p_3	...	p_n	...

Such a table is called the distribution series of a discrete random variable. Events $X = x_1, X = x_2, ..., X = x_n$, which mean that a random value X will acquire the corresponding value $x_1, x_2, ..., x_n$, are incompatible and only possible, that is, they form a complete group. Then,

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1 . \qquad (4.1)$$

A distribution series can be depicted graphically by plotting the values of a random variable on the abscissa axis and the corresponding probabilities on the ordinate axis. By connecting the plotted points, we get a distribution polygon or probability distribution polygon (Fig. 4.1).

◄ **Example 4.1.** The following prizes are awarded in the lottery: a car worth 5,000 dollars, 4 televisions worth 250 dollars, 5 VCRs worth 200 dollars. A total of 1,000 tickets worth 7 dollars each were sold. Write down the law of distribution of net winnings, which will be received by a lottery participant who bought one ticket

The solution. Let the random variable X be the netto winning of the lottery participant who purchased one ticket. Then possible values of this random variable: $x_1 = 0 - 7 = -7$ (ticket without winnings);

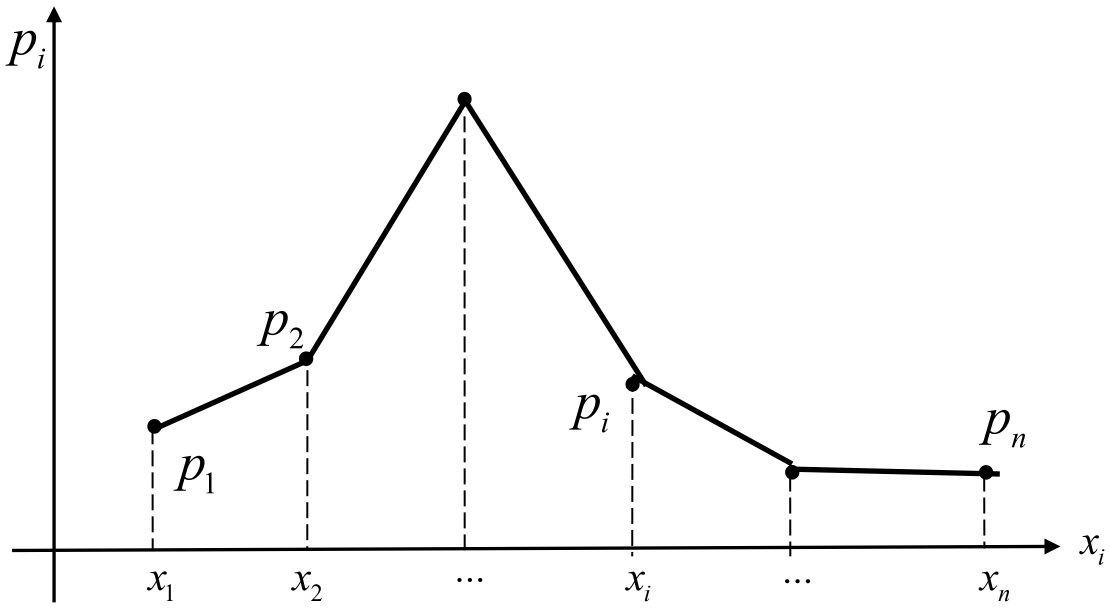


Fig. 4.1. Probability distribution polygon

$x_2 = 200 - 7 = 193$ (the ticket won a VCR); $x_3 = 250 - 7 = 243$ (the ticket won a TV); $x_4 = 5000 - 7 = 4993$ (the ticket won a car). The corresponding probabilities (according to the classical definition of probability) are:

$P(X = 193) = \dfrac{5}{1000} = 0{,}005$; $P(X = 243) = \dfrac{4}{1000} = 0{,}004$;

$P(X = 4993) = \dfrac{1}{1000} = 0{,}001$.

Then the distribution series will look like:

x_i	-7	193	243	4993
p_i	0,990	0,005	0,004	0,001

►

◄ **Example 4.2.** The probabilities that a student will pass a semester exam during a session in subjects *A* and *B* are 0.7 and 0.9, respectively. Write down the law of distribution of the number of semester exams that a student will take.

The solution. Let the random variable X be the number of exams passed by the student. Possible values of this random variable are: 0, 1, 2. Let the event A_i be that the student passed the i-th exam (i=1,2). Then the probabilities that the student will pass 0, 1, 2 exams during the session are, respectively, equal to:

$$P(X=0)=P(\overline{A}_1\overline{A}_2)=P(\overline{A}_1)P(\overline{A}_2)=(1-0,7)(1-0,9)=0.03;$$

$$P(X=1)=P(A_1\overline{A}_2+\overline{A}_1A_2)=P(A_1)P(\overline{A}_2)+P(\overline{A}_1)P(A_2)=0.34;$$

$$P(X=2)=P(A_1A_2)=P(A_1)P(A_2)=0,7\cdot 0,9=0.63.$$

Distribution series:

x_i	0	1	2
p_i	0,03	0,34	0,63

Figure 4.2 shows the resulting series in the form of a polygon probability distribution.

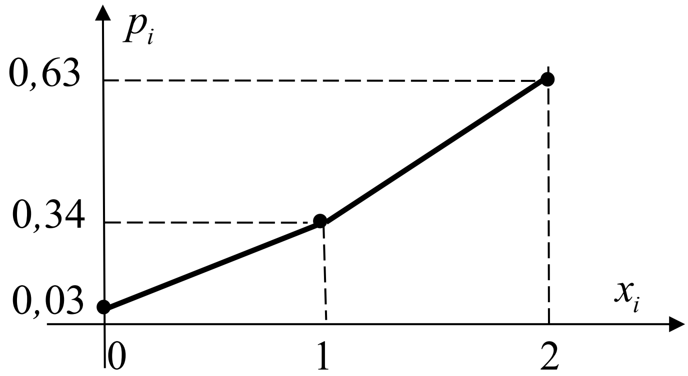


Fig. 4.2. Polygon of probability distribution of random variable X (to example 4.2) ►

4.3. Mathematical operations on random events

Two random variables are called independent if the distribution law of one of them does not change depending on the possible values of the other. That is, independence of discrete random variables X and Y means independence of events $X = x_i$ and $Y = y_j$, $i = \overline{1,n};\ j = \overline{1,m}$.

For example, if there are tickets for two different money lotteries, then the random variables X and Y, which mean the winnings in the first and second lotteries, respectively, will be independent.

Let two random variables be given:

X:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Y:

y_j	y_1	y_2	...	y_m

p_j	p_1	p_2	...	p_m

The product kX of a random variable X by a constant variable k is a random variable that takes values kx_i with the same probabilities p_i.

***m*-th power of a random variable** X $\left(X^m\right)$ is a random variable that takes values x_i^m with the same probabilities p_i.

◄ **Example 4.3.** A random variable X is given:

x_i	-2	1	2
p_i	0,5	0,3	0,2

Find the law of distribution of random variables: $a)$ $Y = 3X$; $b)$ $Y = X^2$.

The solution. $a)$ $Y = 3X$: $b)$: $Y = X^2$

y_i	1	4
p_i	0,3	0,7

y_i	-6	3	6
p_i	0,5	0,3	0,2

Since the value 4 is repeated twice, according to the addition theorem $P(Y = 4) = 0,5 + 0,2 = 0,7$ ►

The sum (difference or product) of random variables X and Y is a random variable that takes all possible values of the form $x_i + y_j$ ($x_i - y_j$ or $x_i \cdot y_j$), where $i = \overline{1,n}$; $j = \overline{1,m}$ with the probabilities p_{ij} that the random variable X will take the value x_i, and the random variable Y will take the value y_j : $p_{ij} = P((X = x_i)(Y = y_j))$.

If random variables X and Y are independent, then according to the theorem of multiplication of probabilities for independent events:

$$p_{ij} = P(X = x_i) \cdot P(Y = y_j) = p_i \cdot p_j.$$

◄ **Example 4.4.** The distribution laws of two independent random variables **are** given: X:

x_i	0	2	4
p_i	0,5	0,2	0,3

Y:

y_j	-2	0	2
p_j	0,1	0,6	0,3

Find the distribution laws of random variables: 1) $Z = X - Y$; 2) $Z = X \cdot Y$.

The solution. 1) For the convenience of calculations, we will compile an auxiliary table: in each cell of the table the value of the random variable $Z = X - Y$ and the product of the corresponding probabilities of the random variables X and Y are written down.

x_i \ y_j	p_i \ p_j	-2	0	2
		0,1	0,6	0,3
0	0,5	2 / 0,05	0 / 0,3	-2 / 0,15
2	0,2	4 / 0,02	2 / 0,12	0 / 0,06
4	0,3	6 / 0,03	4 / 0,18	2 / 0,09

For example, if $X = 4$, and $Y = -2$, then a random variable $Z = X - Y = 4 - (-2) = 6$ with probability:

$$P(Z = 6) = P(X = 4) \cdot P(Y = -2) = 0{,}3 \cdot 0{,}1 = 0{,}03.$$

Since among the values of a random variable Z there are those that are repeated, we add the corresponding probabilities (according to the theorem of addition of probabilities).
Then the law of the distribution of a random variable Z will take the form:

z_i	-2	0	2	4	6
p_i	0,15	0,36	0,26	0,20	0,03

2) The distribution law of a random variable $U = XY$ can be found similarly to item 1.

u_i	-8	-4	0	4	8
p_i	0,03	0,02	0,8	0,06	0,09

▶

4.4. Mathematical expectation of a discrete random variable

The mathematical expectation, or average value, $M(X)$ of a discrete random variable X is called the sum of products of all its values by corresponding probabilities:

$$M(X) = \sum_{i=1}^{n} x_i p_i \ . \qquad (4.2)$$

◀ **Example 4.5.** The laws of distribution of random variables X and Y are known - the number of points scored by the 1st and 2nd athletes.
X:

x_i	0	1	2	3	4	5	6	7	8	9	10

p_i	0,15	0,11	0,04	0,05	0,04	0,1	0,1	0,04	0,05	0,12	0,2

Y:

y_i	0	1	2	3	4	5	6	7	8	9	10
p_i	0,01	0,03	0,05	0,09	0,11	0,24	0,21	0,1	0,1	0,04	0,02

Determine which of the athletes is better.

The solution. Looking at the laws of distribution of these random variables, it is quite difficult to answer this question. For an unambiguous answer, we will find the mathematical expectation of both random variables:

$$M(X)=\sum_{i=1}^{10} x_i p_i = 0\cdot 0{,}15+1\cdot 0{,}11+\ldots+10\cdot 0{,}2=5{,}36;$$

$$M(Y)=\sum_{i=1}^{10} y_i p_i = 0\cdot 0{,}01+1\cdot 0{,}03+\ldots+10\cdot 0{,}02=5{,}36,$$

that is, the average number of points scored by athletes is the same.

Respond: athletes are equal in skill according to average indicators ►

If a discrete random variable X takes an infinite but countable set of values $x_1, x_2, \ldots, x_n, \ldots$, then the mathematical expectation of such a discrete random variable is called the sum of the series (if it coincides absolutely) $M(X)=\sum_{i=1}^{\infty} x_i p_i$.

Properties of mathematical expectation

1. The mathematical expectation of a constant value is equal to this constant value: $M(C) = C$.

2. The constant factor can be put beyond the mathematical expectation sign:

$$M(kX) = kM(X).$$

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

$$M(X \pm Y) = M(X) \pm M(Y).$$

Proof. According to the definition, the sum (difference) of random variables $X+Y$ $(X-Y)$ is a random variable that takes a value $x_i + y_j$ $(x_i - y_j)$ with probabilities $p_{ij} = P(X = x_i) \cdot P(Y = y_j)$. So

$$M(X \pm Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i \pm y_j\right)p_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij} \pm \sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}.$$

Since in the first sum x_i does not depend on the index j, and in the second sum y_j does not depend on the index i, then

$$M(X \pm Y) = \sum_{i=1}^{n}x_i\sum_{j=1}^{m}p_{ij} \pm \sum_{i=1}^{n}y_j\sum_{j=1}^{m}p_{ij} = \sum_{i=1}^{n}x_i p_i \pm \sum_{j=1}^{m}y_j p_j = M(X) \pm M(Y)$$

●

4. Mathematical expectation of the product of a finite number of independent random variables is equal to the product of their mathematical expectations. For two independent variables, we have:

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

5. . If all values of a random variable increase (decrease) by constant variable C, then the mathematical expectation will also increase (decrease) by this variable:

$$M(X \pm C) = M(X) \pm C.$$

6. Mathematical expectation of deviation of a random variable from its mathematical expectation is zero:

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

Proof. Based on the fact that, by definition, a mathematical expectation is a constant value, then let some calculated value be $a = M(X)$. Then

$$M(X - a) = M(X) - a = M(X) - M(X) = 0. \bullet$$

4.5. Dispersion of a discrete random variable

The dispersion $D(X)$ of a random variable X is the mathematical expectation of the square of its deviation from the mathematical expectation:

$$D(X) = M(X - M(X))^2 \ . \qquad (4.3)$$

Dispersion is a measure of the spread of a random variable around its mathematical expectation. If the random variable X is discrete with a finite number of values, then $D(X) = \sum_{i=1}^{n} \left(x_i - a\right)^2 p_i$, де $a = M(X)$.

If the random variable X is discrete with an infinite but countable number of values, then $D(X) = \sum_{i=1}^{\infty} \left(x_i - a\right)^2 p_i$, provided that the series coincides.

More often, a quantity called the mean square deviation and is used to characterize random variables $\sqrt{D(X)}$ is denoted by σ_x.

◄**Example 4.6.** In the problem about athletes (see example 4.5), we calculate the dispersion of random variables X and Y:

The solution. $M(X) = M(Y) = 5{,}36$.

$$D(X)=(0-5{,}36)^2\cdot 0{,}15+(1-5{,}36)^2\cdot 0{,}11+...+(10-5{,}36)^2\cdot 0{,}2=13{,}61$$

$$D(Y)=(0-5{,}36)^2\cdot 0{,}01+(1-5{,}36)^2\cdot 0{,}03+...+(10-5{,}36)^2\cdot 0{,}02=4{,}17$$

$$\sigma_x=\sqrt{D(X)}=3{,}69;\ \sigma_y=\sqrt{D(Y)}=2{,}04.$$

Therefore, if the averages are equal, the dispersion is less for the second athlete. This allows us to claim that the second athlete is better (he shoots "heavier").

Answer: the second athlete shoots better. ►

Properties of the dispersion of a random variable

1. The dispersion of a constant variable is zero: $D(C)=0$.

2. A constant factor can be taken out for the dispersion sign if it is squared:

$$D(kX)=k^2D(X).$$

Proof.

$$D(kX)=M(kX-M(kX))^2=M(k^2(X-M(X))^2=k^2D(X)\ \bullet$$

3. Formula for quick calculation of dispersion:

$$D(X)=M(X^2)-\left(M(X)\right)^2.$$

Proof.

Let $a=M(X)$, then $D(X)=M(X-a)^2=M(X^2-2Xa+a^2)$.

a - value is constant, non-random. Then,

$$D(X)=M(X^2)-2aM(X)+a^2=M(X^2)-2a^2+a^2=M(X^2)-a^2\ \bullet$$

4.. The dispersion of the algebraic sum of a finite number of independent random variables is equal to the sum of their dispersions:

$$D(X\pm Y)=D(X)\pm D(Y).$$

Mathematical expectation, dispersion, mean squared deviation and other numbers, which are used to present in a concise form the most essential features of the distribution, are called numerical characteristics of a random variable.

4.6. Distribution function of a discrete random variable

Until now, for a complete description of a discrete random variable, we used the law of its distribution in an analytical or tabular form, which allows us to find the probability of any values of the random variable X .

However, this description of a random variable X is not universal. It cannot be applied to a continuous random variable, because, firstly, it is impossible to enumerate all the infinitely uncountable set of its values; secondly, the probabilities of each single taken value of a continuous random variable are equal to zero.

There is also another approach to describe the law of distribution of a random variable X: consider not the probabilities of events $X = x$ for different *x* (as is the case in the distribution series), but consider the probabilities of the event $X < x$, where *x* is the current variable. The probability obviously depends on *x*, i.e. is some function of *x*.

The distribution function of a random variable X is a function $F(x)$ equal to the probability that the random variable X will acquire a value less than *x*:

$$F(x) = P(X < x).$$

The function $F(x)$ is sometimes called the **integral distribution function** or the integral distribution law. Geometrically, the distribution function is interpreted as the probability that a random point X is located to the left of a given point *x*.

◄ **Example 4.7.** A distribution series of a random variable X is given::

x_i	1	4	5	7
p_i	0,4	0,1	0,3	0,2

Find and draw a graph of the distribution function of a random variable X..

The solution. We will set different values of x and find for them $F(x)=P(X<x)$.

1. If $x \le 1$, then, obviously, $F(x)=0$ (including when $x=1$ $F(1)=P(X<1)=0$).

2. Suppose that $1<x\le 4$ (for example, x=2); $F(x)=P(X=1)=0{,}4$. It is obvious that $F(4)=P(X<4)=0{,}4$.

3. Suppose that $4<x\le 5$ (for example, x=4.25);

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

It is obvious that $F(5)=0{,}5$.

Suppose that $5<x\le 7$.

$$F(x)=P(X<x)=P(X=1)+P(X=4)+P(X=5)=0{,}5+0{,}3=0{,}8.$$

It is obvious that $F(7)=0{,}8$.

4. Suppose that $x>7$.

$$F(x)=P(X<x)=P(X=1)+P(X=4)+P(X=5)+P(X=7)=$$
$$=0{,}8+0{,}2=1.$$

Let's draw the graph of the function $F(x)$ (Fig. 4.3).

Therefore, $F(x)=\begin{cases} 0 & \text{при } x \le 1, \\ 0{,}4 & \text{при } 1 < x \le 4, \\ 0{,}5 & \text{при } 4 < x \le 5, \\ 0{,}8 & \text{при } 5 < x \le 7, \\ 1{,}0 & \text{при } x > 7. \end{cases}$

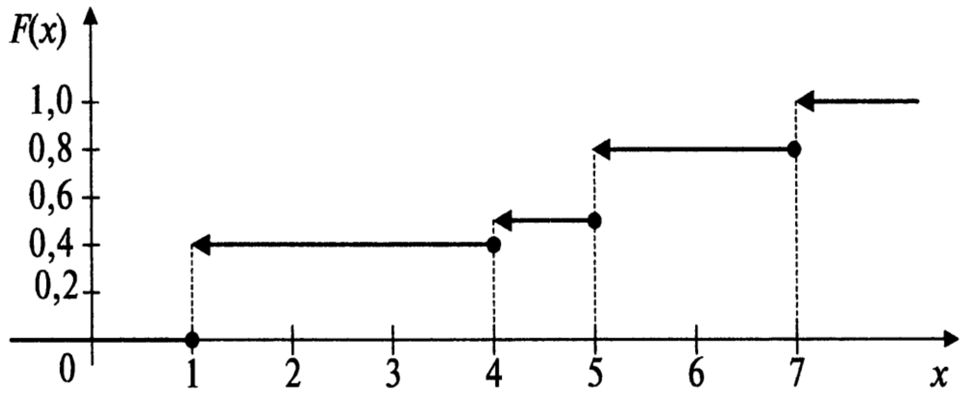


Fig. 4.3. The distribution function of a random variable (to example 4.7.)

When approaching the discontinuity points from the left, the function retains its value (such a function is said to be continuous from the left). These points are highlighted on the graph. ►

The distribution function of any discrete random variable is a discontinuous step function whose values increase at points that correspond to the possible values of the random variable and are equal to the probabilities of these values. The sum of all increases of the function $F(x)$ is equal to 1.

Properties of the distribution function

1. The distribution function of a random variable is a non-negative function whose values are limited to zero and one: $0 \le F(x) \le 1$.

2. The distribution function of a random variable is a nondecreasing function on the entire numerical axis.

Let x_1 and x_2 be the points of the numerical axis, and $x_1 > x_2$. We will show that $F(x_2) \geq F(x_1)$. Consider two incompatible events: $A = (X < x_1)$ and $B = (x_1 \leq X < x_2)$. Then, $A + B = (X < x_2)$ (See Fig. 4.4)

$$A + B = (X < x_2)$$

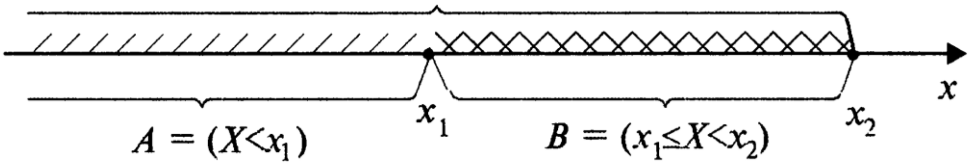


Fig. 4.4. Geometric interpretation of the relationship between events A and B

According to the theorem of the addition of probabilities of incompatible events $P(A+B) = P(A) + P(B)$ або $P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2)$,

from here $$F(x_2) = F(x_1) + P(x_1 \leq X < x_2). \quad (4.4)$$

Since the probability $P(x_1 \leq X < x_2) > 0$, then $F(x_2) \geq F(x_1)$, that is $F(x)$, is a non-decreasing function.

3. At minus infinity, the distribution function is zero, at plus infinity it is equal to one, i.e. $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$, $F(\infty) = \lim\limits_{x \to \infty} F(x) = 1$

$F(-\infty) = P(X < -\infty) = 0$, as the probability of an impossible event $X < -\infty$

$F(+\infty) = P(X < +\infty) = 1$, as the probability of a valid event $X < +\infty$.

4. The probability of a random variable falling into the interval $[x_1, x_2)$ (including x_1) is equal to the increment of its distribution function on this interval, that is, $$P(x_1 \leq X < x_2) = F(x_2) - F(x_1). \quad (4.5)$$

Formula (4.5) follows directly from formula (4.3).

◄ **Example 4.8**. The distribution function of a random variable X has the form:

$$F(x) = \begin{cases} 0, & at\ x \leq 0, \\ x/2, & at\ 0 < x \leq 2, \\ 1, & at\ x \geq 2. \end{cases}$$

Find the probability that the random variable will take a value within the interval [1, 3).

The solution. According to formula (4.5):

$$P(x_1 \leq X < x_2) = F(3) - F(1) = 1 - \frac{1}{2} = \frac{1}{2}.$$ ►

Control questions

1. Will the distribution law be a complete form of the description of discrete random variable?
2. Name the arithmetic operations that can be performed on two independent discrete random variables.
3. Definition of the mathematical expectation of a discrete random variable.
4. What characteristic of a discrete random variable can be analyzed due to dispersion?
5. Basic properties of the distribution function of a random variable.

Chapter 5

Continuous random variables

5.1. Probability density

A random variable X is said to be **continuous** if its distribution function is continuous at any point and differentiable everywhere except, perhaps, at certain points. Figure 5.1 shows the distribution function of a continuous random variable X, which is differentiable at all but three inflection points.

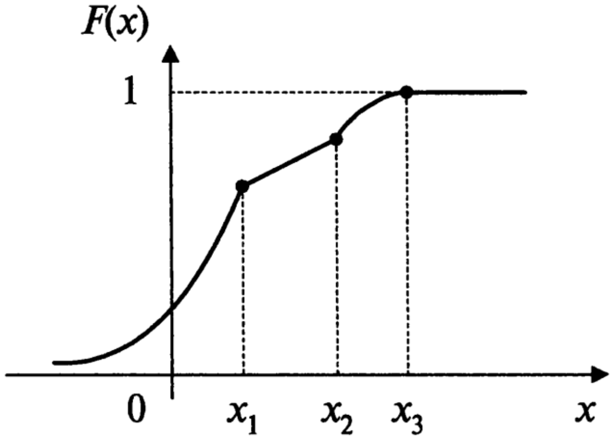


Fig. 5.1. An example of the distribution function of a continuous random variable

Theorem 5.1. The probability of any single value of a continuous random variable is zero.

Proof. We will show that for any value x_1 of the random variable X the probability $P(X = x_1) = 0$. Let's present $P(X = x_1)$ in the form

$$P(X = x_1) = \lim_{x_2 \to x_1} P(x_1 \le X < x_2).$$

Applying property (4.5) of the distribution function of a random variable X and taking into account the continuity of $F(x)$, we obtain

$$P(X = x_1) = \lim_{x_2 \to x_1} \left(F(x_2) - F(x_1) \right) = \lim_{x_2 \to x_1} F(x_2) - F(x_1) =$$
$$= F(x_1) - F(x_1) = 0$$

Up to this point, we have considered trials that were reduced to a scheme of events, and only impossible events had zero probability. It follows from the theorem above that possible events can also have zero probability, since the event, which consists in the fact that a random variable X has assumed a specific value x_1, is possible. At first glance, this conclusion may seem paradoxical. Indeed, if, for example, an event $\alpha \le X \le \beta$ has a non-zero probability, then it turns out that it represents the sum of events, which consist in the fact that the random variable X will acquire any specific values on the segment $[\alpha, \beta]$ and which have zero probability. However, there is no contradiction here, because the theorem of addition (more precisely, the axiom of addition) is valid only for a finite and countable infinite set of events, and the set of events that denotes the specified sum is not such.

The idea of an event having a non-zero probability but consisting of events having a zero probability is no more paradoxical than the idea of a segment having a definite length when no point on the segment has a non-zero length. A segment consists of such points, but its length is not equal to the sum of their lengths ●

Consequence. If X is a continuous random variable, then the probability that the random variable will acquire a value from the interval (x_1, x_2) does not depend on whether this interval is open or closed, i.e.

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

$$P(x_1 \le X \le x_2) = P(X = x_1) + P(x_1 < X < x_2) + P(X = x_2) =$$
$$= 0 + P(x_1 < X < x_2) + 0 = P(x_1 < X < x_2).$$

Other equalities are proved similarly.

A continuous random variable can be specified not only by a distribution function. Let us introduce the concept of **probability density** of a continuous random variable. Consider the probability of a continuous random variable hitting the segment $\left[x,\ x+\Delta x\right]$. According to formula (4.5), the probability

$P(x \le X \le x+\Delta x) = F(x+\Delta x) - F(x)$,i.e. is equal to the increment of the distribution function $F\left(X\right)$ on this segment. Then the probability per unit length, that is, the average probability density on the segment from x to $x+\Delta x$ is equal to

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

Moving to the boundary under the condition that $\Delta x \to 0$, we obtain the probability density at the point x:

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

which is the derivative of the distribution function $F\left(x\right)$ (for a continuous random variable $F\left(x\right)$ it is a differentiable function).

The probability density (**distribution density** or simply **density**) $\varphi(x)$ of a continuous random variable X is called the derivative of its distribution function:

$$\varphi\left(x\right) = F'\left(x\right) . \qquad (5.1)$$

The probability density $\varphi(x)$, like the distribution function $F\left(x\right)$, is a form of the distribution law, but unlike the distribution function, it exists only for **continuous** random variables.

The probability density is sometimes called a **differential function** or a **differential distribution law**.

A probability density $\varphi(x)$ graph is called a **distribution curve**.

◀ **Example 5.1.** The distribution function of a random variable X has the form:

$$F(x)=\begin{cases}0 & at\ x\le 0\\ x/2 & at\ 0<x\le 2\\ 1 & at\ x\ge 2\end{cases}$$

Find the probability density of a random variable X.

The solution. Probability density $\varphi(x)=F'(x)$, i.e

$$\varphi(x)=\begin{cases}0 & at\ x\le 0 \text{ i } x>2\\ 1/2 & at\ 0<x\le 2\end{cases}$$ ▶

5.2. Properties of the probability density of a continuous random variable

1. Probability density is an integral function, i.e

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

Proof. $\varphi(x)\ge 0$ as the derivative of a monotonically nondecreasing function $F(x)$.●

2. The probability of a continuous random variable falling into the interval $[a,\ b]$ is equal to the definite integral of its probability density within the range from *a* to *b*, i.e.

$$P(a\le X\le b)=\int_a^b \varphi(x)dx. \qquad (5.2)$$

Proof. According to the property 4 of the distribution function $P(a \le X \le b) = F(b) - F(a)$. Since $F(x)$ is the primitive for the probability density $\varphi(x)$ $(F'(x) = \varphi(x))$, according to the Newton-Leibnitz formula, the increment of the primitive on the segment $[a,\ b]$ is the definite integral $\int_a^b \varphi(x)dx$, i.e., formula (5.2) is correct.

The geometrically obtained probability is equal to the area of the figure, which is bounded from above by the distribution curve and rests on the segment $[a,\ b]$ (Fig. 5.2) ●

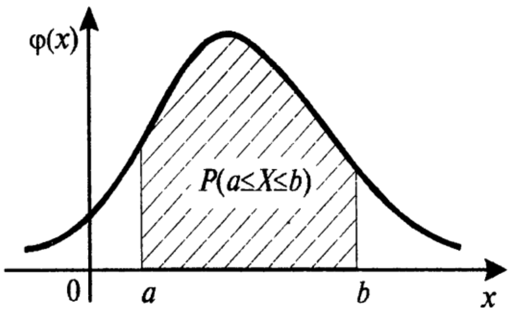


Fig. 5.2. Geometric illustration chance of hitting random values per segment $[a,b]$.

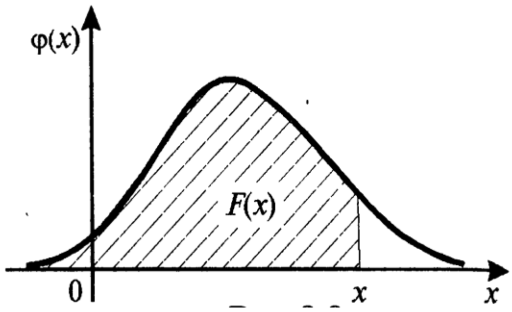


Fig. 5.3. Geometric illustration distribution functions of a random variable.

3. The distribution function of a continuous random variable can be expressed through the probability density by the formula:

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx \tag{5.3}$$

Geometrically, the distribution function is equal to the area of the figure, which is bounded from above by the distribution curve and lies to the left of point x (Fig. 5.3).

4. The integral within infinite limits of the probability density of a continuous

random variable is equal to unity:

$$\int_{-\infty}^{+\infty} \varphi(x)dx = 1. \qquad (5.4)$$

According to the formula (5.3): $F(x) = \int_{-\infty}^{x} \varphi(x)dx$ and at $x \to +\infty$ $F(+\infty) = 1$, that is, equality is verified (5.4).

Geometrically, properties 1 and 4 of the probability density mean that its graph - the **distribution curve** - does not lie below the abscissa axis, and the total area of the figure bounded by the distribution curve and the abscissa axis is equal to one. The concepts of mathematical expectation $M(X)$ and dispersion $D(X)$ introduced earlier for a discrete random variable can be extended to continuous random variables.

To obtain the corresponding formulas for $M(X)$ and $D(X)$ it is sufficient to replace in the formulas for a discrete random variable X:

— the sign of the sum $\sum_{i=1}^{n}$ with the sign of the integral with infinite limits $\int_{-\infty}^{+\infty}$;

— discrete values x_i with x, which changes continuously;

— probability p_i with probability element $\varphi(x)dx$.

By the element of probability we will understand the probability of a random value X hitting a segment $[x,\ x+dx]$ (with precision to infinitely small higher orders); geometrically, the element of probability is approximately equal to the area of the elementary rectangle, which rests on the segment $[x,\ x+dx]$ (Fig. 5.4).

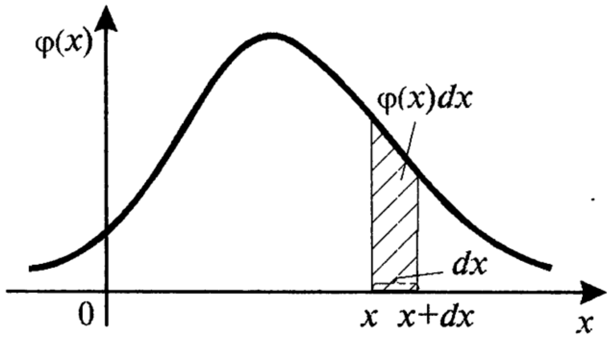


Fig. 5.4. An element of probability

As a result, we have the following formulas for the mathematical expectation and dispersion of a continuous random variable X:

$$a = M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx, \quad (5.5)$$

(if the integral coincides absolutely) and

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 \varphi(x)dx, \quad (5.6)$$

(if the integral converges).

In practice, the range of values of the random variable for which $\varphi(x) \neq 0$, is limited and the indicated integrals coincide, $M(X)$ and $D(X)$ therefore also exist.

All the properties of mathematical expectation and dispersion discussed above for discrete random variables are also valid for continuous variables.

In particular, property 3 of dispersion has the form:

$$D(X) = M(X^2) - a^2 \quad \text{or} \quad D(X) = \int_{-\infty}^{+\infty} x^2 \varphi(x)dx - a^2. \quad (5.7)$$

Remark. Along with discrete and continuous random variables, in practice there are **mixed** random variables, for which the distribution function *F(x)* is continuous on some segments, and has gaps in some points. An example of a mixed random variable can be a worker's earnings, proportional to his output, but not less than the guaranteed amount of payment x_0. (When $x = x_0$, the distribution function *F(x)* has a jump from zero to some value of p_0, and when $x > x_0$, it increases continuously) For mixed random variables, the formula (4.5) of the probability of a random variable falling into any interval $[x_0,\ x_1)$ remains valid.

◄ **Example 5.2.** The function $\varphi(x)$ is given in the form:

$$\varphi(x) = \begin{cases} 0 & at\ x \le 1 \\ \dfrac{A}{x^4} & at\ x > 1 \end{cases}$$

Find: *a*) the value of the constant *A*, at which the function will be the probability density of some random variable X;

b) the expression of the distribution function $F(x)$;

c) calculate the probability that the random variable *X* will take the value on the segment [2; 3];

d) find the mathematical expectation and dispersion of a random variable X.

The solution: *a*) To be the probability density $\varphi(x)$ of some random variable X, it must be non-negative, that is $\varphi(x) \ge 0$, or $\dfrac{A}{x^4} \ge 0$, hence, from here $A \ge 0$, and it must satisfy property 4. Therefore, according to formula (5.4), $\int_{-\infty}^{+\infty} \varphi(x)dx = 1$. In accordance,

$$\int_{-\infty}^{+\infty} \varphi(x)dx = \int_{-\infty}^{1} 0 \cdot dx + \int_{1}^{+\infty} \frac{A}{x^4}dx = 0 + \lim_{b\to+\infty} \int_{1}^{b} \frac{A}{x^4}dx =$$

$$= \frac{A}{3} \lim_{b\to+\infty}\left(-\frac{1}{x^3}\Big|_1^b\right) = \frac{A}{3} \lim_{b\to+\infty}\left(1 - \frac{1}{b^3}\right) = \frac{A}{3} = 1$$

So, from here $A = 3$.

b) Using formula (5.2), we find $F(x)$.

If $x \le 1$, then $F(x) = \int_{-\infty}^{x} \varphi(x)dx = \int_{-\infty}^{x} 0 \cdot dx = 0$.

If $x > 1$, then $F(x) = 0 + \int_{1}^{x} \frac{3}{x^4}dx = -\frac{1}{x^3}\Big|_1^x = 1 - \frac{1}{x^3}$.

So, $F(x) = \begin{cases} 0 & at\ x \le 1 \\ 1 - \frac{1}{x^3} & at\ x > 1 \end{cases}$.

c) Using formula (5.2)

$$P(2 \le X \le 3) = \int_{2}^{3} \frac{3}{x^4}dx = -\frac{1}{x^3}\Big|_2^3 = \frac{1}{2^3} - \frac{1}{3^3} = \frac{19}{216}.$$

The probability $P(2 \le X \le 3)$ could be found directly as the increment of the distribution function:

$$P(2 \le X \le 3) = F(3) - F(2) = \left(1 - \frac{1}{3^3}\right) - \left(1 - \frac{1}{2^3}\right) = \frac{19}{216}.$$

d) Using formula (5.5) let's calculate

$$a = M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx = \int_{-\infty}^{1} 0 \cdot dx + \int_{1}^{+\infty} x\left(\frac{3}{x^4}\right)dx = 0 + 3 \lim_{b \to +\infty} \int_{1}^{b} \frac{dx}{x^3} =$$

$$= 3 \lim_{b \to +\infty}\left(-\frac{1}{2x^2}\Big|_1^b\right) = \frac{3}{2}\lim_{b \to +\infty}\left(1 - \frac{1}{b^2}\right) = \frac{3}{2}.$$

We calculate the dispersion $D(X)$ using the formula (5.7). First we will find

$$M(X^2) = \int_{-\infty}^{+\infty} x^2\varphi(x)dx = \int_{-\infty}^{+\infty} x^2\left(\frac{3}{x^4}\right)dx = 3.$$

(the calculation of the integral is similar to the one above). Therefore,

$$D(X) = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} ►$$

Remark. In some cases, if there is a graph of the distribution function $F(x)$, it is useful to have in mind the geometric interpretation of the mathematical expectation $M(X)$ of a random variable X: $M(X) = S_2 - S_1$, where S_1 and S_2 are the areas of the figures bounded, respectively, by the *Oy* axis, a straight line $y = 1$ and a curve $y = F(x)$ on the interval $(0, +\infty)$ and between the curve $y = F(x)$ and the *Ox* axes and *Oy* on the interval $(-\infty, 0)$ (Fig. 5.5).

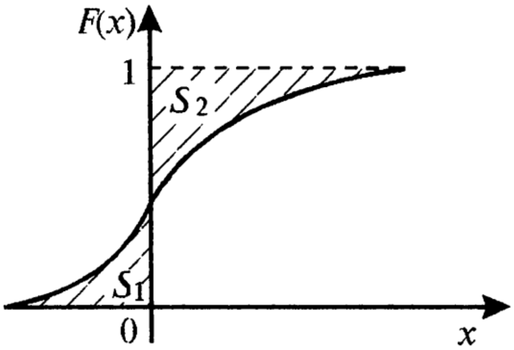


Fig. 5.5. Geometric interpretation mathematical expectation of randomness values X.

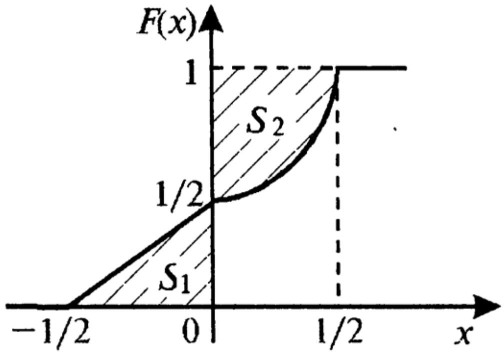


Fig. 5.6. Mathematical expectation randomness values *X*, given by the distribution function.

For example, to find the mathematical expectation $M(X)$ of a random variable *X* given by the distribution function $F(x)$, which consists of straight line segments and an arc of a circle (Fig. 5.6), it is not necessary to find $\varphi(x)$, and then $M(X)$. It is much easier to find $M(X)$, using its geometric interpretation, i.e

$$M(X) = S_2 - S_1 = \frac{1}{4}\pi R^2 - \frac{1}{2}ah = \frac{1}{4}\pi\left(\frac{1}{2}\right)^2 - \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{\pi - 2}{16} \approx 0{,}072.$$

5.3. Mode and median. Quantiles. Moments of random variables. Asymmetry and excess

In addition to mathematical expectation and dispersion, the theory of probability also uses several numerical characteristics that reflect certain properties of the distribution.

The mode $Mo(X)$ of a random variable X is its most probable value (for which the probability p_i or probability density $\varphi(x)$ reaches a maximum).

If the probability or probability density reaches a maximum not at one, but at several points, the distribution is called **polymodal** (Fig. 5.7).

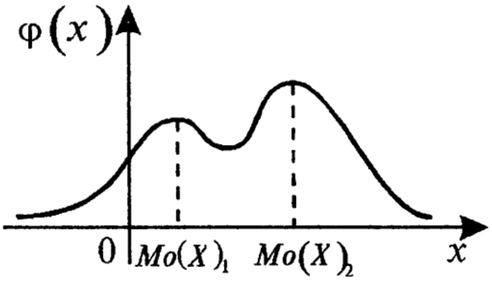


Fig. 5.7. Polymodal distribution

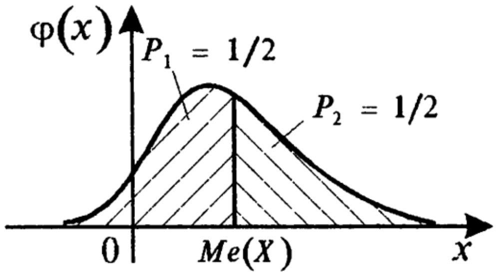


Fig. 5.8. The median is continuous random variable

The median $Me(X)$ of a continuous random variable X is its value for which

$$P(X < Me(X)) = P(X > Me(X)) = \frac{1}{2}, \tag{5.8}$$

that is, the probability that the random variable X will acquire a value less than the median $Me(X)$ or greater than $Me(X)$, is the same and equal to 1/2. Geometrically, a vertical line $x = Me(X)$ passing through a point with an abscissa equal to $Me(X)$, divides the area of the figure under the distribution curve into two equal parts (Fig. 5.8). It is obvious that at the point $x = Me(X)$ the distribution function is equal to 1/2, i.e. $F(Me(X)) = \frac{1}{2}$ (Fig. 5.9).

◄ **Example 5.3.** Find the mode, median, and mathematical expectation of a random variable X with a probability density $\varphi(x) = 3x^2$ of $x \in [0;1]$.

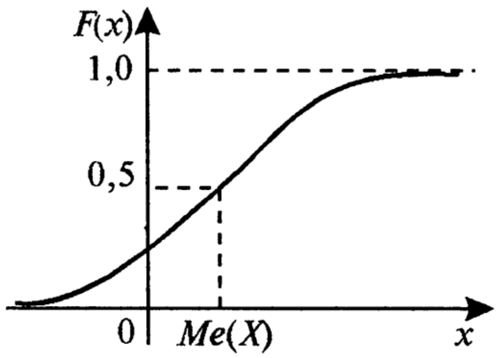


Fig. 5.9. The value of the distribution function in points $x = Me(X)$.

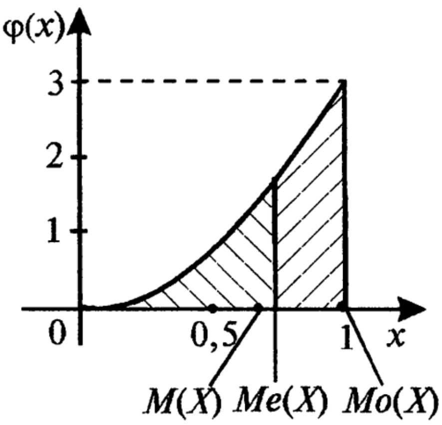


Fig. 5.10. Probability density (for example 5.3.)

The solution. The distribution curve is shown in Fig. 5.10. Obviously, the probability density $\varphi(x)$ is maximal at $x = Mo(X) = 1$. Median $Me(X) = b$ we find from condition (5.8):

$$\int_{-\infty}^{b} \varphi(x)dx = \frac{1}{2} \text{ або } \int_{-\infty}^{b} \varphi(x)dx = \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{b} 3x^2 dx = x^3 \Big|_0^b = b^3 = \frac{1}{2},$$

отже, $b = Me(X) = \sqrt[3]{1/2} \approx 0{,}79$.

We calculate the mathematical expectation using the formula (5.5):

$$M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx = \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{1} x\left(3x^2\right)dx + \int_{1}^{+\infty} 0 \cdot dx = \frac{3}{4}x^4 \Big|_0^1 = 0{,}75.$$

The relative location of the points $M(X), Me(X), Mo(X)$ in ascending order of the abscissa is shown in Fig.5.10 ►

Along with the numerical characteristics mentioned above, the concepts of **quantiles** and **percentage** points are used to describe random variables.

A quantile of level *q* or **(q-quantile)** is a value of a random variable at which its distribution function takes a value x_q equal to q, i.e.

$$F(x_q) = P(X < x_q) = q. \tag{5.9}$$

Some quantiles have received a special name. The median of a random variable is a quantile of the 0.5 level (i.e. $Me(X) = x_{0,5}$). Quantiles $x_{0,25}$ and $x_{0,75}$ received the name of the **upper** and **lower** quantiles, respectively.

Closely related to the concept of quantile is the concept of **percentage point**. The 100q% point is a quantile x_{1-q}, that is, the value of a random variable X at which $P(X \geq x_{1-q}) = q$.

◄ **Example 5.4.** According to the data of example 5.3, find the quantile $x_{0,3}$ and the 30% point of the random variable X.

The solution. According to the formula (5.5) the distribution function

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx = \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{x} 3x^2 dx = x^3.$$

We find the quantile $x_{0,3}$ from equation (5.9), i.e. $x_{0,3}^3 = 0,3$, therefore, $x_{0,3} \approx 0,67$.

Let's find the 30% point of the random variable X, or quantile $x_{0,7}$, from the equation $x_{0,7}^3 = 0,7$, therefore, $x_{0,7} \approx 0,89$ ►

Among the numerical characteristics of a random variable, a special place belongs to the **initial** and **central moments**.

The initial moment of the k-th order of a random variable X is called the mathematical expectation of the k-th power of this variable:

$$\nu_k = M(X^k). \qquad (5.10)$$

The central moment of the k-th order of a random variable X is called the mathematical expectation of the k-th power of the deviation of the random variable from its mathematical expectation:

$$\mu_k = M\left(X - M(X)\right)^k, \qquad (5.11)$$

or $\mu_k = M(X-a)^k$, where $a = M(X)$.

Formulas for calculating moments for discrete random variables (which take values x_i with probability p_i) and continuous variables (with density probabilities $\varphi(x)$) are given in the Table. 5.1. It is not difficult to notice that at $k=1$ the first initial moment, the random variable X turns into a mathematical expectation, that is $\nu_1 = M(X) = a$, at $k=2$ the second

Table 5.1.

Formulas for calculating moments for discrete and continuous random variables

Moment	A random variable	
	Discreet	Continuous
Initial	$\nu_k = \sum_{i=1}^{n} x_i^k p_i$ (5.12)	$\nu_k = \int_{-\infty}^{+\infty} x^k \varphi(x) dx$
Central	$\mu_k = \sum_{i=1}^{n} (x_i - a)^k p_i$ (5.14)	(5.13)

		$\mu_k = \int_{-\infty}^{+\infty} (x-a)^k \varphi(x) dx$ (5.15)

the central moment is in the dispersion, i.e. $\mu_2 = D(X)$. The central moments μ_k can be expressed through the initial moments ν_k according to the formulas:

$$\mu_1 = 0,$$
$$\mu_2 = \nu_2 - \nu_1^2,$$
$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3,$$
$$\mu_4 = \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4 \text{ і т.д.}$$

For example,

$$\mu_3 = M(X-a)^3 = M(X^3 - 3aX^2 + 3a^2X - a^3) =$$
$$= M(X^3) - 3aM(X^2) + 3a^2M(X) - a^3 =$$
$$= \nu_3 - 3\nu_1\nu_2 + 3\nu_1^2\nu_1 - \nu_1^3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3$$

(when proving, it was taken into account that $a = M(X) = \nu_1$, it is a non-random value). It is known that the mathematical expectation $M(X)$, or the first initial moment, characterizes the **average value** or the **position** of the distribution of a random variable X on the numerical axis; dispersion $D(X)$, or the second central moment μ_2, is the **degree of dispersion** of the distribution of X relative to $M(X)$. For a more detailed description of the distribution, moments of higher orders are used.

The third central moment μ_3 is used to characterize the asymmetry (skewness) of the distribution. It has the dimension of a random variable cube. To obtain a dimensionless value, it is divided by σ^3, where σ is the mean square deviation of the random variable X. The obtained value A is called the **coefficient of asymmetry** of the random variable:

$$A = \frac{\mu_3}{\sigma^3}. \tag{5.16}$$

If the distribution is symmetric relative to the mathematical expectation, then the coefficient of asymmetry is $A = 0$.

In the Fig. 5.11 two distribution curves: I and II are shown. Curve I has positive (right-sided) asymmetry ($A > 0$), and curve II - negative (left-sided) ($A < 0$).

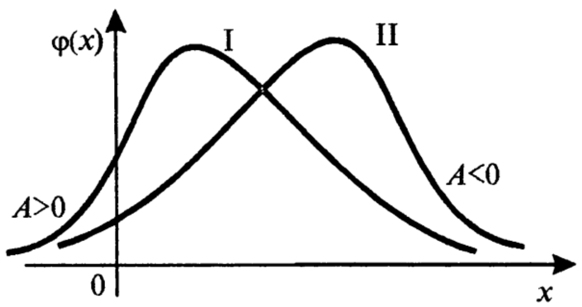


Fig. 5.11. Examples of asymmetry of distributions

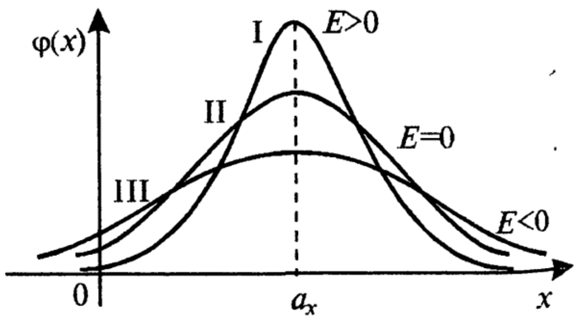


Fig. 5.12. Examples of different distributions "steepness"

The fourth central moment μ_4 is used to characterize the steepness (sharp-top or flat-top) of the distribution.

The **kurtosis** (or kurtosis coefficient) of a random variable is called a number

$$E = \frac{\mu_4}{\sigma_4} - 3. \tag{5.17}$$

The number 3 is subtracted from the ratio μ_4 / σ^4, because for the normal distribution, which occurs quite often, the ratio $\mu_4 / \sigma^4 = 3$. Curves with a sharper top than a normal curve have a positive kurtosis, and those with a flat top have a negative kurtosis (Fig. 5.12).

◄ **Example 5.5.** Find the coefficient of asymmetry and kurtosis of a random variable distributed according to Laplace's law with probability density $\varphi(x)=\frac{1}{2}e^{-|x|}$.

The solution. Since the distribution of the random variable X is symmetric concerning the ordinate axis, all odd (both initial and central) moments are equal to 0, i.e. $\nu_1=0,\ \nu_3=0,\ \mu_3=0$, and according to (5.12) the coefficient of asymmetry is $A=0$.

To find the kurtosis, it is necessary to calculate even the initial moments ν_2 and ν_4:

$$\nu_2=\int_{-\infty}^{+\infty}x^2\varphi(x)dx=\int_{-\infty}^{+\infty}x^2\left(\frac{1}{2}e^{-|x|}\right)dx=2\cdot\frac{1}{2}\int_{0}^{+\infty}x^2e^{-x}dx=2.$$

In accordance,

$$D(X)=\mu_2=\nu_2-\nu_1^2=2-0^2=2 \text{ i } \sigma_x=\sqrt{D(X)}=\sqrt{2}.$$

$$\nu_4=\int_{-\infty}^{+\infty}x^4\varphi(x)dx=\int_{-\infty}^{+\infty}x^4\left(\frac{1}{2}e^{-|x|}\right)dx=2\cdot\frac{1}{2}\int_{0}^{+\infty}x^4e^{-x}dx=24.$$

Now the kurtosis according to the formula (5.13) is

$$E=\frac{\mu_4}{\sigma^4}-3=\frac{24}{\left(\sqrt{2}\right)^4}-3=3.$$

The kurtosis of the distribution is positive, which indicates the sharp peak of the distribution curve $\varphi(x)$ (Fig. 5.13).

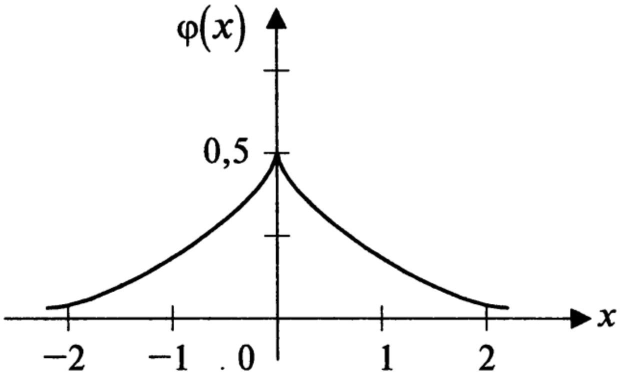


Fig. 5.13. Distribution curve (to example 5.5)

►

Control questions

1. Give examples of continuous random variables.
2. Explain: why the event $\alpha \le X \le \beta$ has a non-zero probability, although it consists of events whose probabilities are zero?
3. How the density of the distribution is related to the derivative of the function distribution?
4. What are moments of random variables used for? What do they characterize?
5. Define the asymmetry and the kurtosis of distributions of continuous random variables.

Chapter 6

Basic laws of random variable distributions

6.1. Binomial law of distribution

A discrete random variable X has a **binomial distribution law** with parameters n and p, if it takes the values 0, 1, 2, ..., *m*, ..., *n* with probabilities

$$P(X = m) = C_n^m p^m q^{n-m}, \tag{6.1}$$

where $0 < p < 1,\ q = 1 - p$.

The probabilities $P(X = m)$ are found according to Bernoulli's formula. So, the binomial law of distribution is the law of distribution of the number $X = m$ of occurrences of an event A in n independent trials, in each of which it can occur with the same probability p.

The distribution series of the binomial law has the form:

x_i	0	1	2	...	m	$\dots$	n
p_i	q^n	$C_n^1 pq^{n-1}$	$C_n^2 p^2 q^{n-2}$	$\dots$	$C_n^m p^m q^{n-m}$	$\dots$	p^n

The definition of the binomial law is correct, since the main property of the distribution series $\sum_{i=0}^{n} p_i = 1$ is fulfilled, because $\sum_{i=0}^{n} p_i$ is the sum of all terms of the Newton binomial:

$$q^n + C_n^1 pq^{n-1} + C_n^2 p^2 q^{n-2} + \dots + C_n^m p^m q^{n-m} + \dots + p^n = (q + p)^n = 1^n = 1$$

(hence the name of the law - binomial).

Theorem 6.1. The mathematical expectation of a random variable X distributed according to the binomial law**,**

$$M(X) = np \ , \tag{6.2}$$

and its dispersion

$$D(X) = npq \ . \tag{6.3}$$

Proof. Need to been done by yourselves.

Consequence. The mathematical expectation of the frequency $\frac{m}{n}$ of an event in n independent trials, in each of which it can appear with the same probability p, is equal to p, i.e.

$$M\left(\frac{m}{n}\right) = p \ , \tag{6.4}$$

and its dispersion

$$D\left(\frac{m}{n}\right) = \frac{pq}{n} \ . \tag{6.5}$$

The frequency $\frac{m}{n}$ of the event is $\frac{X}{n}$, i.e. $\frac{m}{n} = \frac{X}{n}$, where X is a random variable distributed according to the binomial law. Ago

$$M\left(\frac{m}{n}\right) = M\left(\frac{X}{n}\right) = \frac{1}{n} M(X) = \frac{1}{n} \cdot np = p,$$

$$D\left(\frac{m}{n}\right)=D\left(\frac{X}{n}\right)=\frac{1}{n^2}D(X)=\frac{1}{n^2}\cdot npq=\frac{pq}{n}.$$

Remark. Now the meaning of the arguments in the functions $f(x)$ and $\Phi(x)$ contained in the local and integral Moivre-Laplace theorem becomes clear. So, in the function $f(x)$, the argument $x=\frac{m-np}{\sqrt{npq}}$ is the deviation of the number $X=m$ of occurrences of event A in n independent trials, distributed according to the binomial law, from its average value $M(X)=np$, which is expressed in standard deviations $\sigma_x=\sqrt{D(X)}=\sqrt{npq}$. The argument $x=\frac{\Delta\sqrt{n}}{\sqrt{pq}}=\frac{\Delta}{\sqrt{pq/n}}$ in the function $\Phi(x)$ is the deviation Δ of the frequency $\frac{m}{n}$ of event A in n independent trials from its probability p in a separate trial, expressed in standard deviations $\sigma\left(\frac{m}{n}\right)=\sqrt{D\left(\frac{m}{n}\right)}=\sqrt{\frac{pq}{n}}$. The binomial distribution law is widely used in the theory and practice of statistical product quality control, to describe the functioning of mass service systems, in the modeling of asset prices, in the theory of shooting, and in other areas.

◄ **Example 6.1.** The store received shoes from two factories in ratios of 2:3. 4 pairs of shoes were bought. Find the law of distribution of the number of purchased pairs of shoes that were manufactured at the first factory. Find the mathematical expectation and mean square deviation of this random variable.

The solution. The probability that a randomly selected pair of shoes is made by the first factory is $p=\frac{2}{2+3}=0,4$. The random variable X - the number of pairs of shoes among the four produced at the first factory has a binomial distribution law with parameters $n = 4$, $p = 0{,}4$. The distribution series has the form:

x_i	0	1	2	3	4
p_i	0,1296	0,3456	0,3456	0,1536	0,0256

Values $p_i = P(X = m), (m = 0,1,2,3,4)$ calculated by the formula (6.1): $P(X = m) = C_4^m \cdot 0{,}4^m \cdot 0{,}6^{4-m}$.

Let's find the mathematical expectation and dispersion of a random variable X using formulas (6.2) and (6.3):

$M(X) = np = 4 \cdot 0{,}4 = 1{,}6,\ \ D(X) = npq = 4 \cdot 0{,}4 \cdot 0{,}6 = 0{,}96$.►

◄ **Example 6.2.** According to the data of example 6.1, find the mathematical expectation and dispersion of the frequency (proportion) of pairs of shoes made at the first factory among the 4 purchased pairs.

The solution. We have $n = 4$, $p = 0{,}4$. According to formulas (6.4) and (6.5):

$$M\left(\frac{m}{n}\right) = 0{,}4,\ \ D\left(\frac{m}{n}\right) = \frac{0{,}4 \cdot 0{,}6}{4} = 0{,}06.$$ ►

6.2. Poisson's distribution law

A discrete random variable X has the **Poisson's distribution law** with parameter $\lambda > 0$, if it takes the values 0, 1, 2, ..., *m*, ... (an infinite but countable number of values) with probabilities

$$P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!} = P_m(\lambda), \qquad (6.6)$$

The series of the distribution of the Poisson's law has the form:

x_i	0	1	2	...	m	...
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	...	$\frac{\lambda^m e^{-\lambda}}{m!}$	...

The definition of Poisson's law is correct because the basic property of the distribution series $\sum_{i=1}^{\infty} p_i = 1$ is fulfilled because the sum of the series

$$\sum_{i=1}^{\infty} p_i = e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2!} + \ldots + \frac{\lambda^m e^{-\lambda}}{m!} + \ldots =$$

$$= e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^m}{m!} + \ldots) = e^{-\lambda} \cdot e^{\lambda} = 1$$

(it is taken into account that the series expansion of the function e^x at is written in parentheses at $x = \lambda$).

Fig. 6.1 shows the polygon of the random distribution values distributed according to the Poisson's law $P(X = m) = P_m(\lambda)$ with parameters $\lambda = 0{,}5$; $\lambda = 1$; $\lambda = 2$; $\lambda = 3{,}5$.

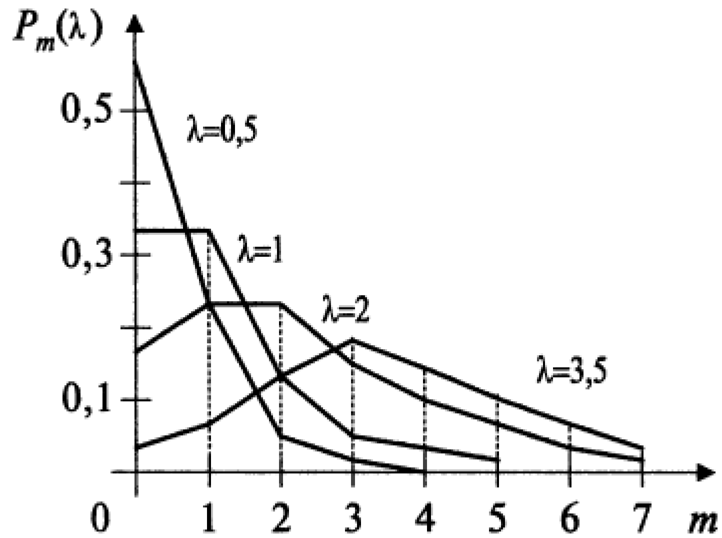


Fig. 6.1. The distribution polygon of a random variable distributed according to the

Theorem 6.2. The mathematical expectation and dispersion of a random variable distributed according to the Poisson's law coincide and are equal to the parameter λ of this law, i.e.

$$M(X) = \lambda, \tag{6.7}$$

$$D(X) = \lambda. \tag{6.8}$$

Proof.

Let's find the mathematical expectation of a random variable X:

$$a = M(X) = \sum_{i=1}^{\infty} x_i p_i = \sum_{m=0}^{\infty} m \frac{\lambda^m e^{-\lambda}}{m!} = \sum_{m=1}^{\infty} \frac{\lambda^m e^{-\lambda}}{(m-1)!} =$$

$$\lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = \lambda e^{-\lambda} (1 + \lambda + \frac{\lambda^2}{2!} + ...) = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

The dispersion of a random variable X can be found using the formula $D(X) = M(X^2) - a^2$. First, we get the formula for

$$M(X^2) = \sum_{i=1}^{\infty} x_i^2 p_i = \sum_{m=0}^{\infty} m^2 \frac{\lambda^m e^{-\lambda}}{m!} = \sum_{m=1}^{\infty} m \frac{\lambda^m e^{-\lambda}}{(m-1)!} =$$

$$= e^{-\lambda} \sum_{m=1}^{\infty} \frac{\big((m-1)+1\big)\lambda^m}{(m-1)!} = \lambda^2 e^{-\lambda} \sum_{m=2}^{\infty} \frac{\lambda^{m-2}}{(m-2)!} + \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} =$$

$$= \lambda^2 e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right) + \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right) =$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda$$

$$D(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda \quad \bullet$$

At sufficiently large n (generally at $n \to \infty$) and small values p $\left(p \to 0\right)$, provided that the product np is a constant value $\left(np \to \lambda = const\right)$, the Poisson's distribution law approaches the binomial distribution, since in this case the Poisson's probability function (6.6) well approximates the probability function (6.1), which is determined by the Bernoulli's formula. Therefore, the Poisson's distribution law is a limiting case of the binomial law at $n \to \infty, p \to 0, np \to \lambda = const$.

According to Poisson's law, for example, the number of births of quadruplets, the number of failures on an automatic line, the number of failures of a complex system in "normal mode", the number of "service requests" received per unit of time in mass service systems, etc. are distributed.

If a random variable is the sum of two independent random variables, each of which is distributed according to the Poisson's law, then that sum is also distributed according to the Poisson's law.

◄ **Example 6.3**. Prove that the sum of two independent random variables distributed according to the Poisson's law with parameters λ_1 and λ_2, is also distributed according to the Poisson's law with parameter $\lambda = \lambda_1 + \lambda_2$.

The solution. Let the random variables $X = m$ and $Y = n$ be distributed according to the Poisson's law, respectively, with the parameters λ_1 and λ_2. Since the random variables X and Y are independent, $Z = X \cdot Y$ takes the value s with probability

$$P(Z = s) = P(X = m) \cdot P(Y = n) =$$

$$= \sum_{m+n=s} \frac{\lambda_1^m e^{-\lambda_1}}{m!} \cdot \frac{\lambda_2^n e^{-\lambda_2}}{n!} = e^{-(\lambda_1+\lambda_2)} \sum_{m+n=s} \frac{\lambda_1^m \lambda_2^n}{m!n!} =$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{n=0}^{s} \frac{\lambda_1^{s-n} \cdot \lambda_2^n}{(s-n)!n!} = \frac{e^{-(\lambda_1+\lambda_2)}}{s!} \sum_{n=0}^{s} \frac{s!}{(s-n)!n!} \lambda_1^{s-n} \lambda_2^n.$$

Given that $\lambda = \lambda_1 + \lambda_2$, and given that

$$\sum_{n=0}^{s} \frac{s!}{(s-n)!n!} \lambda_1^{s-n} \lambda_2^n = \sum_{n=0}^{s} C_s^n \lambda_1^{s-n} \lambda_2^n = (\lambda_1 + \lambda_2)^s = \lambda^s,$$

that is $P(Z = s) = \dfrac{e^{-\lambda} \lambda^s}{s!}$, we get a random variable $Z = XY$ distributed according to the Poisson's law with the parameter $\lambda = \lambda_1 + \lambda_2$.▶

6.3. Geometric distribution

A discrete random variable $X = m$ has a **geometric distribution** with parameter p, if it takes the values 1, 2, ..., *m* ... (an infinite but countable set of values) with probabilities

$$P(X = m) = pq^{m-1}, \tag{6.9}$$

where $0 < p < 1,\ q = 1 - p$.

The series of the geometric distribution of a random variable has the form:

x_i	1	2	3	...	m	…
p_i	p	pq	pq^2	…	pq^{m-1}	…

The probabilities p_i, form a geometric progression with the first term p and the denominator q (hence the name "geometric distribution").

The definition of the geometric distribution is correct, since the sum of the series

$$\sum_{i=1}^{\infty} p_i = p + pq + ... + pq^{m-1} + ... = p(1 + q + ... + q^{m-1} + ...) = p\frac{1}{1-q} = \frac{p}{p} = 1$$

since $\frac{1}{1-q} = \frac{1}{p}$ — is the sum of the geometric series $\sum_{m=1}^{\infty} q^{m-1}$ at $|q| < 1$).

A random variable $X = m$ having a geometric distribution is the number m of trials that are carried out according to the Bernoulli's scheme, with the probability p of an event occurring in each trial until the first positive result.

So, for example, the number of calls by the radio operator on the correspondent until the call is accepted is a random variable that has a geometric distribution with parameter $p = 0,4$.

$$M(X) = \frac{1}{p}, \tag{6.10}$$

and its dispersion

$$D(X) = \frac{q}{p^2}, \tag{6.11}$$

where $q = 1 - p$.

◄ **Example 6.4.** Inspection of a large batch of details is carried out until a defective one is detected (without a limit on the number of inspected details). Draw up the law of distribution of the number of checked details. Find its mathematical expectation and dispersion, if it is known that the probability of failure for each detail is 0.1.

The solution. The random variable X is the number of checked details before detecting a defective one, has a geometric distribution (6.9) with the parameter $p = 0,1$. Therefore, the distribution series has the form:

$X = m$:

x_i	1	2	3	4	...	m	...
p_i	0,1	0,09	0,081	0,0729	...	$0,9^m \cdot 0,1$	...

According to formulas (6.10) and (6.11)

$$M(X) = \frac{1}{p} = \frac{1}{0,1} = 10, \quad D(X) = \frac{q}{p^2} = \frac{0,9}{0,1^2} = 90. \ ►$$

6.4. Hypergeometric distribution

A discrete random variable X has a hypergeometric distribution with parameters n, M, N, if it takes the values 0, 1, 2, …, *m*,…, *min (n, M)* with probabilities

$$P(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \qquad (6.12)$$

where $M \le N, \ n \le N; \ n, M, N -$ natural numbers.

The hypergeometric distribution has a random variable $X = m$ - the number of objects that have a given property among n objects randomly obtained (without return) from the set of N objects that M have this property.

Thus, the distribution of a random variable X - the number of inaccurate devices, among four taken at random - is a hypergeometric distribution with parameters $n = 4, \ M = 3, \ N = 10$.

Theorem 6.3. The mathematical expectation of a random variable X that has a hypergeometric distribution with parameters n, M, N

$$M(X) = n\frac{M}{N}\ , \tag{6.13}$$

and its dispersion

$$D(X) = n\frac{M}{N-1}\left(1-\frac{M}{N}\right)\left(1-\frac{n}{N}\right). \tag{6.14}$$

A random variable $X = m$, distributed according to the binomial law, can be interpreted as the number m of objects having a given property, from the total number of n objects randomly selected from some imaginary infinite set, where p of which objects has this property. Therefore, the hypergeometric distribution can be considered as a modification of the binomial distribution for the case of a finite set consisting of objects N from which M have this property.

The hypergeometric distribution is widely used in the practice of statistical acceptance control of the quality of industrial products, in tasks related to the organization of sample surveys, and other areas.

◄ **Example 6.5.** In the lottery, participants who guessed 3, 4, 5 and 6 sports out of a randomly selected 6 out of 45 receive cash prizes (the amount of the reward increases with the number of sports guessed). Find the law of the distribution of a random variable X - the number of guessed sports among randomly selected from six. What is the probability of receiving a monetary reward? Find the mathematical expectation and dispersion of a random variable**.**

The solution. The number of sports guessed in the 6 out of 45 lottery is a random variable that has a hypergeometric distribution with parameters $n = 6,\ M = 6,\ N = 45$. The series of its distribution:

x_i	0	1	2	3	4	5	6
p_i	0,40056	0,42413	0,15147	0,02244	0,00137	0,00003	0,0000001

Probability of receiving a cash prize:

$P(3 \le x \le 6) = \sum_{i=3}^{6} P(x=i) =$ 0,02244 + 0,00137 + 0,00003 + 0,0000001 = 0,02384≈0,024.

$$M(X) = 6 \cdot \frac{6}{45} = 0{,}8; \; D(X) = 6 \cdot \frac{39}{44}\left(1 - \frac{39}{45}\right)\left(1 - \frac{6}{45}\right) = 0{,}6145.$$

Thus, the average number of guessed sports out of 6 is only 0.8, and the probability of winning is only 0.024. ►

6.5. Uniform distribution law

A continuous random variable X has a uniform distribution law on the interval [*a, b*], if its probability density $\varphi(x)$ is constant on this interval and equal to zero outside its boundaries, i.e.

$$\varphi(x) = \begin{cases} \dfrac{1}{b-a}, & \text{при } a \le x \le b, \\ 0, & \text{при } x < a,\, x > b. \end{cases} \tag{6.15}$$

The distribution curve $\varphi(x)$ and the graph of the distribution function $F(x)$ of a random variable X are shown in Fig. 6.2 *a, b*.

Theorem 6.4. The distribution function of the random variable *X*, which is distributed according to the uniform law, has the form:

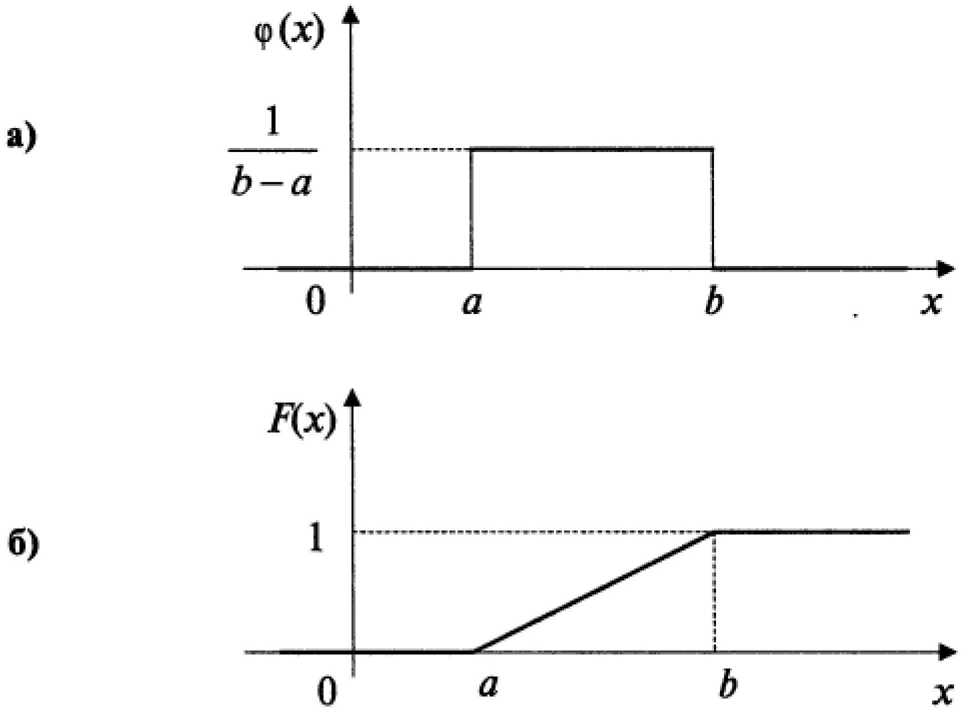


Fig. 6.2. a) a density plot of a uniform distribution; *b*) graph of the distribution function for the uniform law

$$F(x)=\begin{cases}0, & \text{при } x\le a\\ (x-a)/(b-a), & \text{при } a<x\le b\\ 1, & \text{при } x>b\end{cases}.$$

Its mathematical expectation

$$M(X)=\frac{a+b}{2}, \quad (6.17)$$

and dispersion

$$D(X)=\frac{(b-a)^2}{12}. \quad (6.18)$$

Proof. If $x\le a$ then the distribution function $F(x)=0$.

If $a<x\le b$ then:

$$F(x)=\int_a^x \frac{dx}{b-a}=\int_a^x \frac{x}{b-a}=\frac{x-a}{b-a}.$$

If $x>b$ then:

$$F(x)=\int_a^b \frac{dx}{b-a}=\frac{b-a}{b-a}=1.$$

Formula (6.16) is proved.

The mathematical expectation of a random variable X, taking into account its mechanical interpretation as the center of mass, is equal to the abscissa of the middle of the segment, i.e. $M(X)=\frac{a+b}{2}$. You can use general formulas:

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

$$D(X)\int_{-\infty}^{+\infty}\left[X-M(X)\right]^2\varphi(x)dx=\int_a^b\left(x-\frac{a+b}{2}\right)^2\frac{dx}{b-a}=$$

$$\frac{1}{3(b-a)}\left(x-\frac{a+b}{2}\right)^2=\frac{1}{3(b-a)}\left(\frac{(b-a)^2}{8}-\frac{(a-b)^3}{8}\right)=\frac{(b-a)^2}{12}\ \bullet$$

The uniform distribution law is used in the analysis of rounding errors in numerical calculations, in some mass service tasks, in the statistical modeling of observations subject to a given distribution. A random variable distributed according to a uniform law on the interval [0;1] is called a "random number from 0 to 1" and is used to obtain random variables with any distribution law.

◄ **Example 6.6.** Metro trains run at intervals of 2 minutes. A passenger steps onto the platform at a certain point in time. What is the probability that the passenger will have to wait no more than half a minute? Find the mathematical expectation and mean square deviation of a random variable X - train waiting time.

The solution. The random variable X - the waiting time for the train in the time (in minutes) segment [0;2] has a uniform distribution law $\varphi(x)=\frac{1}{2}$

Therefore, the probability that the passenger will have to wait no more than half a minute is equal $1/4$ to the area of the rectangle, which is equal to one (Fig. 6.3), i.e.

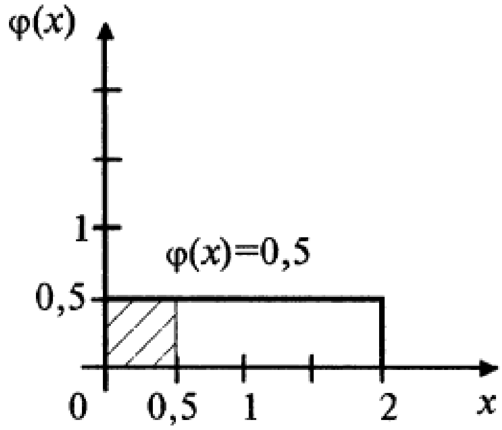


Fig. 6.3. Density function (to example 6.6)

$$P(X\le 0,5)=\int_0^{0,5}\frac{1}{2}dx=\frac{1}{2}x=\frac{1}{4};\qquad M(X)=\frac{0+2}{2}=1(хв);$$

$$D(X)=\frac{(2-0)^2}{12}=\frac{1}{3};\ \sigma_x=\sqrt{D(X)}=\sqrt{\frac{1}{3}}=\frac{\sqrt{3}}{3}\approx 0,58(хв).\blacktriangleright$$

6.6. Exponential distribution law

A continuous random variable X is distributed according to the **exponential law** with parameter $\lambda > 0$, if its probability density has the form:

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{при } x \geq 0 \\ 0, & \text{при } x < 0 \end{cases}. \qquad (6.19)$$

The curve of the distribution density $\varphi(x)$ and the graph of the distribution function $F(x)$ of a random variable X are shown in Fig. 6.4 *a, b*.

Theorem 6.5. The distribution function of a random variable X distributed according to the exponential law has the form:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{при } x \geq 0 \\ 0, & \text{при } x < 0 \end{cases}, \qquad (6.20)$$

its mathematical expectation

$$M(X) = \frac{1}{\lambda}, \qquad (6.21)$$

and dispersion

$$D(X) = \frac{1}{\lambda^2}. \qquad (6.22)$$

Proof. If $x < 0$ then distribution function $F(x) = 0$;

if $x \geq 0$ then

$$F(X) = \int_0^x \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^x = 1 - e^{-\lambda x}$$

Formula (6.19) is proved.

Let's find the mathematical expectation of a random variable X using the method of integration by parts:

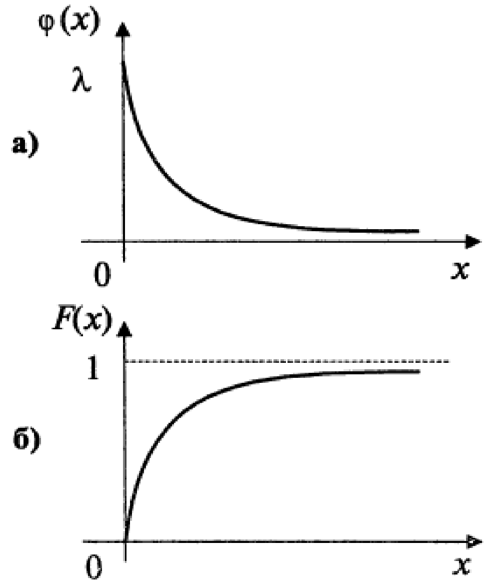


Fig. 6.4. a) a probability density plot of the exponential distribution; б) graph of the distribution function of a random variable distributed according to the exponential law.

$$a = M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx = \lim_{b\to\infty}\int_{0}^{b} x\lambda e^{-\lambda x}dx = \ldots = \frac{1}{\lambda}.$$

To find the dispersion $D(X)$, we first define

$$M(X^2) = \int_{-\infty}^{\infty} x^2\lambda e^{-\lambda x}dx = \ldots = \frac{2}{\lambda^2}.$$

So, $$D(X) = M(X)^2 - a^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \bullet$$

It follows from the proven theorem that for a random variable that has an exponential distribution law, the mathematical expectation is equal to the mean squared deviation, i.e.: $M(X) = \sigma_x = \frac{1}{\lambda}$

The exponential distribution law plays a large role in mass service theory and reliability theory. So, for example, the time interval T between two adjacent events in a simple flow has an exponential distribution with the parameter λ – flow intensity.
The property of the exponential distribution is widely used in Markov random processes.

◄ **Example 6.7.** It was established that the time to repair TVs is a random variable X with an exponential distribution. Determine the probability that it will take at least twenty days to repair the TV, if the average time to repair TVs is 15 days. Find the probability density, distribution function, and mean square deviation of a random variable X.

The solution. Under the condition of mathematical expectation $M(X)=\frac{1}{\lambda}=15$, hence the parameter $\lambda = 1/15$ and the probability density and distribution function have the form:

$$\varphi(x)=\frac{1}{15}e^{-\frac{1}{15}x};\ F(x)=1-e^{-\frac{1}{15}x}\ (x \geq 0).$$

The desired probability $P(X \geq 20)$ can be found by integrating the probability density, i.e. $P(X \geq 20)=P(20 \leq X < +\infty)=\int\limits_{-\infty}^{+\infty}\frac{1}{15}e^{-\frac{1}{15}x}dx$,

but it's easier to do it using the distribution function:

$$P(X \geq 20)=1-P(X<20)=1-F(20)=1-\left(1-e^{-\frac{20}{15}}\right)=e^{-\frac{20}{15}}=0{,}264.$$

Mean square deviation: $\sigma_x = M(X)$ = 15 days ►

6.7. Normal distribution law

The **normal law of distribution** is most often encountered in practice. The feature that sets it apart from other laws is that it is a limiting law to which other distributions approach under certain typical conditions encountered.

A continuous random variable X is distributed according to the **normal law** (Gauss's law) with parameters a and σ^2, if its probability density has the form:

$$\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (6.23)$$

The curve of the normal distribution law is called a **normal** or **Gaussian** curve. In Fig. 6.5. *a, b* show a normal curve $\varphi_N(x)$ with parameters a and σ^2, i.e. $N(a;\sigma^2)$, and a graph of the distribution function of a random variable X that has a normal law.

The normal curve is symmetric with respect to the straight line $x = a$, has a maximum at the point equal to $1/(\sigma\sqrt{2\pi})$, i.e. $f_{max}(a) = \frac{1}{\sigma\sqrt{2\pi}} \approx \frac{0,3989}{\sigma}$ and two inflection points $x = a \pm \sigma$ with ordinates $f_{nep}(a \pm \sigma) = \frac{1}{\sigma\sqrt{2\pi e}} \approx \frac{0,2420}{\sigma}$.

Theorem 6.6. The mathematical expectation of a random variable X, which is distributed according to a normal law, is equal to the parameter a of this law, i.e. $M(X) = a$, and its dispersion is equal to the parameter σ^2, i.e. $D(X) = \sigma^2$.

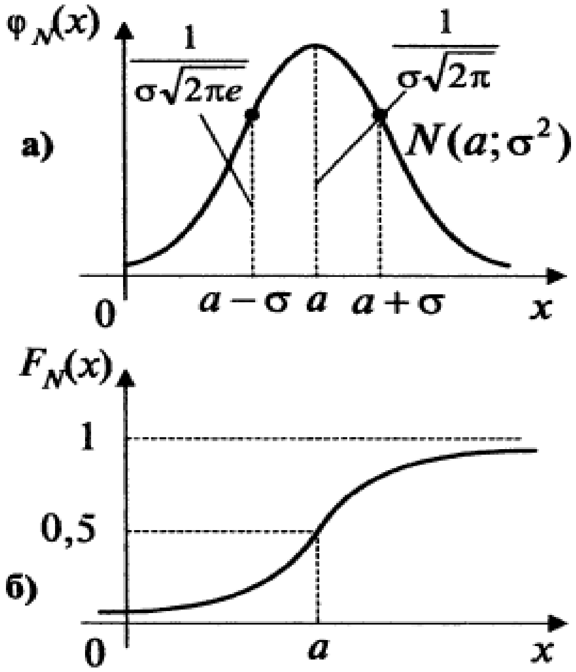


Fig. 6.5. *a*) normal law probability density plot; *b*) graph of the normal distribution function.

Proof. Mathematical expectation of a random variable X:

$$M(X) = \int_{-\infty}^{\infty} x\varphi_N(x)dx = \int_{-\infty}^{\infty} x\frac{1}{\sigma\sqrt{2\pi}}dx$$

Let's replace the variable: $t = \dfrac{x-a}{\sigma\sqrt{2}}$.

Then $x = a + \sigma\sqrt{2t}$ and $dx = \sigma\sqrt{2}dt$, the limits of integration do not change, therefore

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

$$= \frac{\sigma\sqrt{2}}{\sqrt{\pi}}\int_{-\infty}^{\infty} te^{-t^2}dt + \frac{a}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-t^2}dt = 0 + \frac{a}{\sqrt{\pi}}\sqrt{\pi} = a$$

(the first integral is equal to zero as the integral of the odd function that is symmetric with respect to the origin of the interval, and the second integral $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$ is the Euler-Poisson integral).

The dispersion of a random variable X:

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

Let's make a replacement:

$$D(X) = \int_{-\infty}^{\infty} \sigma^2 \cdot 2t^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2} \sigma\sqrt{2} dt = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt = -\frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t de^{-t^2} .$$

Applying the method of integration by parts, we get

$$D(X) = -\frac{\sigma^2}{\sqrt{\pi}} te^{-t^2} + \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = 0 + \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2 \quad \bullet$$

Let's find out how the normal curve will change when the parameters a and σ^2 (or σ) are changed. If $\sigma = const$, and the parameter $a(a_1 < a_2 < a_3)$ changes, that is, the center of symmetry of the distribution, then the normal curve will shift along the abscissa axis without changing its shape (Fig. 6.6).

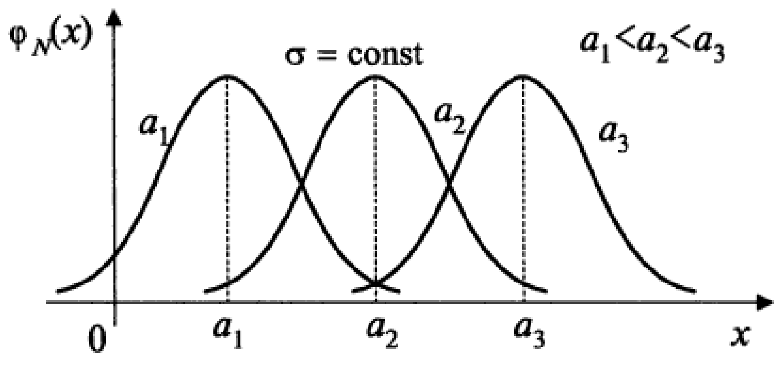


Fig. 6.6. Change of the normal curve when changing the parameter a.

If $a = const$ the parameter σ^2 (or σ) changes, then it changes the ordinate of the maximum $f_{max}(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$ of the curve. As the ordinate of the maximum of the curve increases, it σ decreases, but since the area under any distribution curve must remain equal to unity, the curve becomes flatter, stretching along the abscissa axis; when σ decreases, on the contrary, the normal curve stretches upwards, simultaneously contracting from the sides. Fig. 6.7 shows normal curves with parameters $\sigma_1, \sigma_2, \sigma_3$ where $\sigma_1 < \sigma_2 < \sigma_3$. Thus, the parameter a (aka mathematical expectation) characterizes the position of the center, and the parameter (aka dispersion) characterizes the shape of the normal curve.

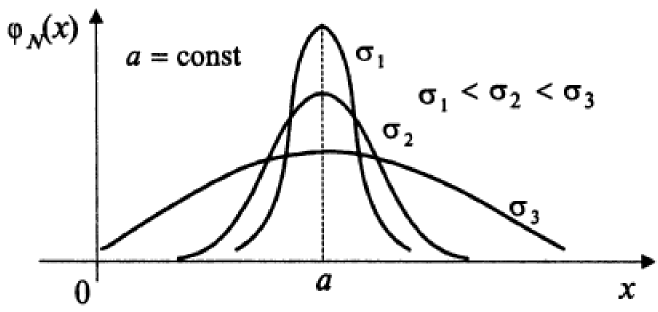


Fig. 6.7. Changing the shape of the normal curve when changing the parameter σ^2.

The normal distribution law of a random variable with parameters a=0, $\sigma^2 = 1$, i.e. $N(0; 1)$, is called **standard** or **normalized**, and the corresponding normal curve is called standard or normalized.

The difficulty of directly finding the distribution function of a random variable that has a normal law and the probability of its hitting a certain interval is because the integral of the function cannot be found in elementary functions. Therefore, they are expressed through a tabulated function:

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \qquad (6.24)$$

$\Phi(x)$ is **function (integral of probabilities) of Laplace**. The Laplace's function was already encountered during the consideration of the Moivre-Laplace's integral theorem, where its properties were also considered. Geometrically, the Laplace's function is the area under the standard normal curve on a segment $[-x;x]$ (Fig. 6.8).

Theorem 6.7. The distribution function of a random variable X, which has a normal distribution law, is expressed through the Laplace's function $\Phi(x)$ by the formula:

$$F_N(x) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right). \qquad (6.25)$$

Proof. By definition, the distribution function:

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

Let's replace the variable: $t = \frac{x-a}{\sigma}$, $x = a + t\sigma$, $dx = \sigma dt$, at $x \to -\infty$ $t \to -\infty$, ago

$$F_N(x) = \int_{-\infty}^{\frac{x-a}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2}\sigma dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-t^2/2}dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-t^2/2}dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x-a}{\sigma}} e^{-t^2/2}dt$$

The first integral

$$\int_{-\infty}^{0} e^{-t^2/2}dt = \frac{1}{2}\int_{-\infty}^{+\infty} e^{-t^2/2}dt = \frac{1}{2}\sqrt{2}\int_{-\infty}^{+\infty} e^{-(t/\sqrt{2})^2} d\left(\frac{t}{\sqrt{2}}\right) = \frac{\sqrt{2}}{2}\cdot\sqrt{\pi} = \sqrt{\frac{\pi}{2}}$$

(due to the parity of the integral function and the fact that the Euler-Poisson's integral is equal to $\sqrt{\pi}$).

The second integral, taking into account (6.24), takes the form: $\frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right)$.

Therefore, $F_N(x) = \frac{1}{\sqrt{2\pi}}\cdot\sqrt{\frac{\pi}{2}} + \frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right)$.●

Geometrically, the distribution function is the area under the normal curve on the interval $(-\infty, x)$ (Fig. 6.9). As you can see, it consists of two parts: the first, on the interval $(-\infty, a)$, which is equal to $1/2$, that is, half of the entire plane under the normal curve, and the second, on the interval (a, x), which is equal to $\frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right)$.

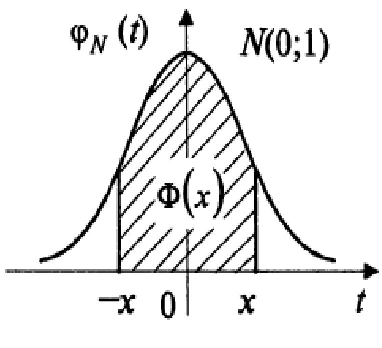


Fig. 6.8. Geometric interpretation Laplace functions

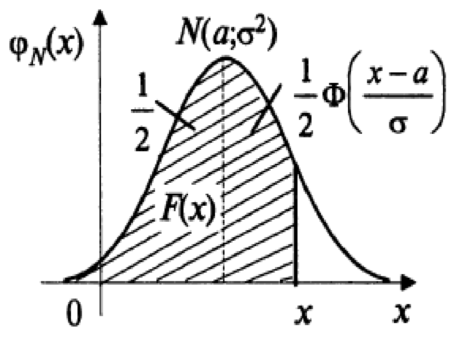


Fig. 6.9. Geometric interpretation distribution functions

Properties of a normally distributed random variable

1. The probability of a random variable X, which has a normal distribution, falling into the interval $[x_1, x_2]$, is equal to

$$P(x_1 \le X \le x_2) = \frac{1}{2}\left(\Phi(t_2) - \Phi(t_1)\right), \qquad (6.26)$$

where $t_1 = \dfrac{x_1 - a}{\sigma}$, $t_2 = \dfrac{x_2 - a}{\sigma}$. (6.26a)

Proof. Probability $P(x_1 \le X \le x_2)$ is the increment of the distribution function on the segment $[x_1, x_2]$, therefore,

$$P(x_1 \le X \le x_2) = F(x_2) - F(x_1) = \left(\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x_2 - a}{\sigma}\right)\right) - \left(\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x_1 - a}{\sigma}\right)\right) =$$

$$= \frac{1}{2}\left(\Phi(t_2) - \Phi(t_1)\right),$$

where t_1 and t_2 are determined by formula (6.26a) (Fig 6.10).●

2. The probability that the deviation of the random variable X, which has a normal distribution, from the mathematical expectation a will not exceed the value $\Delta > 0$ (in absolute value) is equal to

$$P(|X - a| \le \Delta) = \Phi(t), \qquad (6.27)$$

$$\text{where } t = \frac{\Delta}{\sigma} \qquad (6.27a)$$

Proof. $P(|X-a|\le\Delta)=P(a-\Delta\le X\le a+\Delta)$. Taking into account (6.26), as well as the property of the odd Laplace's function, we obtain $P(|X-a|\le\Delta)=\frac{1}{2}\left[\Phi\left(\frac{(a+\Delta)-a}{\sigma}\right)-\Phi\left(\frac{(a-\Delta)-a}{\sigma}\right)\right]=$

$=\frac{1}{2}\left[\Phi\left(\frac{\Delta}{\sigma}\right)-\Phi\left(-\frac{\Delta}{\sigma}\right)\right]=\frac{1}{2}\left[\Phi\left(\frac{\Delta}{\sigma}\right)+\Phi\left(\frac{\Delta}{\sigma}\right)\right]=\Phi\left(\frac{\Delta}{\sigma}\right)=\Phi(t)$, where $t=\Delta/\sigma$ (Fig. 6.11).

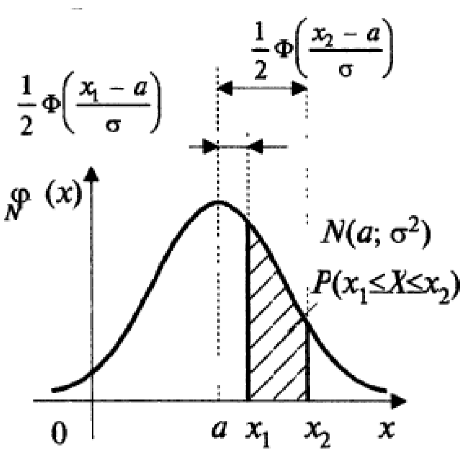


Fig. 6.10. Geometric interpretation probabilities $P(x_1\le X\le x_2)$

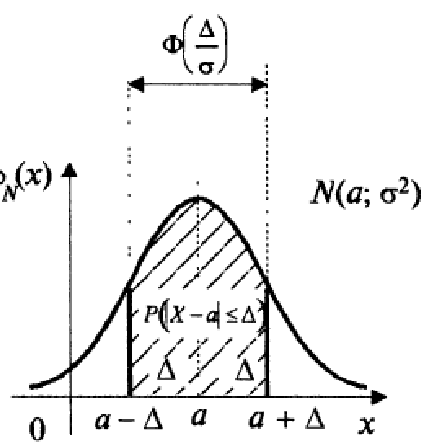


Fig. 6.11. Geometric interpretation probabilities $P(|X-a|\le\Delta)$

Remark. The considered approximate Moivre-Laplace's integral formula follows from the property of a normally distributed random variable at $x_1=a, x_2=b, a=np$ and $\sigma_x=\sqrt{npq}$, since the binomial law of the distribution of the random variable $X=m$ with parameters n and p, for which this formula was obtained, leads to the normal law at $n\to\infty$. Similarly, the consequences of the Moivre-Laplace's integral formula for the number $X=m$ of occurrence of an event in n independent trials and its frequency m/n follow from the properties of the normal law.

Let's calculate using the probability formula $P(|X-a|\leq \Delta)$ for different values of Δ (using the table in Appendix 3). We will get:

$$\Delta = \sigma \qquad P(|X-a|\leq \sigma) = \Phi(1) = 0{,}6827;$$

$$\Delta = 2\sigma \qquad P(|X-a|\leq 2\sigma) = \Phi(2) = 0{,}9545;$$

$$\Delta = 3\sigma \qquad P(|X-a|\leq 3\sigma) = \Phi(3) = 0{,}9973$$ (fig. 6.12).

Hence the "**rule of three sigma**":

if the random variable X has a normal distribution law with parameters a and σ^2, that is $N(a;\sigma^2)$, it is almost certain that its values fall into the interval $(a-3\sigma, a+3\sigma)$.

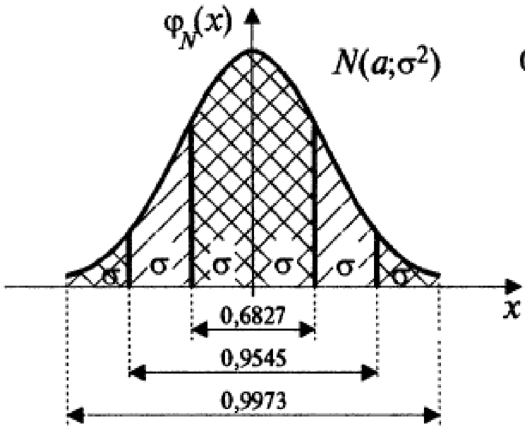


Fig. 6.12. Illustration of "rule of three sigma"

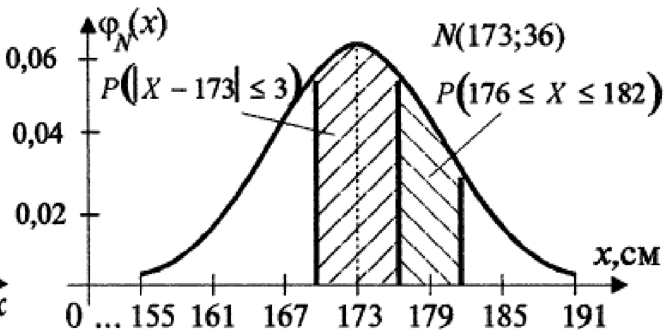


Fig. 6.13. Illustration to example 6.8

Deviation of a normally distributed random variable X by more than 3σ (in absolute value) is an almost impossible event, since its probability is very small:

$$P(|X-a|) > 3\sigma) = 1 - P(|X-a|\leq 3\sigma) = 1 - 0{,}9973 = 0{,}0027.$$

Let's find the **coefficient of asymmetry** and **kurtosis** of a random variable X that has a normal distribution. Due to the symmetry of the normal curve with respect to the vertical line $x = a$ that passes through the center $a = M(X)$ of the distribution, the coefficient of asymmetry of the normal distribution $A = 0$.

The kurtosis of a normally distributed random variable X:

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0,$$

where it is taken into account that the central moment of the 4th order is determined by the corresponding formula, i.e

$$\mu_4 = \int_{-\infty}^{+\infty} (x-a)^4 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = 3\sigma^4.$$

Thus, the kurtosis of the normal distribution equals zero, and the "steepness" of other distributions is determined in relation to the normal.

◄ **Example 6.8.** Assuming that the height of men of a certain age group is a normally distributed random variable X with parameters $a = 173$ and $\sigma^2 = 36$, write:

1. a) expression of the probability density and distribution function of a random variable X;

b) parts of suits of 4th height (176-182 cm) and 3rd height (170-176 cm), which must be provided in the total volume of production for this age group;

c) quantile $x_{0,7}$ and 10% point of random variable X.

2. Formulate the "rule of three sigma" for a random variable X.

The solution. 1. a)

$$\varphi_N(x)=\frac{1}{6\sqrt{2\pi}}e^{-\frac{(x-173)^2}{2\cdot 36}};$$

$$F_N(x)=\frac{1}{6\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-173)^2}{2\cdot 36}}dx=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{x-173}{6}\right);$$

b) the share of suits of the 4th height (176-182 cm) in the total volume of production will be determined by the appropriate formula, as the probability

$$P(176\le X\le 182)=\frac{1}{2}\left[\Phi(t_2)-\Phi(t_1)\right]=\frac{1}{2}\left[\Phi(1{,}50)-\Phi(0{,}50)\right]=$$

$=\frac{1}{2}(0{,}8664-0{,}3829)=0{,}2418$, (fig. 6.13), because

$t_1=\frac{176-173}{6}=0{,}50$; $t_2=\frac{182-173}{6}=1{,}50$.

The share of suits of the 3rd height can be determined similarly, but it is easier to do it using the formula $P(|X-a|\le\Delta)=\Phi(t)$, if we take into account that the given interval is symmetric with respect to the mathematical expectation $a=M(X)=173$, тобто $170\le X\le 176\ \Leftrightarrow\ |X-173|\le 3$,

$$P(170\le X\le 176)=P(|X-173|\le 3)=\Phi(\frac{3}{6})=\Phi(0{,}5)=0{,}3829;$$

c) we find the quantile $x_{0,7}$ of a random variable from the equation

$$F(x_{0.7})=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{x_{0.7}-173}{6}\right)=0.7,$$

where $\Phi\left(\frac{x_{0.7}-173}{6}\right)=\Phi(t)=0{,}4$. According to the table of applications 2 we find $t=0{,}524$ і $x_{0,7}=6t+173=176(см)$. This means that 70% of men in this group are up to 176 *cm* tall.

The 10% point is the quantile (found similarly). This means that 10% of men are at least 181 cm tall.

2. It is practically reliable that the height of men of this group is within the range of $a-3\sigma=173-3\cdot 6=155$ to $a+3\sigma=173+3\cdot 6=191$ (*см*), i.e, $155\leq X\leq 191$ (см).►

The normal law of distribution occupies a significant place in the theory and practice of probabilistic statistical methods.

Control questions

1. Basic properties of the binomial distribution law.
2. Basic properties of the Poisson's distribution law.
3. Uniform and exponential laws of distribution. What random values most often have given distributions?
4. It is known that a normally distributed random variable takes values: a) less than 248 with a probability of 0.975; b) more than 279 z with a probability of 0.005. Find the distribution function of a random variable.
5. We have a normally distributed random variable X with mathematical expectation a and dispersion σ^2. It is necessary to approximately replace the normal law of distribution with a uniform law in the intervals (α,β); select the borders α,β so as to keep mathematical expectation and dispersion of a random variable X unchanged.

Chapter 7

Multidimensional random variables

7.1. The concept of a multidimensional random variable and its distribution law

Quite often, the trial result is characterized by some system of random variables $X_1, X_2, ..., X_n$, which is called a multidimensional random variable or random vector.

Examples.

1. The success of a University graduate is characterized by a system of random variables $X_1, X_2, ..., X_n$ - grades from various disciplines.

2. The weather in this area at a certain time of day can be characterized by a system of random variables: X_1— temperature, X_2— humidity, X_3— pressure, etc.

Using the theoretical set interpretation, it can be asserted that any random variable X_i $\left(i = 1, ..., n\right)$ is a function of elementary events ω included in the space of elementary events $\Omega\left(\omega \in \Omega\right)$. Therefore, a multidimensional random variable can be considered as a function of elementary events ω:

$$\left(X_1, X_2, ..., X_n\right) = f\left(\omega\right).$$

This means that each elementary event ω is matched with several real numbers $x_1, x_2, ..., x_n$ that acquired random values $X_1, X_2, ..., X_n$ as a result of the trial.

◄ **Example 7.1.** Throw two dice at the same time; random variable X - the sum of points resulting from the trial; a random variable Y is their product. Show that the two-dimensional random variable (X,Y) is a function of the elementary events ω.

The solution. The set of elementary events consists of 36 elementary outcomes:

$$\Omega = (\omega_1,\ldots,\omega_{36}) = \{[1,1];\,[1,2];\ldots;[1,6];\,[2,1];\ldots;[6,6]\},$$

where, for example, $\omega_9 = [2,3]$, which means the appearance of a 2 on the first dice and the appearance of a 3 on the second dice. The result of the trial is an elementary event ω_i and the random variables X,Y acquire certain values: for example, when $\omega_9 = [2,3]$ $X = 5$ and $Y = 6$. The sum of all these values (X,Y) is a function of elementary events.►

Geometrically, two-dimensional (X,Y) and three-dimensional (X,Y,Z) random variables can be represented by a random point or a random vector on a plane or in space, while random variables X,Y or X,Y,Z are components of these vectors. In the case of n-dimensional space ($n > 3$) a random point or a random vector is meant, but the geometric interpretation loses its expressiveness. A multidimensional random variable is fully defined by its **distribution law**. In the case of a two-dimensional discrete random variable, its distribution is presented in the form of a distribution table (matrix), each cell (i,j) of which contains probabilities of the product of events

$p_{ij} = P\big((X = x_i)(Y = y_j)\big)$ (table. 7.1).

Table 7.1.

The distribution law of a two-dimensional random variable

y_j \ x_i	y_1	…	y_j	…	y_m	$\sum_{j=1}^{m}$
x_1	p_{11}	…	p_{1j}	…	p_{1m}	p_1
…	…	…	…	…	…	…
x_i	p_{i1}	…	p_{ij}	…	p_{im}	p_i
…	…	…	…	…	…	…
x_n	p_{n1}	…	p_{nj}	…	p_{nm}	p_n
$\sum_{i=1}^{n}$	p_1	…	p_j	…	p_m	1

The events $\left((X = x_i)(Y = y_j)\right)$, $(i = 1,...,n; j = 1,...,m)$ are incompatible and only possible, that is, they form a complete group, therefore, $\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1$. To find from the distribution table the probability that a one-dimensional random variable will acquire a certain value, it is necessary to find the sum of the probabilities p_{ij} of the corresponding row or column of the given table. If you fix the value of one of the arguments, for example, $Y = y_j$, then the resulting distribution of the random variable X will be called a **conditional distribution**. The probabilities $p_j(x_i)$ of this distribution will be the conditional probabilities of the event $X = x_i$, which are determined under the condition that the event $Y = y_j$ has occurred. Based on the definition of conditional probability

$$p_j(x_i) = \frac{P\left[(X = x_i)(Y = y_j)\right]}{P(Y = y_j)} = \frac{p_{ij}}{p_j} \qquad (7.1)$$

Similarly, it is possible to set the conditional distribution of a random variable Y under the condition $X = x_i$, using conditional probabilities:

$$p_i\left(y_j\right) = \frac{P\left[\left(X = x_i\right)\left(Y = y_j\right)\right]}{P\left(X = x_i\right)} = \frac{p_{ij}}{p_i} \tag{7.2}$$

◀ **Example 7.2.** The distribution law of a discrete two-dimensional random variable $\left(X, Y\right)$ is presented in the form of a table:

x_i \ y_j	-1	0	1	2
1	0,10	0,25	0,3	0,15
2	0,10	0,05	0,00	0,05

Find: 1) distribution laws of one-dimensional random variables X and Y;

2) conditional laws of the distribution of a random variable X under the condition $Y = 2$ and a random variable Y under the condition $X = 1$;

3) calculate $P\left(Y < X\right)$.

The solution. 1) A random variable can take the following values:

$X = 1$ with probability $p_1 = 1.10 + 0.25 + 0.30 + 0.15 = 0.8$;

$X = 2$ with probability $p_2 = 1.10 + 0.05 + 0.00 + 0.05 = 0.2$.

So, the law of distribution X has the form:

X :

x_i	1	2
p_i	0,8	0,2

Similarly, the distribution law Y has the form:

$Y:$

y_j	-1	0	1	2
p_j	0,2	0,3	0,3	0,2

2) We obtain the conditional law X of conditional distribution $Y=2$ if the probabilities p_{ij} in the last column of the table are divided by their sum, i.e. by $P(Y=2)=0,2$:

$X_{Y=2}:$

x_i	1	2
$p_j(x_i)$	0,75	0,25

Similarly, to obtain the conditional law Y of distribution under the conditions $X=1$, we divide probability p_{ij}, which are in the first row of the table, by their sum, i.e. by $P(X=1)=0,8$. Then:

$Y_{X=1}:$

y_j	-1	0	1	2
$p_i(y_j)$	0,125	0,3	0,3	0,2

►

7.1.1. The distribution function of a multidimensional random variable

The distribution function of n -dimensional random variable $(X_1, X_2, ..., X_n)$ is a function $F(x_1, x_2, ..., x_n)$ that is a form of expressing the probability of simultaneous fulfillment of inequalities $X_1 < x_1, X_2 < x_2, ..., X_n < x_n$, i.e.

$$F(x_1, x_2, ..., x_n) = P(X_1 < x_1, X_2 < x_2, ..., X_n < x_n). \quad (7.3)$$

In the case of a two-dimensional random variable (X, Y), the distribution function $F(x, y)$ is determined by the equality: $F(x, y) = P(X < x, Y < y)$. Geometrically, the distribution function $F(x, y)$ means the probability of a random point (X, Y) falling into the shaded area - the infinite quadrant located to the left and below the point $M(X, Y)$ (Fig. 7.1). The right and upper bounds of the region are not taken into account - this means that the distribution function is continuous from the left for each of the arguments.

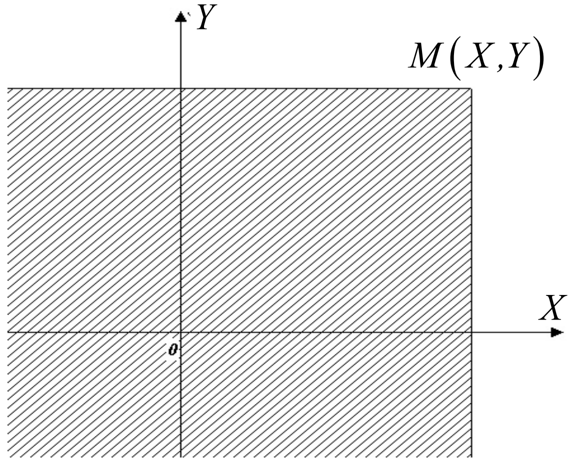


Fig. 7.1. Geometric interpretation of the two-dimensional random distribution function values

In the case of a discrete two-dimensional random variable, its distribution function is determined by the formula:

$$F(x,y) = \sum_i \sum_j p_{ij}, \tag{7.4}$$

where the sum of probabilities is found for all i for which $x_i < x$ and for all j for which $y_i < y$.

Properties of the distribution function of a two-dimensional random variable

1. The distribution function $F(x,y)$ is non-negative and its values are between zero and one, that is,

$$0 \le F(x,y) \le 1. \tag{7.5}$$

This statement follows from the fact that $F(x,y)$ is a form of expression of probability.

2. The distribution function $F(x,y)$ is a non-decreasing function for each of the arguments, that is,

when $x_2 > x_1$, $F(x_2,y) \ge F(x_1,y)$,

when $y_2 > y_1$, $F(x,y_2) \ge F(x,y_1)$.

As the shaded area (Fig. 7.1.) increases with the growth of any argument, the probability of a random point (X,Y) falling into this area also increases, so the distribution function $F(x,y)$ cannot decrease (decrease).

3. If at least one of the arguments turns into $-\infty$, the distribution function $F(x,y)$ is zero, that is,

$$F(x,-\infty) = F(-\infty,y) = F(-\infty,-\infty) = 0. \tag{7.6}$$

The distribution function $F(x,y)$ will be zero because the events $X < -\infty$, $Y < -\infty$ and their product are impossible events.

4. If one of the arguments is converted to $+\infty$, the distribution function $F(x,y)$ is correspondingly converted to the distribution function of the other argument:

$$F(x,+\infty) = F_1(x), \;\; F(y,+\infty) = F_2(y), \qquad (7.7)$$

where $F_1(x)$ and $F_2(y)$ are distribution functions of random variables X and Y , that is,

$$F_1(x) = P(X < x), \qquad F_2(y) = P(Y < y).$$

The product of an event $(X < x)$ and a reliable event $(Y < +\infty)$ is the event itself $(X < x)$, therefore $F(x,+\infty) = P(X < x) = F_1(x)$, Similarly, $F(y,+\infty) = F_2(y)$.

5. If both arguments are converted to , then the distribution function is equal to one:

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

$F(+\infty,+\infty) = 1$ is a consequence of the fact that reliable events took place at the same time: $(X < +\infty)$, $(Y < +\infty)$, and this event is also reliable.

Geometrically, the distribution function is some surface that has given properties. For a discrete two-dimensional random (X,Y), its distribution function is some stepped surface, the steps of which correspond to the discrete values of the function $F(x,y)$. If the distribution function $F(x,y)$ is known, then it is possible to find the probability of a random point (X,Y) hitting the boundary of the rectangle $ABCD$ (Fig. 7.2), i.e. $P[(x_1 \le X < x_2)(y_1 \le Y < y_2)]$.

Since this probability is equal to the probability of hitting the infinite quadrant with a vertex $B(x_2, y_2)$ minus the probability of hitting the quadrants with vertices at the points $A(x_1, y_2)$ and $C(x_2, y_1)$, respectively, plus the probability of hitting the quadrant with the vertex $D(x_1, y_1)$ at the point (based on the fact that this probability was subtracted twice), then

$$P[(x_1 \le X < x_2)(y_1 \le Y < y_2)] =$$
$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \quad (7.8)$$

7.2. Probability density of a two-dimensional random variable

A two-dimensional random variable (X,Y) is called **continuous** if its distribution function $F(x,y)$ is continuous and differentiable concerning each of the arguments, and for which there is a second mixed derivative $F''_{xy}(x,y)$.

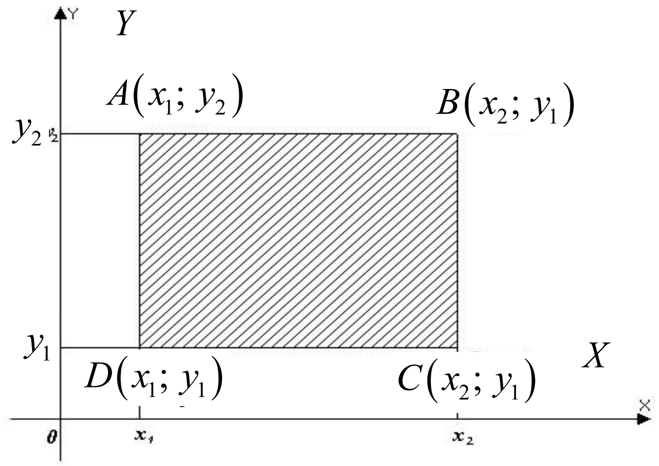


Fig. 7.2. Geometric interpretation of finding the probability of hitting a random point in the border of a rectangle

Let's find the probability of a random point (X,Y) falling into a rectangle with sides Δx and Δy, that is,

$$P = P[(x \le X < x + \Delta x)(y \le Y < y + \Delta y)]. \quad (7.9)$$

Let's mark $x_1 = x$, $x_2 = x + \Delta x$, $y_1 = y$, $y_2 = y + \Delta y$, we will get:

$$P = [F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y)] - [F(x + \Delta x, y) - F(x, y)].$$

The average probability density in a given rectangle is equal to the ratio of the probability P to the area of the rectangle $\Delta x \Delta y$, that is,

$$\varphi_{cp} = \frac{P[(x \le X < x + \Delta x)(y \le Y < y + \Delta y)]}{\Delta x \cdot \Delta y}. \qquad (7.10)$$

Let $\Delta x \to 0$, $\Delta y \to 0$. Then, taking into account (7.10), we find

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \varphi_{cp} = \lim_{\Delta y \to 0} \frac{1}{\Delta y}\left(\lim_{\Delta x \to 0} \frac{F(x+\Delta x, y+\Delta y) - F(x, y+\Delta y)}{\Delta x} - \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x,y)}{\Delta x} \right).$$

Since the function $F(x,y)$ is continuous and differentiated for each of the arguments, the last expression will take the form:

$$\lim_{\Delta x \to 0} \varphi_{cp} = \lim_{\Delta y \to 0} \frac{F_x'(x, y+\Delta y) - F_x'(x,y)}{\Delta y} = [F_x'(x,y)]_y' = F_{xy}''(x,y).$$

The second mixed derivative of its distribution function is called the **probability density** (distribution density) of a continuous two-dimensional random variable (X,Y):

$$\varphi(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = F_{xy}''(x,y). \qquad (7.11)$$

Geometrically, the probability density of a two-dimensional random variable (X,Y) is a distribution surface in space $Oxyz$ (Fig. 7.3).

The probability density $\varphi(x,y)$ has properties similar to the properties of the probability density of a one-dimensional random variable:

1. The probability density of a two-dimensional random variable is a non-negative function: $\varphi(x,y) \ge 0$.

This property follows from the definition of the probability density as the limit of the ratio of two non-negative quantities: $F(x,y)$ — a non-decreasing function for each of the arguments.

2. The probability of a continuous two-dimensional value (X,Y) entering the region D is equal to: $P[(X,Y)\in D]=\iint\limits_D \varphi(x,y)dxdy$.

Let's explain this formula geometrically. The element of probability, equal to $\varphi(x,y)dxdy$, means (with an accuracy of infinitely small higher orders) the probability of a random point (X,Y) hitting an elementary rectangle with sides dx and dy (Fig. 7.4).

This probability is approximately equal to the volume of an elementary parallelepiped with height $\varphi(x,y)$, resting on an elementary rectangle with sides dx and dy.

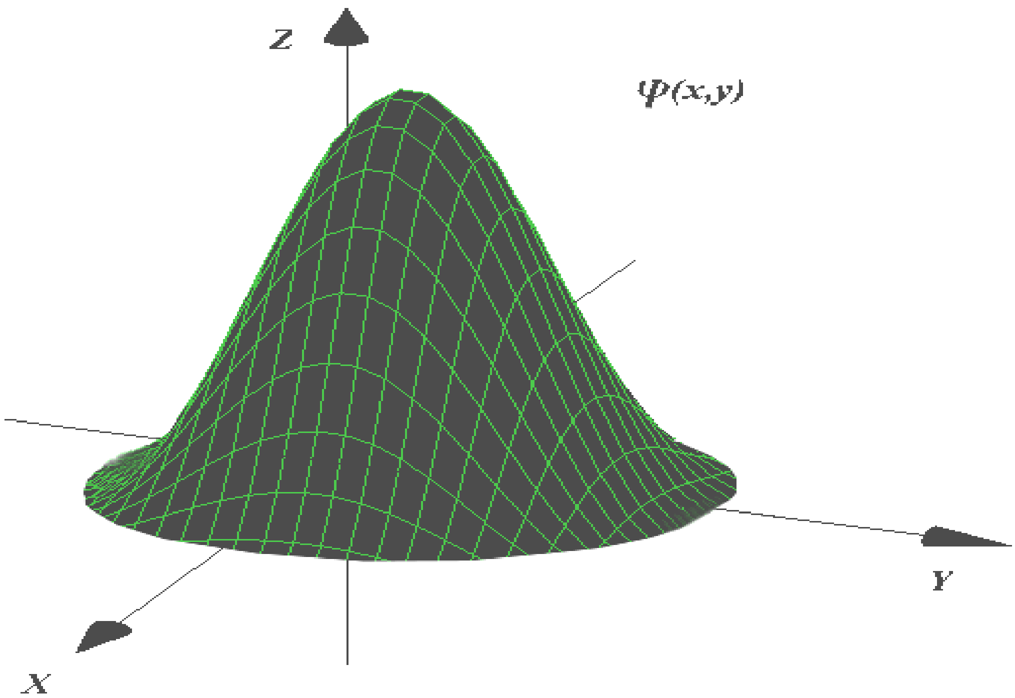


Fig. 7.3. Density distribution of a two-dimensional random variable

The probability of a two-dimensional random variable falling into a region D on a plane Oxy is represented geometrically by the volume of a cylindrical body, bounded from above by the surface of the distribution $\varphi(x, y)$ and resting on the region D, and analytically by the double integral .

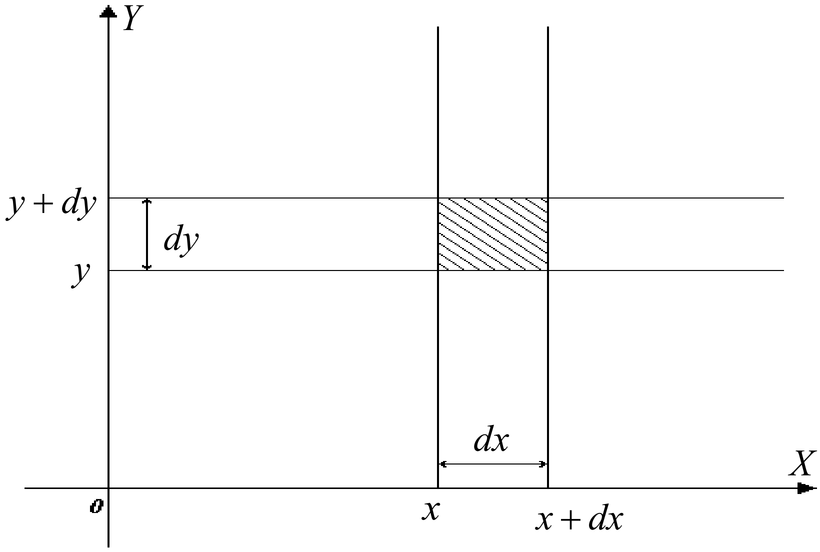


Fig. 7.4. Geometric illustration of an elementary rectangle

3. The distribution function of a continuous two-dimensional random variable can be expressed in terms of its probability density $\varphi(x, y)$ by the formula:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \varphi(x, y) dx dy . \qquad (7.12)$$

The distribution function $F(x, y)$ is the probability of a point falling into the infinite quadrant D, which can be considered as a rectangle bounded by abscissas $-\infty$ and x and ordinates $-\infty$ and y. Ago

$$F(x, y) = \iint_{D} \varphi(x, y) dx dy = \int_{-\infty}^{x} \int_{-\infty}^{y} \varphi(x, y) dx dy .$$

4. The double improper integral of the probability density of a two-dimensional random variable is equal to unity:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x,y)dxdy = 1. \qquad (7.13)$$

Improper integral $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x,y)dxdy = 1$ — is the probability of hitting the entire plane Oxy, that is, this event is reliable and its probability is equal to 1. This means that the total volume of the body bounded by the distribution surface and the plane Oxy is equal to 1.

7.2.1. Finding distribution functions and probability density of one-dimensional components of a two-dimensional random variable

It is known that $F(x,+\infty) = F_1(x)$ and $F(+\infty, y) = F_2(y)$. Denoting $y = +\infty$ and $x = +\infty$ in the corresponding formulas, we obtain the distribution functions of one-dimensional random variables X and Y:

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} \varphi(x,y)dxdy, \; F_2(y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{y} \varphi(x,y)dxdy. \qquad (7.14)$$

We differentiate the distribution functions by the arguments x and y, respectively, and get the probability densities of one-dimensional random variables X and Y:

$$\varphi_1(x) = \int_{-\infty}^{+\infty} \varphi(x,y)dy, \qquad \varphi_2(y) = \int_{-\infty}^{+\infty} \varphi(x,y)dx. \qquad (7.15)$$

The improper integral within infinite limits of the density $\varphi(x,y)$ of a two-dimensional random variable gives the probability density $\varphi_2(y)$ according to the argument x, and the probability density $\varphi_1(x)$ according to the argument y

◀ **Example 7.3.** A two-dimensional random variable is distributed uniformly over the area of a circle of radius $R = 1$ (Fig. 7.5). Determine:

a) the density and distribution function of a two-dimensional random variable (X, Y);

b) density probabilities and the distribution function for each of the one-dimensional components X and Y;

c) the probability that the distance from the point (X, Y) to the origin of coordinates will be less than $1/3$.

The solution. a) Provided: $\varphi(x, y) = \begin{cases} C, \text{ при } x^2+y^2 \le 1 \\ 0, \text{ при } x^2+y^2 > 1 \end{cases}.$

The constant can be found from the relation:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x, y)dxdy = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} Cdxdy = 1.$$

It is easier to do this based on the geometric content of the ratio, which determines that the volume of the body bounded by the distribution surface $\varphi(x, y)$ and the plane Oxy is equal to 1. In this case, it is the volume of a cylinder with the base area $\pi R^2 = \pi \cdot 1^2 = \pi$ and height C (Fig. 7.6), which is equal to $\pi \cdot C = 1$ from here $C = 1/\pi$. Therefore,

$$\varphi(x, y) = \begin{cases} 1/\pi, & \text{при } x^2 + y^2 \le 1 \\ 0, & \text{при } x^2 + y^2 > 1 \end{cases}.$$

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \varphi(x, y)dxdy = \frac{1}{\pi}\int_{-\infty}^{x} dx \int_{-\infty}^{y} dy$$

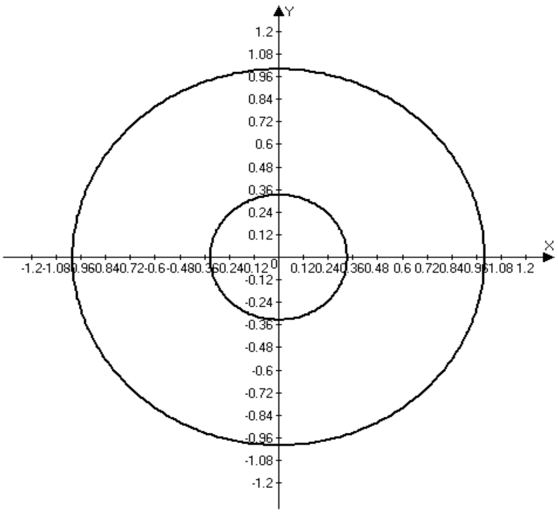


Fig. 7.5. Illustration for an example 7.3.

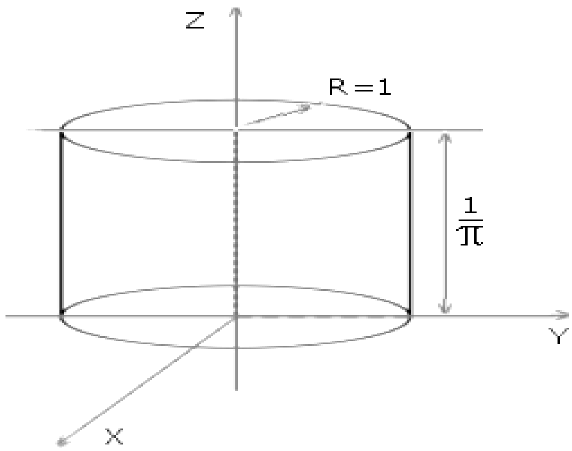


Fig. 7.6. Illustration for an example 7.3.

This integral coincides with the area of the region D with precision up to a factor $1/\pi$ — the area of intersection of the circle $x^2+y^2\le 1$ with the infinite quadrant to the left and below the point $M(x,y)$ (Fig. 7.7). When $x\le -1$, $-\infty<y<+\infty$ or, when $-\infty<x<+\infty$, $y\le -1$ $F(x,y)=0$, because in this case the region D is empty, and when $x>1$, $y>1$ $F(x,y)=1$, because under the given conditions, the region D completely coincides with the circle $x^2+y^2\le 1$, on which the joint density of the distribution $\varphi(x,y)$ is different from zero.

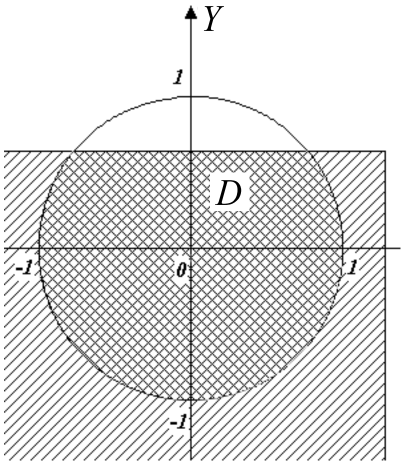
b) Find the distribution functions of one-dimensional components X and Y:

at $-1<x\le 1$

$$F_1(x)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}\varphi(x,y)dxdy=\int_{-1}^{x}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dxdy=$$

$$\frac{1}{\pi}\int_{-1}^{x}\left(y\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\right)dx=\frac{1}{\pi}\int_{-1}^{x}2\sqrt{1-x^2}\,dx=$$

$$\frac{1}{\pi}\left[(x\sqrt{1-x^2})+\arcsin x\,\Big|_{-1}^{x}\right]=\frac{1}{2}+\frac{1}{\pi}(x\sqrt{1-x^2}+\arcsin x)$$



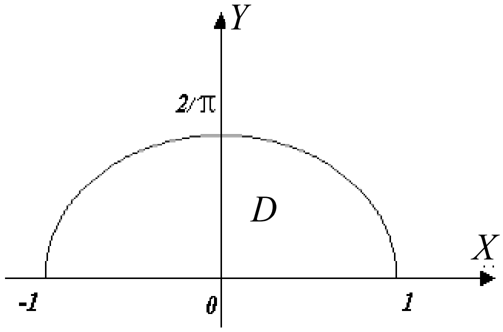


Fig. 7.7. Illustration for an example 7.3. Fig. 7.8. Graph density distribution $\varphi_1(x)$.

Therefore,

$$F_1(x)=\begin{cases} 0, \ at \ x\le 1 \\ \dfrac{1}{2}+\dfrac{1}{\pi}(x\sqrt{1-x^2}+\arcsin x), \ at \ -1<x\le 1. \\ 1, \ at \ x>1 \end{cases}$$

Similarly,

$$F_1(x)=\begin{cases} 0, \ at \ y\le 1 \\ \dfrac{1}{2}+\dfrac{1}{\pi}(y\sqrt{1-y^2}+\arcsin y), \ at \ -1<y\le 1. \\ 1, \ at \ y>1 \end{cases}$$

Let's find the probability densities of one-dimensional components X and Y:

$$\varphi_1(x)=\int_{-\infty}^{+\infty}\varphi(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-x^2}\,(-1\le x\le 1).$$

The density graph $\varphi_1(x)$ is shown in Fig. 7.8.

Similarly, $\varphi_2(y)=\int\limits_{-\infty}^{+\infty}\varphi(x,y)dx=\frac{2}{\pi}\sqrt{1-y^2}\,(-1\le y\le 1)$.

c) The sought probability: $P\left(\sqrt{X^2+Y^2}<\frac{1}{3}\right)=P\left(X^2+Y^2<\frac{1}{9}\right)$,

that is, the probability that a random point (X,Y) will be located on the plane of a circle with a radius $R_1=1/3$ (Fig. 7.5) could be found using the formula:

$$P\left(X^2+Y^2<\frac{1}{9}\right)=\int\limits_{-\frac{1}{3}}^{\frac{1}{3}}\int\limits_{-\sqrt{\frac{1}{9}-x^2}}^{\sqrt{\frac{1}{9}-x^2}}\frac{1}{\pi}dxdy.$$

However, it is easier to do this using the concept of "geometric probability", that is, $P\left(X^2+Y^2<\frac{1}{9}\right)=(\pi R_1^2)/(\pi R^2)=R_1^2/R^2=\left(\frac{1}{3}\right)^2/1^2=\frac{1}{9}$.►

7.3. Conditional laws of distribution. Numerical characteristics of a two-dimensional random variable. Regression.

The conditional distribution law of one of the one-dimensional components of a two-dimensional random variable (X,Y) is its distribution law, calculated under the condition that the second component assumed a certain value (or fell into a certain interval).

For discrete values:

$$p_j(x_i)=\frac{P\left[(X=x_i)(Y=y_i)\right]}{P(Y=y_i)};\ p_j(y_i)=\frac{P\left[(X=x_i)(Y=y_i)\right]}{P(X=x_i)}.$$

In the case of continuous random variables, it is necessary to find probability densities of conditional distributions $\varphi_y(x)$ and $\varphi_x(y)$. To do this, in the above formulas, we replace the probabilities of events with the corresponding "elements of probabilities", i.e. $P\left[(X = x_i)(Y = y_i)\right]$ on $\varphi(x,y)dxdy$; $P(X = x_i)$ on $\varphi(x)dx$; $P(Y = y_i)$ on $\varphi(y)dy$; $p_j(x_i)$ on $\varphi_y(x)dx$; $p_j(y_i)$ on $\varphi_x(y)dy$, and, after reductions, we get:

$$\varphi_y(x) = \frac{\varphi(x,y)}{\varphi_2(y)}; \quad \varphi_x(y) = \frac{\varphi(x,y)}{\varphi_1(x)}.$$

The conditional probability density of one of the components of a two-dimensional random variable is equal to the ratio of the joint probability density to the probability density of the other component.

The ratio $\varphi(x,y) = \varphi_1(x)\varphi_x(y) = \varphi_2(y)\varphi_y(x)$ is called the **multiplication theorem of probability densities**.

Conditional probability densities can be expressed in terms of the joint density:

$$\varphi_y(x) = \frac{\varphi(x,y)}{\int_{-\infty}^{\infty} \varphi(x,y)dy}; \quad \varphi_x(y) = \frac{\varphi(x,y)}{\int_{-\infty}^{\infty} \varphi(x,y)dx}.$$

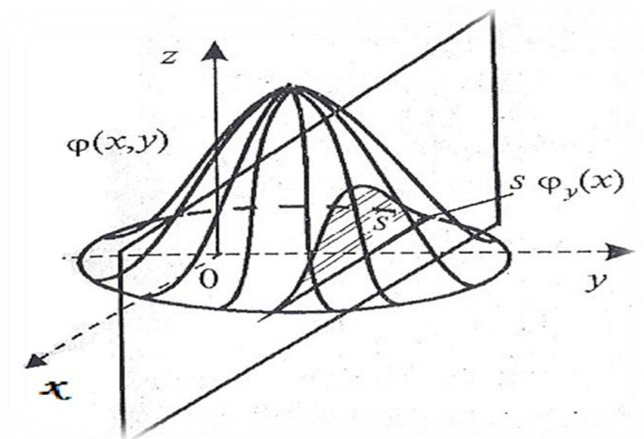


Fig. 7.9. Graphical interpretation of conditional probability density.

If the joint density $\varphi(x,y)$ of a two-dimensional random variable is some geometric distribution surface, then, for example, the conditional density $\varphi_y(x)$ is a distribution curve, which is similar to the intersection of this surface with a plane $Y = y$, parallel to the coordinate plane XOZ and such that cuts off the segment y on the axis Oy (Fig. 7.9), and is obtained by by dividing all ordinates by the area of the given section s (hence, the intersection of the distribution surface is the curve s, where $0 \le s \le 1$).
Similarly, the geometric content of conditional density can be explained $\varphi_x(y)$.
For a continuous random variable (X,Y), mathematical expectations and dispersions are calculated according to the formulas:

$$a_x = M(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\varphi(x,y)dxdy;$$

$$a_y = M(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\varphi(x,y)dxdy;$$

$$D(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-a_x)^2 \varphi(x,y)dxdy$$

$$D(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-a_y)^2 \varphi(x,y)dxdy$$

Conditional mathematical expectations $M_y(X)$ and $M_x(Y)$, as well as conditional dispersions $D_y(X)$ and $D_x(Y)$ are based on the usual formulas for mathematical expectation and dispersion, in which instead of event probabilities or probability densities, conditional probabilities p_i, p_j or conditional probability densities $\varphi_1(x), \varphi_2(y)$ are used conditional probabilities $p_j(x_i)$

or conditional probability densities $\varphi_x(y), \varphi_y(x)$.

For example,

$$M_x(Y)=\int_{-\infty}^{\infty} y\varphi_x(y)dy;\ D_x(Y)=\int_{-\infty}^{\infty}(y-M_x)^2\varphi_x(y)dy.$$

The conditional mathematical expectation of a random variable Y at $X=x$ $(M_x(Y))$ is a function of x and is called a **regression function** or simply a **regression of *Y* on *X***; similarly, $M_y(X)$ is called the **regression function** or simply the **regression of *X* on *Y***. Graphs of these functions are called regression curves or **regression lines**.

The properties of conditional mathematical expectations and dispersions are analogous to the properties of "unconditional" mathematical expectations and dispersions.

Among the additional properties of conditional mathematical expectation, consider the following:

1) If *Z=g(X),* where *g* is some random function of *X,* then:

$M_z(M_x(Y))=M_z(Y)$, for example, $M(M_x(Y))=M(Y)$

(rule of repeated hope).

2) If *Z=g(X),* where *g* is some random function of *X,* then

$M_x(ZY)=ZM_x(Y)$.

3) If random variables X and Y are independent, then $M_x(Y)=M(Y)$.

Control questions

1) What is called a distribution function of a multidimensional random variable?
2) Properties of the distribution function of a multidimensional random variable.
3) The probability density of a multidimensional random variable and its relation to the distribution function.
4) Is it possible to determine the one-dimensional probability density component of a multidimensional random variable? If possible, then in what way?
5) What is a multidimensional random variable regression?

Chapter 8

Dependent and independent random variables. Codispersion and correlation coefficient. Two-dimensional normal distribution law

8.1. Dependent and independent random variables

Random variables X and Y are called **independent** if their aggregate distribution function $F(x,y)$ can be represented as a product of distribution functions $F_1(x)$ and $F_2(y)$ these random variables, i.e., $F(x,y)=F_1(x)\cdot F_2(y)$. Otherwise, random variables X and Y are called **dependent.**

For independent continuous random variables X and Y:

$$\varphi(x,y)=\varphi_1(x)\cdot\varphi_2(y),$$

that is, the conditional probability densities coincide with the corresponding unconditional densities: $\varphi_y(x)=\varphi_1(x)$ and $\varphi_x(y)=\varphi_2(y)$.

In general cases, if the aforementioned conditions are not fulfilled, a different type of dependence is used: the dependence between two random variables will be **probabilistic** (**stochastic** or statistical) if each value of one of them corresponds to a certain (conditional) distribution of the other.

In this case, when the value of one of the random variables is known, it is impossible to unambiguously determine the value of the other, and it is only possible to indicate its distribution. For example, the relationship between the number of equipment repairs and the costs of preventive measures; human weight and height; quantity applied fertilizers and yield is a probabilistic dependence.

Figure 8.1 shows examples of dependent and independent random variables X and Y.

In fig. 8.1 a) the dependence between X and Y is manifested in the fact that with a change in *x,* the distribution of Y and the mathematical expectation change $M_x(Y)$ (*x* increases and $M_x(Y)$ increases). This figure also shows the regression line Y on X.

In fig. 8.1 b) the dependence of random variables is manifested in changes in conditional dispersions (increases x with increase $D_x(Y)$), while $M_x(Y) = const$, that is, the regression line Y on X is parallel to the axis Ox

In fig. 8.1 c) random variables X and Y independent, and therefore, conditional mathematical expectation $M_x(Y)$ and conditional dispersion $D_x(Y)$

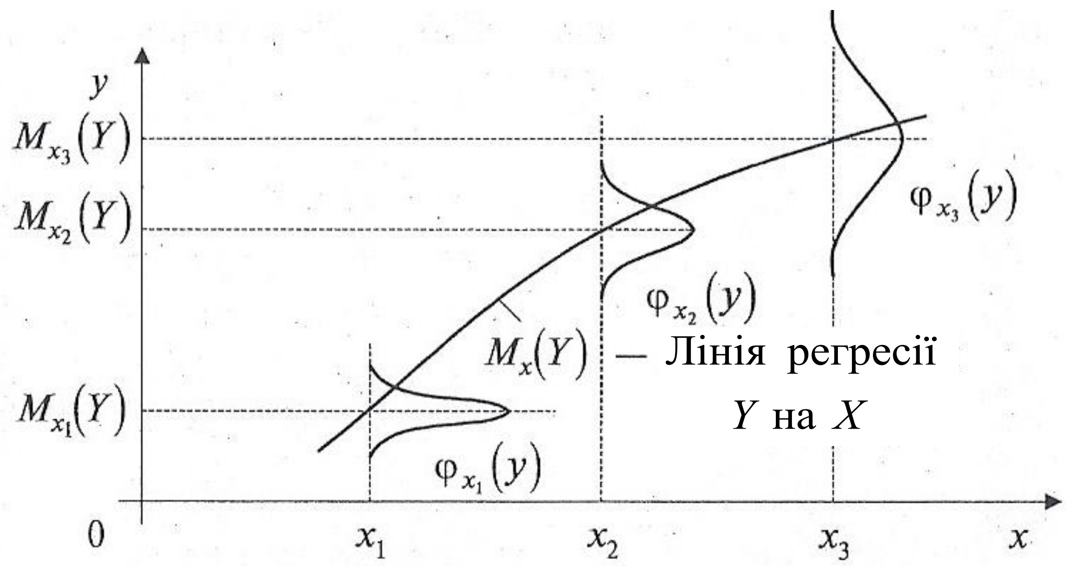


Fig. 8.1. a) Dependence of random variables *X* and *Y*.

do not change. Thus, it can be stated that if the random variables X are Y independent, then the regression lines are parallel to the coordinate axes.

◄ **Example 8.1.** A two-dimensional random variable is distributed uniformly on the plane of a circle with a radius $R=1$ (Fig. 7.5, Chapter 7). Determine:

a) conditional densities of random variables X and Y;

b) if random variables X and Y are dependent;

c) conditional mathematical expectations and dispersions.

The solution. a) Let's find the conditional density $\varphi_y(x)$ ($\varphi_2(y) \neq 0$):

$$\varphi_y(x) = \frac{\varphi(x,y)}{\varphi_2(y)} = \begin{cases} \dfrac{1}{2\sqrt{1-y^2}} & at\ |x| < \sqrt{1-y^2} \\ 0 & at\ |x| > \sqrt{1-y^2} \end{cases}.$$

The graph $\varphi_y(x)$ is shown in Fig. 8.2.

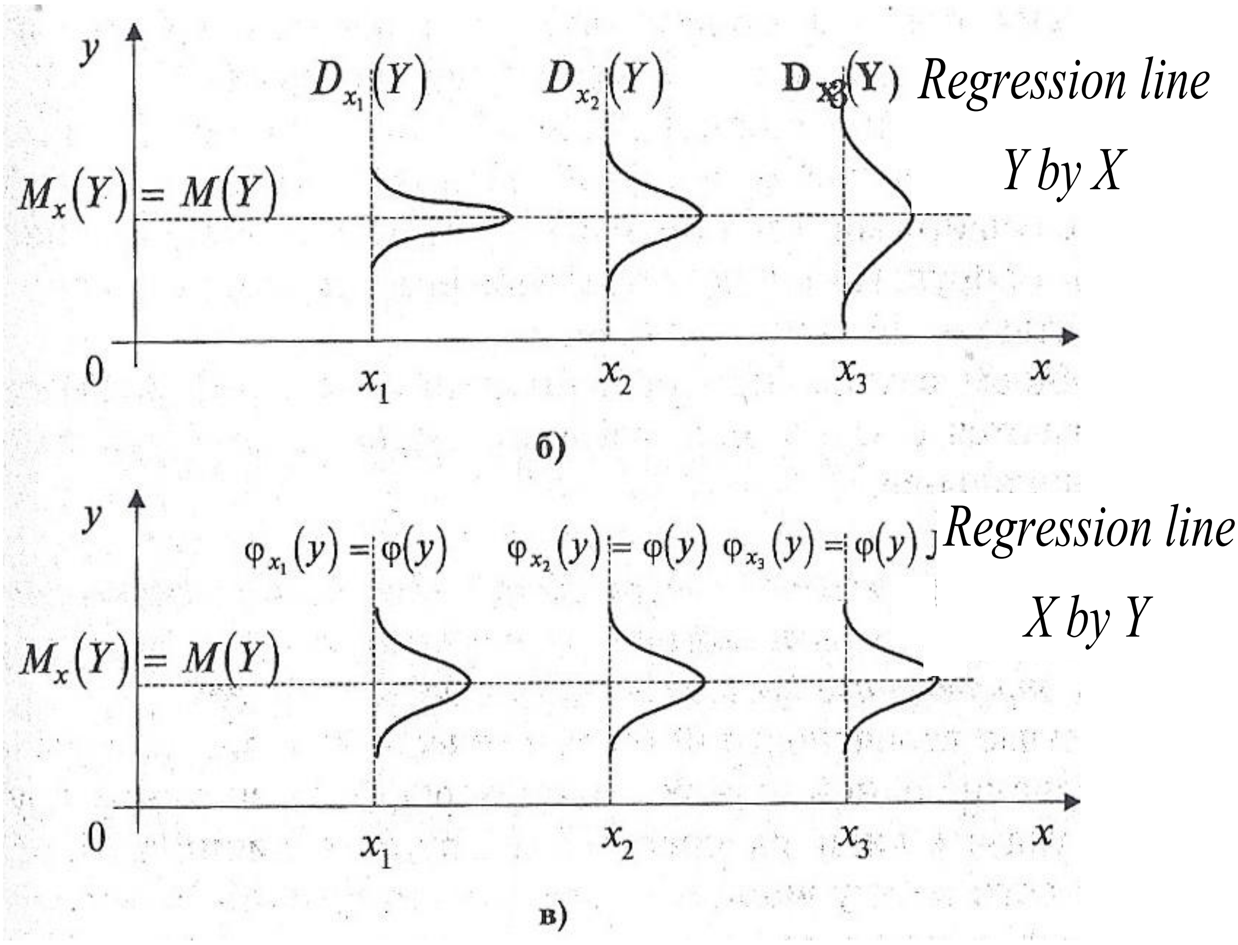


Fig. 8.1. *b)* Regression line Y by X; c) Regression lines of independent random variables X and Y.

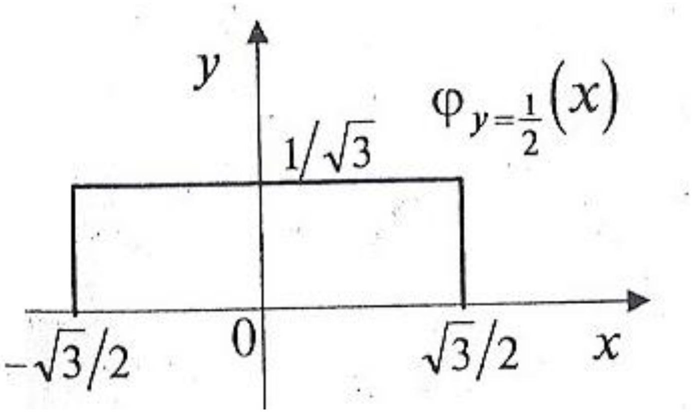


Fig. 8.2. Conditional distribution density for example 8.1.

Similarly,

$$\varphi_x(y)=\frac{\varphi(x,y)}{\varphi_1(x)}=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}} & at\ |y|<\sqrt{1-x^2}\\ 0 & at\ |y|>\sqrt{1-x^2}\end{cases}.$$

b) X and Y are dependent random variables, because

$$\varphi(x,y)\neq\varphi_1(x)\cdot\varphi_2(y).$$

c) Let's find the conditional mathematical expectation $M_x(Y)$, taking into account that $\varphi_x(y)=\dfrac{1}{2\sqrt{1-x^2}}$:

$$M_x(Y)=\int_{-\infty}^{\infty}y\varphi_x(y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}y\frac{1}{2\sqrt{1-x^2}}dy=\frac{1}{2\sqrt{1-x^2}}\cdot\frac{y^2}{2}\Bigg|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}=0.$$

Similarly, $M_y(X)=\int_{-\infty}^{\infty}x\varphi_y(x)dx=0$.

Thus, the regression line Y on X coincides with the *Ox* axis, and the regression line X on Y — with the *Oy* axis.

Let's find the conditional dispersion $D_x(y)$:

$$D_x(Y)=\int_{-\infty}^{\infty}\left(y-M_x(Y)\right)^2\varphi_x(y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}(y)^2\frac{1}{2\sqrt{1-x^2}}dy=$$

$$=\frac{1}{2\sqrt{1-x^2}}\cdot\frac{y^3}{3}\Bigg|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}=\frac{\left(1-x^2\right)}{3}\quad(0\le x\le 1)$$

Similarly, $D_y(x)=\dfrac{\left(1-y^2\right)}{3}\ (0\le y\le 1)$.

Thus, with distance from the origin, the dispersion of conditional distributions decreases from 1/3 to 0. ►

8.2. Codispersion and correlation coefficient

Mathematical expectations and dispersions of random variables do not fully characterize a two-dimensional random variable since they do not express the degree of dependence of its components. Dispersion and correlation coefficient play this role·

The codispersion (or correlation moment) K_{xy} of random variables X and Y is the mathematical expectation of the product of the deviations of these values from their mathematical expectations:

The codispersion (or correlation moment) K_{xy} of random variables is the mathematical expectation of the product of the deviations of these values from their mathematical expectations: $K_{xy} = M\Big(\big(X - M(X)\big)\big(Y - M(Y)\big)\Big)$,

or $K_{xy} = M\Big(\big(X - a_x\big)\big(Y - a_y\big)\Big)$, where $a_x = M(X)$; $a_y = M(Y)$.

It follows from the definition that $K_{xy} = K_{yx}$ i $K_{xx} = M\big(X - a_x\big)^2 = D(X)$, that is, the codispersion of a random variable with itself is its dispersion.

For discrete random variables: $K_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}\big(x_i - a_x\big)\big(y_j - a_y\big)p_{ij}$.

For continuous random variables:

$$K_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(x - a_x\big)\big(y - a_y\big)\varphi(x,y)\,dxdy.$$

The codispersion of two random variables characterizes both the degree of dependence of random variables and their dispersion around a point $\big(a_x, a_y\big)$.

Properties of codispersion of random variables

1. The codispersion of two independent random variables is zero.

For independent random variables $\varphi(x,y)=\varphi_1(x)\cdot\varphi_2(y)$. Therefore, the codispersion formula will have the form:

$$K_{xy}=\int_{-\infty}^{\infty}(x-a_x)\varphi_1(x)dx\int_{-\infty}^{\infty}(y-a_y)\varphi_2(y)dy=0,$$

since each of the obtained integrals is a central moment of the first order, which is equal to zero.

2. The codispersion of two random variables is equal to the mathematical expectation of their product minus the product of mathematical expectations:

$$K_{xy}=M(XY)-M(X)\cdot M(Y), \text{ or } K_{xy}=M(XY)-a_x\cdot a_y.$$

By definition:

$$K_{xy}=M\left((X-a_x)(Y-a_y)\right)=M\left(XY-a_yX-a_xY+a_xa_y\right).$$

Because, $a_x=M(X)$; $a_y=M(Y)$, we will get:

$$K_{xy}=M(XY)-a_ya_x-a_xa_y+a_xa_y=M(XY)-a_xa_y.$$

3. The absolute value of the codispersion of two random variables does not exceed the product of their mean square deviations:

$$|K_{xy}|\le\sigma_x\cdot\sigma_y$$

Consider the obvious inequality: $M\left(\frac{X-a_x}{\sigma_x}\pm\frac{Y-a_y}{\sigma_y}\right)^2\ge 0,$

or $$M\left(\left(\frac{X-a_x}{\sigma_x}\right)^2\pm 2\left(\frac{X-a_x}{\sigma_x}\cdot\frac{Y-a_y}{\sigma_y}\right)+\left(\frac{Y-a_y}{\sigma_y}\right)^2\right)=$$

$$=\frac{1}{\sigma_x^2}M\left(X-a_x\right)^2 \pm \frac{2}{\sigma_x\sigma_y}M\Big(\left(X-a_x\right)\left(Y-a_y\right)\Big)+\frac{1}{\sigma_y^2}M\left(Y-a_y\right)^2=$$

$$=\frac{D(X)}{\sigma_x^2}\pm\frac{2K_{xy}}{\sigma_x\sigma_y}+\frac{D(Y)}{\sigma_y^2}=2\pm\frac{2K_{xy}}{\sigma_x\sigma_y}\geq 0.$$

(Note that σ_x and σ_y are non-random values and their dispersions

$D(X)=M\left(X-a_x\right)^2=\sigma_x^2$, $D(Y)=M\left(Y-a_y\right)^2=\sigma_y^2$)

So, the proof follows from the inequality.

Codispersion is a quantity that has a dimension determined by the product of the dimensions of random variables. This significantly limits the use of codispersion in the study of various random variables. The **correlation coefficient** is devoid of this drawback.

The correlation coefficient of two random variables is the ratio of their codispersion to the product of the mean square deviations of these variables:

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

It follows from the definition that $\rho_{xy}=\rho_{yx}=\rho$. The correlation coefficient is a dimensionless quantity.

Properties of the correlation coefficient

1. The correlation coefficient takes its value on the segment [-1;1]:

$$-1\leq\rho\leq 1.$$

It follows from the inequality $2\pm\frac{2K_{xy}}{\sigma_x\sigma_y}\geq 0$ that

$$2 \pm 2\rho \geq 0 \Rightarrow -1 \leq \rho \leq 1.$$

2. If the random variables are independent, their correlation coefficient is zero. Since for independent random variables $K_{xy} = 0$, then in this case and $\rho = 0$. Random variables are called **uncorrelated** if their correlation coefficient is zero.
3. If the correlation coefficient of two random variables is equal (in absolute value) to unity, then there is a **linear functional relationship** between these random variables.

   Let's use the ratio $M\left(\dfrac{X - a_x}{\sigma_x} \pm \dfrac{Y - a_y}{\sigma_y}\right)^2 = 2 \pm \dfrac{2K_{xy}}{\sigma_x \sigma_y} = 2 \pm 2\rho$.

   If $\rho = \mp 1$, then $2 \pm 2\rho = 0$ i $M\left(\dfrac{X - a_x}{\sigma_x} \pm \dfrac{Y - a_y}{\sigma_y}\right)^2 = 0$.

   The equality of the mathematical expectation of a non-negative random variable of zero means that the random variable itself is identically equal to zero, therefore:

   $\dfrac{X - a_x}{\sigma_x} \pm \dfrac{Y - a_y}{\sigma_y} = 0$, when $\rho = \mp 1$, or $Y = a_y + \dfrac{\sigma_y}{\sigma_x}(X - a_x)$ when $\rho = 1$

   and

   $Y = a_y - \dfrac{\sigma_y}{\sigma_x}(X - a_x)$ when $\rho = -1$, that is, there is a linear functional relationship between the values X and Y.

◀ **Example 8.2.** The distribution law of a discrete two-dimensional random variable (X, Y) is presented in the form of a table (example 7.2, section 7):

x_i \ y_j	-1	0	1	2
1	0,10	0,25	0,3	0,15
2	0,10	0,05	0,00	0,05

Find the codispersion and correlation coefficient of random variables X and Y.

The solution. According to previous results (see example 7.2), the distribution laws of one-dimensional random variables X and Y have the following form:

X:

x_i	1	2
p_i	0,8	0,2

Y:

y_j	-1	0	1	2
p_j	0,2	0,3	0,3	0,2

Let's find the mathematical expectation and the mean square deviation of these random variables:

$$a_x = M(X) = \sum_{i=1}^{2} x_i p_i = 1 \cdot 0{,}8 + 2 \cdot 0{,}2 = 1{,}2;$$

$$M(X^2) = \sum_{i=1}^{2} x_i^2 p_i = 1^2 \cdot 0{,}8 + 2^2 \cdot 0{,}2 = 1{,}6;$$

$$D(X) = M(X^2) - a_x^2 = 1{,}6 - 1{,}2^2 = 0{,}16; \; \sigma_x = \sqrt{D(X)} = 0{,}4.$$

$$a_y = M(Y) = \sum_{j=1}^{4} y_j p_j = (-1) \cdot 0{,}2 + 0 \cdot 0{,}3 + 1 \cdot 0{,}3 + 2 \cdot 0{,}2 = 0{,}5;$$

$$M(Y^2) = \sum_{j=1}^{4} y_j^2 p_j = (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}3 + 1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}2 = 1{,}3;$$

$$D(Y) = M(Y^2) - a_y^2 = 1{,}3 - 0{,}5^2 = 1{,}05; \sigma_y = \sqrt{D(Y)} = 1{,}025$$

The mathematical expectation $M(XY)$ can be found from the distribution table of a two-dimensional random variable using the formula:

$$M(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}=\sum_{i=1}^{2}\sum_{j=1}^{4}x_i y_j p_{ij}=$$
$$=1\cdot(-1)\cdot 0{,}1+0\cdot(-1)\cdot 0{,}25+1\cdot 1\cdot 0{,}3+1\cdot 2\cdot 0{,}15+2\cdot(-1)\cdot 0{,}1+2\cdot 0\cdot 0{,}05+$$
$$+2\cdot 1\cdot 0+2\cdot 2\cdot 0{,}05=0{,}5$$

Let's calculate the codispersion:

$$K_{xy}=M(XY)-a_x a_y=0{,}5-1{,}2\cdot 0{,}5=-0{,}1.$$

Correlation coefficient: $\rho=\dfrac{K_{xy}}{\sigma_x\sigma_y}=-0{,}244$.

The obtained results indicate that there is a negative linear relationship between the random variables X and Y, that is, they are inversely proportional. ►

With the help of correlation, some properties of mathematical expectation and dispersion can be supplemented and clarified.

1. The mathematical expectation of the product of two random variables is equal to the sum of the product of their mathematical expectations and the codispersion of these random variables

values: $$M(XY)=M(X)M(Y)+K_{xy}.$$

Proof. The given ratio follows from the formula

$$K_{xy}=M(XY)-M(X)\cdot M(Y).$$

If $K_{xy}=0$, then $M(XY)=M(X)M(Y)$.●

2. The dispersion of the sum of two random variables is equal to the sum of their dispersions plus their doubled codispersion:

$$D(X+Y)=D(X)+D(Y)+2K_{xy}.$$

Proof. Let $Z=X+Y$. According to the property of mathematical expectation $a_z=a_x+a_y$. Ago $Z-a_z=(X-a_x)+(Y-a_y)$.

According to the definition of dispersion:

$$D(X+Y)=D(Z)=M(Z-a_z)^2=M(X-a_x)^2+$$
$$+2M((X-a_x)(Y-a_y))+M(Y-a_y)^2=D(X)+D(Y)+2K_{xy}\bullet$$

8.3. Two-dimensional normal distribution law

A random variable $(X;Y)$ is said to be distributed according to **the two-dimensional normal law** if its joint probability density has the form:

$$\varphi_N(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-L(x,y)},$$

where

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

The two-dimensional normal distribution law is determined by five parameters: $a_x, a_y, \sigma_x^2, \sigma_y^2, \rho$.

Let's find out the theoretical and probabilistic meaning of these parameters: $M(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\varphi_N(x,y)dx\,dy=a_x$; similarly we calculate and $M(Y)=a_y$. $D(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a_x)^2\varphi_N(x,y)dx\,dy=\sigma_x^2$; similarly we calculate and $D(Y)=\sigma_y^2$.

$$K_{xy}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a_x)(y-a_y)\varphi_N(x,y)dx\,dy=\rho\sigma_x\sigma_y .$$

Thus, the parameters a_x and a_y express the mathematical expectation of the random variables X and Y, the parameters σ_x^2 and σ_y^2 are their dispersions, and $\frac{K_{xy}}{\sigma_x \sigma_y} = \rho$ the correlation coefficient between the random variables X and Y.

Let's find the probability densities of one-dimensional random variables X and Y:

$$\varphi_1(x) = \int_{-\infty}^{\infty} \varphi_N(x, y) dy = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-a_x)^2}{2\sigma_x^2}}$$

and similarly, $\varphi_2(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-\frac{(y-a_y)^2}{2\sigma_y^2}}$.

Each of the distribution laws of one-dimensional random variables X and Y is normal with parameters, respectively (a_x, σ_x^2) i (a_y, σ_y^2).

Let's find conditional probability densities:

$$\varphi_y(x) = \frac{\varphi_N(x, y)}{\varphi_2(y)} = \frac{1}{\sigma_x \sqrt{1-\rho^2} \sqrt{2\pi}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{x-a_x}{\sigma_x} - \rho \frac{y-a_y}{\sigma_y} \right)^2}.$$

and similarly, $\varphi_x(y) = \frac{1}{\sigma_y \sqrt{1-\rho^2} \sqrt{2\pi}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{y-a_y}{\sigma_y} - \rho \frac{x-a_x}{\sigma_x} \right)^2}$.

Each of the conditional laws of the distribution of random variables X and Y is normal with conditional mathematical expectation and conditional dispersion, which are determined by the formulas:

$$M_y(X) = a_x + \rho \frac{\sigma_x}{\sigma_y}(y - a_y), \qquad D_y(X) = \sigma_x^2 (1 - \rho^2),$$

$$M_x(Y) = a_y + \rho\frac{\sigma_y}{\sigma_x}(x - a_x), \qquad D_x(Y) = \sigma_y^2(1-\rho^2).$$

From these formulas it follows that regression lines $M_y(X)$ and $M_x(Y)$ normally distributed random variables are straight lines.

Conditional dispersions $D_y(X)$ and $D_x(Y)$, and therefore and $\sigma_y(X)$ and $\sigma_x(Y)$ are constant and do not depend on the values of y or x. This property is called homoscedasticity or uniformity of conditional normal distributions and is essential for statistical analysis.

Theorem 8.1. If two normally distributed random variables X and Y are uncorrelated, then they are independent.

Proof. By condition, the correlation coefficient is $\rho = 0$. Let's find the expression of the compatible density:

$$\varphi_N(x,y) = \frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-a_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{(y-a_y)^2}{2\sigma_y^2}} = \varphi_1(x)\varphi_2(y),$$

where $\varphi_1(x)$ and $\varphi_2(y)$ — probability density of one-dimensional random variables X and Y. This means independence of random variables X and Y. ●

The surface $\varphi_N(x,y)$ of the normal distribution is a hill-like surface called the "Gaussian tent" (Fig. 8.3).

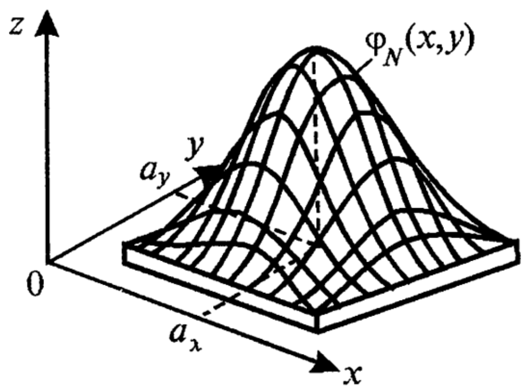


Fig. 8.3. $N(a_x, a_y, \sigma_x^2, \sigma_y^2, \rho)$ "Gaussian tent"

Cross-sections of this surface by planes $x = a$ perpendicular to the *Ox* axis and $y = b$, perpendicular to the *Oy* axis, have the form of normal curves with centers that lie on the regression line Y at X and X at Y, and with mean square deviations equal to $\sigma_x\sqrt{1-\rho^2}$ and $\sigma_y\sqrt{1-\rho^2}$.

The section of the surface of the normal distribution by a plane $z = c$ (where $0 < c < \varphi_N\left(a_x, a_y\right)$), parallel to the *Oxy* plane, is an ellipse, which is called a **scattering ellipse** (Fig. 8.4):

$$\frac{(x-a_x)^2}{\sigma_x^2} - 2\rho\frac{(x-a_x)(y-a_y)}{\sigma_x\sigma_y} + \frac{(y-a_y)^2}{\sigma_y^2} = a^2,$$

where $a^2 = -2(1-\rho^2)ln(2\pi c\sigma_x\sigma_y\sqrt{1-\rho^2})$.

(Due to the limitation for c, the argument of the logarithm is less than 1, and the value of the logarithm itself is negative.)

The center of the ellipse is at the point $\left(a_x, a_y\right)$, and its axes make angles with the axis *Ox* α and $\pi/2 + \alpha$, where α is determined from the condition

$$tg2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}.$$

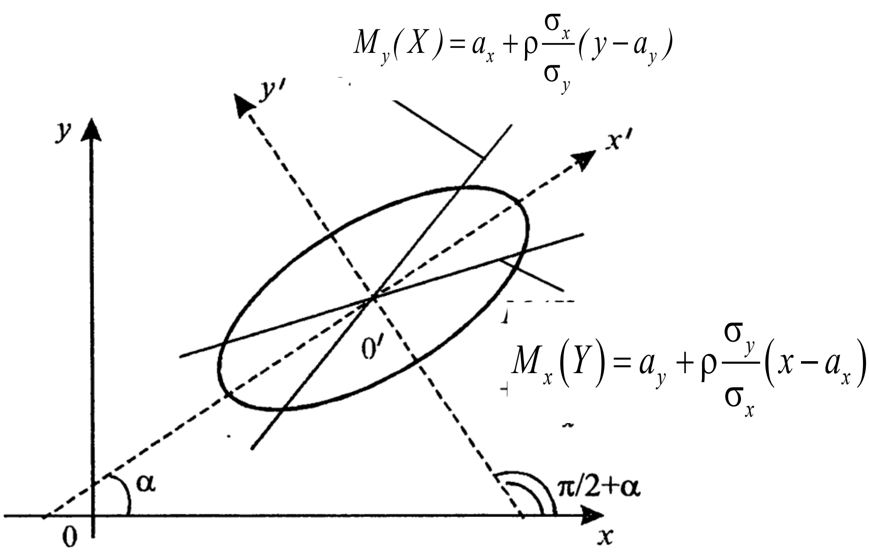


Fig. 8.4. Scattering ellipse

8.4. Functions of random variables. Composition of distribution laws

One of the important tasks of probability theory is to determine the distribution law of the function of one or more random variables if the distributions of one or more arguments are known. The construction of laws of distribution of functions of some discrete random variables was considered earlier. Let us turn to continuous random variables.

Let is a continuous random variable X with probability density $\varphi(x)$, and the random variable Y is a function of X, i.e. $Y = f(X)$. It is necessary to find the distribution law of a random variable Y. Let the function $f(X)$ be strictly monotonic, continuous, and differentiable on the segment $[a,b]$ and $f(a) = c$, $f(b) = d$ (Fig. 8.5).

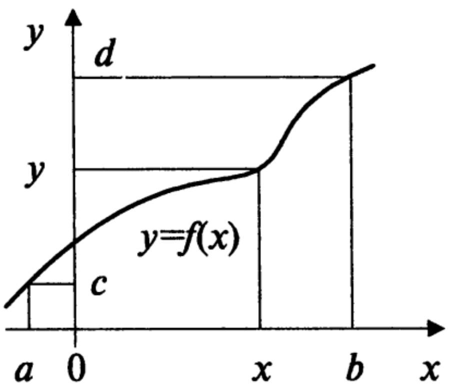


Fig. 8.5. Function $f(X)$ — is strictly monotonic, continuous and differentiable on the interval $[a,b]$.

Let $f'(x)>0$, then the distribution function $G(Y)$ of a random variable $Y=f(X)$:

$$G(y)=P(Y<y)=\begin{cases}0, \text{ when } y<c\\ \int_c^y g(y)dy, \text{ when } c<y<d\end{cases}$$

where $g(y)$ — probability density of a random variable $Y=f(X)$.

If $c\le y\le d$, $G(y)=P(Y<y)=P(f(X)<y)$.

Inequality $f(X)<y$ is equivalent to inequality $X<f^{-1}(y)$, where $f^{-1}(y)$ — a function inverse of a function $f(x)$ on a segment $[a,b]$. Ago

$$G(y)=P\left[X<f^{-1}(y)\right]=\int_d^{f^{-1}(y)}\varphi(x)dx.$$

According to the theorem on the derivative of the integral over a variable upper bound

$$g(y)=G'(y)=\varphi\left(f^{-1}(y)\right)\left|\left(f^{-1}(y)\right)'\right|.$$

(We take the derivative $\left(f^{-1}(x)\right)'$ as an absolute value, because if the function $f(x)$ is decreasing on the segment $[a,b]$, then its inverse function $f^{-1}(y)$ is also decreasing and derivative $\left(f^{-1}(x)\right)' < 0$, and the probability density $g(y)$ cannot be negative).

To find the **numerical characteristics** of a random variable $Y = f(X)$, it is not necessary to know the law of its distribution, it is enough to know the law of the distribution of the argument:

$$a_y = M(Y) = M\left(f(X)\right) = \int_{-\infty}^{\infty} f(x)\varphi(x)dx,$$

$$D(Y) = D\left(f(X)\right) = \int_{-\infty}^{\infty} \left(f(x) - a_y\right)^2 \varphi(x)dx.$$

◄ **Example 8.3.** Find the probability density of a random variable $Y = 1 - X^3$, where the random variable X is distributed according to the Cauchy law with the probability density $\varphi(x) = \dfrac{1}{\pi(1+x^2)}$.

The solution. By condition $y = f(x) = 1 - x^3$, from here $x = f^{-1}(y) = \sqrt[3]{1-y}$. Derivative (by absolute value):

$$\left|\left(f^{-1}(y)\right)'\right| = \frac{1}{3\sqrt[3]{(1-y)^2}}.$$

Probability density: $g(y) = \dfrac{1}{3\pi\left(1+\sqrt[3]{(1-y)^2}\right)\sqrt[3]{(1-y)^2}}$. ►

◄ **Example 8.4.** Find the mathematical expectation and dispersion of the random values $Y = 2 - 3\sin X$ if the probability density of random value X is given by the function $\varphi(x) = \frac{1}{2}\cos x$ on the segment $\left[-\pi/2,\ \pi/2\right]$.

The solution. Mathematical expectation:

$$a_y = M(Y) = \int_{-\pi/2}^{\pi/2} (2 - 3\sin x)\frac{1}{2}\cos x\, dx = 2.$$

Dispersion: $$D(Y) = M(Y^2) - a_y^2.$$

$$M(Y^2) = \int_{-\pi/2}^{\pi/2} (2 - 3\sin x)^2 \frac{1}{2}\cos x\, dx = 7 \quad, \text{i } D(Y) = 7 - 2^2 = 3. \text{ ►}$$

For practice, the task of determining the law of distribution of the sum of two random variables, that is, the law of distribution of a random variable $Z = X + Y$, is of particular importance. If X and Y are independent random variables, then a **composition** (convolution) of distribution laws is used. Let X and Y be random variables with distribution densities $\varphi_1(x)$ and $\varphi_2(y)$, respectively. Let's find the distribution function of the random variable Z:

$$F(z) = P(Z < z) = P(X + Y < z) = \iint_{D_z} \varphi(x, y)dx\, dy,$$

where D_z — the set of all points of the Oxy plane whose coordinates satisfy the inequality $x + y < z$ (fig. 8.6), $\varphi(x, y)$— compatible density of a two-dimensional random variable (X, Y). Since X and Y are independent random variables, then $\varphi(x, y) = \varphi_1(x)\varphi_2(y)$ and

$$F(z)=\int_{-\infty}^{\infty}\int_{-\infty}^{z-x}\varphi_1(x)\varphi_2(y)dxdy=\int_{-\infty}^{\infty}\varphi_1(x)dx\int_{-\infty}^{z-x}\varphi_2(y)dy.$$

Let's find the probability density $\varphi(z)$:

$$\varphi(z)=F'(z)=\int_{-\infty}^{\infty}\varphi_1(x)dx\left(\int_{-\infty}^{z-x}\varphi_2(y)dy\right)',\ \varphi(z)=\int_{-\infty}^{\infty}\varphi_1(x)\varphi_2(z-x)dx.$$

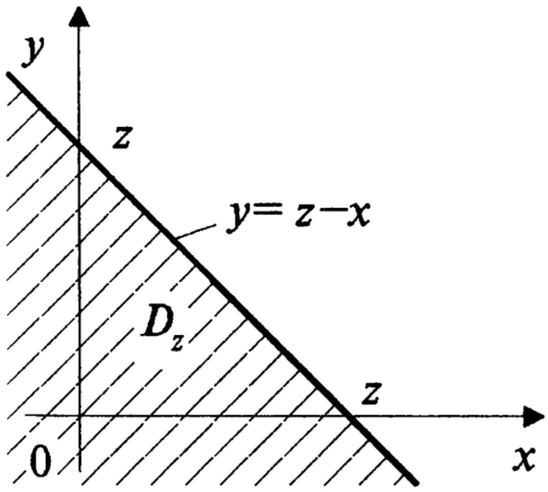


Fig. 8.6. Illustration for the plural D_z

The last formula is called the **formula of the composition** of two distributions or the **convolution formula**, which in abbreviated form has the form:

$$\varphi=\varphi_1*\varphi_2.$$

◄ **Example 8.5.** Find the law of distribution of the sum of two random variables that have a uniform distribution on the segment [0; 1].

The solution. Let $Z=X+Y,$ where $\varphi_1(x)=1$ when $0\le x\le 1$ and $\varphi_2(y)=1$ when $0\le y\le 1$. Probability density

$$\varphi(z) = \int_0^1 1 \cdot \varphi_2(z-x)dx = \int_0^1 \varphi_2(z-x)dx.$$

If $z < 0$, then for $0 \le x \le 1$ $z - x < 0$; if $z > 2$, then for $0 \le x \le 1$ $z - x > 1$, therefore, in these cases $\varphi_2(z-x) = 0$ and $\varphi(z) = 0$.

Let $0 \le z \le 2$. The integrand function $\varphi_2(z-x)$ will differ from zero only for the values of x, for which $0 \le z - x \le 1$ or, for $z - 1 \le x \le z$.

If $0 \le z \le 1$, then $\varphi(z) = \int_0^z 1 \cdot dx = z$.

If $1 \le z \le 2$, then $\varphi(z) = \int_{z-1}^1 1 \cdot dx = 2 - z$.

Combining the cases, we get:

$$\varphi(z) = \begin{cases} 0, & \text{when } z < 0, z > 2 \\ z, & \text{when } 0 \le z \le 1 \\ 2 - z, & \text{when } 1 \le z \le 2 \end{cases}.$$

This law of distribution is called **Simpson's law of distribution** or the law of the isosceles triangle (Fig. 8.7).

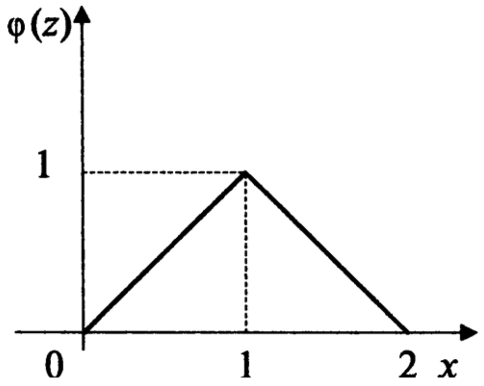


Fig. 8.7. Simpson's law of distribution. ►

A composition of normal distribution laws also has a normal distribution. So, if

X and Y are independent normally distributed random variables, i.e. $X \sim N(a_x, \sigma_x^2)$, $Y \sim N(a_y, \sigma_y^2)$, then the random variable $Z = X + Y$ is also normally distributed: $Z \sim N(a_x + a_y, \sigma_x^2 + \sigma_y^2)$. If the random variables X and Y are normally distributed and dependent (correlation coefficient $\rho \neq 0$), then the random variable $Z = X + Y$ is equally normally distributed with parameters

$$a_z = a_x + a_y, \qquad \sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y.$$

Control questions

1. Which two random variables can be called independent?
2. Why the correlation coefficient better reflects the relationship between random variables?
3. What is a regression line?
4. What are the parameters of the random variable $Z = X + Y$ equal to if $X \sim N(a_x, \sigma_x^2)$, $Y \sim N(a_y, \sigma_y^2)$?

Chapter 9

The law of large numbers. Markov's, Chebyshov's, Bernoulli's inequalities. Central limit theorem

In a broad sense, the law of large numbers is a general principle according to which, according to the formulation of Academician A.N. Kolmogorov, the combined effect of a large number of random factors (under some fairly general conditions) contributes to the emergence of a result that is practically independent of randomness. In other words, with a large number of random variables, their average result ceases to be random and can be predicted with a high degree of probability.

The law of large numbers is a series of mathematical theorems that determine the conditions under which the arithmetic mean of many random variables practically ceases to be random.

9.1. Markov's inequality (Chebyshov's lemma)

Theorem 9.1. If the random variable X takes only non-negative values and has a mathematical expectation, then for any positive number A the inequality holds:

$$P(x > A) \le \frac{M(X)}{A}.$$

Proof. For a discrete random variable X. We arrange the values of the random variable X in the sequence of growth: part of the values $x_1, x_2, ..., x_k$ will not exceed the number A, and the second part $x_{k+1}, ..., x_n$ will be more than the number A, i.e. $x_1 \le A, x_2 \le A, ..., x_k \le A; x_{k+1} > A, ..., x_n > A$ (Fig. 9.1)

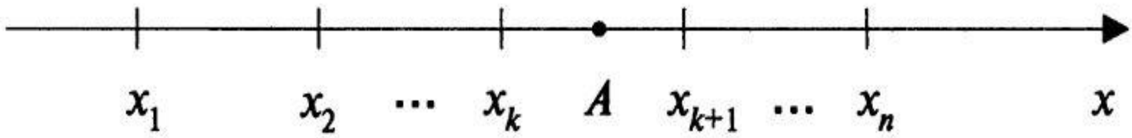


Fig. 9.1. A sequence of values of a discrete random variable X

Let's write down the expression for the mathematical expectation $M(X)$:

$$x_1 p_1 + x_2 p_2 + \ldots + x_k p_k + x_{k+1} p_{k+1} + \ldots + x_n p_n = M(X)$$

where $p_1, p_2, \ldots, p_n$ - probabilities that the random variable X will take a value, respectively $x_1, x_2, \ldots, x_n$. Discarding the first k integral terms (all $x_i \geq 0$), we get $x_{k+1} p_{k+1} + \ldots + x_n p_n \leq M(X)$. Substituting in this inequality the values $x_{k+1} + \ldots + x_n$ with a smaller number A, we will strengthen the inequality

$$A(p_{k+1} + \ldots + p_n) \leq M(X) \text{ or } p_{k+1} + \ldots + p_n \leq \frac{M(X)}{A}$$

The sum of the probabilities on the left side of the obtained inequality is the sum of the probabilities of the events $X = x_{k+1}, \ldots, X = x_n$, that is, the probability of the event $X > A$.

Ago $P(X > A) \leq \dfrac{M(X)}{A}$. Since the events $X > A$ and $X \leq A$ are opposite, replacing $P(X > A)$ by the expression $1 - P(X \leq A)$, we get another form of Markov's inequality:

$$P(X \leq A) \leq 1 - \frac{M(X)}{A}. \bullet$$

Markov's inequality can be applied only to non-negative random variables.

◄ **Example 9.1.** The average number of calls received at the switchboard of the factory during an hour is 300. Estimate the probability that during the next hour the number of calls received at the switchboard: *a*) will exceed 400, *b*) will not be more than 500.

The solution. *a*) By condition $M\left(X\right)=300$. Therefore, $P(X>400)\le\frac{300}{400}$, that is, the probability that the number of calls will exceed 400 will not exceed 0.75.

b) $P(X\le 500)\le 1-\frac{300}{500}$, that is, the probability that the number of calls will not exceed 500 will be at least 0.4. ►

◄ **Example 9.2.** The sum of all deposits in a bank branch is UAH 2 million, and the probability that a randomly selected deposit does not exceed UAH 10,000 is 0.6. What can be said about the number of depositors?

The solution. Let X is the value of a randomly selected contribution, and n is the number of all contributions. Then, under the condition of the problem, the average amount of the contribution: $M(X)=\frac{2000}{n}$ (thousand hryvnias). According to Markov's inequality: $P(X\le 10)\le 1-\frac{M(X)}{10}$, or

$$P(X\le 10)\le 1-\frac{2000}{10n}.$$

Considering that $P(X\le 10)\le 0{,}6$, we get $1-\frac{200}{n}\le 0{,}6$, from here $n\le 500$, that is, the number of depositors does not exceed 500.►

9.2. Chebyshov's inequality

Theorem 9.2. For any random variable with mathematical expectation and dispersion, Chebyshov's inequality is valid:

$$P\left(|X-a|>\varepsilon\right)\le\frac{D(X)}{\varepsilon^2},$$

where $a=M(X),\ \varepsilon>0$.

Proof. Let's apply Markov's inequality $P(x>A)\le\frac{M(X)}{A}$ to a random variable $X'=(X-a)^2$. Let's enter the notation: $A=\varepsilon^2$. We will get $P\left((X-a)^2>\varepsilon^2\right)\le\frac{M(X-a)^2}{\varepsilon^2}$. Since the inequality $(X-a)^2>\varepsilon^2$ is equivalent to inequality $|X-a|>\varepsilon$, and $M(X-a)^2$ is the dispersion of a random variable X, then from inequality $P\left((X-a)^2>\varepsilon^2\right)\le\frac{M(X-a)^2}{\varepsilon^2}$ we get the desired inequality:

$$P\left(|X-a|>\varepsilon\right)\le\frac{D(X)}{\varepsilon^2}.$$

Considering that the events $|X-a|>\varepsilon$ and $|X-a|\le\varepsilon$ are opposite, inequality Chebyshov can be written in another form:

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

Chebyshov's inequality applies to any random variables. In the form $P\left(|X-a|>\varepsilon\right)\le\frac{D(X)}{\varepsilon^2}$, it sets the upper limit, and in the form

$P(|X-a|\le \varepsilon)\ge 1-\dfrac{D(X)}{\varepsilon^2}$ - the lower limit of the probability of the event under consideration. Let's write Chebyshov's inequality for some random variables:

a) for a random variable $X=m$ having a binomial distribution law with mathematical expectation $a=M(X)=np$ and dispersion $D(X)=npq$:

$$P(|m-np|\le \varepsilon)\ge 1-\frac{npq}{\varepsilon^2};$$

b) for the frequency $\dfrac{m}{n}$ of the event in n independent trials, in each of which it can occur with the same probability $a=M\left(\dfrac{m}{n}\right)=p$, and has the dispersion

$D\left(\dfrac{m}{n}\right)=\dfrac{pq}{n}$: $\quad P\left(\left|\dfrac{m}{n}-p\right|\le \varepsilon\right)\ge 1-\dfrac{pq}{n\varepsilon^2}$.

◄ **Example 9.3.** On a livestock farm, the average daily water withdrawal is 1000 l, and the mean square fluctuation of this random value does not exceed 200 l. Estimate the probability that water consumption on the farm on any selected day will not exceed 2000 L using: a) Markov's inequality, b) Chebyshov's inequality.

The solution. *a*) Let the random variable X be water consumption on the livestock farm (l). By condition $M(X)=1000$. Using Markov's inequality, we get $P(X\le 2000)\ge 1-\dfrac{1000}{2000}=0{,}5$, i.e. not less than 0,5.

b) Dispersion $D(X)=\sigma^2\le 200^2$. Since the limits of the interval $0\le X\le 2000$ are symmetrical with respect to the mathematical expectation $M(X)=1000$, the Chebyshov inequality can be applied to estimate the probability of a given event:

$$P(X \leq 2000) = P(0 \leq X \leq 2000) = P\left(|X - 1000| \leq 1000\right) \geq 1 - \frac{200^2}{1000^2} = 0{,}96,$$

i.e. not less than 0,96.

In this example, the estimate of the probability of the event, found using the Markov inequality ($P \geq 0{,}5$), was refined using the Chebyshov inequality ($P \geq 0{,}96$). ►

◄ **Example 9.4**. The probability of manufacturing a standard detail on the machine is 0.96. Using Chebyshov's inequality, estimate the probability that the number of defective details among 2000 parts is between 60 and 100 (inclusive). Specify the probability of a given event using the Moivre-Laplace integral theorem. Explain the difference between the obtained results.

The solution. According to the condition of the problem, the probability that the detail is defective is equal to $p = 1 - 0{,}96 = 0{,}04$. The number of defective details $X = m$ has a binomial distribution law, and its limits of 60 and 100 are symmetrical relative to mathematical expectation

$$a = M(X) = np = 2000 \cdot 0{,}04 = 80.$$

Therefore, an estimate of the probability of a given event

$$P(60 \leq m \leq 100) = P(-20 \leq m - 80 \leq 20) = P\left(|m - 80| \leq 20\right)$$

can be found by the formula $P\left(|m - np| \leq \varepsilon\right) \geq 1 - \frac{npq}{\varepsilon^2}$:

$$P\left(|m - 80| \leq 20\right) \geq 0{,}808.$$

Therefore, the probability is at least 0.808. Applying a corollary from the Moivre-Laplace integral theorem, we obtain

$$P\left(\left|m-80\right| \le 20\right) \approx \Phi\left(\frac{20}{\sqrt{76{,}8}}\right) = 0{,}979,$$

that is, the sought probability is approximately equal to 0.979. The obtained result $P \approx 0{,}979$ does not contradict the estimate found using Chebyshov's inequality — $P \ge 0{,}808$. The difference in results is explained by the fact that Chebyshov's inequality gives only the lower limit of the probability estimate of the desired event for any random variable, and the Moivre-Laplace integral theorem gives a fairly accurate value of the probability P itself (the more accurate, the larger n), since it is applied only to a random variable , which has a definite, namely, binomial distribution law. ►

◄ **Example 9.5.** Estimate the probability that the deviation of any random variable from its mathematical expectation will be no more than three root mean square deviations (by absolute value) - (rule of three sigma).

The solution. According to the formula $P\left(\left|m-np\right| \le \varepsilon\right) \ge 1-\frac{npq}{\varepsilon^2}$, considering that $D(X) = \sigma^2$, we will get:

$$P\left(\left|X-a\right| \le 3\sigma\right) \ge 1-\frac{\sigma^2}{(3\sigma)^2} = \frac{8}{9} = 0{,}889,$$

that is, not less than 0.889. (For a normal law, the rule of three sigma is fulfilled with a probability P equal to 0.9973, i.e. $P = 0{,}9973$) It can be shown that for a uniform distribution law $P = 1$, for an exponential $P = 0{,}9827$, etc. Thus, the rule of three sigma (with a fairly high probability of its implementation) can be applied to most random variables encountered in practice. ►

◄ **Example 9.6.** According to statistics, on average, 87% of newborns live to be

50 years old. Using Chebyshov's inequality, estimate the probability that out of 1000 newborns, the proportion of those who lived to age of 50 will differ from the probability of this event by no more than 0.04 (by absolute value).

The solution. Provided $n = 1000, p = 0,87,$ in accordance $q = 1 - p = 0,13$,

$P\left(\left|\frac{m}{n} - p\right| \leq 0,04\right) \geq 1 - \frac{0,87 \cdot 0,13}{1000 \cdot 0,04^2} = 0,929$, that is, not less than 0.929. (A sufficiently accurate value of the probability of this event was obtained when using a consequence of the Moivre-Laplace integral theorem, which was equal to 0.9998).

Remark. If the mathematical expectation $M(X) > A$ or dispersion of a random variable is $D(X) > \varepsilon^2$, then the right-hand sides of the Markov and Chebyshov inequalities will be negative, or will be greater than 1. This means that the application of these inequalities in these cases will lead to a trivial result: the probability of an event is greater than a negative number or less than a number greater than 1. This conclusion is obvious even without using these inequalities. This circumstance reduces the value of the Markov and Chebyshev inequalities when solving practical problems, but does not reduce their theoretical value.

9.3. Chebyshov's theorem

Chebyshov's theorem. If the dispersions of n independent random variables $X_1, X_2, ..., X_n$ are limited by one and the same constant value, then with an unlimited increase in the number n, the arithmetic mean of the random variables goes straight (probably coincides) to the arithmetic mean of their mathematical expectations $a_1, a_2, ..., a_n$, i.e.

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{a_1 + a_2 + ... + a_n}{n}\right| \leq \varepsilon\right) = 1,$$

or $$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{\mathfrak{R}} \frac{\sum_{i=1}^{n} a_i}{n}$$

Proof. Let's prove the formula:

$$\lim_{n\to\infty} P\left(\left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n} \right| \le \varepsilon \right) = 1 .$$

By condition $M(X_1) = a_1, M(X_2) = a_2, \ldots, M(X_n) = a_n$,

$D(X_1) \le C, D(X_2) \le C, \ldots, D(X_n) \le C,$ where C is constant.

We obtain Chebyshov's inequality for the arithmetic mean of random variables, that is, for $X = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

Let's find the mathematical expectation $M(X)$ and dispersion estimate $D(X)$:

$$M(X) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \right) = \frac{1}{n}\big(M(X_1) + M(X_2) + \ldots + M(X_n) \big) =$$

$$= \frac{a_1 + a_2 + \ldots + a_n}{n};$$

$$D(X) = D\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \right) =$$

$$= \frac{1}{n^2}\big(D(X_1) + D(X_2) + \ldots + D(X_n) \big) \le \frac{nC}{n^2} = \frac{C}{n};$$

Let's write Chebyshov's inequality for a random variable

$X = \big(X_1 + X_2 + \ldots + X_n \big) / n$:

$$P\left(\left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n} \right| \le \varepsilon \right) \ge 1 - \frac{D(X)}{\varepsilon^2} .$$

Since it has been proven that $D(X) \le \dfrac{C}{n}$, then

$$1-\frac{D(X)}{\varepsilon^2}\geq 1-\frac{C/n}{\varepsilon^2}=1-\frac{C}{n\varepsilon^2}$$

and strengthen the inequality:

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\leq\varepsilon\right)\geq 1-\frac{C}{n\varepsilon^2}.$$

If $n\to\infty$, then the value $\frac{C}{n\varepsilon^2}$ goes to zero. Therefore. we will get the required formula $\lim\limits_{n\to\infty}P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\leq\varepsilon\right)=1$.

Now let's find out the expression - "convergence by probability". The concept of the limit of a variable value $X(\lim\limits_{n\to\infty}X=a$ or $X\to a$ when $n\to\infty)$ means that starting from some moment of its change for any (even very small) number $\varepsilon>0$ will satisfy the inequality $|X-a|<\varepsilon$. In the round brackets of the expression $\left|\frac{\left(\sum_{i=1}^{n}X_i\right)}{n}-\frac{\left(\sum_{i=1}^{n}a_i\right)}{n}\right|<\varepsilon$ contains a similar expression, where $\left(\sum_{i=1}^{n}X_i\right)\Big/n$ - random variable, and $\left(\sum_{i=1}^{n}a_i\right)\Big/n$ - a constant number. However, this does not mean that this inequality will always hold starting at some point. Because $\left(\sum_{i=1}^{n}X_i\right)\Big/n$ - random variable, it is possible that in some cases the inequality will not hold. However, with an increase in the number n, the probability of inequality $\left|\left(\sum_{i=1}^{n}X_i\right)\Big/n-\left(\sum_{i=1}^{n}a_i\right)\Big/n\right|\leq\varepsilon$

then it is possible that in some cases the inequality will not be fulfilled. However, as the number increases, the probability of inequality $\left|\left(\sum_{i=1}^{n} X_i\right)\Big/ n - \left(\sum_{i=1}^{n} a_i\right)\Big/ n\right| \le \varepsilon$ leads to 1, that is, this inequality will be fulfilled in the vast majority of cases. In other words, with sufficiently large implementations of the considered inequality, the event is practically reliable, and inequalities of the opposite meaning are practically impossible.

Thus, the approximation $\left(\sum_{i=1}^{n} X_i\right)\Big/ n$ to $\left(\sum_{i=1}^{n} a_i\right)\Big/ n$ should be understood not as a categorical statement, but as a statement whose correctness is guaranteed with a probability as close to 1 as $n \to \infty$. This circumstance is reflected in the formulation of the theorem "coincides by probability" and in the expression $\xrightarrow[n\to\infty]{\Re}$.

With a large number n of random variables $X_1, X_2, ..., X_n$, it is almost certain that their random average value $X = \left(\sum_{i=1}^{n} X_i\right)\Big/ n$, differs arbitrarily little from a non-random value $\left(\sum_{i=1}^{n} a_i\right)\Big/ n$, that is, it practically ceases to be random. ●

A consequence of Chebyshov's theorem. If independent random variables $X_1, X_2, ..., X_n$ have the same mathematical expectation equal to a, and their dispersions are bounded by the same constant value, then:

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|\le\varepsilon\right)\ge 1-\frac{C}{n\varepsilon^2}$$

, $$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|\le\varepsilon\right)=1,$$

$$\frac{\sum_{i=1}^{n}X_i}{n}\xrightarrow[n\to\infty]{\Re}a.$$

These formulas follow from Chebyshov's theorem, since $M(X)=M\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{1}{n}\left(M(X_1)+M(X_2)+\ldots+M(X_n)\right)=$

$$=\frac{a_1+a_2+\ldots+a_n}{n}=\frac{na}{n}=a.$$

●

Chebyshov's theorem and its corollary are of great practical importance. For example, the insurance company needs to establish the amount of the insurance premium that the insured must pay; at the same time, the insurance company undertakes to pay a certain insurance amount in the event of an insured event. Then, on the basis of these data and the estimated insurance amount, the amount of the insurance premium is determined. Without taking into account the effect of the law of large numbers (Chebyshov's theorem), large losses of the insurance company are possible (if the amount of the insurance premium is underestimated), or the loss of attractiveness of insurance services (if the amount of the premium is overestimated). Another example. If it is necessary to measure some quantity that has a specific, certain value, independent measurements of this quantity are carried out. Let the result of each measurement be a random variable $X_i, (i=1,2,\ldots,n)$. If there are no systematic errors during the measurements (the measurement results do not deviate in the same direction), then it can be assumed that $M\left(X_i\right)=a$ at any i. Then, on the basis of a consequence of Chebyshov's theorem, the arithmetic mean of n the measurement results $\left(\sum_{i=1}^{n} X_i\right)\Big/ n$ coincides with probability to the true value. This is the basis for choosing the arithmetic mean as an estimate of the true value a.

This is the basis for choosing the arithmetic mean as an estimate of the true value a. If all measurements are carried out with the same accuracy, characterized by dispersion $D\left(X_i\right)=\sigma^2$, then the dispersion of their average

$$D\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)=\frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right)=\frac{\sigma^2}{n},$$

and its mean square deviation is accordingly determined by the ratio: $\frac{\sigma}{\sqrt{n}}$.

The resulting relation is known as the "rule of the root of *n*". Thus, by increasing the number of measurements, it is possible to reduce the impact of random errors (but not systematic) as desired, that is, to increase the accuracy of determining the true value a.

Remark. If the measuring device has accuracy δ (for example, δ — half the width of the division of the uniform scale of the device, for which the reading is made), then the above-mentioned method cannot obtain the accuracy of measuring a value a greater than δ. Each measurement gives a result with uncertainty δ and, obviously, their arithmetic mean will have the same uncertainty δ. Thus, trying to use the law of large numbers to obtain the value a of a with a greater degree of accuracy than the instrument allows for a single measurement is erroneous.

◄ **Example 9.7.** To determine the average burning time of electric lamps in a batch of 200 identical boxes, one lamp was taken (from each box). Estimate the probability that the average burning duration of the selected 200 electric lamps differs from the average burning duration of the lamps in the entire batch by no more than 5 hours (in absolute value), if it is known that the average square deviation of the burning duration of the lamps in each box is less than 7 hours.

The solution. Let X_i - duration of burning of the electric lamp taken from the - ith box (hours). By the condition of dispersion $D\left(X_i\right) < 7^2 = 49$. It is obvious that the average burning time of the selected lamps is equal $\frac{X_1 + X_2 + ... + X_{200}}{200}$, and the average duration of lamp burning in the entire batch

$$\frac{M\left(X_1\right) + M\left(X_2\right) + ... + M\left(X_{200}\right)}{200} = \frac{a_1 + a_2 + ... + a_{200}}{200}.$$

Then the probability of the desired event:

$$P\left(\left|\frac{X_1 + X_2 + ... + X_{200}}{200} - \frac{a_1 + a_2 + ... + a_{200}}{200}\right| \le 5\right) \ge 1 - \frac{49}{200 \cdot 25} \approx 0{,}9902$$

i.e. not less than 0,9902. ►

◄ **Example 9.8.** How many measurements of a given quantity must be carried out to ensure, with a probability of at least 0.95, that the deviation of the arithmetic mean of these measurements from the true value of the quantity is no more than 1 (by absolute value), if the mean square deviation of each of the measurements does not exceed 5?

The solution. Let X_i - the result of the - i th measurement $(i=1,2,...,n)$;

a - the true value of the quantity, i.e $M\left(X_i\right)=a$ at any i.

It is necessary to find n at which

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-a\right|\le 1\right)\ge 0{,}95 .$$

This inequality will hold if $1-\frac{C}{n\varepsilon^2}=1-\frac{25}{n1^2}\ge 0{,}95$, from here $\frac{25}{n}\le 0{,}05$

and $n\ge\frac{25}{0{,}05}=500$, that is, at least 500 measurements are required.►

9.4. Bernoulli's theorem

Bernoulli's theorem. The frequency of the event in n repeated independent trials, in each of which it can occur with the same probability p, with an unlimited increase in the number n, coincides in probability with the probability p of this event in a separate trial:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=1,$$

or

$$\frac{m}{n}\xrightarrow[n\to\infty]{\mathfrak{R}} p .$$

Proof. The proof follows directly from Chebyshev's inequality for the event frequency at $n\to\infty$. The essence of Bernoulli's theorem is that with a large number n of repeated independent trials, it is practically certain that the frequency (or statistical probability) of an event $\frac{m}{n}$ is a random quantity, no matter how little it differs from a non-random quantity p - the probability of an event, that is, it practically ceases to be random.●

Remark. Bernoulli's theorem is a consequence of Chebyshov's theorem, because the frequency of an event can be represented as the arithmetic mean of n independent alternative random variables having the same distribution law. The proof of the theorem is also possible without reference to Chebyshov's theorem (inequality). Historically, this theorem was proved much earlier than the more general Chebyshov theorem. Bernoulli's theorem provides a theoretical rationale for replacing the unknown probability of an event with its frequency, or statistical probability obtained in n repeated independent trials conducted under the same set of conditions.

So, for example, if we do not know the probability of the birth of a boy, then as its value we can accept the frequency (statistical probability) of this event, which, as is known from long-term statistical data, is approximately 0.515. Bernoulli's theorem is a link that allows you to connect the formal axiomatic definition of probability with the empirical law of constancy of relative frequency. The theorem makes it possible to justify the wide application of probabilistic research methods in practice.

A direct generalization of Bernoulli's theorem is Poisson's theorem, when the probabilities of the event in each trial are different.

Poisson's theorem. The frequency of the event in n repeated independent trials, in each of which it can occur according to the probabilities $p_1, p_2, \ldots, p_n$ with an unlimited increase in the number n, coincides in probability with the arithmetic mean of the event probabilities in individual trials, i.e.

$$\lim_{n\to\infty} P\left(\left| \frac{m}{n} - \frac{p_1, p_2, \ldots, p_n}{n} \right| \le \varepsilon \right) = 1, \qquad \text{or}$$

$$\frac{m}{n} \xrightarrow[n\to\infty]{\mathfrak{R}} \frac{\sum_{i=1}^{n} p_i}{n}.$$

Proof. Poisson's theorem follows directly from Theorem Chebyshov, if as random variables $X_1, X_2, ..., X_n$ we consider alternative random variables having distribution laws with parameters $p_1, p_2, ..., p_n$.

Since the mathematical expectations of random variables $X_1, X_2, ..., X_n$ are equal to $p_1, p_2, ..., p_n$, respectively, and their dispersions $p_1q_1, p_2q_2, ..., p_nq_n$ are limited to one number. ●

9. 5. Central limit theorem

The central limit theorem is a group of theorems that establish the conditions under which the normal distribution law is widely used.

Lyapunov's theorem. If $X_1, X_2, ..., X_n$ are independent random variables, each of which has a mathematical expectation $M(X_i) = a_i$, $i = \overline{1 \div n}$, and a dispersion $D(X_i) = \sigma_i^2$, an absolute central moment of the third order $M\left(\left|X_i - a_i\right|^3\right) = m_i$ and $\lim\limits_{n\to\infty} \frac{\sum_{i=1}^{n} m_i}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}} = 0$, then the law of distribution of the sum $Y_n = X_1 + X_2 + ... + X_n$ at $n \to \infty$ indefinitely approaches the normal distribution with a mathematical expectation $a=\sum_{i=1}^{n} a_i$ and dispersion $D(Y_n)=\sigma^2=\sum_{i=1}^{n}\sigma_i^2$.

Let's accept the theorem without proof. ●

The unbounded approximation of the sum $Y_n = X_1 + X_2 + ... + X_n$ distribution law to the normal distribution law means that

$$\lim_{n\to\infty} P\left(\left|\frac{Y_n-\sum_{i=1}^{n}a_i}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}}\right|\le z\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt=\frac{1}{2}+\frac{1}{2}\Phi(z),$$ where $\Phi(z)$

— the Laplace function. Content of the condition $\lim_{n\to\infty}\frac{\sum_{i=1}^{n} m_i}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^{3/2}}=0$ is that in sum $Y_n = X_1 + X_2 + ... + X_n$ there should be no terms whose influence on the scattering Y_n is quite large compared to the other terms, and there should also be no large number of random terms whose influence is very small compared to the total influence of the other terms.

For example, the consumption of electricity for household purposes for a month in each apartment of a high-rise building can be represented in the form of *n* various random variables. If the value of the consumed electricity in each apartment does not stand out sharply among the rest, then based on Lyapunov's theorem, it can be assumed that the consumption of electricity in the whole house will be a random value that has a normal distribution. And if, for example, a computer center is placed in one of the apartments, then the conclusion about a normal distribution will be wrong.

Consequence. If $X_1, X_2, ..., X_n$ are independent random variables, each of which has a mathematical expectation $M(X_i) = a,\ i = \overline{1 \div n}$, dispersion $D(X_i) = \sigma^2$, absolute central moment of the third order $M\left(\left|X_i - a_i\right|^3\right) = m_i$, then the law of distribution of the sum $Y_n = X_1 + X_2 + ... + X_n$ at $n \to \infty$ indefinitely approaches the normal law.

Proof comes down to checking the condition:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} m}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}} = \lim_{n\to\infty} \frac{mn}{\left(n\sigma^2\right)^{3/2}} = \lim_{n\to\infty} \frac{m}{\sigma^3 \sqrt{n}} = 0. \bullet$$

If all random variables X_i have the same distribution law, then the distribution law of their sum indefinitely approaches the normal law when $n \to \infty$.

We will prove the local and integral theorems of Moivre-Laplace.

Consider a random variable $Z = \frac{m - np}{\sqrt{npq}}$, where $X = m$ — the number of occurrences of some event in n independent trials, given that in each trial it can appear with probability p. A random variable $X = m$ has a binomial distribution for which the mathematical expectation $M(X) = np$ and $D(X) = npq$.

The random variable Z is discrete, but with a large number of trials, its values are located very closely on the abscissa axis, and it can be considered as a continuous random variable with probability density $\varphi(Z)$. Numerical characteristics of a random variable Z:

$$a_z = M(Z) = \frac{M(X) - np}{\sqrt{npq}} = \frac{np - np}{\sqrt{npq}} = 0;$$

$$D(Z) = \frac{D(X) - 0}{npq} = \frac{npq}{npq} = 1.$$

Using the central limit theorem, it can be asserted that the random variable Z has a distribution close to normal with parameters $a = 0$ and $\sigma^2 = 1$.
We use the property of the normal distribution:

$$P(z_1 \le Z \le z_2) \approx \frac{1}{2}\left(\Phi(z_2) - \Phi(z_1)\right).$$

Let's mark $z_1 = \frac{a - np}{\sqrt{npq}}$, $z_2 = \frac{b - np}{\sqrt{npq}}$, considering that $Z = \frac{m - np}{\sqrt{npq}}$. We get the inequality $z_1 \le Z \le z_2$, which is equivalent to the inequality $a \le m \le b$. As a result, we get the Moivre-Laplace integral formula:

$$P(a \le m \le b) \approx \frac{1}{2}\left(\Phi(z_2) - \Phi(z_1)\right).$$

Similarly, the local Moivre-Laplace formula is proved.

Remark. When using the central limit theorem for statistical research, it is necessary to take into account the speed of convergence to the normal distribution law of the sum of values and the type of distribution.

Control questions

1. Markov's inequality and conditions of its application.
2. Chebyshov's inequality and conditions of its application.
3. The central limit theorem and its advantages in research of random variables.

Appendix 1

The values of the Poisson's function $P_m(\lambda)=\frac{\lambda^m}{m!}e^{-\lambda}$

m \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2223	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1216	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

m \ λ	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
0	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0001
1	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,1805	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0902	0,1681	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171	0,0901
8	0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318	0,1251
10	0,0000	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11	0,0000	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12	0,0000	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728	0,0948
13	0,0000	0,0000	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14	0,0000	0,0000	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15	0,0000	0,0000	0,0000	0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16	0,0000	0,0000	0,0000	0,0000	0,0003	0,0015	0,0045	0,0109	0,0217
17	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0021	0,0058	0,0128
18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0009	0,0029	0,0071
19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0037
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0006	0,0019
21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0009
22	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004
23	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002
24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Appendix 2

The values of the Gaussian's function $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Wholes and tenths of fates x	Hundred destinies x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0041	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018

Continuation of the table of values of the Gaussian's function

Wholes and tenths of fates x	Hundred destinies x									
	0	1	2	3	4	5	6	7	8	9
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,0001338									
4,5	0,0000160									
5,0	0,0000015									

Appendix 3

The values of the Laplace's function $\Phi(x)=\frac{2}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$

Wholes and tenths of fates *x*	Hundred destinies *x*									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7984	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9392	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

REFERENCES

1. Siegmund, David O. Probability theory. Encyclopedia Britannica, 27 Jan. 2023, https://www.britannica.com/science/probability-theory.
2. Philip B. Stark. Electronic summary of lectures. Philip B. Stark is Professor of Statistics and Associate Dean of Mathematical and Physical Sciences at the University of California, Berkeley, URL: https://www.stat.berkeley.edu/~stark/SticiGui/Text/reasoning.htm
3. Hartigan, J.A. (1983). Theories of Probability. In: Bayes Theory. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-8242-3_1
4. Овчинников П.П., Михайленко В.М. Вища математика. У 2 ч. Ч 2: Диференціальні рівняння. Операційне числення. Ряди та їх застосування. Стійкість за Ляпуновим. Рівняння математичної фізики. Оптимізація і керування. Теорія ймовірностей. Числові методи – К.: Техніка, 2004. – 792с.

Електронне мережеве навчальне видання

Онай Микола Володимирович, доцент, к.т.н.

Сущук-Слюсаренко Вікторія Ігорівна

Дичка Андрій Іванович, доктор філософії

Основи теорії ймовірностей

Навчальний посібник

(Англійською мовою)

Навчальний посібник розроблено для ознайомлення студентів з теоретичними відомостями та практичними прийомами теорії ймовірностей, а також як посібник для використання на практичних заняттях. Навчальний посібник призначений для студентів, які навчаються за спеціальністю 121 (F2) «Інженерія програмного забезпечення» факультету прикладної математики «КПІ» ім. Ігоря Сікорського.

Реєстр. № НП Обсяг 4,9 авт. арк.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Проспект Берестейський, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.