Ministry of Ukraine Transport and Communication
State Department of Communication and Informatization

Odessa National Academy of Communication after A.Popov

Department of Higher Mathematics

Theory of Probability
Educational Aid

For Students Studying a Course of Higher Mathematics in English

Odessa – 2008

УДК 510(083) План НМВ 2007/2008 уч. г.

Составители:
доц. Гавдзинский В. Н., ст. преподаватель Коробова Л. Н.

В пособии в краткой форме представлены основные сведения по теории вероятностей для студентов академии, изучающих высшую математику на английском языке. Основные теоремы и формулы приведены с доказательством, а также даны решения типовых примеров и задания для самостоятельного решения.

CONTENTS

CHAPTER 1
FUNDAMENTAL DEFINITIONS AND THEOREMS OF PROBABILITY

§ 1.1. Combinatorial Analysis. Arrangements. Permutations. Combinations

Definition. A set consisting of m elements chosen from n giving elements and put in the definite order is called an **arrangement** of n elements taken m and denoted by A_n^m.

Arrangements can differ one from another either by elements or by their order.

Arrangements are calculated by the formula

$$A_n^m = n(n-1)\ldots(n-m+1) = \frac{n!}{(\mathrm{n - m})!} \qquad (1.1.1)$$

where

$$n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n.$$

Example 1.1.1. In how many ways it is possible to choose the trade-union group organizer and his deputy from the group of seven students?

Solution. $N = A_7^2 = 7 \cdot 6 = 42.$

Definition. Any ordered set which consists of m elements is called a **permutation** and denoted by P_m.
It is read: permutation of m digits at a time.

Permutations are calculated by the formula

$$P_m = m! \qquad (1.1.2)$$

Example 1.1.2. In how many ways six different textbooks can be planted on one book shelf?

Solution. $N = P_6 = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720.$

Definition. An arrangement of objects in which the order of the selection in not important is a **combination** and denoted by C_n^m.
It is read: combination of n digits taken m at a time.

Combinations are calculated by the formula

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m!(\mathrm{n - m})!}. \qquad (1.1.3)$$

Note. In practice it is better to use the next formula:

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

Example1.1.3. A standard deck of playing cards consists of 52 cards. How many five-card hands can be chosen from this deck?

Solution. We have $N = C_{55}^5 = \frac{52!}{5!(52-5)!} = \frac{55 \cdot 54 \cdot 53 \cdot 52 \cdot 51}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 2\,598\,960$. There are $2\,598\,960$ five- cards hands.

The following formulas are valid

1. $C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$ (consequence of the Newton binomial formula).
2. $C_n^m = C_n^{n-m}$ (symmetry property).
3. $C_n^n = 1,\ C_n^1 = n.$
4. $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$, where $0 \le m \le n$ (a recurrent formula).

§ 1.2. Events. Classification of events

Definition. Realization of some definite complex of conditions is called a **trial** (or **test**). A possible result of trial is called an **event**.

Definition. An event that can occur or can not occur in a given trial is termed a **random event**.

Definition. An event which can not come in a given trial is called an **impossible event**.

Definition. An event which necessarily comes in a given trial is termed a **certain event**.

Definition. Events $A_1, A_2, \ldots, A_n$ that can not occur at the same time are **mutually exclusive** or **incompatible**. Otherwise events are named **joint**.

Definition. Events $A_1, A_2, \ldots, A_n$ are said to be **uniquely possible** if one of them necessarily occurs in a given trial.

Definition. Events $A_1, A_2, \ldots, A_n$ form a **total group** if they are unique and incompatible.

Definition. Two events A and $\overline{A}$ which form a total group are called **mutually opposite**.

Definition. We will say that an event *B* implies an event *A* if in the given trial the occurrence of the event *B* implies that of the event *A*.

§ 1.3. Algebra of Events

Union of two events. A complex event which consists of the occurrence of at least one of the events *A* and *B* is called the union or the sum of two events *A* and *B*. The union of the events *A* and *B* is denoted $A \cup B$. For exclusive events *A* and *B* we also use the notation $A + B$.

The union of several event $A_1, A_2, \ldots, A_n$ is defined in the same way, and is denoted by $\underset{k}{\cup} A_k$, $k = 1, 2, ..., n$.

Intersection of two events. The joint appearance of two events *A* and *B* is called the intersection or the product of these two events and is denoted $A \cap B$ or *AB*.

By the intersection of several events $A_1, A_2, \ldots, A_n$ denoted by $\underset{k}{\cup} A_k$ $(k = 1, 2, ..., n)$ we mean the event consisting of the occurrence of all events.

Given two events *A* and *B*, by the **difference** $A - B$ we mean the event in which *A* occurs but not *B*.

Properties of unions and intersections. The operations of unions and intersections of events posses some properties which are similar to those of addition and multiplication of numbers.

1. The union and intersection of events are commutative: $A \cup B = B \cup A$, $AB = BA$.

2. The union and intersection of events are associative: $(A \cup B) \cup C = A \cup (B \cup C) = (A \cup C) \cup B = A \cup B \cup C$, $(AB)C = A(BC) = (AC)B = ABC$.

3. The union and *intersection* of events are distributive: $(A \cup B)C = AC \cup BC$.

All these properties follow directly from the definition of union and intersection. Thus, $(A \cup B)C$ means the joint occurrence of the event *C* with the event *A*, or with the event *B*, or with *A* and *B* together. The event $AC \cup BC$ also means the occurrence of either C together with A, or C together with B, or C together with AB.

But not all laws of addition and multiplication of numbers are valid for union and intersection of events. Thus, for instance, the events $A \cup A$ and AA evidently coincide with A. Therefore $A \cup A = AA = A$.

Complementary events. Non-occurrence of A, which is denoted by $\overline{A}$, is the event complementary to the event A.

It is easy to see that the event A is complementary to the event $\overline{A}$: $\overline{\overline{A}} = A$.

Gain and loss in a game, failure of a device in a given time interval and its faultless functioning in the same time interval, are examples of complementary events.

It is evident that complementary events are exclusive, and their union is a certain event:

$$A\overline{A} = \acute{\varnothing}, \qquad A \cup \overline{A} = \Omega,$$

where Ω is the certain event and $\acute{\varnothing}$ is the impossible event.
It is also clear that

$$A \cup \acute{\varnothing} = A, \quad A\acute{\varnothing} = \acute{\varnothing}, \quad A \cup \Omega = \Omega, \quad A\Omega = A.$$

§ 1.4. Classical Definition of Probability

Definition. The **probability** of an event A is defined by the formula

$$P(A) = \frac{m}{n}, \tag{1.4.1}$$

where n is the total number of outcomes uniquely possible, equiprobable and incompatible, m is the number of outcomes leading to the occurrence of the event A.

Example 1.4.1. In throwing a single unbiased die, there $n = 6$ mutually exclusive, equiprobable and unique outcomes, namely getting a number of spots equal to each of the numbers 1 through 6.

Solution. Let A be the event consisting of getting an even number of spots. Then there are $m = 3$ outcomes leading to the occurrence of A, and hence $P(A) = \frac{3}{6} = \frac{1}{2}$.

Example 1.4.2. From a lot of n items, k are defective. Find the probability that l items out of a random sample of size m selected for inspection are defective.

Solution. The number of possible ways to choose m items out of n is C_n^m. The favorable cases are those in which l defective items among the k defective items are selected (this can be done in C_k^l ways) and the remaining $m - l$ are nondefective, i.e., they are chosen from the total number $n - k$ (in C_{n-k}^{m-l} ways). Thus number of favorable cases is $C_k^l \cdot C_{n-k}^{m-l}$. The required probability will be

$$P(A) = \frac{C_k^l \cdot C_{n-k}^{m-l}}{C_n^m}.$$

Properties of Probabilities.

1. $0 \le P(A) \le 1$ for any event;

2. $P(\Omega) = 1$ for a certain event;

4. $P(Ø) = 0$ for an impossible event.

§ 1.5. Statistical Definition of Probability

Let n be the total number of experiments in the whole series of trials and m be a number of experiments in which A occurs. Then the ratio $W(A) = \frac{m}{n}$ is called the **relative frequency** of the event (in the given series of trials).

It turns out that the relative frequencies $\frac{m}{n}$ observed in different series of trials are virtually the same for large n, clustering about some constant

$$P(A) \sim \frac{m}{n}, \qquad (1.5.1)$$

called the probability of the event A.

More exactly, (1.5.1) means that

$$P(A) = \lim_{n \to \infty} \frac{m}{n}.$$

Roughly speaking, the probability $P(A)$ of the event A equals the fraction of experiments leading to the occurrence of A in a large series of trials.

Example (De Mere's paradox).

As a result of extensive observation of dice games, the French gambler de Mere noticed that the total number of spots showing on three dice thrown simultaneously turns out to be 11 (the event *A*) more often than it turns out to be 12 (the event *B*), although from his point of view both events should occur equally often. De Mere reasoned as follows: *A* occurs in just six ways (6:4:1, 6:3:2, 5:5:1, 5:4:2, 5:3:3, 4:4:3), and *B* also occurs in just six ways (6:5:1, 6:4:2, 6:3:3, 5:5:2, 5:4:3, 4:4:4). Therefore *A* and *B* have the same probability $P(A) = P(B)$.

The fallacy in this argument was found by Pascal, who showed that the outcomes listed by de Mere are not actually equiprobable. In fact, one must take account not only of the number of spots showing on the dice, but also of the particular dice on which the spots appear. For example, numbering the dice and writing the number of spots in the corresponding order, we find that there are six distinct outcomes leading to the combination 6:4:1, namely (6,4,1), (6,1,4), (4,1,6), (1,6,4) and (1,4,6), whereas there is only one outcome leading to the combination 4:4:4, namely (4,4,4). The appropriate equiprobable outcomes are those described by triples of numbers *(a,b,c)*, where *a* is the number of spots on the first die, *b* the number of spots one second die, and *c* the number of spots on the third die. It is easy to see that there are then precisely $n = 6^3 = 216$ equiprobable outcomes. Of these, $m(A) = 27$ are favorable to the event *A* (in which the sum of all the spots equals 11), but only $m(B) = 25$ are favorable to the event *B* (in which the sum of all the spots equals 12). To see this, note that a combination *a:b:c* occurs in 6 distinct ways if *a, b* and *c* are distinct, in 3 distinct ways if two (and only two) of the numbers *a, b* and *c* are distinct, and if only 1 way if *a* = *b* = *c*. Hence *A* occurs in 6+6+3+6+3+3=27 ways, while *B* occurs in 6+6+3+3+6+1=25 ways. This fact explains the tendency observed by de Mere for 11 spots to appear more often than 12.

§ 1.6. Dependent Events. Conditional Probability. Multiplication Theorem of Probabilities

Definition. Two events *A* and *B* are said to be **independent** if occurrence of one is in no way affected by the occurrence or nonoccurrence of the other. Otherwise event are called **dependent**.

Example 1.5.1. An urn contains 5 white and 3 black balls. At random from an urn a person draws a ball without returning it to the urn. The second ball is taken out of the urn. What is the probability that this ball is black (event *B*).

Solution. The occurrence of the event *B* depends on of what color was the first ball. If the first ball is white (event *A*), then the probability of the event *B* provided that the event *A* has come is

$$P_A(B) = P(B / A) = \frac{3}{7}.$$

If the first ball is black then the probability of the event *B* provided that the event *A* has not come (event $\overline{A}$ has come) is

$$P_{\overline{A}}(B) = P(B / \overline{A}) = \frac{2}{7}.$$

$P_A(B)$ is called the **conditional probability** of the event *B* relative to the event *A*. In this case the events *A* and *B* are **dependent** events.

Theorem. The probability of the product of two dependent events is equal to the probability of one of them multiplied by the conditional probability of the other

$$P(AB) = P(A)P_A(B) = P(B)P_B(A) \tag{1.6.1}$$

Proof. Let *n* be the total number of uniquely possible, equiprabable and incompatible outcomes; *m* a number of outcomes leading to the occurrence of the event *A*; *l* the number of outcomes leading to the occurrence of the event *B* and *k* the number of outcomes leading to the occurrence of both *A* and *B*. Then

$$P(AB) = \frac{k}{n} = \frac{km}{nm} = \frac{m}{n} \cdot \frac{k}{m},$$

but $\frac{m}{n} = P(A)$, $\frac{k}{m} = P_A(B)$ and hence

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

It follows from (1.6.1) that the probability of joint appearance of any number of events is equal to the probability of one of them multiplied by:
the conditional probability of another relative to the first event,
the conditional probability of the third event relative to the intersection of two first events, etc. by the conditional probability of the last event relative to the intersection of all preceding ones:

$$P(A_1 A_2 ... A_n) = P(A_1)\, P(A_2 / A_1)\, P(A_3 / A_1 A_2) \ldots P(A_n / A_1 A_2 ... A_{n-1}) \tag{1.6.2}$$

It is easy to see that conditional probabilities have properties analogous to those of ordinary probabilities. For example,

1. $0 \le P_A(B) \le 1$

2. If *A* and B are incompatible, so that *AB* = Ǿ, then $P_A(B) = 0$

3. If *A* implies, so that $A \subset B$ then $P_A(B) = 1$.

If the events A and B are independent then the probability for their product is defined by the formula

$$P(AB) = P(A)P(B) \tag{1.6.3}$$

Example1.6.2. The break in an electric circuit occurs when at least one out of three elements connected in series is out of order. Compute the probability that the break in the circuit will not occur if the elements may be out of order with the respective probabilities 0.3, 0.4 and 0.6. How does the probability change if the first element in never out of order?

Solution. The required probability equals the probability that all three elements are working. Let A_k $(k = 1,2,3)$ denote the event that the kth element functions. Then $p = P(A) = P(A_1 A_2 A_3)$. Since the events may be assumed independent, $p = P(A_1) = P(A_2) = P(A_3) = 0.7 \cdot 0.6 \cdot 0.4 = 0.168.$

If the first element is not out of order, then $p = P(A_2 A_3) = 0.24.$

Example1.6.3. A lot of 100 items undergoes a selective inspection. The entire lot us rejected if there is at least one defective item if five items checked. What is the probability that the given lot will be rejected if it contains 5 % defective items?

Solution. Find the probability q of the complementary event A consisting of the situation in which the lot will be accepted. The given event is an intersection of five events $A = A_1 A_2 A_3 A_4 A_5$, where A_k $(k = 1,2,3,4,5)$ means that the kth item checked is good.

The probability of the event A_1 is $P(A_1) = \frac{95}{100}$ since there are only 100 items, of which 95 are good. After the occurrence of the event A_1, there remain 99 items, of which 94 are good and, therefore, $P(A_2 / A_1) = \frac{94}{99}$. Analogously, $P(A_3 / A_1 A_2) = \frac{93}{98}$, $P(A_4 / A_1 A_2 A_3) = \frac{92}{97}$, $P(A_5 / A_1 A_2 A_3 A_4) = \frac{91}{96}$. According to the formula (1.6.2), we find that $q = \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{93}{98} \cdot \frac{92}{97} \cdot \frac{91}{96} = 0.77.$

The required probability $p = 1 - q = 0.23.$

§ 1.7. The Addition Theorem

Theorem. The probability of the sum of two events is equal to the sum of their probabilities minus the probability of their intersection, that is

$$P(A + B) = P(A) + P(B) - P(AB) \tag{1.7.1}$$

Proof. Let n be the total number of uniquely possible, equiprobable and incompatible outcomes; m the number of outcomes leading to the occurrence of the event A; l the number of outcomes leading to the occurrence of the event B and k the number of outcomes leading to the occurrence of both A and B. Then

$P(A+B)=\dfrac{m+l-k}{n}=\dfrac{m}{n}+\dfrac{l}{n}-\dfrac{k}{n}$, but $\dfrac{m}{n}=P(A)$, $\dfrac{l}{n}=P(B)$, $\dfrac{k}{n}=P(AB)$, and hence

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

Observe, that $k=0$ if the events A and B are incompatible.

Consequence 1. If the events A and B are incompatible then

$$P(A+B)=P(A)+P(B) \tag{1.7.2}$$

since $P(AB)=0$ for incompatible events.

Consequence 2. The sum of probabilities of events $A_1, A_2, \ldots, A_n$, which form a total set is equal to 1, that is

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1 \tag{1.7.3}$$

Since the events $A_1, A_2, \ldots, A_n$ form a total set, their sum is a certain event $\sum_{k=1}^{n} A_k=\Omega$. On using the formula (1.7.2) we can show that

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

and $P\left(\sum_{k=1}^{n} A_k\right)=P(\Omega)$ that is

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1, \text{ since } P(\Omega)=1.$$

Complementary events are incompatible and form a total set. There it follows from (1.7.3) that the sum of the probabilities of complementary events is unity:

$$P(A)+P(\overline{A})=1 \tag{1.7.4}$$

This formula is very important for practice. In many problems it is difficult to calculate the probability of an event while the probability of the complementary event may be easily calculated. In such cases formula (1.7.4) is useful.

Example. The scheme of the electric circuit between two points M and N is given in Figure 1. Malfunctions during an interval of time T of different elements of the circuit represent independent events with the following probabilities (Table 1)

Element	K_1	K_2	L_1	L_2	L_3
Probability	0.6	0.5	0.4	0.7	0.9

Table 1.7.1

Find the probability of a break in the circuit during the indicated interval of time.

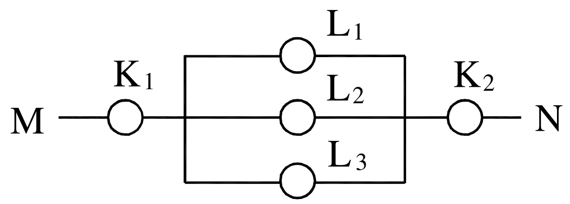

Fig.1.7.1

Solution. Denote by A_j $(j=1,2)$ the event meaning that an element K_j is out of order, by *A* that at least one element K_j is out of order and by *B* that all three elements L_i $(i=1,2,3)$ are out of order. Then, the required probability is

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

Since

$$P(A) = P(A_1) + P(A_2) - P(A_1)P(A_2) = 0.8,$$
$$P(B) = P(L_1)P(L_2)P(L_3) = 0.252,$$

we get $p \approx 0.85$.

§ 1.8. The Total Probability Formula

Suppose that an event *A* can occurs only together with one of events $H_1, H_2, \ldots, H_n$, which form a total group.

Events $H_1, H_2, \ldots, H_n$ are called **hypotheses**.

Probabilities $P(H_1), P(H_2), \ldots, P(H_n)$ are known and conditional probabilities $P_{H_i}(A), (i=1,2,\ldots,n)$ also are known. Find the probability $P(A)$.

Since the events $H_1, H_2, \ldots, H_n$ form a total group their union (sum) is a certain event.

The event *A* may appear only together with some event H_i. Thus the event *A* is the union of the events $AH_1, AH_2, \ldots, AH_n$. As by condition the events $H_1, H_2, \ldots, H_n$ are exclusive (incompatible) the events AH_1, $AH_2, \ldots$, AH_n are also exclusive and we can use the addition theorem

$$P(A) = P(AH_1) + P(AH_2) + \ldots + P(AH_n) = \sum_{k=1}^{n} P(AH_k).$$

On using the product theorem we obtain

$$P(AH_k) = P(H_k)P_{H_k}(A)$$

and therefore

$$P(A) = \sum_{k=1}^{n} P(H_k)P_{H_k}(A) \qquad (1.8.1)$$

Example. Electric bulbs are produced at two plants. Products of the first plant contains 70% standard bulbs, the second – 80%. The bulbs were sent for sale to shops. To the first shop were sent 60% of the general amount of bulbs, to the second – 40%.

What is the probability, that at random purchased bulb at the shop is a standard bulb?

Solution. Denote by *A* the event meaning that a standard bulb was bought, by H_1 – a bulb was produced at the first plant, by H_2 – a bulb was produced at the second plant.

We have $P(H_1) = 0.6$, $P(H_2) = 0.4$, $P_{H_1}(A) = 0.7$, $P_{H_2}(A) = 0.8$.

By the total probability formula we have

$$P(A) = P(H_1)\cdot P_{H_1}(A) + P(H_2)\cdot P_{H_2}(A) = 0.6 \cdot 07 + 0.4 \cdot 0.8 = 0.74.$$

§ 1.9. Bayes Formula

In practice we are often interested in a total group of incompatible events $H_1, H_2, \ldots, H_n$ whose probabilities $P(H_i)(i = 1,2,\ldots n)$ are known. These events are not observable but one may observe some event *A* whose conditional probabilities $P_{H_i}(A)(i = 1,2,\ldots n)$ are known. Assume that a trial was performed resulting in the appearance of the event *A*. Using this result of the trial it is required to make some inferences about the events $H_1, H_2, \ldots, H_n$, namely to determine their probabilities after the trial. In other words, it is necessary to find the conditional probabilities of the events $H_1, H_2, \ldots, H_n$ with respect to the event *A*.

From the probability multiplication theorem (1.6.1) follows

$$P(AH_k) = P(A)P_A(H_k) = P(H_k)P_{H_k}(A)$$

whence

$$P_A(H_k) = \frac{P(H_k)P_{H_k}(A)}{P(A)}.$$

Substituting the expression of the probability of the event *A* from the formula of total probability (1.8.1) we obtain

$$P_A(H_k)=\frac{P(H_k)P_{H_k}(A)}{\sum_{i=1}^{n}P(H_i)P_{H_i}(A)}\,(k=1,2,...,n) \qquad (1.9.1)$$

This formula which solves the problem is called Bayes formula.
The probabilities $P(H_k)(k=1,2,...n)$ of the events H_1, H_2,..., H_n before the trial are usually called **prior probabilities**, from the Latin **a priori**, which means "primary" i.e. in our case before the trial was performed. The probabilities $P_{H_k}(A)(k=1,2,...,n)$ of the same events after the trial are called **posterior** probabilities, from the Latin **a posteriori**, which means "after", i.e. after the trial was performed.

*Example.*A telegraphic communications system transmits the signals dot and dash. Assume that the statistical properties of obstacles are such that an average of $\frac{2}{5}$ of the dots and $\frac{1}{3}$ of the dashes are changed. Suppose that the ratio between the transmitted dots and the transmitted dashes is 5:3. What is the probability that a received signal will be the same as the transmitted signal if

a) the received signal is dot,
b) the received signal is a dash.

Solution. Let *A* be the event that a dot is received, and *B* that a dash is received. One can make two hypotheses: H_1 that the transmitted signal was a dot; and H_2 that the transmitted signal was a dash. By assumption, $\frac{P(H_1)}{P(H_2)}=\frac{5}{3}$.
Moreover, $P(H_1)+P(H_2)=1$. Therefore $P(H_1)=\frac{5}{8}$, $P(H_2)=\frac{3}{8}$.

One knows that

$$P_{H_1}(A)=\frac{3}{5},\ P_{H_2}(A)=\frac{1}{3},\ P_{H_1}(B)=\frac{2}{5},\ P_{H_2}(B)=\frac{2}{3}.$$

The probabilities of *A* and *B* are determined from the total probability formula:

$$P(A)=\frac{5}{8}\cdot\frac{3}{5}+\frac{3}{8}\cdot\frac{1}{3}=\frac{1}{2},\ P(B)=\frac{5}{8}\cdot\frac{2}{5}+\frac{3}{8}\cdot\frac{2}{3}=\frac{1}{2}.$$

The required probabilities are:

$$P_A(H_1)=\frac{P(H_1)P_{H_1}(A)}{P(A)}=\frac{\frac{5}{8}\cdot\frac{3}{5}}{\frac{1}{2}}=\frac{3}{4};$$

$$P_A(H_2)=\frac{P(H_2)P_{H_2}(B)}{P(B)}=\frac{\frac{3}{8}\cdot\frac{2}{3}}{\frac{1}{2}}=\frac{1}{2}.$$

CHAPTER II
RANDOM VARIABLES

§ 2.1. General Definitions. Discrete Random Variables

Definition. A **random variable** is a variable, which assumes as a result of a trial depending on casual circumstances.

Definition. **A random variable is said to be discrete** if it takes only a finite or countable infinite number of values.

Definition. **A random variable is said to be continuous** if it takes all values in some finite or infinite interval, that is values completely fill some interval (finite or infinite).

A set of possible values of a discrete random variable X and the probabilities of these values refers to as the law of the distribution of a discrete random variable. *DRV* is set by the table

X	x_1	x_2	$\dots$	x_n
P	p_1	p_2	$\dots$	p_n

Where x_i $(i = 1,2,...,n)$ a values of *DRV* X and $p_i = P(X = x_i)$ – corresponding probabilities.
$p_1 + p_2 + \dots + p_n = 1$, since the events $X = x_i$ form a total group.

Example2.1.1. A total number of lottery tickets is 100. This lottery distributes 1 prize worth 50 dollars, 10 prizes worth 1 dollar each. Find the law of the distribution of the *RV X* as a volume of a winning for one lottery ticket.

Solution.RV X can takes the following values: $x_1 = 50$, $x_2 = 1$, $x_3 = 0$.

Then $p_1 = P(X = x_1) = \frac{1}{100} = 0.01$,

$p_2 = P(X = x_2) = \frac{10}{100} = 0.1$,

$p_3 = P(X = x_3) = 1 - (p_1 + p_2) = 1 - 0.11 = 0.89$.

So, the law of the distribution of this *DRV* is

X	50	1	0
P	0.01	0.1	0.89

Table 2.1.1.

§ 2.2. Arithmetical Operations with *DRV*

Definition. Two ***DRV*** are called **independent** if the law of distribution of one of them does not depend on what possible values were accepted by other *DRV*.

Let independent *DRV* X and Y be denoted by the distribution tables

X	x_1	x_2	$\ldots$	x_m
P	p_1	p_2	$\ldots$	p_m

Y	y_1	y_2	$\ldots$	y_n
P	q_1	q_2	$\ldots$	q_n

Definition. By the **sum of two independent *DRV*** X and Y we mean a *DRV* which accepts all possible values $x_i + y_j$ which probabilities p_{ij} defined by the product theorem.

In the case of independent variables we have

$$p_{ij} = p_i\, q_j \tag{2.2.1}$$

Definition. By the **product of two independent *DRV*** X and Y we mean a *DRV* which accepts all possible values $x_i\, y_j$ with probabilities p_{ij} defined by the product theorem.

Definition. By **the *k*-th degree of a *DRV*** X we mean a *DRV* which takes values x_i^k with the same probabilities p_i.

Definition. By a product of a *DRV* X and a constant *A* we mean a *DRV* which accepts values Ax_i with the same probabilities.

Remark. For *DRV* $X \cdot X \neq X^2$, $X + X \neq 2X$

§ 2.3. Numerical Characteristics of *DRV*

In what follows we consider such numerical characteristics:

- **A.** Mathematical expectation or mean value.
- **B.** Dispersion or variance.
- **C.** Mean square deviation.

A. Mathematical Expectation and its Properties

Definition. By the mathematical expectation of a *DRV X* we mean the sum of products of all its values and corresponding probabilities, that is

$$M(X) = \sum_{i=1}^{n} x_i p_i \tag{2.3.1}$$

Let us calculate a mathematical expectation of the *DRV* and the average win for one ticket in the example of § 2.1 (Table 2.1.1.)

$$M(X) = 1 \cdot \frac{10}{100} + 50 \cdot \frac{1}{100} + 0 \cdot \frac{89}{100} = 0.6.$$

Average win is $\frac{10+50}{100} = 0.6.$

So we can see that the probability meaning of a mathematical expectation is the mean value of a random variable.

Using the formula (2.3.1) it is possible to prove the following properties of mathematical expectation (*ME*).

1. $M(C) = C$,
2. $M(CX) = C\,M(X)$ for an arbitrary constant *C*;
3. $M(X \pm Y) = M(X) \pm M(Y)$ for arbitrary random variables *X* and *Y*;
4. $M(XY) = M(X) \cdot M(Y)$ for independent random variables *X* and *Y,*
5. $M(X - M(X)) = 0.$

B. Dispersion and its Properties

Definition. By the dispersion of a random variable X denoted by $D(X)$ we mean the mean square value $M(X-M(X))^2$ of the difference $X-M(X)$.

So

$$D(X)=M(X-M(X))^2 \qquad (2.3.2)$$

Theorem. The dispersion also can be found by the formula

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

Proof. $D(X)=M(X-M(X))^2=M(X^2-2M(X)\cdot X+M^2(X))=$
$=M(X^2)-2M(X)\cdot M(X)+M(M(X^2))=M(X^2)-M^2(X).$

The following properties of a dispersion are valid.

1. A dispersion of a constant is equal to zero: $D(C)=0$.

2. $D(CX)=C^2D(X)$ for an arbitrary constant C.

3. $D(X\pm Y)=D(X)+D(Y)$ for independent random variables *X* and *Y*.

4. $D(XY)=M(X^2)M(Y^2)-M^2(X)M^2(Y)$ for independent random variables *X* and *Y*.

C. *Definition.* By mean **square deviation** we mean

$$\sigma(X)=\sqrt{D(X)}. \qquad (2.3.3)$$

Example. Find the mean square deviation of the data in the following table

X	-2	-1	0	1	2
P	0.1	0.1	0.5	0.1	0.2

For *DRV* X^2 we get

X^2	4	1	0	1	4
P	0.1	0.1	0.5	0.1	0.2

Hence $M(X)=0.2$, $M(X^2)=1.4$ and

$D(X)=M(X^2)-M^2(X)=1.4-0.04=1.36,$ so

$\sigma(X)=\sqrt{D(X)}=\sqrt{1.36}.$

§ 2.4. Bernoulli Trials. The Binomial and Poisson Distribution

By Bernoulli trials we mean identical independent experiments in each of which an event *A*, say, may occur with probability

$$p=P(A)\ (p\neq 0)$$

or fail to occur with probability

$$q=1-p.$$

Occurrence of the event *A* is called a "**success**" and nonoccurrence of *A* (i.e., occurrence of the complementary event $\overline{A}$) is called "**failure**".

Suppose that *n* independent trials are performed in every one of which the probability of the event *A* is equal to *p*. It is required to find the probability $P_n(m)$ that the event *A* will appear *m* times.

Example 2.4.1. A marksman fires 5 shots at a target. The probability of hitting the target is one and the same $p=0.8$ for all shots. Find the probability that the target is hit one time, two times, …, five times.

Solution. $P_5(0)=P(\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot\overline{A})=(0.2)^5$,

$$P_5(1)=P(A\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}+\overline{A}\cdot A\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}+...+\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot A)=$$
$$=5\cdot 0.8\cdot(0.2)^4,$$

$$P_5(2)=P(A\cdot A\cdot\overline{A}\cdot\overline{A}\cdot\overline{A}+...+\overline{A}\cdot\overline{A}\cdot\overline{A}\cdot A\cdot A)=10\cdot(0.8)^2\cdot(0.2)^3.$$

Hence we can write

$$P_5(0)=C_5^0 q^5,\ P_5(1)=C_5^1 pq^4,\ P_5(2)=C_5^2 p^2 q^3, \ldots,\ P_5(5)=C_5^5 p^5.$$

Generalize these formulas we get

$$P_n(m)=C_n^m p^m q^{n-m}, \qquad (2.4.1)$$

where

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

The formula (2.4.1) is called Bernoulli formula and corresponding probability distribution of the random variable *m* given by the formula (2.4.1) is known as the **binomial distribution**. The binomial distribution is specified by two parameters, the probability *p* of a single success and the number of trials *n*. Suppose the number of trials is large while the probability of success *p* is relatively small, so that each success is a rather rare event while the average number of successes *np* is appreciable. Then it is a good approximation to write

$$P_n(m)\approx \frac{a^m e^{-a}}{m!}, \text{ where } a=np \qquad (2.4.2)$$

A random variable *m* taking only the integral values 0, 1, 2,… is said to be a **Poisson distribution** if $P_n(m)=\frac{a^m e^{-a}}{m!}$.

Example 2.4.2. How many lottery tickets must be bought to make the probability of winning at least *P* ?

Solution. Let *N* be the total number of lottery tickets and *M* the total number of winning tickets. Then $\frac{M}{N}$ is the probability that a bought ticket is one of the winning tickets. The purchase of each ticket can be regarded as a separate trial with probability of "success" $p=\frac{M}{N}$ in a series of *n* independent trials, where *n* is the number of tickets bought. If *p* is relatively small, as is usually the case, and the given probability *P* is relatively large, then it is clear that a rather large number of tickets must be bought to make the probability of buying at least one winning ticket no smaller than *P*. Hence the number of winning tickets among those purchased is a random variable with an approximately Poisson distribution, i.e. the probability that there are precisely *m* winning tickets among *n* purchased tickets is

$$P_n(m) \approx \frac{a^m e^{-a}}{m!}, \text{ where } a = n\frac{M}{N}.$$

The probability that at least one of the tickets is a winning ticket is just

$$1 - P_n(0) = 1 - e^{-a}.$$

Hence n must be at least as large as the smallest positive integer satisfying the inequality

$$e^{-a} = e^{-\frac{nM}{N}} \le 1 - P.$$

§ 2.5. The Local Laplace Theorem

Suppose the number of trials n is large and m is also large and determining the binomial distribution become very cumbersome. In this case we make use of Laplace formula

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x), \quad \text{where}$$

$$\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \text{ – Gauss function and } x = \frac{m - np}{\sqrt{npq}}.$$

Gauss function $\varphi(x)$ satisfies the following conditions:

1. $\varphi(x) > 0$ for all x;

2. function $\varphi(x)$ is even function, that is $\varphi(-x) = \varphi(x)$;

3. $\lim\limits_{x \to \infty} \varphi(x) = 0$. Practically for $|x| > 4$ $\varphi(x) \approx 0$.

Example 2.5.1. Given $p = 0.6, n = 15.$ Find $P_{15}(7)$.

Solution. $q = 1 - p = 0.4,\ np = 15 \cdot 0.6 = 9,\ x = \frac{m - np}{\sqrt{nmp}} = -1.05,$

$$P_{15}(7) \approx \frac{1}{1.8970} \cdot \frac{1}{\sqrt{6.28}} \cdot e^{\frac{-(1.05)^2}{4}} = \frac{0.2290}{1.8970} \approx 0.1207.$$

By Bernoulli formula we get

$$P_{15}(7) = \frac{15!}{7! \cdot 8!}(0.6)^7 \cdot (0.4)^8 = 0.1183.$$

§ 2.6. Integral Function of Distribution

*Definition.*The probability of the inequality $X < x$ considered as a function of the variable x is called the **distribution integral function** of the random variable X that is

$$F(x) = P(X < x) \qquad (2.6.1)$$

Example 2.6.1. Find the integral function of the following distribution of a random variable and form a graph of this function.

X	0	1	3
P	0.2	0.3	0.5

Solution. Using the formula (2.6.1) we obtain

$$F(x) = \begin{cases} 0, & \text{if } x \le 0 \\ 0.2, & \text{if } 0 < x \le 1 \\ 0.5, & \text{if } 1 < x \le 3 \\ 1, & \text{if } x > 3 \end{cases}$$

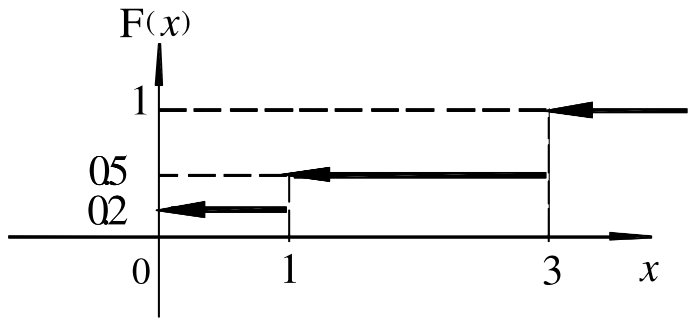

Fig. 2.6.1

*Conclusion.*In the case of *DRV* a graph of integral function of distribution is represented in a staircase form.

Definition. A random variable is called **continuous** if its integral function of distribution is continuous for all *x*.

Now we consider the properties of an integral function of distribution.

1. $0 \le F(x) \le 1;$

2. $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=\lim\limits_{x\to-\infty}P(X<x)=0,$

$$F(-\infty)=\lim_{x\to-\infty}F(x)=\lim_{x\to-\infty}P(X<x)=0,$$

3. An integral distribution function is a non-decreasing function of *x*.

Proof. Let $x_1 < x_2$ then
$$P(X<x_2)=P(X<x_1)+P(x_1\le X<x_2),$$
but
$$F(x_2)=P(X<x_2),\ \ F(x_1)=P(X<x_1),$$
hence
$$F(x_2)-F(x_1)=P(x_1\le X<x_2). \qquad (2.6.2)$$
A probability of any event is non-negative, so

$$F(x_1)\le F(x_2).$$

Rewrite the equality (2.6.2) in such a form to have

$$P(x_1\le X<x_2)=F(x_2)-F(x_1) \qquad (2.6.3)$$

Thus, the probability of the occurrence of a random variable in a given interval is equal to the increment of its distribution function on this interval.

4. The probability of occurrence of a continuous random variable in a specific point is equal to zero.

Proof. On setting in the formula (2.6.3) $x_2 = x_1 + \Delta x$ we get

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

Let $\Delta x \to 0$ then

$\lim\limits_{\Delta x\to0}P(x_1\le X<x_2)=\lim\limits_{\Delta x\to0}(F(x_1+\Delta x)-F(x_1))=\lim\limits_{\Delta x\to0}\Delta F(x_1)=0$ since a function $F(x)$ is continuous.

*Corollary.*The following equalities are valid for a continuous random variable

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

§ 2.7. Differential Distribution Function

Definition. By **differential distribution function (density)** we mean a derivative of an integral distribution function, that is

$$\varphi(x) = F'(x) \tag{2.7.1}$$

A density $\varphi(x)$ satisfies the following conditions:

1. $\varphi(x) \geq 0$ since $\varphi(x) = F'(x)$ and $F(x)$ is a non-decreasing function.
2. $P(a < X < b) = \int_a^b \varphi(x)dx.$ (2.7.2)

Since $F'(x) = \varphi(x)$ then $F(x)$ is an antiderivative of a function $\varphi(x)$. Using the Newton-Leibnitz formula we have

$$P(a < X < b) = F(b) - F(a) = \int_a^b \varphi(x)dx.$$

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

Really

$$\int_{-\infty}^{+\infty} \varphi(x)dx = P(-\infty < X < +\infty) = 1.$$

Corollary 2. $P(a < X < b) = \varphi(c) \cdot (b - a),$ where $c \in [a, b].$

3. Let a density $\varphi(x)$ be given. Find $F(x)$.

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

$P(-\infty < X < x) = \int_{-\infty}^{x} \varphi(x)dx.$

From this it follows

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx. \tag{2.7.3}$$

§ 2.8. Numerical Characteristics of Continuous Random Variables

Let a continuous random variable *X* be given by distribution density $\varphi(x)$. Take its possible values in a closed interval $[a, b]$.

Divide the interval $[a, b]$ into *n* subintervals $\Delta x_1, \Delta x_2, ..., \Delta x_n$ and choose points x_i $\left(i = \overline{1, n}\right)$ between *a* and *b*.

Define *ME* by analogy with that of the case of *DRV* in such a way $P(x_i < X < x_i + \Delta x_i) \approx \varphi(x_i)\Delta x_i$:

$$\sum_{i=1}^{n} x_i \varphi(x_i)\Delta x_i \ . \tag{2.8.1}$$

Passing to the limit, when $\max_i \Delta x_i \to 0$ we get

$$\lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} x_i \varphi(x_i) \Delta x_i = \int_a^b x\varphi(x)dx \qquad (2.8.2)$$

Definition. By the **mathematical expectation** of a continuous random variable we mean the definite integral

$$M(X) = \int_a^b x\varphi(x)dx \qquad (2.8.3)$$

If values of a continuous random variable belong to infinite interval $(-\infty, +\infty)$ then a mathematical expectation is defined by the improper integral

$$M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx \qquad (2.8.4)$$

Since by the dispersion of a random variable we mean the mean square value $M(X - M(X))^2$ of the difference $X - M(X)$ then

$$D(X) = M(X - M(X))^2 = \int_{-\infty}^{+\infty} (x - M(X))^2 \varphi(x)dx \qquad (2.8.5)$$

If values of a continuous random variable belong to the close interval $[a, b]$ then

$$D(X) = \int_a^b (x - M(X))^2 \varphi(x)dx \qquad (2.8.6)$$

Example 2.8.1. Given integral distribution function $F(x)$:

$$F(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \dfrac{x}{2} & \text{if } 0 < x \le 2, \\ 1 & \text{if } x > 2. \end{cases}$$

Find $M(X)$, $D(X)$ and $\sigma(X)$.

Solution. First of all we find the density distribution function $\varphi(x) = F'(x)$.

$$\varphi(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \dfrac{1}{2} & \text{if } 0 < x \le 2, \\ 0 & \text{if } x > 2. \end{cases}$$

$$M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx = \int_{-\infty}^{0} x \cdot 0dx + \int_0^2 x \cdot \frac{1}{2}dx + \int_2^{+\infty} x \cdot 0dx =$$

$$= \frac{1}{2}\int_0^2 xdx = \left.\frac{x^2}{4}\right|_0^2 = 1.$$

$$D(X)=\int_{-\infty}^{+\infty}(x-1)^2\varphi(x)dx = \int_0^2 (x-1)^2\cdot\frac{1}{2}dx = \left.\frac{(x-1)^3}{6}\right|_0^2=\frac{1}{3}.$$

$$\sigma(X)=\sqrt{D(X)}=\frac{1}{\sqrt{3}}.$$

§ 2.9. A Uniform Distribution

The distribution of a continuous random variable whose density is constant in some closed interval $[a,b]$ and is equal to zero outside this interval is called a uniform distribution:

$$\varphi(x)=\begin{cases}\frac{1}{b-a}, & \text{if } x\in[a,b],\\ 0, & \text{if } x\notin[a,b]\end{cases}$$

that is $\varphi(x)=c$ for $x\in[a,b]$ but

$$M(X)=\int_{-\infty}^{+\infty}cdx=1,$$

and hence $\left.cx\right|_a^b=1$, $c=\frac{1}{b-a}$. Instead of a closed interval $[a,b]$ it is possible to take (a,b), or $[a,b)$, $(a,b]$ since a random variable is continuous.

The graph of the density $\varphi(x)$ has the form

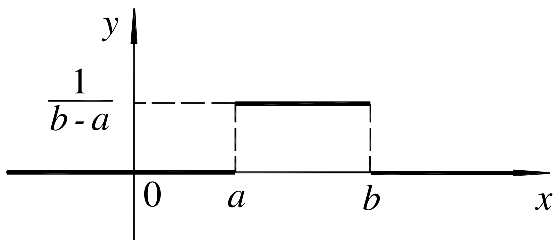

Fig.2.9.1

A uniform distribution is characteristic for the phase of random oscillations. In practice we have to consider harmonic oscillations with random amplitude and phase. In such cases the phase is often a random variable uniformly distributed over the period of oscillations.

Let us find the integral distribution function $F(x)$.

Since $F(x)=\int_{-\infty}^{x}\varphi(t)dt$, then

1. if $x<a$ we have

$$\int_{-\infty}^{x}\varphi(t)dt = \int_{-\infty}^{x}0dt = 0,$$

2. if $a \le x \le b$ we obtain

$$\int_{-\infty}^{x}\varphi(t)dt = \int_{-\infty}^{a}\varphi(t)dt + \int_{a}^{x}\varphi(t)dt = \int_{-\infty}^{a}0\cdot dt + \int_{a}^{x}\frac{dt}{b-a} = \frac{x-a}{b-a},$$

3. if $x > b$ we get

$$\int_{-\infty}^{x}\varphi(t)dt = \int_{-\infty}^{a}\varphi(t)dt + \int_{a}^{b}\varphi(t)dt + \int_{b}^{x}\varphi(t)dt =$$

$$= \int_{-\infty}^{a}0dt + \int_{a}^{b}\frac{dt}{b-a} + \int_{b}^{x}0dt = \frac{b-a}{b-a} = 1.$$

Therefore

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

The graph of this function is of the form:

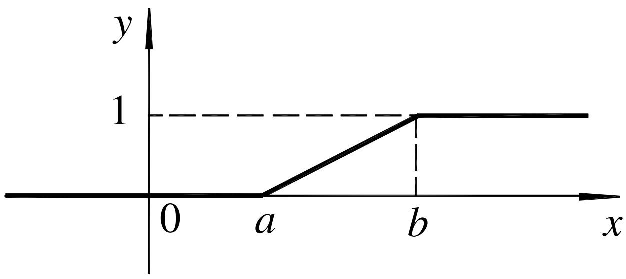

Fig.2.9.2

Let us define $M(X)$ and $D(X)$.

$$M(X) = \int_{-\infty}^{+\infty}x\varphi(x)dx = \int_{a}^{b}\frac{x\cdot dx}{b-a} = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{a+b}{2}.$$

$$D(X) = \int_{-\infty}^{+\infty}(x-M(X))^2\varphi(x)dx = \int_{a}^{b}\left(x-\frac{a+b}{2}\right)^2\frac{dx}{b-a} =$$

$$= \frac{1}{3(b-a)}\left.\left(x-\frac{a+b}{2}\right)^3\right|_a^b = \frac{(b-a)^2}{12}.$$

Example 2.9.1. Find the probability that a random variable *X* assumes values on the closed interval $[\alpha, \beta] \subset [a, b]$.

Solution. $P(\alpha \le x \le \beta) = \int_{\alpha}^{\beta} \varphi(x) dx = F(\beta) - F(\alpha) = \frac{\beta - a}{b - a} - \frac{\alpha - a}{b - a} = \frac{\beta - \alpha}{b - a}$.

§ 2.10. An Exponential Distribution

An exponential distribution is determined by the density:

$$\varphi(x) = \begin{cases} 0, & \text{if } x < 0, \\ ke^{-kx}, & \text{if } x \ge 0, \end{cases} \quad \text{where } k>0 \qquad (2.10.1)$$

The graph of this function is shown in *Fig.2.10.1.*

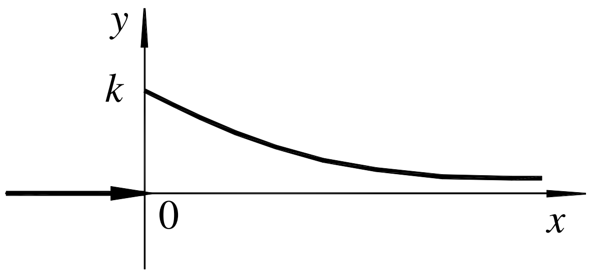

Fig.2.10.1

The integral distribution function $F(x)$ is of the form

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

and is shown in *Fig.2.10.2*.

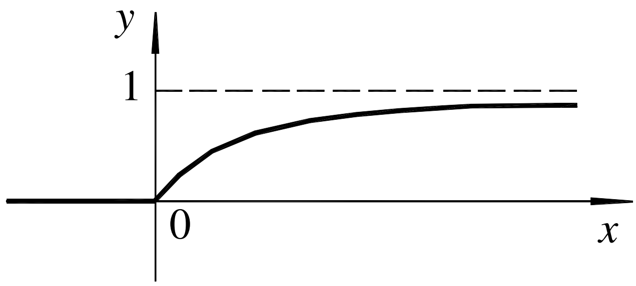

Fig.2.10.2.

Now we find the mathematical expectation and the dispersion

$$M(X)=\int_0^{+\infty} xke^{-kx}dx=\lim_{b\to+\infty}\int_0^{b} xke^{-kx}dx=\begin{bmatrix}\text{integrating}\\ \text{by parts}\end{bmatrix}=$$

$$=\lim_{b\to+\infty}\left(-xe^{-kx}-\frac{1}{k}e^{-kx}\right)\Big|_0^b=0-\frac{1}{k}(0-1)=\frac{1}{k}.$$

$$D(X)=\int_{-\infty}^{+\infty} x^2e^{-kx}dx-M^2(X)=k\int_0^{+\infty} x^2e^{-kx}dx-\frac{1}{k^2}=\begin{bmatrix}\text{integrating by}\\ \text{parts twice}\end{bmatrix}=$$

$$=k\lim_{b\to+\infty}\left(-\frac{x^2}{k}e^{-kx}+\frac{2}{k}\left(-\frac{x}{k}e^{-kx}-\frac{1}{k^2}e^{-kx}\right)\right)\Big|_0^b-\frac{1}{k^2}=$$

$$=k\left(0-\frac{2}{k}\left(0+0-\frac{1}{k^2}(0-1)\right)\right)-\frac{1}{k^2}=\frac{2}{k^2}-\frac{1}{k^2}=\frac{1}{k^2}.$$

Thus

$$M(X)=\frac{1}{k},\ D(X)=\frac{1}{k^2},\sigma(X)=\frac{1}{k}. \qquad (2.10.2)$$

Example 2.10.1. Let T be a time period of operating a radio bulb, which has an exponential distribution. Find the probability of operating a radio bulb not less than 800 hours if an average time period is 400 hours.

Solution.

$$M(T)=400,\ k=\frac{1}{400},\ P(T\ge 800)=1-P(T<800)=1-F(800)=$$

$$=1-\left(1-e^{-\frac{800}{400}}\right)=e^{-2}\approx 0.135.$$

§ 2.11. A Normal Distribution

A normal distribution is determined by the density:

$$\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}},\ x\in R, \qquad (2.11.1)$$

in which a is the mathematical expectation and σ is the mean square deviation.

Consider some properties of the function $\varphi(x)$ given by the formula (2.11.1)

1. $\varphi(x)>0$ for all $x\in R$.
2. The function $\varphi(x)$ has the maximum at $x=a$, equals $\frac{1}{\sigma\sqrt{2\pi}}$. Really $\varphi'(x)=-\frac{x-a}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. Therefore $\varphi'(x)=0$ if $x=a$, hence $\varphi'(x)>0$ for $x<a$, and $\varphi'(x)<0$ for $x>a$,. This means that $x=a$ is the point of maximum and $\varphi_{\max}=\varphi(a)=\frac{1}{\sigma\sqrt{2\pi}}$.
3. Calculating the second derivative of $\varphi(x)$ we can verify that the points $x=a\pm\sigma$ are points of inflexion.
4. The graph of the function $\varphi(x)$ is symmetric with respect to the vertical straight line $x=a$.
5. $\lim_{x\to\infty}\varphi(x)=0,$ that is OX axes is a horizontal asymptote of the function $\varphi(x)$.

   Using these properties we can sketch graph of the function $\varphi(x)$.

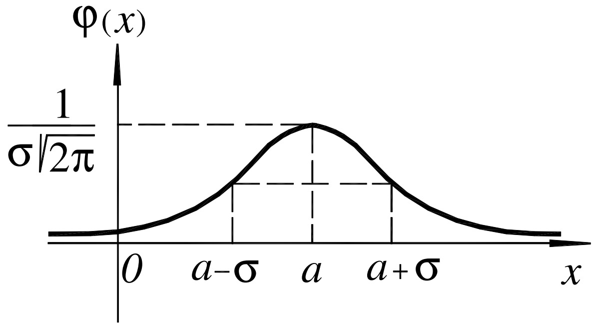

Fig.2.11.1

The graph of the density $\varphi(x)$ is the “bell-shaped” curve.

The integral distribution function $F(x)$ we find by the formula:

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

Setting $\frac{x-a}{\sigma}=t$; $x=a+\sigma t$, $dx=\sigma dt$ we obtain

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\frac{x-a}{\sigma}}e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt+\int_{0}^{(x-a)/\sigma}e^{-\frac{t^2}{2}}dt\right).$$

But

$$\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt=\frac{1}{2}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\frac{\sqrt{2\pi}}{2}\quad\text{(Poisson integral)}$$

and using Laplace function

$$\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt \tag{2.11.2}$$

we finally have

$$F(x)=\frac{1}{2}\left(1+\Phi\left(\frac{x-a}{\sigma}\right)\right). \tag{2.11.3}$$

Now let us derive a formula for the probability of the occurrence of a normally distributed random variable *X* in a given closed interval $[x_1,x_2]$.

We get

$$P(x_1\le X\le x_2)=F(x_2)-F(x_1)=\frac{1}{2}\left(1+\Phi\left(\frac{x_2-a}{\sigma}\right)\right)-$$

$$-\frac{1}{2}\left(1+\Phi\left(\frac{x_1-a}{\sigma}\right)\right)=\frac{1}{2}\left(\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right)\right). \tag{2.11.4}$$

In the special case when the interval $[x_1,x_2]$ is symmetrical with respect to the expectation a, $x_1=a-\varepsilon$, $x_2=a+\varepsilon$, the formula (2.11.4) gives

$$P(a-\varepsilon\le X\le a+\varepsilon)=P(|X-a|\le\varepsilon)=\frac{1}{2}\left(\Phi\left(\frac{\varepsilon}{\sigma}\right)-\Phi\left(-\frac{\varepsilon}{\sigma}\right)\right).$$

Hence taking into account the oddness of the function $\Phi(x)$ we obtain

$$P(|X-a|\le\varepsilon)=\Phi\left(\frac{\varepsilon}{\sigma}\right). \tag{2.11.5}$$

Example 2.11.1. The measurement of the distance to a certain object is accompanied by systematic and random errors. The systematic error equals 50 m. in the direction of decreasing distance. The random errors obey the normal distribution law with the mean square deviation $\sigma=100\,m$. Find

(1) the probability of measuring the distance with an error not exceeding 150 *m*. in absolute value,

(2) the probability that the measured distance does not exceed the actual one.

Solution. Let X denote the total error made in measuring the distance. Its systematic component is a = - 50 *m.* Consequently, the probability density of the total errors has the form

$$\varphi(x) = \frac{1}{100\sqrt{2\pi}} \exp\left(-\frac{(x+50)^2}{20.000}\right).$$

(1) According to the formula (2.11.4), we have

$$P(|X| < 150) = P(-150 < X < 150) = \frac{1}{2}\left(\Phi\left(\frac{150+50}{100}\right) - \Phi\left(\frac{-150+50}{100}\right)\right) =$$

$$= \frac{1}{2}(\Phi(2) - \Phi(-1)).$$

The probability integral $\Phi(x)$ is an odd function and hence, $\Phi(-1) = -\Phi(1)$. From this we get

$$P(|X| < 150) = \frac{1}{2}(\Phi(2) + \Phi(1)).$$

From the table list, we find $\Phi(2) = 0.9545$, $\Phi(1) = 0.6827$; and finally, $P(|X| < 150) = 0.8186$.

(2) The probability that the measured distance will not exceed the actual one is

$$P(-\infty < X < 0) = \frac{1}{2}(\Phi(0.5) + \Phi(\infty)).$$

Since $\Phi(\infty) = \lim_{x\to\infty} \Phi(x) = 1$ and from the table list we find $\Phi(0.5) = 0.3829$, it follows that $P(-\infty < X < 0)$ = 0.6914.

CHAPTER III
The Law of Large Numbers and Limit Theorem of the Theory of Probability

§ 3.1. The Law of Large Numbers

Markov's inequality. If a random variable *X* accepts only non-negative values and has a finite mathematical expectation, then for any positive number α the following inequality holds

$$P(X \geq \alpha) \leq \frac{M(X)}{\alpha} \qquad (3.1.1)$$

Proof. Since *X* accepts only non-negative values, then $M(X) \geq \alpha$ and

$$M(X) = \int_{0}^{+\infty} x\varphi(x)dx \geq \int_{\alpha}^{+\infty} x\varphi(x)dx \geq \alpha \int_{\alpha}^{+\infty} \varphi(x)dx = \alpha P(X \geq \alpha).$$

Therefore $P(X \geq \alpha) \leq \dfrac{M(X)}{\alpha}$.

Corollary3.1.1. $P(X < \alpha) > 1 - \dfrac{M(X)}{\alpha}$.

Really, since the events $X \geq \alpha$ and $X < \alpha$ are complementary, then $P(X \geq \alpha) + P(X < \alpha) = 1,$

hence

$$P(X < \alpha) = 1 - P(X \geq \alpha) > 1 - \frac{M(X)}{\alpha}.$$

Chebyshev's inequality. If a random variable *X* has a finite dispersion, then for any $\varepsilon > 0$ the following inequality holds:

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2} \qquad (3.1.2)$$

Proof. On applying Markov's inequality to the random variable $(X - M(X))^2$ and taking $\alpha = \varepsilon^2$ we get

$$P\left((X - M(X))^2 \geq \varepsilon^2\right) \leq \frac{M(X - M(X))^2}{\varepsilon^2} = \frac{D(X)}{\varepsilon^2},$$

since the inequality $(X - M(X))^2 \geq \varepsilon^2$ is equivalent to the inequality $|X - M(X)| \geq \varepsilon.$

Corollary3.1.2. $P(|X - M(X)|) < \varepsilon > 1 - \dfrac{D(X)}{\varepsilon^2}.$

Since the events $|X - M(X)| \geq \varepsilon$ and $|X - M(X)| < \varepsilon$ are complementary we have

$P(|X-M(X)|\geq \varepsilon)+P(|X-M(X)|<\varepsilon)=1.$ Whence

$P(|X-M(X)|<\varepsilon)=1-P(|X-M(X)|\geq \varepsilon)=1-\frac{D(X)}{\alpha^2}.$

Now we consider some special cases of Chebyshev's inequality.

Let p be a probability of some event A in n repeated independent trials, m is a frequency of the event A; $\frac{m}{n}$ is relative frequency.

1). For a random variable $X = m$ having a binomial law of distribution we have

$M(X) = M(m)=np$, $D(X) = D(m)=npq$:

$$P(|m-np|<\varepsilon)>1-\frac{npq}{\varepsilon^2}. \tag{3.1.3}$$

2). For a random variable $X = \frac{m}{n}$ having a binomial law of distribution we get

$M(X) = M\left(\frac{m}{n}\right)=p$, $D(X) = D\left(\frac{m}{n}\right)=\frac{pq}{n}$:

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)>1-\frac{pq}{n\varepsilon^2}. \tag{3.1.4}$$

Chebyshev's theorem. If X_1, X_2, …, X_n,… is a sequence of random variables, pairwise independent with mathematical expectations a_1, a_2,…, a_n,… whose dispersions are bounded by the constant $D(X_k)\leq C, k=1,2,...,$ then for any constant $\varepsilon>0$,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\leq\varepsilon\right)=1 \tag{3.1.5}$$

Proof. Consider the random variable $\overline{X_n}=\frac{X_1+X_2+...+X_n}{n}$.

Since X_k $(k=1,2,...n)$ are independent random variables then by the corresponding properties of mathematical expectation and dispersion we have

$M(\overline{X_n})=M\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{a_1+a_2+...+a_n}{n}$;

$D(\overline{X_n})=\frac{1}{n^2}(D(X_1)+D(X_2)+...+D(X_n)).$

Apply Chebyshev's inequality to the random variable $\overline{X}_n$ to obtain

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\leq\varepsilon\right)\geq$$
$$\geq 1-\frac{D(X_1)+D(X_2)+...+D(X_n)}{n^2\varepsilon^2}.$$

Since $D(X_k) \le C\,(k = 1,2,...,n)$, then

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{a_1 + a_2 + ... + a_n}{n}\right| \le \varepsilon\right) \ge 1 - \frac{C}{n^2\varepsilon^2},$$

$$\lim_{n\to\infty}\left(1 - \frac{C}{n^2\varepsilon^2}\right) = 1.$$

The theorem has been proved.

Bernoulli's theorem. Let p be a probability of some event A in n repeated independent trials, m is a frequency of the event A, then for any constant $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1 \qquad (3.1.6)$$

Proof. Passing in the inequality (3.1.4) to a limit as $n \to \infty$ we arrive at the formula (3.1.6).

Lyapunov's central limit theorem. If X_1, X_2, …, X_n are independent random variables with mathematical expectation $a_k\ (k = 1,2,...,n)$ and dispersion $D(X_k)(k = 1,2,...,n)$, and also $|X_k - a_k \le \delta|\,(k = 1,2,...,n)$, and dispersions are bounded by one and the same number, that is $D(X_k) \le C, (k = 1,2,...,n)$. Then for $n \to \infty$ the sum $\sum_{k=1}^{n} X_k$ infinitely approaches the normal distribution with mathematical expectation $\sum_{k=1}^{n} a_k$ and dispersion $\sum_{k=1}^{n} \sigma_k{}^2$.

This theorem we accept without proof.

Corollary 3.1.1. If random variables $X_k\ (k = 1,2,...,n)$ are equally distributed then the law of distribution of their sum as $n \to \infty$ approaches the normal law of distribution.

Corollary 3.1.2. If X_1, X_2, …, X_n satisfy conditions of central limit theorem, then applying the formula (2.11.4) to their sum $\sum_{k=1}^{n} X_k$ we obtain the approximate formula

$$P\left(\alpha \le \sum_{k=1}^{n} X_k \le \beta\right) \approx \frac{1}{2}\left(\Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)\right). \qquad (3.1.7)$$

Moivre-Laplace theorem. Let p be a probability of some event A in n repeated independent trials, m is a frequency of the event A, then the following approximate formula is valid

$$P(m_1 \le m \le m_2) \approx \frac{1}{2}\left(\Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)\right) \qquad (3.1.8)$$

Proof. A random variable m can be represented in the form $m = \sum_{k=1}^{n} X_k$, where $X_k = 1$ if the event A occurs in the kth trial, and $X_k = 0$ if it does not occur.

Since $M(X_k) = p,\ D(X_k) = pq,$ then

$$M(m) = M\left(\sum_1^n X_k\right) = np,\ D(m) = D\left(\sum_1^n X_k\right) = npq.$$

On substituting these values in the formula (3.1.7) we obtain

$$P(m_1 \le m \le m_2) \approx \frac{1}{2}\left(\Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)\right).$$

This theorem can be used for the determination of a probability of a deviation of the frequency m from the mathematical expectation np and also for the probability of the deviation of the relative frequency $\frac{m}{n}$ from the probability p.

In the first case we have

$$P(|m - np| \le \varepsilon) = P(np - \varepsilon \le m \le np + \varepsilon) \approx$$

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

Thus

$$P(|m - np| \le \varepsilon) \approx \Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right). \qquad (3.1.9)$$

In the second case we get

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = P(|m - np| \le n\varepsilon) \approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \qquad (3.1.10)$$

Example 3.1.1. The probability that an item will fail during reliability tests is $p = 0{,}05$. What is the probability that during tests with 100 items, the number failing will be

a) at least five;
b) less than five;
c) between five and ten.

Solution. By the de Moivre-Laplace theorem,

$$P(m_1 \le m \le m_2) \approx \frac{1}{2}\left(\Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)\right),$$

if n is sufficiently large. By assumption, $n = 100$, $p = 0.05$, $q = 1 - p = 0.95$.

a) The probability that at least five items fails is

$$P(m \ge 5) = P(5 \le m < 100) \approx \frac{1}{2}\left(\Phi\left(\frac{100-5}{\sqrt{4.75}}\right) - \Phi\left(\frac{5-5}{\sqrt{4.75}}\right)\right) =$$

$$= \frac{1}{2}(\Phi(43.6) - \Phi(0)) = 0.5.$$

b) The probability that less than five items fail is

$$P(m < 5) = P(0 \le m < 5) \approx \frac{1}{2}\left(\Phi\left(\frac{5-5}{\sqrt{4.75}}\right) - \Phi\left(\frac{0-5}{\sqrt{4.75}}\right)\right) =$$

$$= \frac{1}{2}(\Phi(0) - \Phi(2.29)) = 0.489.$$

c) The probability that five to ten items fail is

$$P(5 \le m \le 10) \approx \frac{1}{2}\left(\Phi\left(\frac{10-5}{\sqrt{4.75}}\right) - \Phi\left(\frac{5-5}{\sqrt{4.75}}\right)\right) =$$

$$= \frac{1}{2}(\Phi(2.29) - \Phi(0)) = 0.489.$$

Example 3.1.2. How many independent trials should be performed so that at least five occurrences of an event A will be observed with probability 0.8, if the probability of A in one trial is $P(A) = 0.05$.

Solution. From the Moivre-Laplace theorem, we see that

$$P(m \ge 5) \approx \frac{1}{2}\left(\Phi\left(\frac{n - 0.05}{\sqrt{0.0475n}}\right) - \Phi\left(\frac{5 - 0.05n}{\sqrt{0.0475n}}\right)\right) =$$

$$= \frac{1}{2}\left(\Phi\left(4.36\sqrt{n}\right) - \Phi\left(\frac{5 - 0.05n}{\sqrt{0.0475n}}\right)\right).$$

For $n = 1$, we have $\Phi\left(4.36\sqrt{n}\right) \approx 1;$ therefore, substituting $P(m \ge 5) = 0.8$ we obtain

$$\frac{1}{2}\left(1 - \Phi\left(\frac{5 - 0.05n}{\sqrt{0.0475n}}\right)\right) \approx 0.8,$$

or

$$\Phi\left(\frac{5 - 0.05n}{\sqrt{0.0475n}}\right) = -0.6.$$

From the table we find the argument $x = -0.8416$ corresponding to the value of the function $\Phi(x) = -0.6$.

Solving the equation

$$\frac{5 - 0.05n}{\sqrt{0.0475n}} = -0.8416,$$

we find the unique root $n = 144$. Thus, in order that A occur at least five times with probability 0.8, 144 trials are necessary.