A TABLE OF CONTENTS

Lecture 1

1.1. Random variables. Probability distribution law

A ***random variable*** (RV) is a variable, which resulting from the experiment with a chance outcome, takes on this or that value, therefore being unknown in advance, before an experiment, which value exactly.

We denote random variables by capital letters: *X*, *Y*, *Z*; their values – by the corresponding small letters: *x*, *y*, *z*, and Ω_X – is a set of possible values of variable *X*.

Examples of random variables:

1. Experiment – roll dice; random variable *X* – is number of dot on cube face; $\Omega_X = \{0,1,2,3,4,5,6\}$.

2. Experiment – computer operating up to the first failure; random variables *X* – are time between failures; $\Omega_X = (0,\infty]$.

Depending on the kind of set Ω_X , random variables can be discrete and continuous.

Random variable *X* is called ***discrete*** (DRV), if set Ω_X is enumerable, i.e. all its elements can be arranged in a certain order and be enumerated.

Random variable *X* is called ***continuous,*** indiscrete (CRV), if set Ω_X – is innumerable set.

Distribution law of the random variable *X* is any function (rule, table, etc.), establishing a correspondence between random variable values and probabilities of their occurrence, and allowing to find probabilities of the every possible event $p\{a \le X < b\},\ \forall a,b$, related to a random variable.

1.2. Cumulative distribution function

Cumulative distribution function *F (x)* of the random variable *X* is a function whose value is the probability that a random variable *X* has a value less than or equal to the argument of the function *x*:

$$F(x) = p(X < x). \tag{1.1}$$

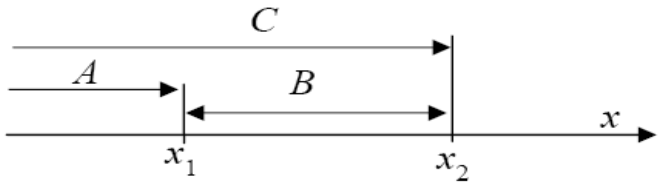
Properties of *X* cumulative distribution function:

1. $F(-\infty) = 0$.

2. $F(+\infty) = 1$.

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

Proving.



$A = \{X < x_1\}$, $B = \{x_1 \le X < x_2\}$, $C = \{X < x_2\}$, then

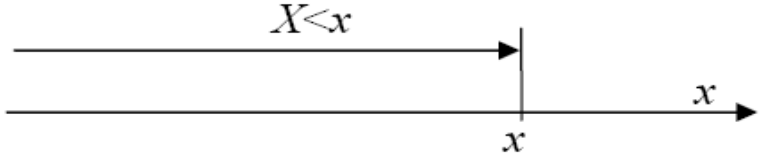
$C=A+B$, $p(C) = p(A) + p(B)$, $p(C) = F(x_2)$, $p(A) = F(x_1)$, $F(x_2) = F(x_1) + p(B)$, $p(B) \le 0 \Rightarrow F(x_1) \le F(x_2)$.

4. Probability of the random variable *X* value getting to the interval:

$$p(x_1 \le X < x_2) = F(x_2) - F(x_1). \tag{1.2}$$

Proving. $p(x_1 \le X < x_2) = p(B) = p(C) - p(A) = F(x_2) - F(x_1)$.

We illustrate these properties by means of the visual geometrical interpretation. For this purpose, we consider a random variable as random point *X* on the axis *OX*, which, as a result of the experiment, can take this or that position.



Then cumulative distribution function $F(x)$ is the probability that random point X as an experimental result will get to the left of the point x. We increase x, moving the point to the right along the abscissa axis, it is obvious therefore, that the probability of fulfilling inequality $X < x$ cannot be decreased (property 3). With reducing x to $-\infty$ – the event $X < x$ becomes impossible, i.e. $F(-\infty) = 0$ (property 1). With increasing x to $+\infty$ - becomes certain, i.e. $F(+\infty) = 1$ (property 2).

Cumulative distribution function is used when considering both discrete and continuous random variables.

1.3. Distribution set

For describing discrete random variables, along with cumulative distribution function $F(x)$, the probabilities distribution set (probability mass function) is used.

The distribution set of the discrete RV X is a table, in the top line of which all possible values of RV $x_1, x_2..., x_n$ ($x_{i-1} < x_i$) are listed, and in the bottom line — probabilities of their occurrence $p_1, p_2..., p_n$, where $p_i = p\{X = x_i\}$.

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

Since events $\{X = x_1\}..., \{X = x_n\}$ are incompatible events and form a entire group, the control relationship is valid

$$p_1 + p_2 + ... + p_n = 1. \qquad (1.3)$$

The ***polygon of probabilities*** is a graphic representation of the probabilities distribution series. All possible random variable values are put on the abscissa axis, and probabilities of these values are put on the axis of ordinates. For visualization, straight-line segments connect the obtained points. The distribution polygon, like the distribution set, completely characterizes a random variable and is one of the distribution law form.

Cumulative distribution function of any discrete random variable is the discontinuous step function, the jumps of which occur at the points corresponding to possible values of the random variable, and are equal to probabilities of these values:

$$F(x) = \sum_{x_i < x} p(X = x_i), \qquad (1.4)$$

where summation is taken over all values of x_i which are less than x.

1.4. Probability density function

The random variable X is called ***continuous***, if its cumulative distribution function $F(x)$ – is continuous and differentiated function for all values of argument.

For continuous cumulative distribution function $F(x)$ *the* probability of any separate value of the random variable should be equal to zero, i.e. there should be no jumps at any point. Such events – possible, but with zero probability – appear only when considering experiments, which are not reduced to the occurrence scheme. This is similar to a body having a certain weight, but not one of the points inside the body possess a finite mass. The small volume possesses a finite mass, but it approaches to zero with volume decreasing, and in the limit is equal to zero for a point. That is, for continuous probabilities distribution, the probability of getting to the arbitrarily small section differs from zero, then the probability of getting to a strictly certain point exactly equals to zero.

The probability of the continuous random variable X getting to the section from x to $x+\Delta x$ is equal to the cumulative distribution function increment on this section: $p\{x \le X < x+\Delta x\} = F(x+\Delta x) - F(x)$. Then the probability density on this section is equal to $\frac{p\{x \le X < x+\Delta x\}}{\Delta x}$. Going to the limit at $\Delta x \to 0$, we obtain the probability density at the point x:

$$\lim_{\Delta x \to 0} \frac{p\{x \le X < x+\Delta x\}}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{dF(x)}{dx} = F'(x) = f(x).$$

The obtained function is one of the forms of the of continuous random variables distribution law forms.

The ***probability density function*** (distribution density) $f(x)$ of the continuous random variable X characterizes the probability density in the neighborhood of point x and is equal to its cumulative distribution function derivative

$$f(x) = \frac{dF(x)}{dx} = F'(x), \qquad (1.5)$$

and the probability density plot is called a distribution curve.

Let there is point x and adjacent to it segment dx. The probability of the random variable X getting to this interval equals $f(x)\,dx$. This variable is called a ***probability element.*** The probability of the random variable X getting to the arbitrary section $[a, b)$ is equal to the sum of probability elements on this section:

$$p\{a \le X < b\} = \int_a^b f(x)dx. \qquad (1.6)$$

In geometrical interpretation $p\{a \le X < b\}$ is equal to the area limited from the top by the probability density function curve $f(x)$ and section $[a, b)$.

Relationship (4.6) allows to express the cumulative distribution function $F(x)$ of the random variable X in terms of its density:

$$F(x) = p\{X < x\} = p\{-\infty < X < x\} = \int_{-\infty}^{x} f(t)dt. \qquad (1.8)$$

The basic properties of the probability density function:

1. Probability density function is non-negative $f(x) \ge 0$, since its primitive $F(x)$ is non- decreasing function (see property 3 $F(x)$, section 1.2). The probability density function is equal to zero ($f(x) = 0$) for those values of x, which the random variable X never assumes in an experiment.

2. The normalizing condition: $\int_{-\infty}^{\infty} f(x)dx = p(-\infty \le X < +\infty) = 1.$ (1.9)

The total area, limited by the distribution curve and the abscissa axis, is equal to 1.

Lecture 2

2.1. Numerical characteristics of the random variable

The distribution law of the random variable is a mathematical form, which completely describes a random variable from probabilistic point of view. However, in many practical problems, there is no need for such complete description, and it is enough to specify only certain numerical parameters, characterizing significant distribution features. Such numbers are called ***numerical characteristics of a random variable***.

2.1.1. Mathematical expectation characterizes the random variable weighed mean value and is defined by formulas:

$$m_X = \mathrm{M}[X] = E\left[X\right] = \begin{cases} \sum_i x_i \cdot p\{X = x_i\}, \text{for } DRV \\ \int_{-\infty}^{\infty} x \cdot f(x)dx, \text{for } CRV. \end{cases} . \qquad (2.1)$$

where m_x denotes the number obtained after calculations by equation (5.1);

DRV - discrete random variable;

CRV - continuous random variable;

M [X] – is the expectation operator.

As it is seen from (2.1), as mathematical expectation, the «weighed mean value» is used, therefore each of the random variable values is considered with the "weight" proportional to the probability of this value.

The physical meaning of the mathematical expectation – is the mean value of the random variable, i.e. that value, which can be used instead of the random variable in approximate calculations or estimations.

The mathematical expectation possesses the following properties:

1. M [c] = c.

Proving. We consider constant ***c*** as a random discrete variable which assumes one value ***c*** with probability p = 1.

2. M [X+c] = M [X] +$c = m_X + c$.

Proving. $\mathrm{M}[X + c] = \int_{-\infty}^{\infty} (x + c) \cdot f(x)dx = \int_{-\infty}^{\infty} x \cdot f(x)dx + \int_{-\infty}^{\infty} c \cdot f(x)dx = m_X + c$

3. M [c·X] = c·M [X] $= c \cdot m_X$.

Proving. $\mathrm{M}[cX] = \int_{-\infty}^{\infty} cx \cdot f(x)dx = c\int_{-\infty}^{\infty} x \cdot f(x)dx = c \cdot m_X$

2.1.2. Ordinary moments. The k-order ordinary moment $\alpha_k(x)$ of the random variable X is the k-degree mathematical expectation of this random variable:

$$\alpha_k(x) = \mathrm{M}[X^k] = \begin{cases} \sum_{i=1}^{N} x_i^k \cdot p_i, & \text{for } DRV; \\ \int_{-\infty}^{\infty} x^k \cdot f(x)dx, & \text{for } CRV. \end{cases} . \qquad (2.2)$$

We consider several of the first k ordinary moments:

$k = 0$ $\alpha_0(x) = M[X^0] = M[1] = 1$;

$k = 1$ – a $\alpha_1(x) = M[X^1] = M[X] = m_X$ – mathematical expectation;

$k = 2$ – $\alpha_2(x) = M[X^2]$ – is used for dispersion calculation.

Centered random variable $\overset{\circ}{X}$ is the random variable, mathematical expectation of which is at the beginning of the origin of coordinates (at the number axis center), i.e $M[\overset{\circ}{X}] = 0$.

The centering operation (transfer from the non-centered variable X to the centered $\overset{\circ}{X}$) looks like

$$\overset{\circ}{X} = X - m_X \quad .$$

2.1.3. Central moments. *The* k-order central moment $\mu_k(x)$ *of the* random variable X is the k-degree mathematical expectation of the centered random variable $\overset{\circ}{X}$:

$$\mu_k(x) = \mathrm{M}[\overset{\circ}{X}^k] = \begin{cases} \sum_{i=1}^{N} (x_i - m_X)^k \cdot p_i, & \text{for } DRV, \\ \int_{-\infty}^{\infty} (x - m_X)^k \cdot f(x)dx, & \text{for } CRV. \end{cases} . \qquad (2.3)$$

We consider several of the first *k* ordinary moments:

$k = 0$ $\mu_0(x) = M[\overset{\circ}{X}^0] = M[1] = 1$;

$k = 1$ $\mu_1(x) = M[\overset{\circ}{X}^1] = M[\overset{\circ}{X}] = 0$;

$k = 2$ – a $\mu_2(x) = M[\overset{\circ}{X}^2] = M[(X - m_X)^2] = M[X^2] - 2m_X M[X] + m_X^2 = \alpha_2(x) - m_x^2 = D_X$ – dispersion or variance.

2.1.4. The random variable ***dispersion*** (variance) characterizes the degree of dispersion (scattering) of random variable values with respect to its mathematical expectation and is determined by formulas:

$$D_x = D[X] = \mu_2(x) = \alpha_2(\mathrm{x}) - m_X^2 = \begin{cases} \sum_{i=1}^{N} (x_i - m_X)^2 p_i = \sum_{i=1}^{N} x_i^2 p_i - m_X^2, & \text{for } DRV; \\ \int_{-\infty}^{\infty} (x - m_X)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - m_X^2, & \text{for } CRV. \end{cases} \qquad (2.4)$$

Dispersion properties:

1. $D[c] = 0$.

Proving. $D[c] = M\left[(c - M[c])^2\right] = M\left[(c-c)^2\right] = M[0] = 0$

2. $D[X+c] = D_X$.

Proving.

$$D[X + c] = M\left[(X + c - M[X + c])^2\right] = M\left[(X + c - m_X - c)^2\right] = M[(X - m_X)^2] = D_X$$

follows from property 2 (items 2.1.1) of the mathematical expectation. It becomes clear, if it is taken into account, that variables X and $X+c$ differ only by the reference and are equally dispersed about their mathematical expectations. It is obvious, that the centering operation does not change the random variable dispersion: $D[\overset{\circ}{X}] = D[X - m_X] = D[X]$.

3. $D\,[c \cdot X] = c^2 \cdot D_X$.

Proving. $D[cX] = M[c^2 X^2] - \left(M[cX]\right)^2 = c^2 (M[X^2] - m_X^2) = c^2 D_X$

The random variable dispersion has the dimension of the random variable square, therefore, dispersion is not quite convenient for analysing the value range of the variable *X*. The root-mean-square deviation (RMSD), the dimension of which matches the random variable dimension, does not feature this disadvantage.

2.1.5. The root-mean-square deviation (standard deviation) of the random variable *X* characterizes the range width of values *X* and is equal to

$$\sigma_X = \sigma[X] = +\sqrt{D[X]}\,. \qquad (2.5)$$

RMSD is measured in the same physical units, as the random variable.

Rule 3σ. Practically, all random variable values are in within the interval

$$[m_X - 3\sigma_X;\ m_X + 3\sigma_X;]. \qquad (2.6)$$

Mathematical expectation and dispersion (or RMSD) – are most often used characteristics of a random variable. They characterize the most important distribution features: its position and degree values dispersion. For more detailed description, the ordinary and central moments of the higher orders are used. Beside the mathematical expectation, in practice other characteristics of the value distribution position are also often applied.

2.1.6. The random variable ***mode*** is equal to its most probable value, i.e. that value for which probability p_i (for a discrete random variable) or $f(x)$ (for a continuous random variable) reaches the maximum:

$$p(X = Mo) = \max,\quad f(Mo) = \max.$$

Distribution with one maximum of the probability density function is called "unimodal". If the distribution polygon or distribution curve have more than one maximum, distribution is called "polymodal". If in the middle the distribution features not the maximum, but the minimum, it is called "antimodal".

2.1.7. The median of the random variable *X* is equal to such its value, for which the condition $p\{X < Me\} = p\{X \geq Me\}$ is fulfilled. The median, as a rule, exists only for continuous random variables. The value of *Me* can be determined as the solution of one of the following equations:

$$\int_{-\infty}^{Me} f(x)dx = 0,5;\ \int_{Me}^{+\infty} f(x)dx = 0,5;\ F(Me) = 0,5. \qquad (2.7)$$

In point *Me the* area limited by the distribution curve is halved.

3.1.8. Fractile χ_p of the random variable *X* - is such its value, for which the condition is fulfilled

$$p\{X < \chi_p\} = F(\chi_p) = p. \qquad (2.8)$$

It is obvious, that the median is the fractile $\chi_{0,5}$.

Lecture 3

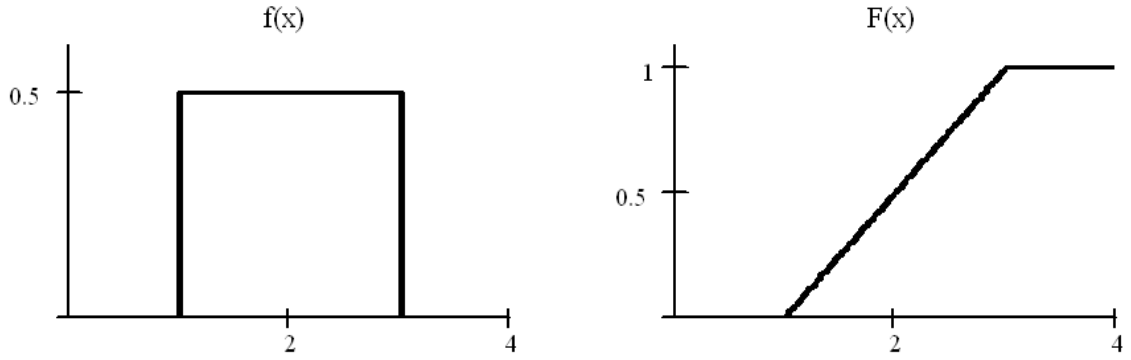
3.1. Uniform distribution

A continuous random variable X ***has uniform distribution*** if its probability density in a certain interval [*a; b*] is constant, i.e. if all values of X in this interval are equiprobable:

$$f(x)=\begin{cases}0, x<a,\\ \dfrac{1}{b-a}, a\le x\le b,\\ 0, x>b.\end{cases}\quad F(x)=\begin{cases}0, x<a,\\ \dfrac{x-a}{b-a}, a\le x\le b,\\ 1, x>b.\end{cases} \tag{3.1}$$

The random variable with uniform distribution is designated as $U(a,b)$.

The density and of uniform cumulative distribution function plots are more presented below at $b = 3$ and $a = 1$.



Numerical characteristics of the uniformly distributed random variable:

$$m_X=\frac{a+b}{2},\ D_X=\frac{(b-a)^2}{12}. \tag{3.2}$$

If necessary to determine parameters *a* and *b* by known m_X, D_X , the following formulas are used:

$$a=m_X+\sigma_X\sqrt{3},\ b=m_X-\sigma_X\sqrt{3}. \tag{3.3}$$

Occurrence conditions:

1. Random variable X – round-off errors at the limited digit grid:

- rounding off to minor integer $X\in[-1,0],\ m_X=-0,5$,

- rounding off to major integer $X \in [0,1]$, $m_X = 0,5$,

- rounding off to the nearest integer $X \in [-0,5; 0,5]$, $m_X = 0$,

where 1 – is the digital resolution.

2. Random variable X – is the value readout error from the measuring device analog indicating scale, $X \in [-0,5; 0,5]$, $m_X = 0$, where 1 – is the scale division value.

3. Pseudo-random variables generators, for example RANDOM, RND built-in to high level programming languages.

3.2. Exponential distribution

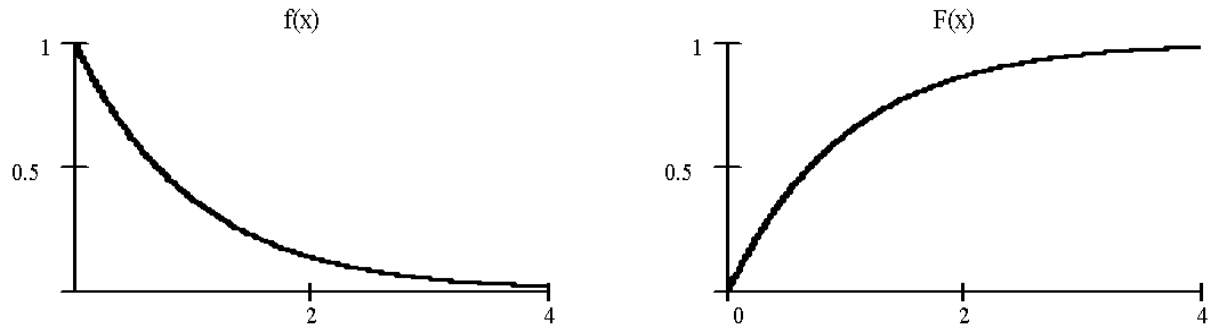
The continuous random variable X assuming only positive values has ***exponential distribution,*** if its probability density and cumulative distribution function are equal to:

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, x \geq 0, \\ 0, \ x < 0; \end{cases} \quad F(x) = \begin{cases} 1 - e^{-\frac{x}{\lambda}}, x \geq 0, \\ 0, \ x < 0, \end{cases} \qquad (3.4)$$

where λ – distribution parameter ($\lambda > 0$).

The random variable with exponential distribution is denoted as $E(\lambda)$.

The density and exponential cumulative distribution function plots are presented below at λ=1.



Numerical characteristics of the exponential random variable:

$$m_X = \lambda,\ D_X = \lambda^2 . \qquad (3.5)$$

Occurrence conditions. Random variable T – *is the* time interval between two adjacent events in the simple or Poisson stream of random events, therefore, the distribution parameter – $\frac{1}{\lambda}$ is the stream intensity.

3.3 Normal (Gauss) distribution

Probability density function $N(a,\sigma^2)$ of normal random variable:

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad a \in R,\ \sigma^2 > 0 .$$

The mathematical expectation of normal random variable

$$E(N(a,\sigma^2)) = a ,$$

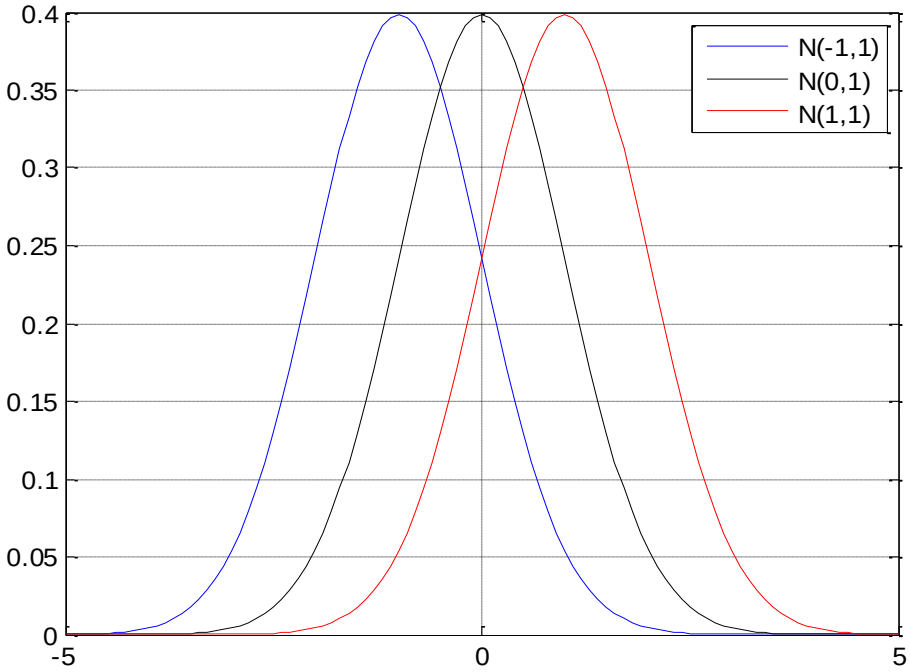


Fig. 3.1

Dispersion of normal random variable

$$D(N(a,\sigma^2)) = \sigma^2$$

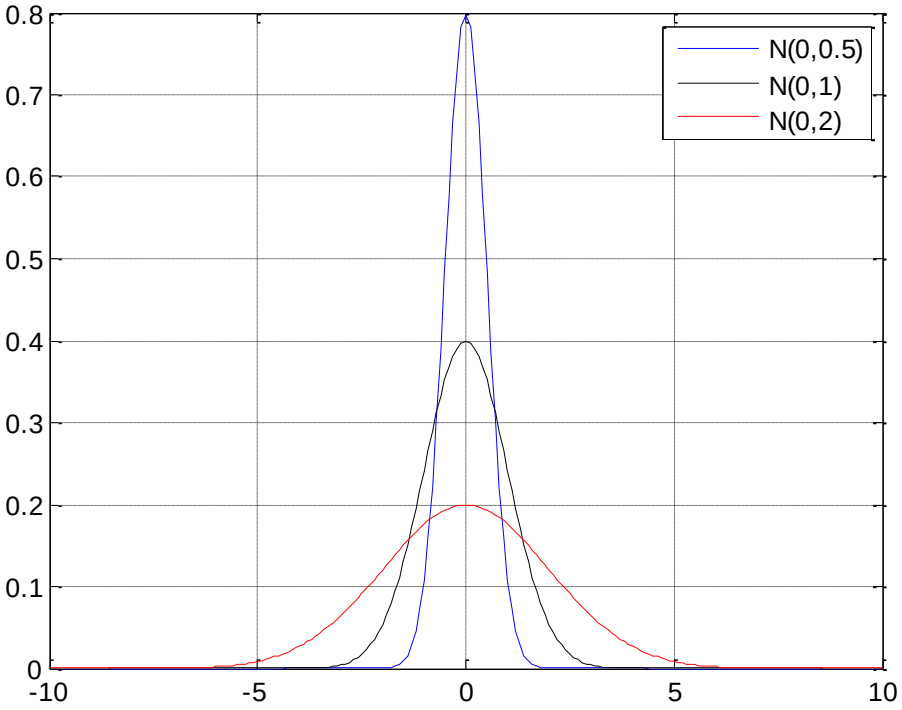


Fig. 3.2

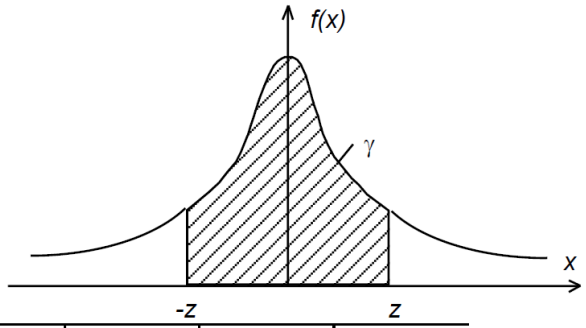
Cumulative distribution functions of normal random variable:

$$F(x) = 0.5 + \Phi\left(\frac{x-a}{\sigma}\right),$$

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

The table 3.1

$$\Phi(z) = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt = \gamma$$



z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,00	0,0000	0,66	0,4907	1,32	0,8132	1,98	0,9523
0,02	0,0160	0,68	0,5035	1,34	0,8198	2,00	0,9545
0,04	0,0319	0,70	0,5161	1,36	0,8262	2,05	0,9596
0,06	0,0478	0,72	0,5285	1,38	0,8324	2,10	0,9643
0,08	0,0638	0,74	0,5407	1,40	0,8385	2,15	0,9684
0,10	0,0797	0,76	0,5527	1,42	0,8444	2,20	0,9722
0,12	0,0955	0,78	0,5646	1,44	0,8501	2,25	0,9756
0,14	0,1113	0,80	0,5763	1,46	0,8557	2,30	0,9786
0,16	0,1271	0,82	0,5878	1,48	0,8611	2,35	0,9812
0,18	0,1428	0,84	0,5991	1,50	0,8664	2,40	0,9836
0,20	0,1585	0,86	0,6102	1,52	0,8715	2,45	0,9857
0,22	0,1741	0,88	0,6211	1,54	0,8764	2,50	0,9876
0,24	0,1897	0,90	0,6319	1,56	0,8812	2,55	0,9892
0,26	0,2051	0,92	0,6424	1,58	0,8859	2,60	0,9907
0,28	0,2205	0,94	0,6528	1,60	0,8904	2,66	0,9920
0,30	0,2358	0,96	0,6629	1,62	0,8948	2,70	0,9931
0,32	0,2510	0,98	0,6729	1,64	0,8990	2,75	0,9940
0,34	0,2661	1,00	0,6827	1,66	0,9031	2,80	0,9949
0,36	0,2812	1,02	0,6923	1,68	0,9070	2,85	0,9956
0,38	0,2961	1,04	0,7017	1,70	0,9109	2,90	0,9963
0,40	0,3108	1,06	0,7109	1,72	0,9146	2,95	0,9968
0,42	0,3255	1,08	0,7199	1,74	0,9181	3,00	0,9973
0,44	0,3401	1,10	0,7287	1,76	0,9216	3,10	0,9981
0,46	0,3545	1,12	0,7373	1,78	0,9249	3,20	0,9986
0,48	0,3688	1,14	0,7457	1,80	0,9281	3,30	0,9990
0,50	0,3859	1,16	0,7540	1,82	0,9312	3,40	0,9993
0,52	0,3969	1,18	0,7620	1,84	0,9342	3,50	0,9995
0,54	0,4108	1,20	0,7699	1,86	0,9371	3,60	0,9997
0,56	0,4245	1,22	0,7775	1,88	0,9399	3,70	0,9998
0,58	0,4381	1,24	0,7850	1,90	0,9426	3,80	0,9999
0,60	0,4515	1,26	0,7923	1,92	0,9451	3,90	0,9999
0,62	0,4647	1,28	0,7995	1,94	0,9476	4,00	0,9999
0,64	0,4778	1,30	0,8064	1,96	0,9500		

3.4 Chi-square distribution

The random variable

$$\chi_k^2 = \sum_{i=1}^{k} x_i^2 ,$$

where – the $x_1, x_1, \ldots, x_k$ independent normal distributed random variables $N(0,1)$,

have χ^2 (Chi-square) distribution with ***k*** degrees of freedom and notated as $H_1(k)$.

From definition of χ^2 (Chi-square) distribution obviously property:

$$\chi_p^2 + \chi_q^2 \in H_1(p+q). \tag{3.6}$$

where $\chi_p^2 \in H_1(p)$, $\chi_q^2 \in H_1(q)$, and χ_p^2, χ_q^2 are independent,

Probability density function of χ^2 distribution with ***k*** degrees of Freedom $H_1(k)$:

$$f_\xi(x) = \begin{cases} \dfrac{1}{2\Gamma\left(\dfrac{k}{2}\right)}\left(\dfrac{x}{2}\right)^{\frac{k}{2}-1} e^{-\frac{x}{2}}, & x > 0,\ k \in Z, \\ 0, & x \le 0. \end{cases}$$

where $\Gamma\left(\dfrac{k}{2}\right)$ – Gamma function which is defined by expression

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy .$$

Gamma function possesses following properties:

$$\Gamma(x+1) = x\Gamma(x),\ \Gamma(k+1) = k!,\ \Gamma(1) = \Gamma(2) = 1,\ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} .$$

Curve probability density of this distribution are represented on fig. 3.3.

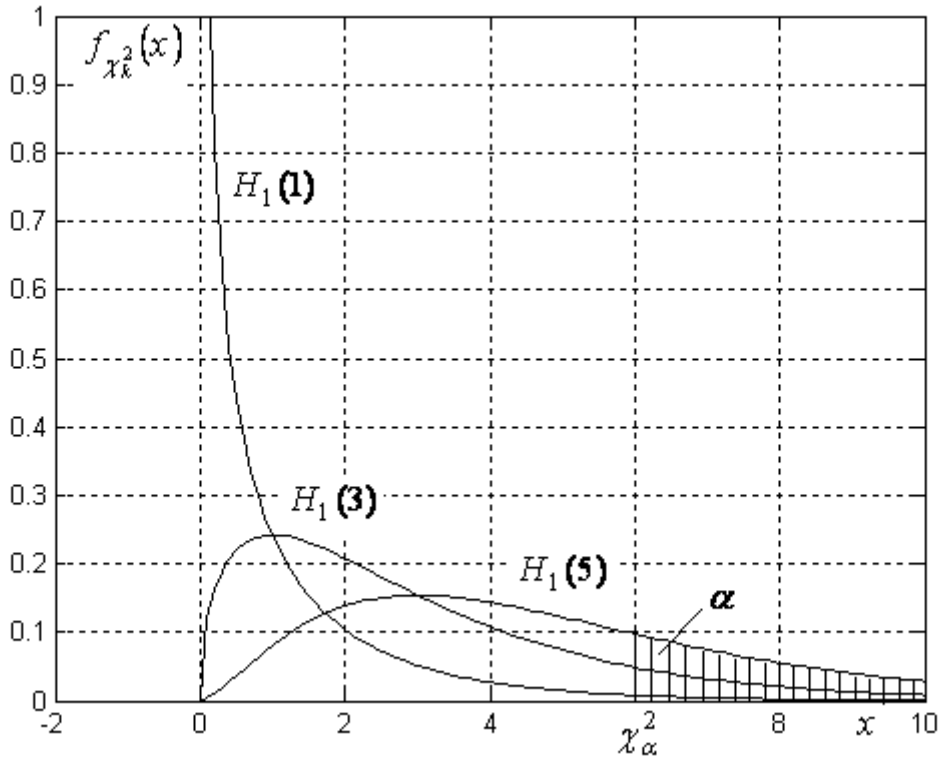


Fig. 3.3

These are the asymmetrical curves located on a positive semiaxis of abscisses. Curves have on one maximum in a point $x = k - 2$.

The mathematical expectation of χ^2 (Chi-square) distribution:

$$E(\chi_k^2) = k.$$

The dispersion of distribution of χ^2 (Chi-square) distribution:

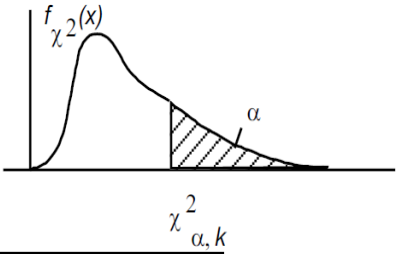
$$D(\chi_k^2) = 2k.$$

The table below allow to solve the equations

$$P(\chi_k^2 > \chi_\alpha^2) = \alpha \text{ , } 0 \le \alpha \le 1.$$

The table 3.2

$$P(\chi^2 > \chi^2_{\alpha,k}) = \alpha$$



k	α					
	0,01	0,02	0,05	0,95	0,98	0,99
1	6,64	5,41	3,84	0,004	0,001	0,000
2	9,21	7,82	5,99	0,103	0,040	0,020
3	11,34	9,84	7,82	0,352	0,185	0,115
4	13,28	11,67	9,49	0,711	0,429	0,297
5	15,09	13,39	11,07	1,145	0,752	0,554
6	16,81	15,03	12,59	1,635	1,134	0,872
7	18,48	16,62	14,07	2,17	1,564	1,239
8	20,10	18,17	15,51	2,73	2,03	1,646
9	21,07	19,68	16,92	3,32	2,53	2,09
10	23,20	21,2	18,31	3,94	3,06	2,56
12	26,2	24,1	21,0	5,23	4,18	3,57
14	29,1	26,9	23,7	6,57	5,37	4,66
16	32,0	29,6	26,3	7,96	6,61	5,81
18	34,8	32,3	28,9	9,39	7,91	7,02
20	37,6	35,0	31,4	10,85	9,24	8,26
22	40,3	37,7	33,9	12,34	10,60	9,54
24	43,0	40,3	36,4	13,85	11,99	10,86
26	45,6	42,9	38,9	15,38	13,41	12,20
28	48,3	45,4	41,3	16,93	14,85	13,56
30	50,9	48,0	43,8	18,49	16,31	14,95

3.5 Student distribution

Random variable

$$t = \frac{u}{\sqrt{v}}\sqrt{n},$$

where u and v –independent random variables, and $u \in N(0,1)\ v \in H_1(n)$, have Student distribution of with k degrees of freedom and notated as $T_1(n)$.

Probability density function of Student distribution

$$f_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Curve probability density of Student distribution are represented on fig. 3.4.

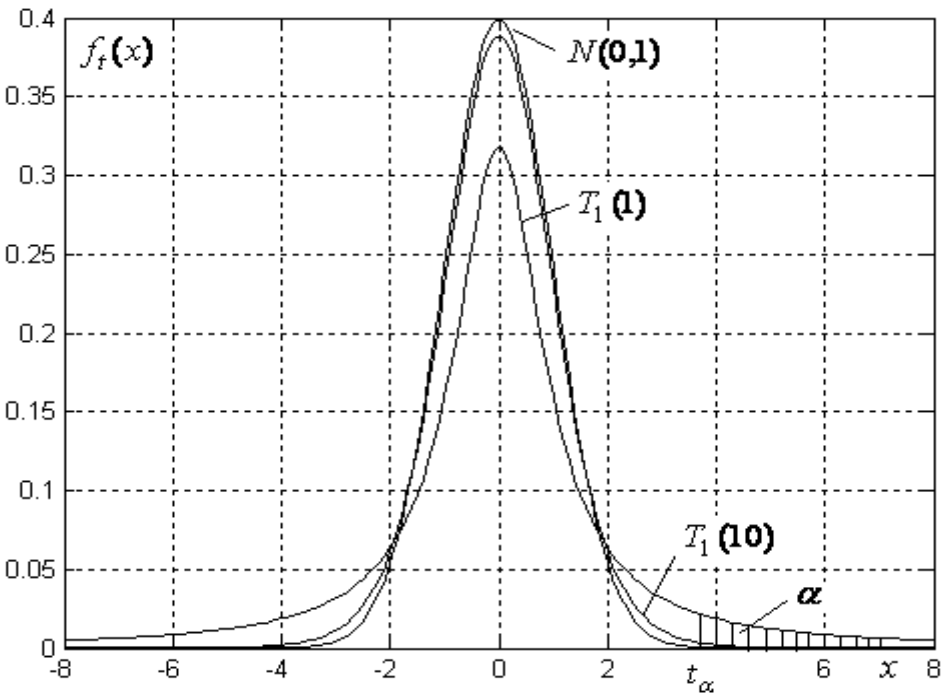


Fig. 3.4

At $n \to \infty$ Student distribution converge to normal distribution $N(0,1)$. However at small n ($n < 30$) it considerably differs from the normal.

The mathematical expectation of Student distribution ($n > 2$):

$$E(t) = 0.$$

The dispersion of distribution of Student distribution ($n > 2$):

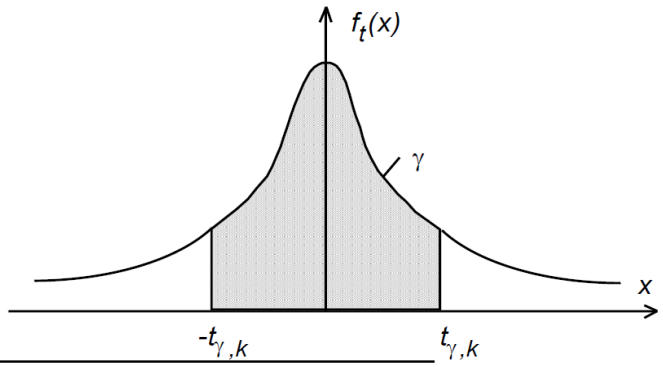
$$D(t) = n/(n-2).$$

There are table satisfying to equality

$$P(-t_{\gamma,k} < t < t_{\gamma,k}) = \gamma,$$

The table 3.3

$$\gamma = \int_{-t_{\gamma,k}}^{t_{\gamma,k}} f_t(x)dx$$



k	γ			
	0,90	0,95	0,98	0,99
1	6,31	12,71	31,8	63,7
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,77	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,943	2,45	3,14	4,71
7	1,895	2,36	3,00	3,50
8	1,860	2,31	2,90	3,36
9	1,833	2,26	2,82	3,25
10	1,812	2,23	2,76	3,17
12	1,782	2,18	2,68	3,06
14	1,761	2,14	2,62	2,98
16	1,746	2,12	2,58	2,92
18	1,734	2,10	2,55	2,88
20	1,725	2,09	2,53	2,84
22	1,717	2,07	2,51	2,82
24	1,711	2,06	2,49	2,80
30	1,697	2,04	2,46	2,75
40	1,684	2,02	2,42	2,70

3.6 Fisher distribution

Random variable

$$f = \frac{nv}{mw},$$

where v and w –independent random variables, $v \in H_1(m)\, w \in H_1(n)$

have Fisher distribution with $m,\ n$ degrees of freedom and notated as $F_1(m,n)$.

Probability density function of Fisher distribution

$$f\ (x,m,n) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \dfrac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{n+m}{2}}}, & x \geq 0, \\ 0, \quad x < 0. \end{cases}$$

where n, m - whole numbers ($n > 0$, m >0)

Curve probability density of Fisher distribution are represented on fig. 3.5.

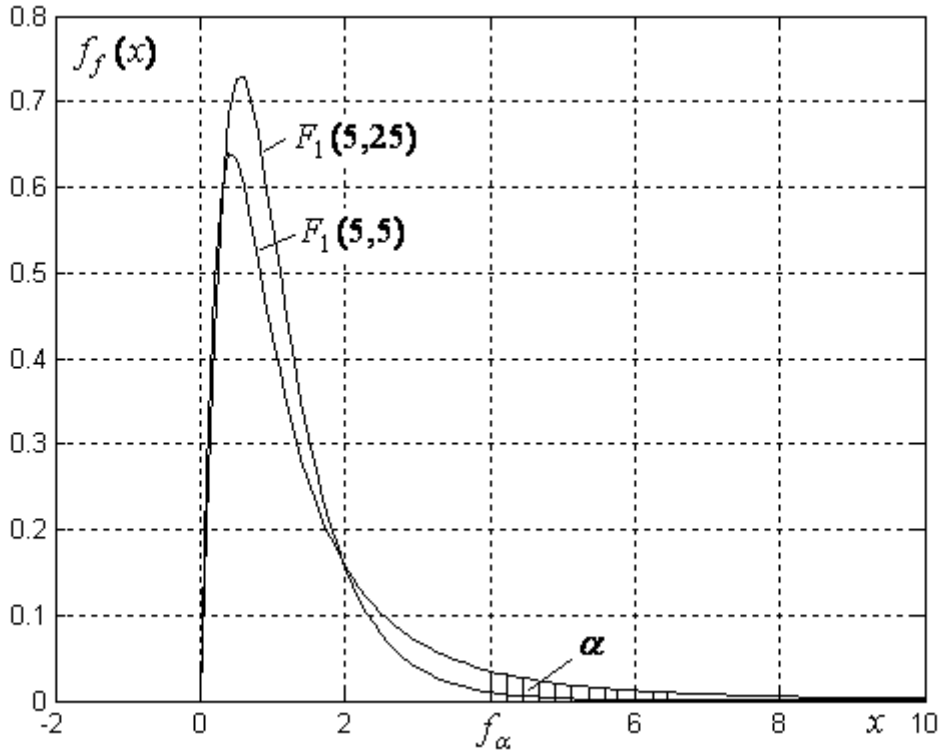


Fig. 3.5

These are the asymmetrical curves located on a positive X-semiaxis which reach a maximum near to a point $x = 1$.

The mathematical expectation of random variable with Fisher distribution

$$m_X = \frac{n}{n-2},$$

The dispersion random variable with Fisher distribution:

$$D_X = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}.$$

3.7 Gamma distribution

Probability density function of random variable with Gamma distribution:

$$f\ (x,a,b)=\begin{cases}\dfrac{1}{\Gamma(a)b^{a}}\,x^{a-1}e^{-\frac{x}{b}}, & x>0,\ b>0,\ a>0,\\ 0, & x\le 0.\end{cases},$$

where a, b - distribution parameter (a > 0, b >0)

The mathematical expectation of random variable with Gamma distribution

$$m_X = ab\,,$$

The dispersion random variable with Gamma distribution:

$$D_X = ab^2\,.$$

Lecture 4

Let a certain random variable X is exposed to the determinate transformation φ, as the result of which the variable Y will appear, i.e. $Y = \varphi(x)$. It is obvious that variable Y will be random, and, as a rule, it is necessary to determine the ***distribution law*** and/or ***numerical characteristics*** of the random variable Y under the known distribution law of variable X and the form of transformation φ.

4.1. The distribution law of the random argument function

In case that X – is a discrete random variable with a known probabilities distribution series:

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_i

then it will be easy to determine the probabilities series of Y. Since $Y = \varphi(x)$, then the value $y_i = \varphi(x_i)$ will appear with probability p_i:

y_i	$\varphi(x_1)$	$\varphi(x_2)$	…	$\varphi(x_n)$
p_i	p_1	p_2	…	p_n

From the given series, by ordering and combining identical values, we obtain the distribution series of random variable Y:

y_i	y_1	y_2	…	y_m
p_j	p_1	p_2	…	p_m

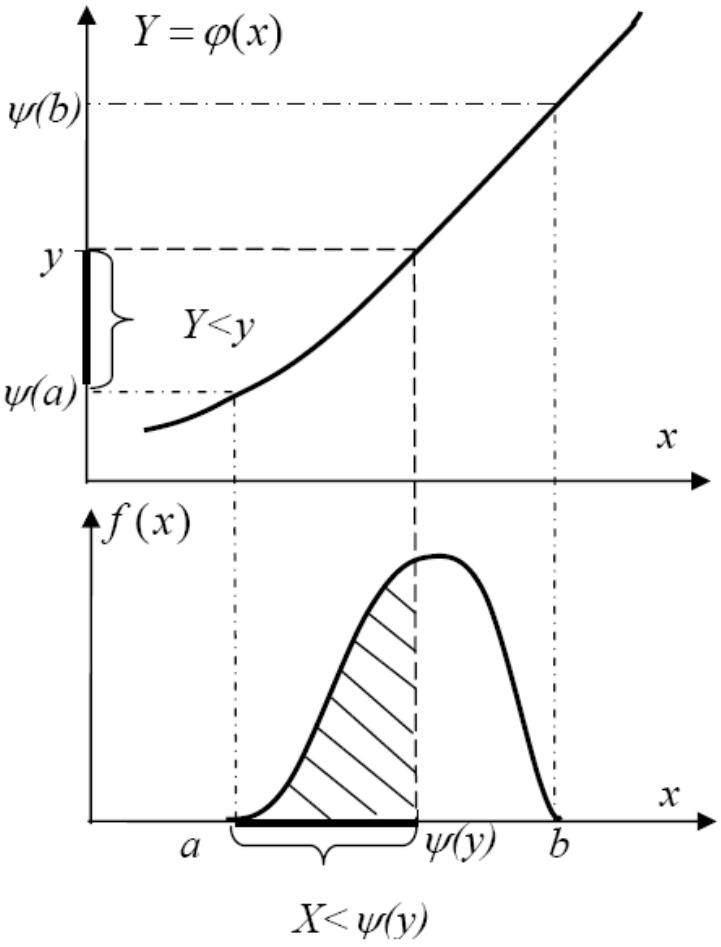
If X – is a continuous random variable with a known probability density $f(x)$, then the algorithm of obtaining the distribution law $Y = \varphi(x)$ depends on the kind of φ. We consider the abscissa axis section [a, b], on which all possible values of variable X are put, i.e. $p(a \le X \le b) = 1$, in a special case $a = -\infty, b = +\infty$. The way of solving this problem depends on the behavior of function φ on the section [*and, b*]: whether it is monotonic on this section or not.

4.1.1. Monotonically increasing function

Let $Y = \varphi(x)$ – is a monotonically increasing function. We determine the cumulative distribution function $G(y)$ of the random variable *Y*. By definition, it is equal to

$$G(y) = p(Y < y) = p(\varphi(x) < y) = p(X < \psi(y)) = \int_{-\infty}^{\psi(y)} f_X(x)dx,$$

where $\psi(y)$ – is the inverse function of $\varphi(x)$.



To fulfill the condition $Y < y$, it is necessary and sufficient, that the random variable *X* will get on the abscissa axis section from *a* to $\psi(y)$. Thus, the cumulative distribution function *Y* for the argument *X* distributed in the interval [*a*, *b*], equals:

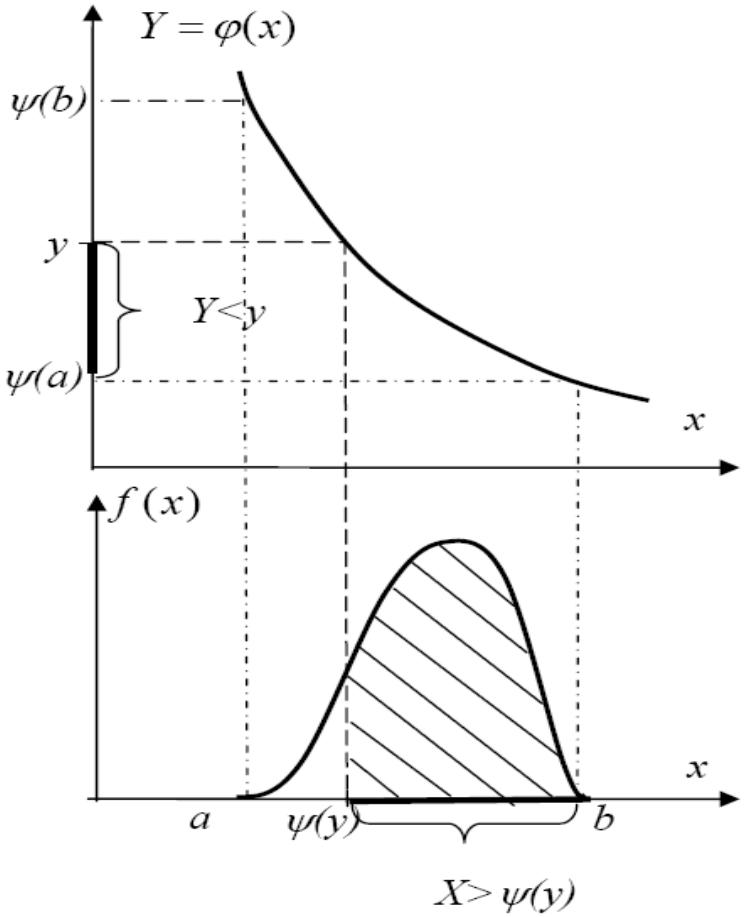
$$G(y) = \begin{cases} 0, y < \psi(a), \\ \int_{a}^{\psi(y)} f_X(x)dx, \psi(a) \le y \le \psi(b), \\ 1, y > \psi(b). \end{cases} \qquad (4.1)$$

4.1.2 Monotonically decreasing function

Let $Y = \varphi(x)$ – is a monotonically decreasing function. We determine the cumulative distribution function $G(y)$ of the random variable *Y*. By definition, it is equal to

$$G(y) = p(Y < y) = p(\varphi(x) < y) = p(X > \psi(y)) = \int_{\psi(y)}^{\infty} f_X(x)dx,$$

where $\psi(y)$ – id the inverse function of $\varphi(x)$.



To fulfill the condition $Y < y$, it is necessary and sufficient, that the random variable *X* will get on the abscissa axis section from $x = \psi(y)$ to *b*. Thus, the cumulative distribution function *Y* for the argument *X* distributed in the interval [*a*, *b*], equals

$$G(y) = \begin{cases} 0, y < \psi(b), \\ \int_{\psi(y)}^{b} f_X(x)dx, \psi(b) \le y \le \psi(a), \\ 1, y > \psi(a). \end{cases} \qquad (4.2)$$

The ***probabilities density*** of the random variable $Y = \varphi(x)$ for any monotonic case looks like:

$$g(y) = G'(y) = \begin{cases} 0, y < y_{\min}, \\ f_X(\psi(y)) \cdot |\psi'(y)|, y_{\min} \le y \le y_{\max}, \\ 0, y > y_{\max}. \end{cases} \qquad (4.3)$$

Example. Let the random variable X has the normal distribution law $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$, $Y = X^3$. Find $g(y)$.

Function $Y = \varphi(x)$ is strictly monotonic, differentiable and has inverse $X = \psi(y) = \sqrt[3]{Y}$. We apply formula (7.1). Since

$$f_X(\psi(y)) = f_X(y^{1/3}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y^{2/3}/2\sigma^2},$$

$$|\psi'(y)| = |(y^{1/3})'| = \frac{1}{3y^{2/3}},$$

then the desired probability density function of function $Y = X^3$:

$$g(y) = \frac{1}{3\sigma y^{2/3}\sqrt{2\pi}} e^{-y^{2/3}/2\sigma^2}.$$

4.2. Simulating of the random variable with the given distribution law by functional transformation

Here we consider an important for the practice problem about simulating the continuous random variable Y with the given distribution law by the functional transformation of other random variable with the known distribution law.

The problem is put as follows: there is a random variable X with uniform distribution in the range [0; 1]

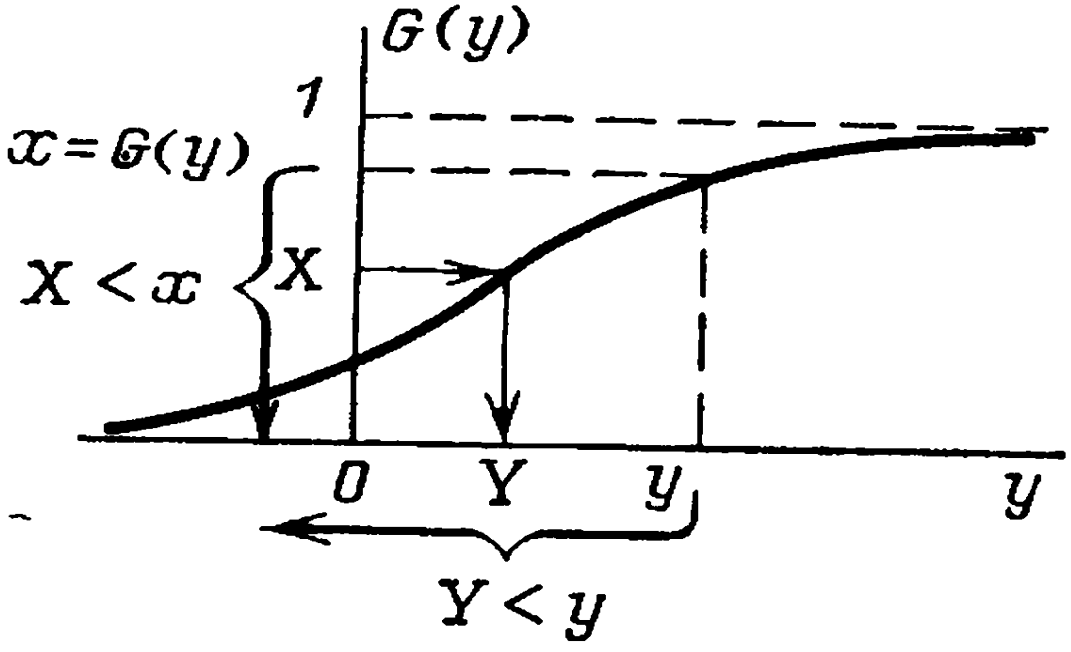
$$f(x)=\begin{cases}0, x<0,\\ 1, 0\le x\le 1,\\ 0, x>1.\end{cases}$$

Question: to which functional transformation $Y=\varphi(x)$ it should be subjected, so that the random variable Y had the given cumulative distribution function $G(y)$?

We prove, that for this purpose it is necessary to subject the random variable X to functional transformation $Y=G^{-1}(x)$,

where $G^{-1}(x)$ - is the inverse function of the required cumulative distribution function $G(y)$.

Let's represent the plot of the cumulative distribution function $G(y)=p(Y<y)$:



Since the cumulative distribution function $G(y)$ of the continuous random variable Y is a continuous monotonically increasing function, then the inverse function $G^{-1}(x)$ will also be a continuous monotonically increasing function, and equations from paragraph ***4.1.1*** can be applied:

$$p(Y < y) = p(G^{-1}(x) < y) = p(X < G(y)) = \int_{-\infty}^{G(y)} f_X(x)dx = \int_{0}^{G(y)} 1dx = G(y)$$

Which was to be proved (Q.E.D. (quod erat demonstrandum).

Example. Let it is required to obtain a random variable *Y* ,which has the exponential distribution law

$$G(y) = \begin{cases} 1 - e^{-\frac{y}{\lambda}}, y \geq 0, \\ 0, \ y < 0, \end{cases}$$

We find $G^{-1}(x)$ inverse function of the desired cumulative distribution function $G(y)$:

$$x = G(y) = 1 - e^{-\frac{y}{\lambda}} \Rightarrow y = G^{-1}(x) = -\lambda \ln(1 - x) .$$

Lecture 5

5.1. The law of large numbers

Let the experiment is carried out, in which we are interested to know the value of the random variable X. When single replication, it is impossible to tell in advance, which value the variable will assume. But in case of *n*-fold (*n*> 100... 1000) experiment repetition, the "mean" (arithmetic mean) value of variable *X* loses a random character and becomes close to a certain constant.

The ***law of large numbers*** – is the set of theorems determining the conditions of tending of random variables arithmetic mean values to a certain constant when carrying out a great number of experiments.

5.1.1. Chebychev's inequality

Chebychev's inequality. For any random variable *X* with mathematical expectation m_X and dispersion D_X the following inequality is fulfilled:

$$p(|X - m_X| \geq \varepsilon) \leq \frac{D_X}{\varepsilon^2}, \qquad (5.1)$$

where $\varepsilon > 0$.

Proving. We consider probability $p(|X| \geq \varepsilon)$:

$$p(|X| \geq \varepsilon) = \int_{|x|\geq\varepsilon} f(x)dx = \int_{|x|\geq\varepsilon} \frac{x^2}{\varepsilon^2}\cdot\frac{\varepsilon^2}{x^2} f(x)dx \leq \int_{|x|\geq\varepsilon} \frac{x^2}{\varepsilon^2} f(x)dx = \frac{1}{\varepsilon^2}\int_{|x|\geq\varepsilon} x^2 f(x)dx \leq$$

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{M[X^2]}{\varepsilon^2}.$$

Thus $p(|X| \geq \varepsilon) \leq \frac{M[X^2]}{\varepsilon^2}$. Having replaced the non-centered variable *X* by the centered $\overset{\circ}{X} = X - m_X$, we will obtain $p(|X - m_X| \geq \varepsilon) \leq \frac{M[(X - m_X)^2]}{\varepsilon^2} = \frac{D_X}{\varepsilon^2}$.

Example. We determine the probability that the random variable will assume the value outside the interval $3\sigma_X$. We suppose in Chebychev's inequality $\varepsilon = 3\sigma_X$, we have:

$$p(|X - m_X| \geq 3\sigma_X) \leq \frac{D_X}{9\sigma_X^2} = \frac{1}{9} \approx 0{,}11.$$

Chebychev's inequality gives only the top border of probability of the given deviation. The probability value cannot exceed this border (0,11) at any distribution law. Therefore, the rule $3\sigma_X$ is fulfilled with probability not less than 0,89.

Probability convergence. *The* sequence of random variables X_n converges in probability to variable *a*, $X_n \xrightarrow[n\to\infty]{p} a$, if with increasing of *n* the probability of that X_n an *a* will be arbitrary close, unrestrictedly verge towards unity:

$$p(|X_n - a| < \varepsilon) > 1 - \delta,$$

where ε, δ – are arbitrary small positive numbers.

One of the most important forms of the law of large numbers – Chebychev's theorem, it establishes relation between the arithmetic mean of the observable random variable values and its mathematical expectation.

5.1.2 Chebychev's theorem

Let *n* identical independent experiments have been carried out, in each of which a random variable *X* has assumed values X_1, X_2, ..., X_n. At a great enough number of independent experiments, the arithmetic mean of the random variable *X* values converges in probability to its mathematical expectation:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{p} m_X. \qquad (5.2)$$

Proving. We consider variable $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$. We determine numerical characteristics of *Y* (see (11.5), (11.7)):

$$m_Y = M[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n} n m_X = m_X;$$

$$D_Y = D[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2} n D_X = \frac{D_X}{n}.$$

Let's write down Chebychev's inequality for variable *Y*:

$$p\left(|Y - m_Y| \geq \varepsilon\right) = p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_X\right| \geq \varepsilon\right) \leq \frac{D_Y}{\varepsilon^2} = \frac{D_Y}{\varepsilon^2 n}.$$

No matter how small is number ε, it is possible to take such a large *n* , so that inequality $\frac{D_X}{\varepsilon^2 n} < \delta$ will be fulfilled, where – δ is an arbitrary small number. Then $p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_X\right| \geq \varepsilon\right) < \delta$. When transition to complementary event $p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_X\right| < \varepsilon\right) > 1 - \delta$, i.e. *Y* converges in probability to m_X.

12.1.3. Bernoulli theorem

Let *n* identical independent experiments have been carried out, in each of which event *A* is possible with probability *p*. Then the *frequency* of occurrence of event *A* in *n* experiments converges in probability to probability of occurrence of *A* in one experiment:

$$p^*(A) \xrightarrow[n\to\infty]{p} p(A), \qquad (5.3)$$

where $p^*(A)$ – is frequency of event *A* in *n* experiments $p^*(A) = \frac{m}{n}$;

m – is the number of experiments, in which event *A* has occurred;

n – is the number of the conducted experiments.

Let random variable X – is the indicator of event A:

$$X = \begin{cases} 1, A \\ 0, \overline{A} \end{cases},$$

then X_i – the indicator of event A in i-experiment.

Numerical characteristics of indicator X of random event (see (2.1)):

$$m_X = p, D_X = qp,$$

where q– is the probability of realization $\bar{A}$, $q = 1 - p$.

Let's apply Chebychev's theorem:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{m}{n} = p^*(A) \underset{n\to\infty}{\overset{p}{\to}} m_X = p = p(A).$$

12.2. The central limit theorem

The given theorem determines conditions at which a random variable with the normal distribution law appear. Various forms of the central limit theorem differ between themselves by the conditions imposed on distributions of the sum of random summands $X_1, X_2, \ldots, X_n$. The more strict are these conditions, the more easier the theorem is proved; the wider conditions, the more difficult is the proving. Here we consider one of the most simple forms of this theorem, namely, the central limit theorem for equally distributed summands.

Theorem. If $X_1, X_2, \ldots, X_n$ – are independent random variables having one and the same distribution with mathematical expectation m and dispersion σ^2 t, then with unbounded increasing of n ($n \to \infty$) the distribution law of their sum $Y = \sum_{i=1}^{n} X_i$ is unboundedly approaching to the normal law with parameters

$$m_Y = n \cdot m, \ \sigma_Y = \sigma\sqrt{n}. \qquad (5.4)$$

A more general form of the central limit theorem we present without proving.

Lyapunov's theorem. If $X_1, X_2, \ldots, X_n$ – are independent random variables having similarly identical dispersions $D_i \approx D$ для $\forall i$, then with unbounded increasing of n ($n \to \infty$) the distribution law of their sum $Y = \sum_{i=1}^{n} X_i$ is unboundedly approaching to the normal law with parameters

$$m_Y = \sum_{i=1}^{n} m_i, \ \sigma_Y = \sqrt{\sum_{i=1}^{n} D_i} \ . \qquad (5.5)$$

The requirement $D_i \approx D, \forall i$ means that no one of the summands is not dominant (the effect of all X_i on the sum Y is approximately equal).

Thus, normal distribution occur when a lot of independent (or weakly dependent) random variables are summed, which are comparable by their influence on the sum dispersion. In practice, such conditions are quite often. Let we consider deviation Y of a certain parameter, for example, of a radio-electronic device from the face value. This deviation (at known assumptions) can be presented as the sum n *of* elementary deviations associated with separate reasons:

$$Y = \sum_{i=1}^{n} X_i \ ,$$

where, for example:

X_1 – deviation caused by the temperature effect;

X_2 – deviation caused by the air humidity effect;

…

X_n – deviation, caused by insufficient product material purity.

The number n of these elementary deviations is quite great, as well as the number n *of the* reasons causing the total deviation Y. Usually the summands $X_1, X_2, \ldots, X_n$ are comparable in the effect on the sum dispersion. Really, if any of the random variables $X_1, X_2, \ldots, X_n$ will produce significantly greater influence on the sum dispersion than all the others, it would be natural to take special measures to eliminate the main reason

of dispersion; if no such measures are undertaken, it can be assumed that remaining random summands are comparable in their effect (uniformly small) on the sum dispersion.

The normal law is widely spread in engineering. In most cases parameter measurement errors, control commands errors and an input errors of various quantities into the technical device are distributed by normal (or close to normal) law. Such error can be usually presented in the form of the sum of many «elementary errors» X_i, each of which is associated with a separate reason almost independent from others. Laplace and Gauss substantiated the normal law for the first time exactly in application to the theory of errors.

In practice, when summing up variables with similar distribution law, the distribution law can be considered normal, if n>10... 20.

Example. Let X – is the random variable uniformly distributed within the interval [0, 1], and is generated, for example, by the pseudo-random variables generator. Based on the central limit theorem, the variable

$$Y = \sigma(\sum_{i=1}^{12} x_i - 6) + m \qquad (5.6)$$

will have almost normal distribution law $N(m, \sigma)$ with parameters m, σ.

Lecture 6

6.1. Mathematical statistics. Basic concepts

Mathematical statistics is the science dealing with methods of processing the experimental data, obtained as a result of random phenomena observation. Each such result can be presented as a set of values assumed by one-dimensional or multidimensional random variable resulting from *n* experiments.

The experiment (***general) population*** is the set of objects from which the sample is made. Each of the objects gives a fixed value of the random variable *X*. The number of objects *N* included into the general population is called *population size.* It can consist of uncountable set of objects.

Sample – is the set $\{x_1, x_2, ..., x_n\}$ of randomly selected objects (values) from the general population. The sample size *n* is the number of objects included into the sample.

The sample features the requirement: it should adequately represent the general population, i.e. to be *to representative*. By the law of large numbers it can be stated that the sample will be *representative*, if it is drawn randomly and each of the general population objects has the same probability to get to the sample.

The elementary processing of sample consists in its sorting, that is in an arrangement of sample values as their increase: $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$. *The* sample located in increasing order of values

$$\{\hat{x}_1, \hat{x}_2, ..., \hat{x}_n\},$$

is called *as a variational series*. The element $\hat{x}_k$ of a variational series is called variants or *k*-th order statistics. So, minimum sample values is called as first order statistics,

$$\min(x_1, x_2, ..., x_n) = \hat{x}_1,$$

Maximum – *n*-th order statistics,

$$\max(x_1, x_2, ..., x_n) = \hat{x}_n.$$

Example 6.1. For studying of growth of 5 men in centimeters has appeared equal are at random selected $x_1 = 175$ $x_2 = 190$. $x_3 = 180$ $x_4 = 173$ $x_5 = 178$. This data makes sample in size $n = 5$. A variational series for the given sample looks like : $\hat{x}_1 = 173$, $\hat{x}_2 = 175$, $\hat{x}_3 = 178$, $\hat{x}_4 = 180$, $\hat{x}_5 = 190$.

One of the main tasks of mathematical statistics is to determine the distribution law of random variable *X*.

6.2. The distribution law estimate

6.2.1. Empirical cumulative distribution function

Empirical cumulative distribution function of the random variable *X* is equal to the frequency that *X* will assume the value smaller, than the function *x* argument, and is determined by the formula

$$F^*(x) = p^*(X < x) = \begin{cases} 0, & x \le \hat{x}_1, \\ \vdots & \\ \dfrac{i}{n}, & \hat{x}_i < x \le \hat{x}_{i+1} \\ \vdots & \\ 1, & x > \hat{x}_n. \end{cases} \qquad (6.1)$$

At $n \to \infty$ the empirical cumulative distribution function $F^*(x)$ converges in probability to the theoretical cumulative distribution function and is its consistent estimate, that is for all x

$$F^*(x) \xrightarrow[n\to\infty]{P} F(x).$$

The proof. We will consider event $A = (X < x)$. Then $P(A) = p = F(x)$. As

$$F^*(x) = \frac{m}{n} = p^*(A),$$

where *m* –quantity of sample values which it is less *x*,

$p^*(A)$ - frequency of event *A*

and under the Bernoulli theorem it is received

$$p^*(A) \xrightarrow[n\to\infty]{P} p(A),$$

Or

$$F^*(x) \xrightarrow[n\to\infty]{P} F(x).$$

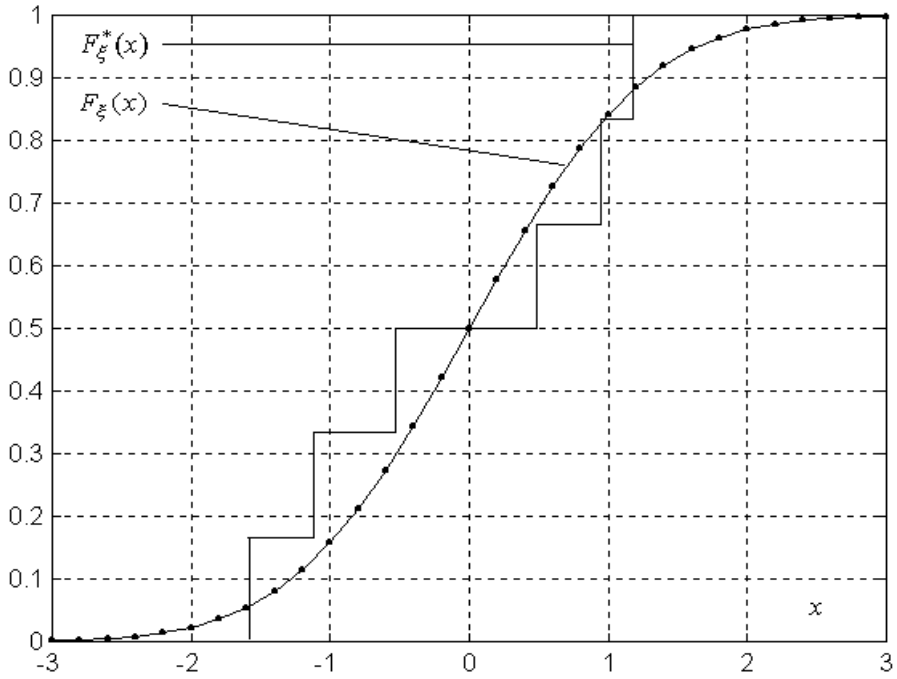


Fig. 6.1 Empirical $F^*(x)$ and theoretical $F(x)$ of cumulative distribution functions

Example 6.2. We will use the data about growth of men from an example 6.1 sections 1.2 for obtaining the empirical distribution function. Since the variational series looks like $\hat{x}_1 = 173, \hat{x}_2 = 175, \hat{x}_3 = 178, \hat{x}_4 = 180, \hat{x}_5 = 190$, then by formula (6.1) we will obtain the function presented on fig. 6.2.

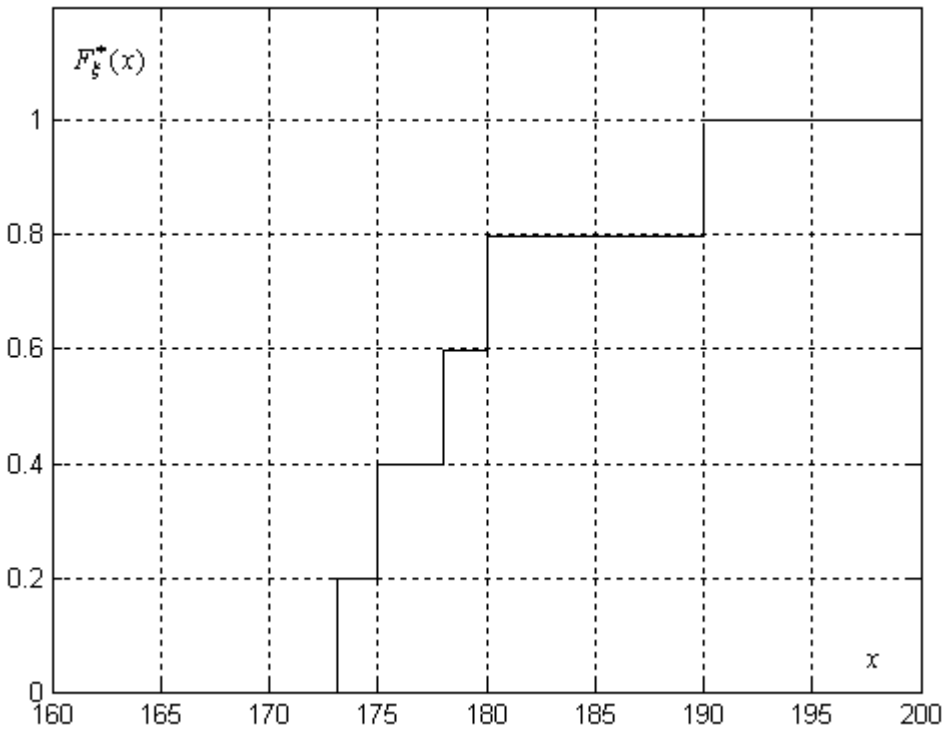


Fig. 1.2 Empirical cumulative distribution function for an example 6.2

The basic properties of function $F^*(x)$:

1. $0 \le F^*(x) \le 1$

2. $F^*(x)$ - non-decreasing step function.

3. $F^*(x) = 0$, for $x < \hat{x}_1$.

4. $F^*(x) = 1$, for $x > \hat{x}_n$.

The empirical cumulative distribution function is the best estimate of the distribution law (unbiased, consistent, efficient). The disadvantage of function $F^*(x)$ consists in its low visualization: it is difficult to determine the distribution law of the random variable *X* visually.

6.2.2. The interval statistical series

The probabilities ***interval statistical series*** is a next table:

j	A_j	B_j	h_j	ν_j	p_j^*	f_j^*
1	A_1	B_1	h_1	ν_1	p_1^*	f_1^*
⋮						
M	A_M	B_M	h_M	ν_M	p_M^*	f_M^*

Were j – is the interval number;

M – is the number of non-overlapping and adjacent to each other intervals, to which the range of values $\left[\hat{x}_1, \hat{x}_n\right]$ is divided:

$$\mathrm{M} \approx \begin{cases} \mathrm{int}\left(\sqrt{n}\right), & n \le 100, \\ \mathrm{int}\left(\left(2 \div 4\right) \cdot \lg\left(n\right)\right), & n > 100, \end{cases} \tag{6.2}$$

where int (x) – is the integer part of number x (it is desirable that n will be divided without remainder by M);

A_j, B_j - are the left and right boundaries of j-interval ($B_j = A_{j+1}$ – intervals adjacent to each other), therefore $A_1 = \hat{x}_1$ $B_M = \hat{x}_n$

h_j – is the length of j- interval $h_j = B_j - A_j$;

ν_j – is the quantity of numbers in the sample, getting to j-interval, $\sum_{j=1}^{M} \nu_j = n$;

p_j^* – is the frequency of hitting to j- interval; $p_j^* = \frac{\nu_j}{n}$ $\sum_{j=1}^{m} p_j^* = 1$

f_j^* – is the statistical probability density in j-interval $f_j^* = \frac{p_j^*}{h_j} = \frac{\nu_j}{n h_j}$.

When constructing the interval statistical probabilities array, the following methods of dividing the value range to intervals are used:

1) ***equal-interval,*** i.e. all intervals are of the same length:

$$h_j = h = \frac{\hat{x}_n - \hat{x}_1}{M}, \ \forall j; \tag{6.3}$$

$$A_j = A_{j-1} + h = \hat{x}_1 + (j-1) \cdot h, \quad j = \overline{2, M}; \tag{6.4}$$

2) ***equiprobable,*** i.e. interval boundaries are chosen so, that in each interval will be the same number of sample values (it is necessary that n will be divided without remainder by M):

$$\nu_j = \nu = \frac{n}{M},\ p_j^* = \frac{1}{M} \forall j; \tag{6.5}$$

$$A_j = \frac{\widehat{x}_{(j-1)\nu} + \widehat{x}_{(j-1)\nu+1}}{2},\ j = \overline{2, M}. \tag{6.6}$$

6.2.4. The histogram

The ***histogram*** – is the statistical image of the of probability density plot $f^*(x)$ of the random variable, it is constructed by the interval statistical array. The *histogram* represents the set of rectangles constructed, as on the base, on intervals h_j of the interval statistical array with the height equal to the statistical probability density f_j^* in the corresponding interval.

For equal-interval method, all histogram rectangles have identical width, and for equiprobable method – the identical area. The areas sum of all histogram rectangles equals to 1. The advantages of the histogram: simplicity of construction, high visualization.

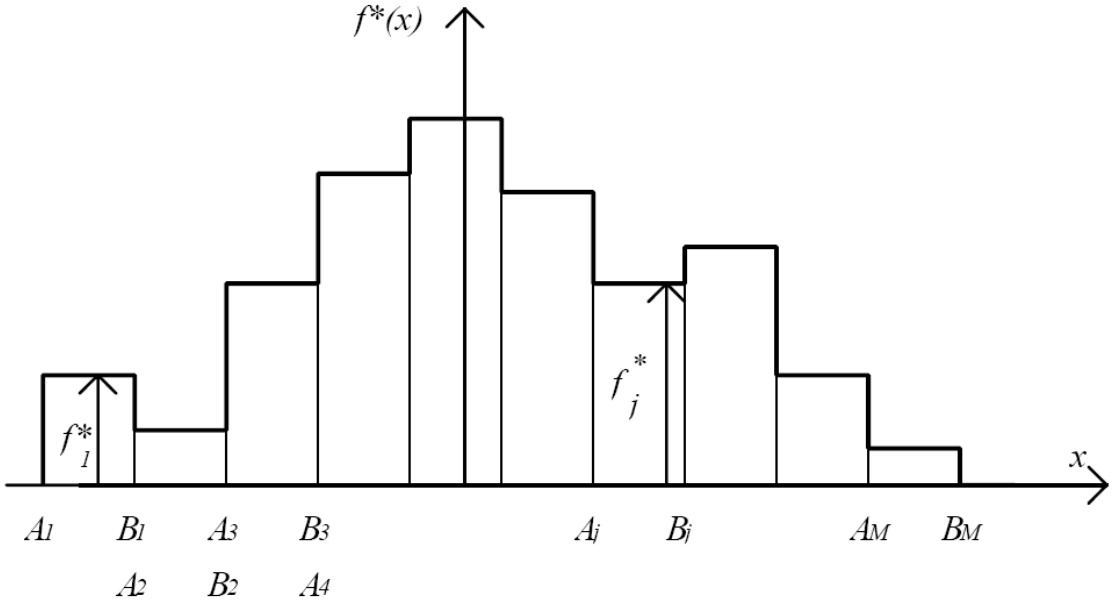


Fig. 6.3 Histogram and corresponding to it theoretical
Probability density

It is easier to determine the distribution law of the random variable *X* by the histogram shape, than by the plot of the empirical cumulative distribution function $F^*(x)$.

Let's consider properties of the histogram as $f^*(x)$ estimations of theoretical density of probability $f(x)$.

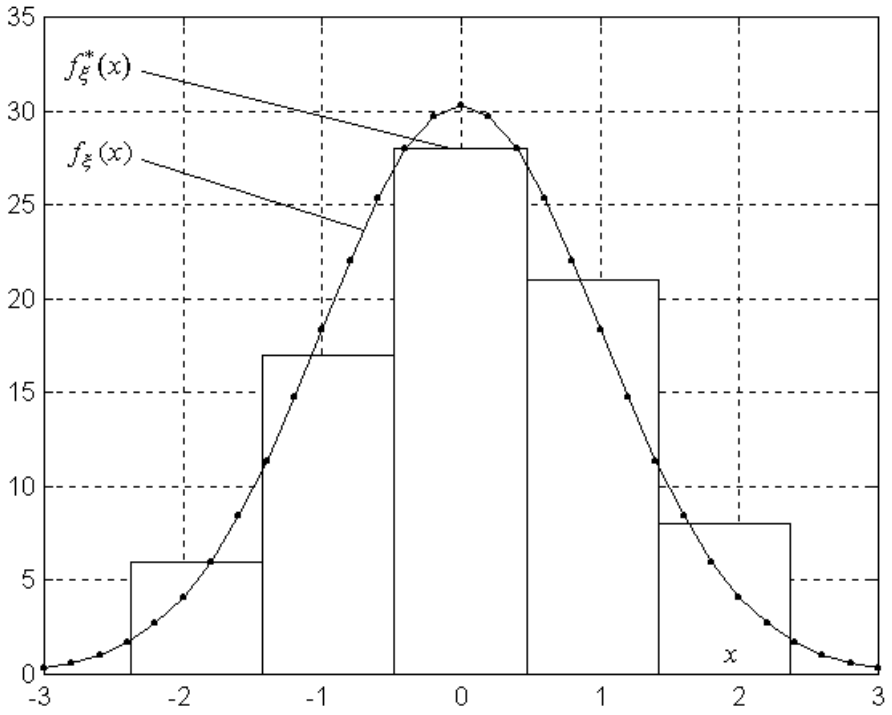


Fig. 6.3 Histogram and corresponding to it theoretical
probability density

Let $z_1, z_2, \ldots, z_l$ –points of splitting of an interval of selective values of a random variable *X* on intervals l, m_i –quantity of the sample values which have got to i th interval (z_i, z_{i+1}), n –sample size. If maximum from splitting intervals aspires to zero at increase sample size *n* the histogram converges $f^*(x)$ in probability to theoretical probability density $f(x)$ and is its consistent estimate, that is for any $\varepsilon > 0$

$$P(|\ f^*(x) - f\ (x)| > \varepsilon) \underset{n\to\infty}{\to} 0 \cdot$$

The proof. For empirical cumulative distribution function random variable ξ it is possible to write down two conditions:

$$P(|\ F_\xi^*(z_{i+1}) - F_\xi(z_{i+1})| > \varepsilon_1) \underset{n\to\infty}{\to} 0,$$

$$P(|\ F_\xi^*(z_i) - F_\xi(z_i)| > \varepsilon_2) \underset{n\to\infty}{\to} 0.$$

Then. $P(|\ F_\xi^*(z_{i+1}) - F_\xi^*(z_i) - F_\xi(z_{i+1}) + F_\xi(z_i)| > \varepsilon_3) \underset{n\to\infty}{\to} 0$ $\varepsilon_3 = \varepsilon_1 + \varepsilon_2$ Having chosen, $\varepsilon_3 = \varepsilon(z_{i+1} - z_i)\ \varepsilon > 0$ we will receive

$$P\left(\left| \frac{F_\xi^*(z_{i+1}) - F_\xi^*(z_i)}{z_{i+1} - z_i} - \frac{F_\xi(z_{i+1}) - F_\xi(z_i)}{z_{i+1} - z_i} \right| > \varepsilon \right) \underset{n\to\infty}{\to} 0.$$

If $z_{i+1} \to z_i$,

$$\frac{F_\xi(z_{i+1}) - F_\xi(z_i)}{z_{i+1} - z_i} \to F'(z_i) = f_\xi(z_i),$$

$$\frac{F_\xi^*(z_{i+1}) - F_\xi^*(z_i)}{z_{i+1} - z_i} = \frac{m}{n(z_{i+1} - z_i)} = f_\xi^*(z_i),$$

$$P(|\ f_\xi^*(z_i) - f_\xi(z_i)| > \varepsilon) \underset{\substack{n\to\infty \\ z_{i+1}\to z_i}}{\to} 0.$$

The remark. If splitting intervals do not decrease in process of increase in sample size the histogram will not be a well-founded estimation of density of probability.

Lecture 7

7.1. Point estimates of numerical characteristics

Statistical estimate Q^* of distribution parameter Q is an approximate value of the parameter calculated by the results of the experiment (by the sample). Statistical estimates are divided into point and interval ones.

The ***point*** estimate is the estimate determined by one number. The point estimate Q^* of parameter Q of random variable X is generally equal to

$$Q^* = \varphi(x_1, x_2, ..., x_n), \tag{7.1}$$

where x_i – are sample values.

It is obvious that estimate Q^* is a random variable, since it is the function of n-dimensional random variable (X_1..., X_n), where X_i – is the value of variable X in i-experiment, and values Q^* will change from sample to sample at random. Estimates should meet a number of requirements.

1. The estimate Q^* is called ***consistent,*** if with increasing of sample n size it converges in probability to the value of parameter Q:

$$Q^* \xrightarrow[n\to\infty]{p} Q \Rightarrow \lim_{n\to\infty}(P(\left|Q^* - Q\right| < \varepsilon)) = 1, \forall \varepsilon > 0. \tag{7.2}$$

The consistency is the minimal requirement to estimates.

2. The consistent estimate Q^* is called ***unbiased,*** if its mathematical expectation is precisely equal to parameter Q for any sample size:

$$\mathrm{M}[Q^*] = Q, \forall n. \tag{7.3}$$

3. The consistent unbiased estimate Q^* is ***efficient,*** if its dispersion is minimal regarding dispersion of any other estimate of this parameter:

$$\mathrm{D}\left[Q^*\right] = \min. \tag{7.4}$$

7.1.1. Mathematical expectation estimate

Based on Chebychev's theorem, as a consistent estimate of mathematical expectation, the arithmetic mean values of sample $\overline{x}$ can be used, which are called sample mean:

$$m_X^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \qquad (7.5)$$

Let's determine numerical characteristics of estimate $\overline{x}$.

$$M[\overline{x}] = M[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n}\sum_{i=1}^{n} m_X = m_X \,,$$

i.e. the estimate is unbiased.

$$D[\overline{x}] = D[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} D_X = \frac{1}{n} D_X \,. \qquad (7.6)$$

Estimate (14.5) is efficient, i.e. its dispersion is minimum, if variable *X* is distributed under the normal law.

7.1.2. The ordinary moment estimate

The consistent ***ordinary moment estimate of k***-order based on Chebychev's theorem is determined by equation

$$\alpha_k^*(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i\right)^k . \qquad (7.7)$$

7.1.3. The dispersion estimate

As the dispersion consistent estimate, the deviations squares arithmetic mean of the sample values from the sample mean can be used:

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 . \qquad (7.8)$$

Let's determine mathematical expectation of estimate S^2. Since the dispersion does not depend on the choice of the coordinate origin, we will choose it at point m_X, i.e. we pass to centered variables:

$$M[S^2] = M[\frac{1}{n}\sum_{i=1}^{n} \overset{\circ}{X}_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} \overset{\circ}{X}_i\right)^2] = M[\frac{n-1}{n^2}\sum_{i=1}^{n} \overset{\circ}{X}_i^2 - \frac{2}{n^2}\sum_{i<j}^{n} \overset{\circ}{X}_i \overset{\circ}{X}_j] =$$

$$= \frac{n-1}{n^2}\sum_{i=1}^{n} M[\overset{\circ}{X}_i^2] - \frac{2}{n^2}\sum_{i<j}^{n} M[\overset{\circ}{X}_i \overset{\circ}{X}_j] = \frac{n-1}{n^2}\sum_{i=1}^{n} D_X - \frac{2}{n^2}\sum_{i<j}^{n} K_{ij} = \frac{n-1}{n} D_X .$$

Covariance $K_{ij} = 0$, since experiments and consequently, and X_i (the value of variable X in *i-experiment*) are independent. Thus, variable S^2 is the biased estimate of dispersion, and the unbiased consistent estimate of dispersion equals:

$$S_0^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2. \qquad (7.9)$$

The variable dispersion S_0^2 is equal to

$$D[S_0^2] = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)} D^2 . \qquad (7.10)$$

For the normal distribution law of variable X formula (14.10) will be of the form

$$D[S_0^2] = \frac{2}{(n-1)} D^2 , \qquad (7.11)$$

for the uniform distribution law –

$$D[S^2] \approx \frac{0,8n+1,2}{n(n-1)} D_X^2 . \qquad (7.12)$$

The consistent unbiased estimate of ***RMSD*** is determined by equation:

$$S_0 = \sqrt{S_0^2} . \qquad (7.13)$$

If sample is taken from normal distribution it is possible to show, as S_0^2 will be ***asymptotic*** effective estimations of a dispersion σ^2.

7.1.4. The central moment estimate

The consistent estimate ***of the central moment of k***-order equals:

$$\mu_k^*(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^k . \qquad (7.14)$$

7.1.5. The probability estimate

Based on Bernoulli theorem, the unbiased consistent and efficient probability estimate of the random event *A* in the scheme of independent experiments is equal to the frequency of this event:

$$p^*(A) = \frac{m}{n}, \qquad (7.15)$$

where *m* – is the number of experiments, in which event *A* occurred;

n – is the number of the conducted experiments.

Numerical characteristics of the probability estimate $p^*(A) = p^*$ are equal to:

$$\mathrm{M}[p^*] = p(A) = p, \ \mathrm{D}[p^*] = \frac{p(1-p)}{n}. \qquad (7.16)$$

Lecture 8

8.1. The distribution parameters estimate

To calculate the distribution parameters estimates, most often method of moments and method of maximum likelihood are used.

Method of moments. Let there is sample $\{x_1..., x_n\}$ of independent random variable values with the known distribution law $f(x,Q_1,...,Q_m)$ and *m* unknown parameters $Q_1,...,Q_m$. It is necessary to calculate estimates $Q_1^*,...,Q_m^*$ of parameters $Q_1,...,Q_m$. The calculations sequence is the following:

1. To obtain analytical expressions *m* of ordinary and/or central theoretical moments:

$$\begin{aligned} \alpha_k(x,Q_1,..,Q_m) &= M\left[X^k\right], \\ \mu_k(x,Q_1,..,Q_m) &= M\left[\left(X-m_x\right)^k\right]. \end{aligned} \qquad (8.1)$$

2. To determine *m* corresponding estimates of ordinary $\alpha_k^*(x)$ and/or central $\mu_k^*(x)$ moments under formulas (7.7, 7.14).

3. To set up and solve with respect to unknown parameters $Q_1,...,Q_m$ the system from *m* equations:

$$\begin{aligned} &\alpha_k(x,Q_1,..,Q_m) = \alpha_k^*(x), \\ &\ldots \qquad\qquad\qquad\qquad , \\ &\mu_k(x,Q_1,...,Q_m) = \mu_k^*(x) \end{aligned}$$

in which theoretical moments are equated to point estimates of the corresponding moments. Each equation looks like $\alpha_k(x) = \alpha_k^*(x)$ or $\mu_k(x) = \mu_k^*(x)$. Solving this system, we obtain estimates $Q_1^*,...,Q_m^*$ of unknown parameters.

Note. The part of the equations can contain ordinary moments, and the rest – central ones.

*Example 8.1.*Find the estimates of parameters ***a*** and σ^2 of the normal general population $N(a,\sigma^2)$ by the method of moments.

Problem solution(Answer). Theoretical moments are equal $\alpha_1 = E(X) = a \;\; \mu_2 = D(X) = \sigma^2$. By equating them to the corresponding sample moments, as estimates, we obtain the known to us the sample mean and sample dispersion:

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

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

*Example 8.2.*Find the estimates of parameters ***a***, ***b*** of the uniform distribution $U(a,b)$ by the method of moments.

Solution. Theoretical moments of the uniform in distribution $U(a,b)$ are equal to $\alpha_1 = E(X) = \frac{a+b}{2} \;\; \mu_2 = D(X) = \frac{(b-a)^2}{12}$. By equating them to the corresponding sample moments, we obtain the system of the equations

$$\frac{a+b}{2} = \bar{x},$$

$$\frac{(b-a)^2}{12} = s^2 .$$

Rewriting this system in the form

$$b + a = 2\bar{x},$$

$$b - a = 2\sqrt{3}\,s,$$

We obtain estimates

$$\hat{a} = \bar{x} - \sqrt{3}\,s, \;\; \hat{b} = \bar{x} + \sqrt{3}\,s .$$

Method of maximum likelihood. According to the given method, estimates $Q_1^*,...,Q_m^*$ are obtained from the condition of the maximum for parameters $Q_1,...,Q_m$ of the positive likelihood function $L(x_1,...x_n,Q_1,...,Q_m)$.

If random variable *X* is continuous, and values x_i are independent, then

$$L(x_1,...x_n,Q_1,...,Q_m)=\prod_{i=1}^{n} f(x_i,Q_1,...,Q_n)$$

If the random variable *X* is discrete and assumes independent values x_i with probabilities $p(X=x_i)=p(x_i,Q_1,...,Q_m)$, then the likelihood function is equal to

$$L(x_1,...x_n,Q_1,...,Q_m)=\prod_{i=1}^{n} p(x_i,Q_1,...,Q_m)$$

According to this method, the system of equations can be written in two forms:

$$\frac{\partial L(x_1,...x_n,Q_1,...,Q_m)}{\partial Q_j}=0,\ j=1,2,...,m, \qquad (8.2)$$

Often this system is simplified by means of the following technique. Since any function and its logarithm reach the extremum at the same argument values, then often not the likelihood function is maximized, but its natural logarithm:

$$\frac{\partial \ln\left(L(x_1,...x_n,Q_1,...,Q_m)\right)}{\partial Q_j}=0,\ j=1,2,...,m. \qquad (8.3)$$

If one takes into account, that the likelihood function is represented in the form of the product, then its logarithm is represented in the form of the sum

$$\ln L(Q_1,...,Q_m)=\sum_{i=1}^{n}\ln f(x_i,Q_1,...,Q_m),$$

and the system of the equations will be transformed to the form

$$\sum_{i=1}^{n}\frac{\partial}{\partial Q_j}\ln f_{\xi}(x_i,Q_1,...,Q_m)=0,\ j=\overline{1,m}. \qquad (8.4)$$

Solving this system, we obtain estimates $Q_1^*,...,Q_m^*$ of unknown parameters.

*Example 8.3.*Find estimates of parameters a and σ^2 of the normal general population $N(a,\sigma^2)$ by the method of maximum likelihood.

Solution. The general population probability density is known accurate within two parameters a , σ^2 :

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

Let us maximize the logarithmic likelihood function, for what we find

$$\ln f_\xi(x_i,a,\sigma^2)=\ln(\frac{1}{\sqrt{2\pi}})-\frac{1}{2}\ln\sigma^2-\frac{(x_i-a)^2}{2\sigma^2}.$$

We find partial derivatives for the estimated parameters

$$\frac{\partial}{\partial a}(\ln f_\xi(x_i,a,\sigma^2)=\frac{x_i-a}{\sigma^2},$$

$$\frac{\partial}{\partial\sigma^2}\ln f_\xi(x_i,a,\sigma^2)=-\frac{1}{2\sigma^2}+\frac{(x_i-a)^2}{2\sigma^4}.$$

To obtain estimates, it is required to solve the system of equations (2.3) which looks as follows

$$\begin{cases}\sum_{i=1}^{n}\frac{x_i-a}{\sigma^2}=0,\\ \sum_{i=1}^{n}\frac{(x_i-a)^2-\sigma^2}{2\sigma^4}=0.\end{cases}$$

From the first equation we find

$$\hat{a}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

Substituting $\bar{x}$ for a to the second equation, we obtain

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

We see that the maximum plausible (likely) estimate of parameters a and σ^2 of the normal distribution are sample average $\bar{x}$ and the sample dispersion $\bar{s}^2$.These estimates not differ from the estimates obtained in Example 8.1by the method of moments.

Let us pay attention to the estimate structure of the parameter a :

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}\frac{1}{n}x_i .$$

This is the linear function of the sample values taken with identical weight factors $\frac{1}{n}$.

Such structure of the estimate is quite clear, since all sample values have the same dispersion, and there is no reason for weighting to them various weight factors.

*Example 8.4.*Find the estimates of parameters ***a***, ***b*** of the uniform distribution $U(a,b)$ by the method of maximum likelihood .

Solution. Since the probability density of the uniform distribution $U(a,b)$ looks as follows

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b, \\ 0, & x \le a, x \ge b, \end{cases}$$

then the likelihood function is equal to

$$L(a,b) = \begin{cases} \prod_{i=1}^{n}\dfrac{1}{b-a}, & a < x_1, x_2, \ldots, x_n < b, \\ 0, & \textit{иначе.} \end{cases}$$

Since $b > a$, and the likelihood function increases with increasing of***b*** and decreasing of***a***, then with the account for the restrictions in the likelihood function expression, we obtain the following estimates:

$$\hat{a} = x_{(1)} = min(x_1, x_2, \ldots, x_n), \quad \hat{b} = x_{(n)} = max(x_1, x_2, \ldots, x_n).$$

These estimates differ from the estimates obtained in Example 8.2 by the method of moments.

8.2 Parameter estimation by the results of unequal measurements

In a number of practical problems, it is required to solve the following problem. A certain physical value(for example, voltage), is measured by measuring devices having various accuracy of measurement. It is required to define the true voltage value by the results of these measurements.

Since more often the errors of metering devices measurements are considered to be distributed by the normal law, then the given problem can be formulated as follows. The parameter a is the average value n of normal general populations with the probability density

$$f_i(x,a,\sigma_i^2)=\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{(x-a)^2}{2\sigma_i^2}},\ i=\overline{1,n}.$$

There is the sample $x_1,x_2,\ldots,x_n$ from these populations, containing one sample value from each population. It is required to find the estimate $\widehat{a}$ of the parameter a according to the available data.

For the solution of the given problem can be used the method of maximum likelihood. The logarithmic likelihood function for one sample value x_i looks as follows

$$\ln f_i(x_i,a)=\ln(\frac{1}{\sqrt{2\pi}})-\frac{1}{2}\ln\sigma_i^2-\frac{(x_i-a)^2}{2\sigma_i^2}.$$

Hence we obtain the derivative

$$\frac{d}{da}(\ln f_i(x_i,a)=\frac{x_i-a}{\sigma_i^2}$$

and the equation for estimating

$$\sum_{i=1}^{n}\frac{x_i-a}{\sigma_i^2}=0.$$

After transforming this equation to the form

$$\sum_{i=1}^{n}\sigma_i^{-2}x_i-a\sum_{i=1}^{n}\sigma_i^{-2}=0$$

weobtain the estimate

$$\widehat{a} = \frac{1}{\sum_{j=1}^{n} \sigma_j^{-2}} \sum_{i=1}^{n} \sigma_i^{-2} x_i \ .$$

This is also a linear combination of sample values, as in example 8.3 for uniformly precise measurements, but with different weights of σ_i^{-2}. In this case, the weight of observation x_i is inversely proportional to the observation dispersion σ_i^2. At identical dispersions of observations, such estimate is reduced to the estimate, obtained in example 8.3.

Lecture 9

9.1 Interval estimate of numerical characteristics

A confidence interval for a certain parameter *Q* is the interval $(Q_L; Q_U)$ covering the value of parameter *Q* with probability γ:

$$p(Q_L < Q < Q_U) = \gamma . \tag{9.1}$$

The value γ is called the confidence probability,

$(Q_L; Q_U)$, –are the lower and upper confidence limits, respectively,

$(Q_U - Q_L)$–is the length of the confidence interval.

The confidence probability γ is selected close to 1 from the set of numbers: 0,9; 0,95; 0,975; 0,99. The confidence interval for parameter *Q* is also called the interval estimate of this parameter.

The problem of constructing the confidence interval is formulated as follows. The probability density of the general population is known within the parameter *Q*, that is, $f(x, Q)$ is known. It is required to find the interval $(Q_L; Q_U)$ of the form (9.1) by the sample $x_1, x_2, ..., x_n$ from this this population.

Since two limits of the confidence interval are determined from one equation (9.1),then there is an infinite set of intervals, satisfying this equation.

To solve the unambiguity problem, from the equation (9.1) we pass to two equations:

$$\begin{cases} P(Q > Q_U) = \alpha_1, \\ P(Q < Q_L) = \alpha_2, \end{cases} \tag{9.2}$$

where $\alpha_1 + \alpha_2 = 1 - \gamma$.

The confidence interval is called symmetric, if $\alpha_1 = \alpha_2 = \alpha = \dfrac{1-\gamma}{2}$.

The symmetric interval is constructed on the basis of the following system of the equations:

$$\begin{cases} P(Q > Q_U) = \dfrac{1-\gamma}{2}, \\ P(Q < Q_L) = \dfrac{1-\gamma}{2}, \end{cases} \tag{9.3}$$

Usually, constructing the confidence interval is imply constructing symmetric confidence interval.

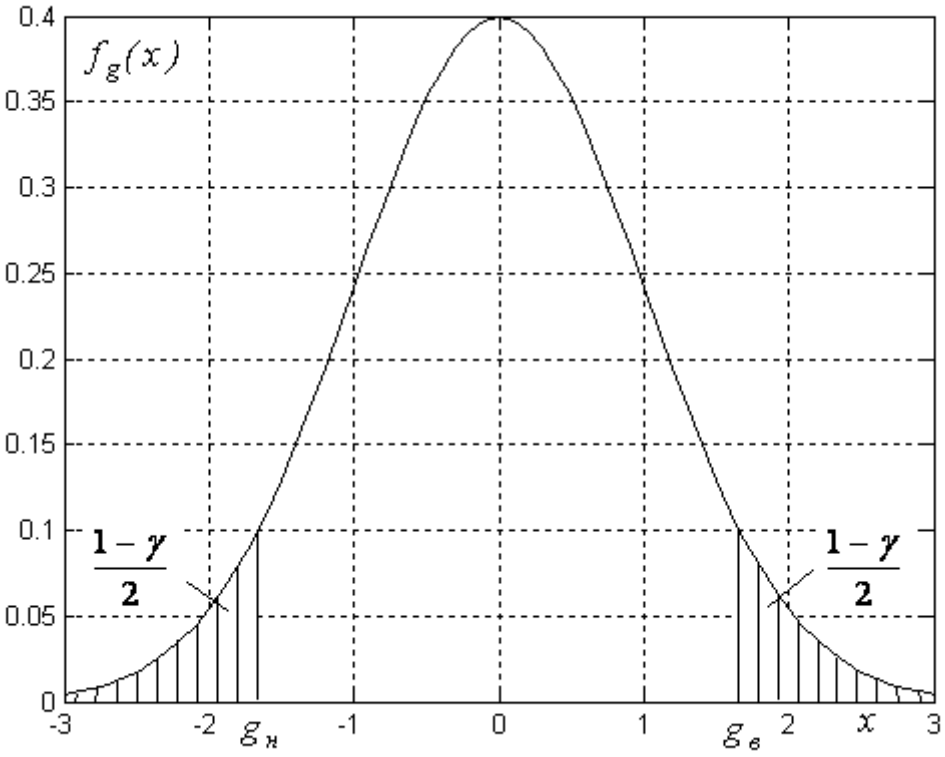


Fig.. 9.1 Illustration of the symmetric confidence interval construction technique

For constructing the symmetric confidence interval for unknown parameter *Q*, the point estimate $Q^*$is usually used. It is obvious, that for constructing the confidence interval the distribution law $f(Q^*)$ of random variable Q^* should be known. The difficulty consists in the fact that the distribution law of point estimate Q^* depends on the distribution law of variable *X* and, hence, on its unknown parameters (in particular, also on parameter *Q* itself). To solve this problem, we use the fact, that variable Q^* represents, as a rule, the sum of *n* independent equally distributed random variables

and, according to the central limit theorem, at large enough sample size n ($n > 10...20$), its distribution law can be considered asymptotically normal. In this case, the confidence interval based on law normalization of the point estimate Q^* ($f(Q^*) \approx N(a, \sigma^2)$, where $a = Q^*, \sigma^2 = D[Q^*]$), will look like:

$$I_\gamma(Q) = \left(Q^* - z_\gamma \cdot \sqrt{D\left[Q^*\right]}; Q^* + z_\gamma \cdot \sqrt{D\left[Q^*\right]} \right),$$

where z_γ is the argument value of Laplace function, $z_\gamma = \arg \Phi(\gamma)$ i.e. $\Phi(z_\gamma) = \gamma$ (see table 3.1, Lecture 3).

If the random variable distribution law is normal and this is known before the experiment, then for constructing confidence intervals for mathematical expectation and dispersion, exact formulas exist, which can be used at any sample size.

9.2 The confidence interval for mathematical expectation

The interval $I_\gamma(m_X)$ for mathematical expectation of random variable X with unknown distribution law with a large enough sample size n ($n > 10...20$) looks like

$$I_\gamma(m_X) = \left[\overline{x} - z_\gamma \frac{S_0}{\sqrt{n}}; \overline{x} + z_\gamma \frac{S_0}{\sqrt{n}} \right]. \qquad (9.4)$$

If the random variable X is distributed under the normal law with parameters m_x and σ_x, then variable $T = \frac{(\overline{x} - m_X)\sqrt{n}}{S_0}$ is distributed under the Student's law $T_1(n-1)$ with (n - 1) degree of freedom.

Student's distribution with k degrees of freedom has the following probability density:

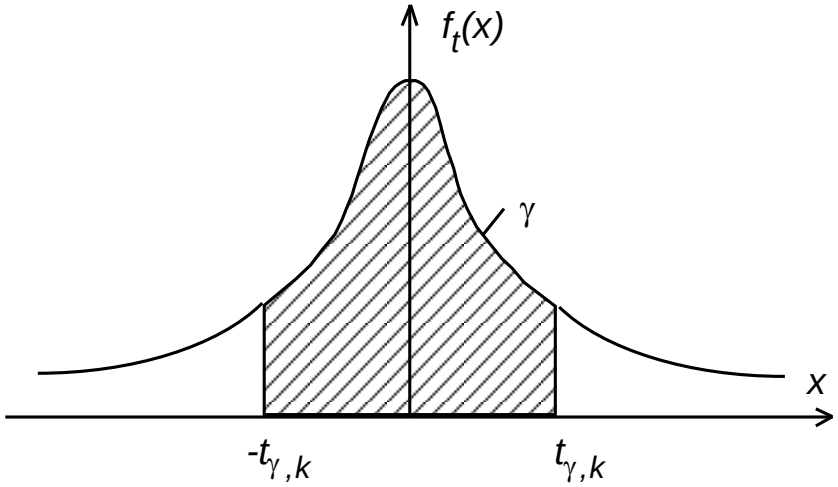
$$f_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \cdot \Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \qquad (9.5)$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ –is gamma function. For

In this case, the exact confidence interval with reliability γ for mathematical expectation of the normal random variable has the form:

$$I_\gamma(m_X) = \left[\bar{x} - t_{\gamma,n-1} \cdot \frac{S_0}{\sqrt{n}}; \bar{x} + t_{\gamma,n-1} \cdot \frac{S_0}{\sqrt{n}} \right], \qquad (9.6)$$

where the $t_{\gamma,n-1}$ – is the value taken from the Student's distribution table(see table 3.3, Lecture 3).



9.3 The confidence interval for dispersion

The interval $I_\gamma(D_X)$ for dispersion of the random variable *X* with unknown distribution law for a large enough sample size *n* ($n > 10...20$) looks like

$$I_\gamma(D_X) = \left[S_0^2 - z_\gamma \sqrt{\frac{2}{n-1}} S_0^2; S_0^2 + z_\gamma \sqrt{\frac{2}{n-1}} S_0^2 \right]. \qquad (9.7)$$

If the random variable *X* is distributed under the normal law with parameters m_x and σ_x, then variable $v = \frac{(n-1)S_0^2}{\sigma_X^2}$ is distributed by the χ^2 (Chi-square) law $H_1(n-1)$

with (*n*–1) degree of freedom, and the exact confidence interval with reliability for dispersion has the form

$$I_\gamma(D_X)=\left[\frac{(n-1)S_0^2}{\chi^2_{\frac{1-\gamma}{2},n-1}};\frac{(n-1)S_0^2}{\chi^2_{\frac{1+\gamma}{2},n-1}}\right], \qquad (9.8)$$

where $\chi^2_{\frac{1-\gamma}{2},n-1}$, $\chi^2_{\frac{1+\gamma}{2},n-1}$ – are the values taken from the table of distribution χ^2 (see table 3.2, Lecture 3).

Formulas (9.6), (9.8) can be used at any sample size *n* (for example, *n* can be less then 10), since these intervals I_γ are constructed based on the knowledge of exact distribution laws of the variables relating Q and Q^*. Besides, if the random variable *X* is distributed under the normal law and its dispersion σ_X^2 is known, then the exact interval I_γ for mathematical expectation for any sample size *n* is determined by equation (9.4), having replaced in it the estimate of RMSD S_0 by its exact value σ_X.

9.4Theconfidence interval for probability

The interval I_γ for probability of event *A* in the scheme of independent Bernoulli experiments looks like

$$p^* - z_\gamma \cdot \sqrt{\frac{p^*(1-p^*)}{n}} < p(A) < p^* + z_\gamma \cdot \sqrt{\frac{p^*(1-p^*)}{n}}, \qquad (9.10)$$

where p^* – is the frequency of occurrence of event *A*in *n* experiments $p^* = p^*(A) = \frac{m}{n}$;

m – is the number of experiments, in which event A occurred;

n – is the number of conducted experiments;

z_γ –is the argument value of Laplace function, $z_\gamma = \arg\Phi(\gamma)$ i.e. $\Phi(z_\gamma) = \gamma$ (see table 3.1, Lecture 3).

Lecture 10

10.1 Statistical check of hypotheses. Classification of hypotheses

Statistical hypothesis is any consistent set of assertions $\{H_0, H_1 \ldots, H_k\}$ concerning the random variable distribution properties. Any of assertions H_i, $i=1, \ldots, k$ is called the hypothesis alternative. The most simple hypothesis is two-alternative: $\{H_0, H_1\}$. In this case, alternative H_0 is called a null hypothesis, and H_1 – a competing hypothesis.

To test hypothesis – means to make the reasonable decision about which of the alternatives is true by the sample $x_1, x_2, \ldots, x_n$ from the general population.

If as the result of the hypothesis testing any of the alternatives is accepted, other alternatives are rejected, that is, are considered false.

The hypothesis is tested, based on the so-called hypothesis test criterion. The criterion is the rule, allowing to accept or to reject this or that alternative by the available sample. Usually they accept or reject the null hypothesis H_0.

The alternative H_i is called parametric, if it specifies the value of the certain distribution parameter θ. Otherwise, it is called non-parametric.

The multi-alternative hypothesis H is called parametric, if all its alternatives are parametric, and non-parametric, if at least one alternative is non-parametric.

The alternative H_i is called simple, if it unambiguously determines the general population distribution, and otherwise, it is called complex.

The multi-alternative hypothesis H is called simple, if all its alternatives are simple, and complex, if at least one of the alternatives is complex.

Let us present examples of hypotheses with their classification. Let the sample is taken from the normal population $N(a, \sigma^2)$ and a_0, a_1, σ_0^2 – are certain fixed numbers. We will formulate the following hypotheses:

1. $\{H_0 : a = a_0;\ H_1 : a = a_1\}$.

This is a two-alternative parametric simple hypothesis about parameter a of the normal general population.

2. $\{H_0 : a = a_0;\ H_1 : a \neq a_1\}$.

This is a two-alternative parametric complex hypothesis, since H_1 –is complex.

3. $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 > \sigma_0^2\}$.

This is a two-alternative parametric complex hypothesis about parameter σ^2 of the normal general population.

Let $f_0(x)$ – is certain completely known probability density and $f_\xi(x)$ –is the probability density of the general population. The hypothesis of the form

4. $\{H_0 :\ f_\xi(x) = f_0(x);\ H_1 :\ f_\xi(x) \neq f_0(x)\}$

is a two-alternative non-parametric complex hypothesis – the so-called hypothesis about the distribution law. Here the assertion is checked, that our sample is taken from distribution $f_0(x)$.

10.2 Significance criterion

Let the two-alternative complex hypothesis $\{H_0, H_1\}$ is under the test, where H_0 – is a simple hypothesis, and H_1 –is complex. Such hypothesis is tested by means of the so-called significance criterion.

On the basis of the significance criterion there is a certain scalar statistics $g = g(x_1,...,x_n)$, which represents deviation of the empirical (sample) data from the hypothetical ones. The significance criterion allows to establish, whether the deviation of the empirical data from the hypothetical is significant, that is, whether the value of the statistics g is significant. From here is the criterion name.

The error of the first kind (alpha error) means that hypothesis H_0 will be rejected, if it is true ("target drop-out"). The probability to make the *error of the first kind* is designated α and is called the *significance level*. In practice, most often significance

level α is chosen from the following set of small numbers: $\{0{,}1;\ 0{,}05;\ 0{,}025;\ 0{,}01\}$. The event with such a probability, can be considered almost impossible, that is, not appearing as the result of one experiment.

The error of the second kind (beta error) means that hypothesis H_0 is accepted, if it is false ("false alarm"). The probability of this kind of error is designated β. The probability not to commit the error of the second sort $(1-\beta)$ is called the *power of test*. To find the power of test, it is necessary to know the criterion probability density of the alternative hypothesis.

Let $f_g(x)$ – is the probability density of statistics g. This probability density is assumed known (provided that H_0 is true). The significance criterion looks as follows

$$P(|g| > g_{\alpha/2}) = \alpha, \tag{10.1}$$

or

$$P(g > g_{\alpha}) = \alpha, \tag{10.2}$$

or

$$P(g < g_{1-\alpha}) = \alpha, \tag{10.3}$$

where α – is significance level;

$g_{\alpha/2}$ g_{α} $g_{1-\alpha}$ are significance limits, or critical values.

The areas determined by conditions $|g| > g_{\alpha/2}$, or $g > g_{\alpha}$, or $g < g_{1-\alpha}$ are called critical areas. These areas are denoted in Fig. 10.1 – 10.3 by hatching.

Let us note, that since we use statistics distribution g provided that H_0 is true, then α represents the probability of deviation of the true hypothesis H_0 and is also called the error probability of 1st kind. The probability of acceptance of the true hypothesis H_0 is equal to $(1-\alpha)$.

Criterion (5.1) is called bilateral or criterion with bilateral critical area. Criterion (5.2) – right-sided. Criterion (5.3) – left-sided. The hypothesis is tested by the following way. The significance level α is chosen. By distribution tables of statistics g, the significance limit $g_{\alpha/2}$, or g_{α} $g_{1-\alpha}$ is selected, depending on the criterion kind. Then by the available sample and the formula for statistics g the empirical value of the statistics $g_э$ is calculated. If $|g_э| > g_{\alpha/2}$ for bilateral criterion (5.1), or for $g_э > g_{\alpha}$ for right-sided criterion (5.2), or for $g_э < g_{1-\alpha}$ for left-sided criterion (5.3) , then the tested hypothesis H_0 is rejected.

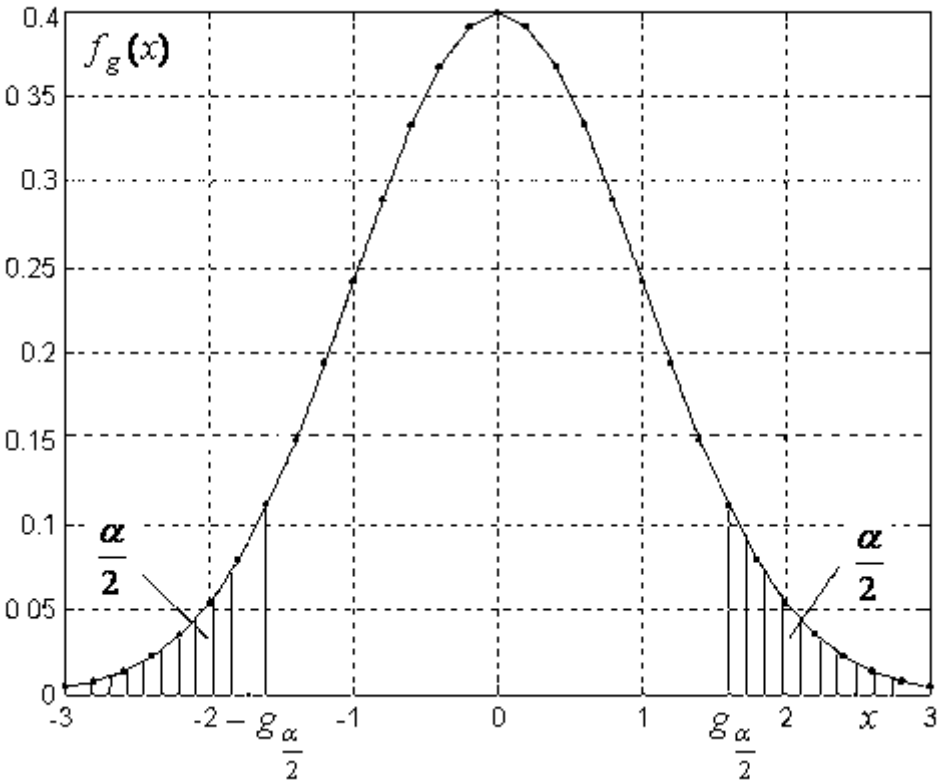


Fig. 10.1 Critical areas for bilateral significance criterion

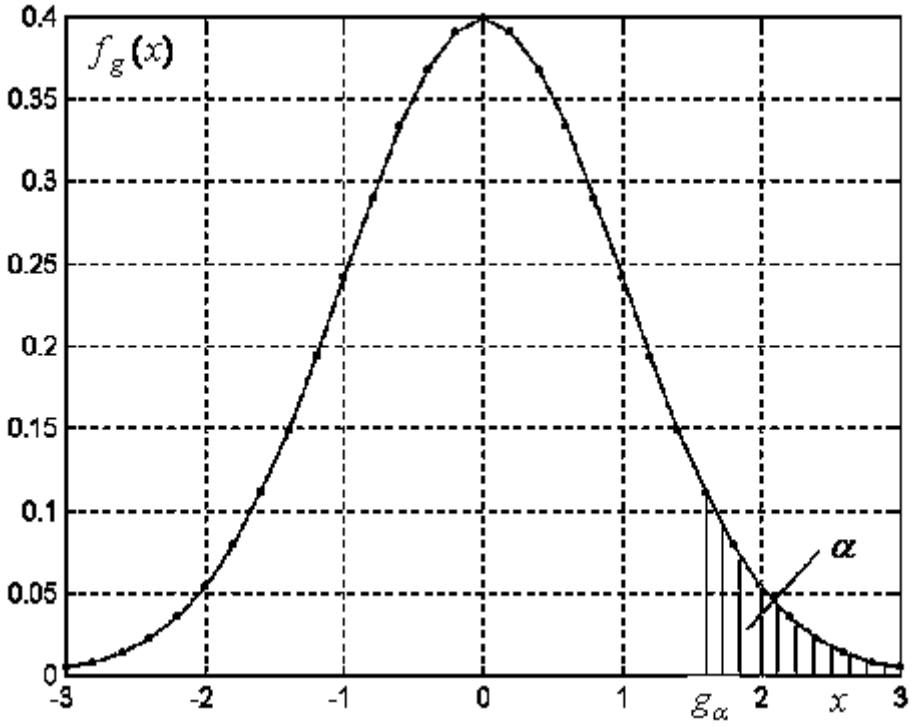


Fig. 10.2 Critical area for right-sided significance criterion

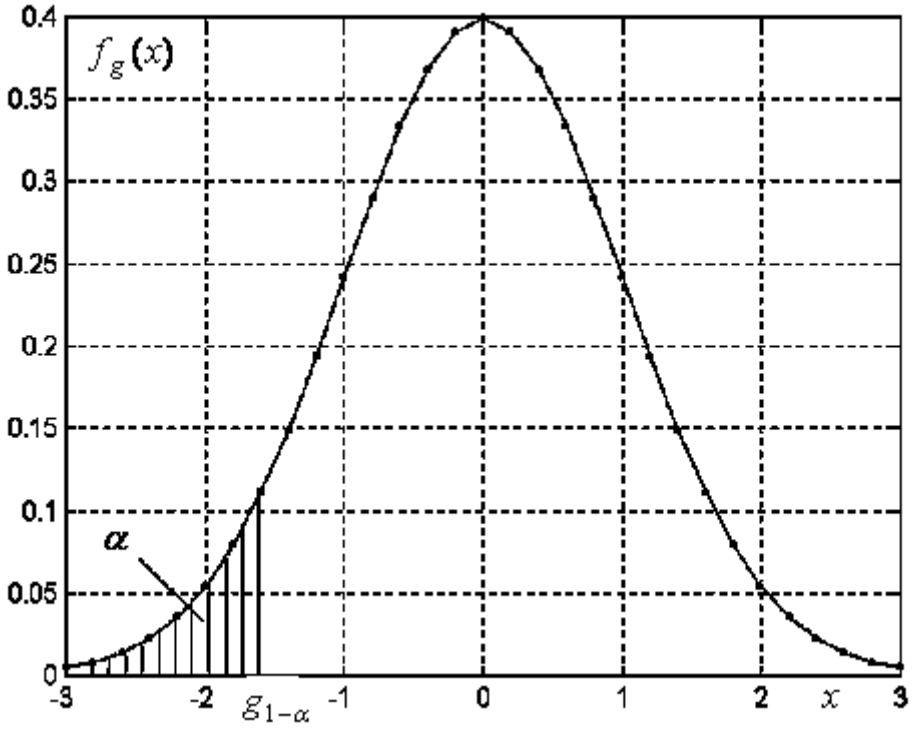


Fig. 10.3 Critical area for left-sided significance criterion

In other words, if the empirical value of statistics $g_э$ gets to the critical area, then the hypothesis under test H_0 is rejected.

Hypothesis H_0 is rejected due to the fact, that there is a contradiction between hypothetical and empirical data, which was revealed by the fact, that there was an event which would not have resulted from the single experiment.

10.3. Hypothesis testing about probabilities equality

Let two series of experiments were conducted, consisting of n_1 and n_2 experiments, respectively. In each of them, the occurrence of the same event A was registered. In the first series event A appeared in k_1 experiments, in the second — in k_2 experiments, therefore, the frequency of event A in the first series has turned out to be higher, than in the second: $p_1^* = \frac{k_1}{n_1} > p_2^* = \frac{k_2}{n_2}$. The difference between two frequencies was equal to

$$U = p_1^* - p_2^*. \qquad (10.4)$$

Question: is this discrepancy significant or not significant? Whether it specifies that in the first experiments series event A was really more probable, than in the second, or discrepancy between the frequencies should be considered random?

We put forward two-alternative hypothesis $\{H_0, H_1\}$, where:

H_0 – distinctions in probabilities do not exist, i.e. both series of experiments are conducted in identical conditions, and discrepancy U is due to random reasons,

H_1 – distinction in probabilities exists, i.e. both series of experiments are made not in identical conditions.

In this, case, null-hypothesis H_0 consists in that both series of experiments are homogeneous, and that probability p of occurrence of event A in them is the same, approximately equal to the frequency, which will be obtained, if both series are mixed

into one: $p \approx p^* = \frac{k_1 + k_2}{n_1 + n_2}$.

At large enough n_1 and n_2, each of random variables p_1^* and p_2^* is distributed almost normally, with the same mathematical expectation $m = p \approx p^*$. As for dispersions D_1 and D_2 in the first and in the second series, - they are different and equal, respectively, (14.16)

$$D_1 \approx \frac{p_1^*(1-p_1^*)}{n_1},\ D_2 \approx \frac{p_2^*(1-p_2^*)}{n_2}.$$

As a criterion, we will use random variable $U = p_1^* - p_2^*$, which also has approximately normal distribution with mathematical expectation $m_U = 0$ and dispersion $D_U = D_1 + D_2 \approx \frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}$, Whence $\sigma_U = \sqrt{D_U} \approx \sqrt{\frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}}$.

We determine critical point U_α for the given significance level α form the equation:

$$\alpha = p(U \geq U_\alpha) = 0.5 - \Phi(\frac{U_\alpha}{\sigma_U}),\ \text{I.e}\ U_\alpha = \sigma_U \cdot \arg\Phi(0.5-\alpha).$$

If the value calculated by equation (15.1) is larger than the critical value, i.e. $U > U_\alpha$, then hypothesis H_0 is rejected, otherwise there is no reason to reject it.

Lecture 11

11.1. Hypothesis test about the distribution law. Fitting criteria

Let it is required to test the hypothesis that the general population has the specified distribution. The criteria for testing such a hypothesis have obtained the name of the fitting criterion.

The hypothesis about the distribution law is put forward and tested in the following way.

1. Construct the empirical distribution function plot $F^*(x)$ by the variational series, and histograms by the interval statistical arrays (equal-interval and/or equiprobable).

2. Based on the graphs forms, put forward a two-alternative hypothesis about the intended (hypothetical) distribution law:

H_0 – variable X is distributed under a certain law:

$$f(x) = f_0(x), \qquad F(x) = F_0(x);$$

H_1 – variable X is not distributed under a certain law:

$$f(x) \neq f_0(x), \qquad F(x) \neq F_0(x),$$

where $f_0(x)$, $F_0(x)$ – is the density and cumulative distribution function of the hypothetical distribution law.

The graph of the empirical cumulative distribution function $F^*(x)$ should be similar to the graph of the cumulative distribution function $F_0(x)$ of the hypothetical law, and the histogram – to the density graph of the hypothetical distribution $f_0(x)$.

3. Calculate point estimates of mathematical expectation $\overline{x}$ and dispersion S_0^2, and, using the method of moments or method of maximum likelihood, determine the estimates of unknown parameters $Q_1^*, ..., Q_m^*$ of the hypothetical distribution law, where $s \leq 2$ – is the number of unknown parameters of the hypothetical distribution law.

The estimates of unknown parameters *a, b* of the ***uniform*** distribution can be determined by the formulas

$$a^* = \bar{x} - \sqrt{3} \cdot S_0, \qquad b^* = \bar{x} + \sqrt{3} \cdot S_0$$

or

$$a^* = \hat{x}_1, \qquad b^* = \hat{x}_n$$

where $\hat{x}_1$, $\hat{x}_n$ – are the first and last values of the variational series, respectively.

The estimate of unknown parameter λ *of the* ***exponential*** distribution can be determined by equation

$$\lambda^* = \frac{1}{\bar{x}}.$$

Estimates of unknown parameters m, σ of the ***normal*** distribution can be determined by formulas:

$$m^* = \bar{x}, \quad \sigma^* = S_0.$$

4. Test the hypothesis about the intended distribution law by means of fitting criterion.

11.2. Pirson criterion

Pirson fitting criterion (χ^2) is one of the most frequently used criteria. The hypothesis testing algorithm about the distribution law is the following.

1. Calculate the value of criterion χ^2 by the interval statistical array (equal-interval or equiprobable) by the equation:

$$\chi^2 = n\sum_{j=1}^{M} \frac{\left(p_j - p_j^*\right)^2}{p_j} = \sum_{j=1}^{M} \frac{\left(\nu_j - np_j\right)^2}{np_j}, \qquad (15.2)$$

where n – is the sample size;

M – is the number of intervals of the interval statistical array;

p_j^* – is the hitting frequency to *j*-interval;

ν_j – is the quantity of numbers in the sample, hitting to *j*-interval;

p_j – theoretical probability of the random variable hitting to *j*-interval, provided that hypothesis H_0 is true:

$$p_j = p(A_j \le X < B_j) = \int_{A_j}^{B_j} f_0(x)dx = F_0(B_j) - F_0(A_j), \qquad (15.3)$$

where $f_0(x)$, $F_0(x)$ – is the density and cumulative distribution function of the hypothetical distribution law.

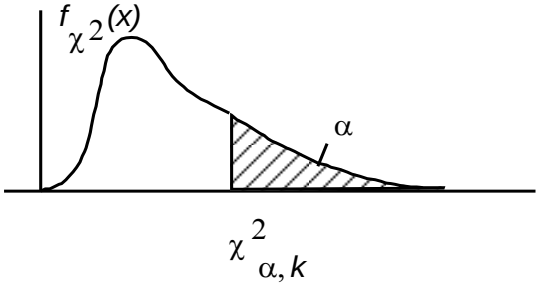
When calculating p_1 and p_M as extreme boundaries of first and last intervals A_1, B_M, theoretical boundaries of the hypothetical distribution law should be used.

The variable χ^2 is distributed under the law, which is called distribution χ^2. The given distribution does not depend on the distribution law of variable *X*, but depends on parameter *k*, which is called the number of degrees of freedom:

$$f_k(u) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\dfrac{k}{2}\right)} u^{\frac{k}{2}-1} e^{-\frac{u}{2}}, u \ge 0, \\ 0, u < 0, \end{cases} \qquad (15.4)$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$ – is gamma function.

Since the analytical expression of the probability density χ^2 is rather complicated, in practice the table of values $\chi^2_{\alpha,k}$, calculated from equation $p(\chi^2 > \chi^2_{\alpha,k}) = \alpha$ for various values *k*, is used.



2. From the distribution table of χ^2 the value $\chi^2_{\alpha,k}$ is chosen, where α – is the set significance level (α = 0,05 or α = 0,01), and k – is the number of degrees of freedom, which is determined by equation

$$k = M - 1 - s,$$

where M – is the number of summands in formula (15.2), i.e. the number of intervals of the interval statistical array;

s – *is the* number of unknown parameters of the hypothetical distribution law, the estimate of which were determined by the original sample.

3. If the value χ^2 calculated by equation (15.2) is larger than the critical value, i.e. $\chi^2 > \chi^2_{\alpha,k}$, then hypothesis H_0 is rejected, otherwise there is no reason to reject it.

11.3. Kolmogorov criterion

The hypothesis is tested

$$H_0 : F(x) = F_0(x)$$

against the alternative

$$H_1 : F(x) \neq F_0(x),$$

where $F(x)$ – is cumulative distribution function of the general population, $F_0(x)$ – is the continuous hypothetical distribution function (completely known function).

The testing algorithm for the hypothesis about the distribution law by means of Kolmogorov fitting criterion is the following.

1. Calculate the Kolmogorov criterion value, based on the empirical cumulative distribution function $F^*(x)$

$$\lambda = \sqrt{n} \cdot Z, \qquad (15.5)$$

where n – is the sample size;

Z – is the maximum deviation module of the empirical cumulative distribution function $F^*(x)$ from the hypothetical distribution function $F_0(x)$, determined by ***all*** n of values x_i of the original sample $Z = \max_{i=1}^{n} \left| F^*(x_i) - F_0(x_i) \right|$

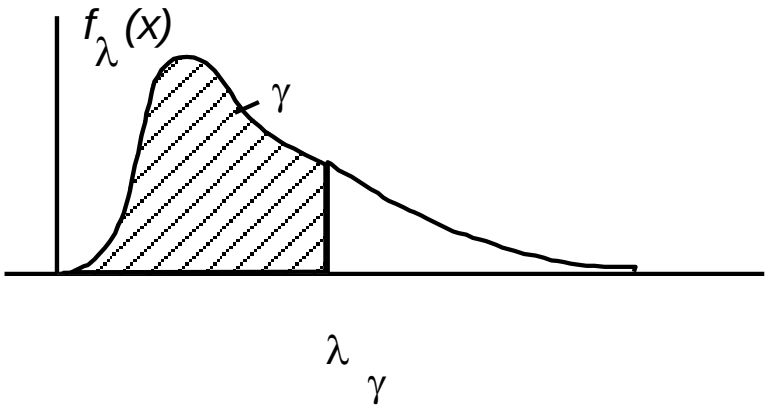
.

The value Z can be determined with sufficient accuracy by graphs of functions $F^*(x)$ and $F_0(x)$, which stand in one coordinate system.

The variable λ is distributed under Kolmogorov's law, which does not depend on the distribution law of variable X:

$$F(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2} . \qquad (15.6)$$

Since the analytical expression of the cumulative distribution function $F(\lambda)$ is rather complicated, in practice they use the table of values of λ_γ calculated from the equation $p(0 \le \lambda < \lambda_\gamma) = \gamma$.



2. Choose the critical value λ_γ, $\gamma = 1 - \alpha$, where α – is the set significance level (α = 0,05 or α = 0,01) from the Kolmogorov distribution table.

Kolmogorov distribution table

$$p(0 \le \lambda < \lambda_\gamma) = \gamma$$

λ_γ	γ
0,50	0,0361
0,54	0,0675
0,58	0,1104
0,62	0,1632
0,66	0,2236
0,70	0,2888
0,74	0,3560
0,78	0,4230
0,82	0,4880
0,86	0,5497
0,90	0,6073
0,94	0,6601
0,98	0,7079
1,02	0,7500
1,06	0,7889
1,10	0,8223
1,14	0,8514
1,18	0,8765
1,22	0,8981
1,26	0,9164
1,30	0,9319
1,34	0,9449
1,38	0,9557
1,42	0,9646
1,46	0,9718
1,50	0,9778
1,54	0,9826
1,58	0,9864
1,62	0,9895
1,66	0,9918
1,70	0,9938
1,74	0,9953
1,78	0,9965
1,82	0,9973
1,86	0,9980
1,90	0,9985
1,94	0,9989
1,98	0,9992

3. If the value λ calculated at step 1 is larger than the critical value, i.e. $\lambda > \lambda_\gamma$, then hypothesis H_0 is rejected, otherwise there is no reason to reject it.

The advantages of Kolmogorov criterion compared to criterion χ^2, are the possibility to apply it for very small sample sizes (n <20), higher "sensitivity", and therefore, less laborious calculations. The disadvantage is that empirical cumulative distribution function $F^*(x)$ should be constructed by ungrouped sampled data, what is inconvenient at large sample sizes. Besides, it should be noted, that Kolmogorov criterion can be applied only in the case when hypothetical distribution is completely known in advance from any theoretical reasons i.e. when not only the kind of cumulative distribution function $F_0(x)$ is known, but also all parameters comprised by it $Q_1,...,Q_m$. Such case rather seldom occur in practice. Usually, only the general form of function $F_0(x)$ is known from theoretical considerations, and the numerical parameters included into it are determined by the given statistical material. When applying criterion χ^2, this circumstance is taken into account by corresponding decreasing of the number of degrees of freedom of k distribution. Kolmogorov's criterion does not provide this matching. If this criterion is still applied in those cases when parameters of hypothetical distribution are determined by statistical data, it gives knowingly low estimated λ values; therefore we in some cases, we risk to accept hypothetical distribution as a plausible hypothesis, which actually is ill-agreed with experimental data.

Lecture 12

12.1 Fitting criterion Mises-Smirnov criterion

The two-alternative hypothesis is under test

$$\left\{H_0 : F_\xi(x) = F_0(x);\ \ H_1 : F_\xi(x) \neq F_0(x)\right\}.$$

The quantitative measure of the deviation of the empirical data from the hypothetical is quantity

$$\omega^2 = \int_{-\infty}^{\infty} \left(F_\xi^*(x) - F_0(x)\right)^2 dF_0(x), \qquad (12.1)$$

where $F_\xi^*(x)$ – is the empirical distribution function . We obtain the expression for the numerical calculation of statistics ω^2 in the assumption that the hypothetical distribution function $F_0(x)$ is continuous and derivative $F_0'(x)$ exists (the probability density). Taking into account expression for the empirical distribution function, we will break the real straight line into the intervals

$$(-\infty, x_{(1)}),\ [x_{(1)}, x_{(2)}),\ \ldots,\ [x_{(n-1)}, x_{(n)})\ [x_{(n)}, +\infty)$$

where $x_{(k)}$ – is the order statistics, and we calculate integral (5.6) by these intervals. We obtain

$$\omega^2 = \int_{-\infty}^{x_{(1)}} \left(F_0(x)\right)^2 dF_0(x) + \sum_{k=1}^{n-1} \int_{x_{(k)}}^{x_{(k+1)}} \left(\frac{k}{n} - F_0(x)\right)^2 dF_0(x) + \int_{x_{(n)}}^{\infty} \left(1 - F_0(x)\right)^2 dF_0(x) =$$

$$= \frac{\left(F_0(x)\right)^3}{3}\Bigg|_{-\infty}^{x_{(1)}} + \sum_{k=1}^{n-1} \frac{\left(F_0(x) - k/n\right)^3}{3}\Bigg|_{x_{(k)}}^{x_{(k+1)}} - \frac{\left(1 - F_0(x)\right)^3}{3}\Bigg|_{x_{(n)}}^{\infty} =$$

$$= \frac{F_0^3(x_{(1)})}{3} + \sum_{q=1}^{n-1} \frac{\left(F_0(x_{(q+1)}) - q/n\right)^3}{3} - \sum_{k=1}^{n-1} \frac{\left(F_0(x_{(k)}) - k/n\right)^3}{3} + \frac{\left(1 - F_0(x_{(n)})\right)^3}{3}.$$

It is easy to notice that the first summand in the last expression can be included into the first sum, and the last summand – into the second sum, that is to write down

$$\omega^2 = \sum_{q=0}^{n-1} \frac{\left(F_0(x_{(q+1)}) - q/n\right)^3}{3} - \sum_{k=1}^{n} \frac{\left(F_0(x_{(k)}) - k/n\right)^3}{3}.$$

If to introduce a new variable of summation $k = q+1$ in the first sum, we will obtain

$$\omega^2 = \sum_{k=1}^{n} \frac{\left(F_0(x_{(k)}) - (k-1)/n\right)^3}{3} - \sum_{k=1}^{n} \frac{\left(F_0(x_{(k)}) - k/n\right)^3}{3}.$$

Having denoted $V = F_0(x_{(k)}) - k/n$, we will have

$$\omega^2 = \sum_{k=1}^{n} \frac{\left(V + 1/n\right)^3}{3} - \sum_{k=1}^{n} \frac{V^3}{3} = \sum_{k=1}^{n} \frac{\left(V + 1/n\right)^3 - V^3}{3} =$$

$$= \sum_{k=1}^{n} \frac{3V^2/n + 3V/n^2 + 1/n^3}{3} = \frac{1}{n} \sum_{k=1}^{n} \left(V^2 + \frac{V}{n} + \frac{1}{3n^2} \right).$$

Supplementing the expression under the sum sign to the full square by V, we will obtain

$$\omega^2 = \frac{1}{n} \sum_{k=1}^{n} \left(V^2 + 2\frac{V}{2n} + \frac{1}{4n^2} - \frac{1}{4n^2} + \frac{1}{3n^2} \right) = \frac{1}{n} \sum_{k=1}^{n} \left[\left(V + \frac{1}{2n} \right)^2 + \frac{1}{12n^2} \right].$$

Since

$$V + \frac{1}{2n} = F_0(x_{(k)}) - \frac{k}{n} + \frac{1}{2n} = F_0(x_{(k)}) - \frac{2k-1}{2n},$$

then we finally have

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^{n} \left(F_0(x_{(k)}) - \frac{2k-1}{2n} \right)^2.$$

The statistics of criterion ω^2 looks as follows

$$z = n\omega^2 = \frac{1}{12n} + \sum_{k=1}^{n} \left(F_0(x_{(k)}) - \frac{2k-1}{2n} \right)^2. \qquad (12.2)$$

For statistics z (5.7) at $n \to \infty$ the limiting distribution exists for which Tables of percentage deviations (Tab. 5.2) are made. The criterion ω^2 is right-sided:

$$P(z > z_\alpha) = \alpha .$$

Table 12.1.

Percentage deviations of the statistics limiting distribution z,

$$P(z > z_\alpha) = \alpha$$

α	0.01	0.02	0.03	0.04	0.05
z_α	0.74	0.62	0.55	0.50	0.46

12.2 Neumann-Pirsona criterion

Let us consider the two-alternative hypothesis $\{H_0, H_1\}$. Let S – is the sample space $X = (x_1, x_2, ..., x_n)$. The test of the formulated hypothesis is reduced to the sample space S partition to two areas G_0 and G_1. If the concrete sample $X = (x_1, x_2, ..., x_n)$ gets to the area G_0, the hypothesis H_0 is accepted, and otherwise, the hypothesis H_1 is accepted. When making the decision, the following errors are possible: *the error of first kind*, when the hypothesis H_0 is true, but is rejected, and *the error of second kind*, when the hypothesis H_0 is not true, but is accepted.

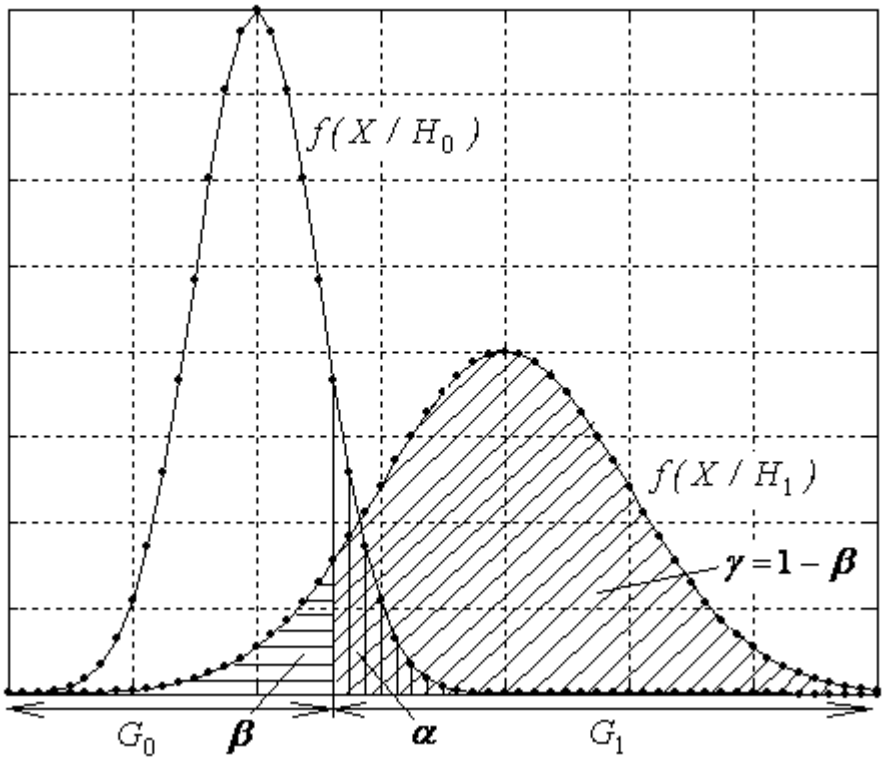


Fig. 12.1 Illustration of probabilities of errors of first and second kind

The problem of testing the two-alternative hypothesis is actual in the radar-location at detecting of air targets when the hypothesis H_0 – the target is present, and the hypothesis H_1 – the target is absent. In this case *the error of first kind* is called the error of false clear (target drop out), and the error of second kind – the false alarm error.

Let α and β – are probabilities of errors of first and second kind, respectively, and $f(X/H_0)$, $f(X/H_1)$ – are the sample probability density, provided the condition of the validity of hypotheses H_0 and H_1, respectively. Then these probabilities are determined by formulas

$$\alpha = P(X \in G_1 / H_0) = \int_{G_1} f(X/H_0)dX,$$

$$\beta = P(X \in G_0 / H_1) = \int_{G_0} f(X/H_1)dX.$$

The quantity $\gamma = 1 - \beta$ is called the strength of criterion, corresponding to partition of $G_0\, G_1$. It represents the probability to reject the false hypothesis and is determined by the formula

$$\gamma = 1 - \beta = P(X \in G_1 / H_1) = \int_{G_1} f(X / H_1) dX \quad .$$

The illustration of probabilities of errors of first and second kind and the strength of criterion is presented in Fig. 12.1.

When testing the two-alternative hypothesis by the significance criterion, we specify the small error probability of first kind (the significance value α, see section 5.2), and we do not control the error probability of second kind β. At the same time, both errors of first and second kind are undesirable. Neumann and Pirson have offered the approach to hypothesis testing, according to which the certain small error probability of the first kind is specified, and the error probability of second kind β is minimized (the strength of criterion $\gamma = 1 - \beta$ is maximized). At such approach, the following optimizing problem is solved:

$$\int_{G_1} f(X / H_0) dX = \alpha = const \quad ,$$

$$\gamma = \int_{G_1} f(X / H_1) dX \to max \quad .$$

The criterion of hypothesis testing obtained as the result of this problem solving, is called (and is) the most powerful in comparison with other criteria.

In *Neumann-Pirson* lemma it is proved, that the optimum area G_1 is the area, for which

$$\{X : f(X / H_1) \geq k f(X / H_0)\}, \qquad (12.3)$$

Criterion (5.14) we can write down in the form of the significance criterion

$$P(l(X) \geq k / H_0) = \alpha, \qquad (12.4)$$

where the statistics $l(X)$ is determined by the expression

$$l(X)=\frac{f(X/H_1)}{f(X/H_0)}$$

and is called the likelihood ratio. Criterion (5.15) is called the Neumann-Pirson criterion for testing the two-alternative simple hypothesis. For hypothesis test it is necessary to obtain statistics distribution $l(X)$ provided the condition of the validity of hypothesis H_0 and to find the significance limit k for this statistics by the Table of percentage deviations of statistics distribution $l(X)$ at the significance level α. If the empirical value $l_э(X)$ of the statistics satisfies the inequality $l_э(X)\geq k$, then hypothesis H_0 is rejected.

Instead of the likelihood ratio it is possible to use the logarithmic likelihood ratio, since if $P(l(X)\geq k/H_0)=\alpha$, then $P(\ln l(X)\geq k'/H_0)=\alpha$, where k' –is a certain new threshold.

Example 5.1. We will test the hypothesis $\{H_0: a=a_0;\ H_1: a=a_1\}$ about the mathematical expectation a of the normal general population $N(a,\sigma^2)$ at the known dispersion σ^2 by the Neumann-Pirson criterion, where a_0, a_1 – are certain numbers. We will use the logarithmic relation of credibility. Since

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

(see example 2.3 of section 2.2), then

$$f(X/H_0)=\frac{1}{\sqrt{(2\pi\sigma^2)^n}}exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x-a_0)^2\right),$$

$$\ln l(X)=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_1)^2+\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_0)^2=\frac{a_1-a_0}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{n(a_1^2-a_0^2)}{2\sigma^2}=$$

$$=\frac{n(a_1-a_0)}{\sigma^2}\bar{x}-\frac{n(a_1^2-a_0^2)}{2\sigma^2},$$

where $\bar{x} = (\sum_{i=1}^{n} x_i) / n$.

Since the statistics $\bar{x}$ under the condition of validity of hypothesis H_0 has normal distribution $N(a_0, \frac{\sigma^2}{n})$ (see section (3.4)), then to determine the significance limit k it is required to use the Tables of this distribution. Since the Tables are usually made for distribution $N(0,1)$, it is expedient to pass to the statistics $u = \frac{\bar{x} - a_0}{\sigma}\sqrt{n}$ having distribution $N(0,1)$, and to use the criterion

$$P(\bar{x} \geq h) = \alpha,$$

where

$h = a_0 + \frac{\sigma}{\sqrt{n}} u_\alpha$,

u_α – is the 100α percent deviation of distribution ($N(0,1)$).

If the empirical value of statistics $\bar{x}$ satisfies to the inequality $\bar{x} \geq h$, then hypothesis H_0 is rejected.

Lecture 13

13.1 Hypothesis test about the mathematical expectation at the known dispersion

There is the sample $x_1, x_2, \ldots, x_n$ from the normal distribution $N(a, \sigma^2)$ and the dispersion σ^2 is known, and it is required to test the hypothesis about the mathematical expectation $H_0 : a = a_0$, where a_0 – is a known number.

For the hypothesis testing the statistics is used

$$u = \frac{\bar{x} - a}{\sigma} \sqrt{n} .$$

If H_0 is true, that is $a = a_0$, then $u \in N(0,1)$. For testing the hypothesis of the kind $\{H_0 : a = a_0;\ H_1 : a \neq a_0\}$ the bilateral significance test is used $P(|u| > u_{\alpha/2}) = \alpha$ (Fig. 13.1),

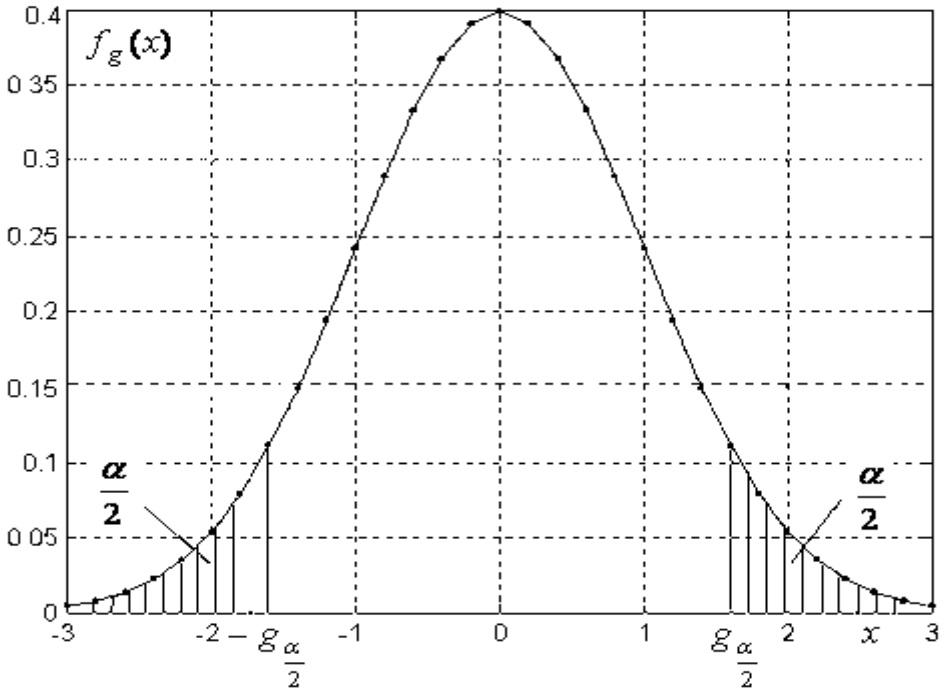


Fig. 13.1 Critical areas for bilateral significance criterion

for the hypotheses $\{H_0 : a = a_0;\ H_1 : a > a_0\}$ – the right-sided criterion $P(u > u_\alpha) = \alpha$ (Fig. 13.2),

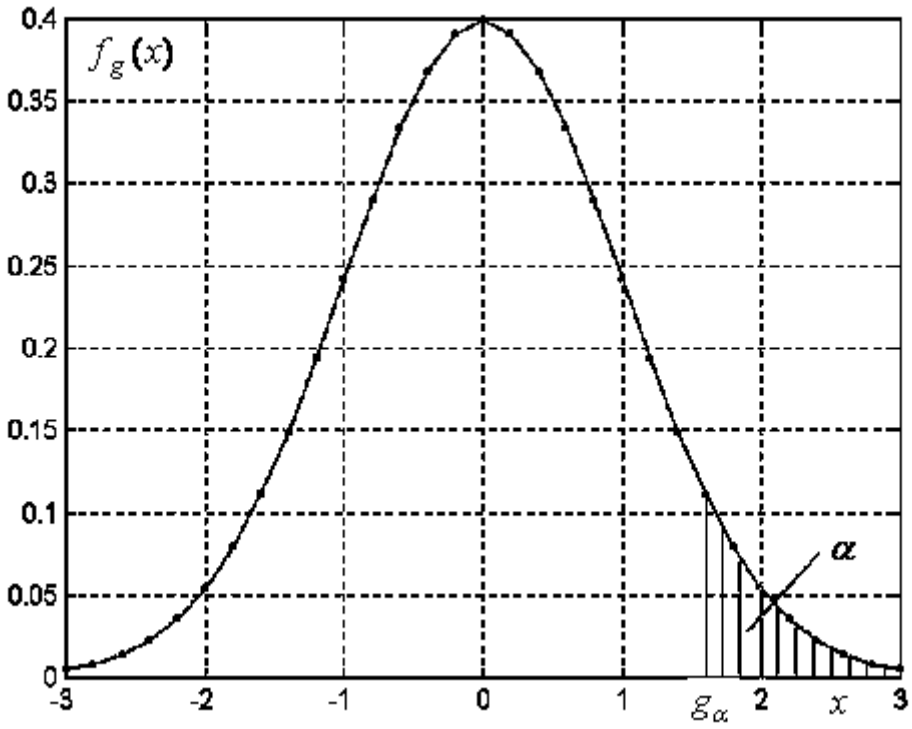


Fig. 12.2 Critical area for right-sided significance criterion

for the hypotheses $\{H_0 : a = a_0;\ H_1 : a < a_0\}$ – the left-sided criterion $P(u < u_{1-\alpha}) = \alpha$ (Fig. 13.3).

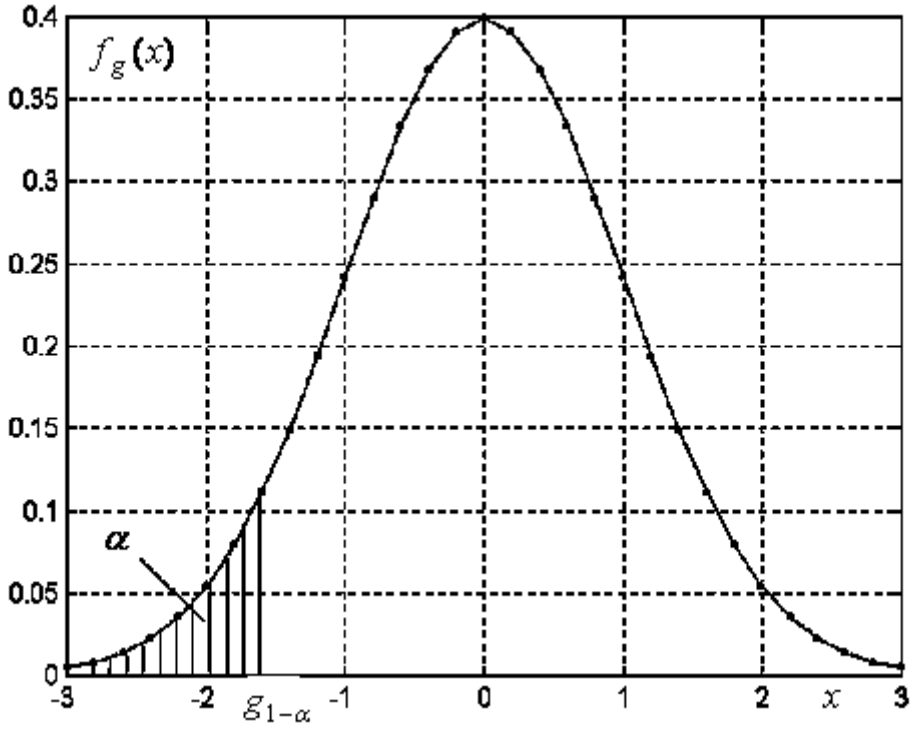


Fig. 13.3 Critical area for left-sided significance criterion

Here $u_{\alpha/2}$, u_{α}, $u_{1-\alpha}$ – $100\dfrac{\alpha}{2}$ -, 100α -, $100(1-\alpha)$ - are percentage distribution deviations $N(0,1)$, respectively (Table 13.1).

There are table satisfying to equality $P(u > u_{\alpha}) = \alpha$,

Table 13.1

Percentage deviations of normal distribution $N(0,1)$,

α	0,0010	0,005	0,010	0,015	0,020	0,030	0,040	0,050
u_{α}	3,0902	2,5758	2,3264	2,1701	2,0938	1,8808	1,7507	1,6449

13.2 Hypothesis test about the mathematical expectation at the unknown dispersion

There is the sample $x_1, x_2, \ldots, x_n$ from the normal distribution $N(a, \sigma^2)$ and the dispersion σ^2 is unknown, and it is required to test the hypothesis about the mathematical expectation $H_0 : a = a_0$, where a_0 – is the known number.

For the hypothesis test the statistics is used

$$t = \frac{\bar{x} - a}{\bar{s}} \sqrt{n-1} = \frac{\bar{x} - a}{s} \sqrt{n}.$$

If H_0 is true, then $t \in T_1(n-1)$. For testing the hypothesis $\{H_0 : a = a_0;\ H_1 : a \neq a_0\}$ the bilateral significance test is applied $P(|t| > t_{\alpha/2}) = \alpha$ (Fig. 13.1), for the hypotheses $\{H_0 : a = a_0;\ H_1 : a > a_0\}$ – the right-sided criterion $P(t > t_\alpha) = \alpha$ (Fig. 13.2), for the hypotheses $\{H_0 : a = a_0;\ H_1 : a < a_0\}$ – the left-sided criterion $P(t < t_{1-\alpha}) = \alpha$ (Fig. 13.3). Here $t_{\alpha/2}$, t_α, $t_{1-\alpha}$ – $100\frac{\alpha}{2}$ -, 100α - and $100(1-\alpha)$ - are percentage distribution deviations $T_1(n-1)$, respectively. In Fig. 13.1, 13.2, 13.3 instead of the statistics g it is required to consider the statistics t.

13.3 Hypothesis test about the dispersion at the known mathematical expectation

The sample $x_1, x_2, \ldots, x_n$ is taken from distribution $N(a, \sigma^2)$, mathematical expectation a is known, and it is required to test the hypothesis about dispersion $H_0 : \sigma^2 = \sigma_0^2$, where σ_0^2 – is the known number.

For hypothesis testing the statistics is used

$$v = \frac{n\bar{s}_0^2}{\sigma^2}.$$

If H_0 is true $v \in H_1(n)$. For testing the hypothesis $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 \neq \sigma_0^2\}$ the bilateral significance test is used, which due to asymmetry of the distribution $H_1(n)$ looks like

$$\begin{cases} P(v > v_{\alpha/2}) = \dfrac{\alpha}{2}, \\ P(v < v_{(2-\alpha)/2}) = \dfrac{\alpha}{2}. \end{cases}$$

The criterion critical region is denoted by dash lines in Fig. 13.4. The quantities $v_{\frac{\alpha}{2}}$, $v_{\frac{(2-\alpha)}{2}}$ – $100\dfrac{\alpha}{2}$ - and $100\dfrac{2-\alpha}{2}$ - *are p*ercentage distribution deviations $H_1(n)$.

For testing the hypothesis $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 > \sigma_0^2\}$ the right-sided criterion is used $P(v > v_\alpha) = \alpha$, for the hypotheses $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 < \sigma_0^2\}$ – the left-sided criterion $P(v < v_{1-\alpha}) = \alpha$,

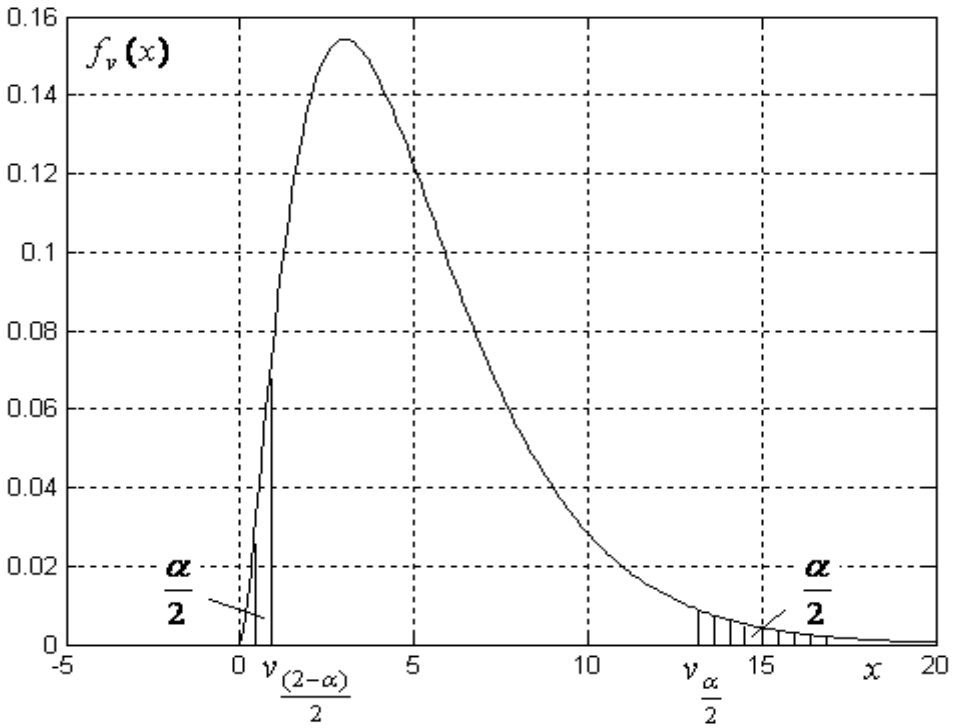


Fig. 13.4 The critical region for the hypothesis about the normal general population dispersion

13.4 Hypothesis test about dispersionat the unknown mathematical expectation

The sample $x_1, x_2, \ldots, x_n$ is taken from the distribution $N(a, \sigma^2)$, mathematical expectation a is unknown, and it is required to test the hypothesis about the dispersion $H_0 : \sigma^2 = \sigma_0^2$, where σ_0^2 – is the known number.

For hypothesis test the statistics is used

$$w = \frac{n\bar{s}^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2}.$$

If H_0 is true, then $w \in H_1(n-1)$. For testing hypothesis $\{H_0 : \sigma^2 = \sigma_0^2;$ $H_1 : \sigma^2 \neq \sigma_0^2\}$ the bilateral significance test is used, which due to asymmetry of the distribution $H_1(n-1)$ looks like

$$\begin{cases} P(w > w_{\alpha/2}) = \alpha/2, \\ P(w < w_{(2-\alpha)/2}) = \alpha/2. \end{cases}$$

The illustration of the given criterion is the same, as in Fig. 13.4, with replacement of v by w. The quantities $w_{\alpha/2}$, $w_{(2-\alpha)/2}$ – $100\frac{\alpha}{2}$ - and $100\frac{2-\alpha}{2}$ - *are p*ercentage distribution deviations $H_1(n-1)$.

For testing the hypothesis $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 > \sigma_0^2\}$ the right-sided criterion is used $P(w > w_\alpha) = \alpha$, for the hypotheses $\{H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 < \sigma_0^2\}$ – the left-sided criterion $P(w < w_{1-\alpha}) = \alpha$.

13.5 Hypothesis test about mathematical expectation equality

There are two samples of different volumes $x_1,\dots,x_m$ and $y_1,\dots,y_n$ from two normal general population $N(a_1,\sigma^2)$ and $N(a_2,\sigma^2)$ respectively, on the assumption that dispersions of general populations are equal to one another, but are unknown. It is required to test the hypothesis that mathematical expectations of these distributions are equal, that is, to test the two-alternative parametrical complex hypothesis

$$\{H_0 : a_1 = a_2;\ H_1 : a_1 \neq a_2\}.$$

To solve this problem, we consider t-Student's criterion, which is based on the comparison of sample mean. Let $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ – is the sample mean of the first general population, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – is the sample mean of the second general population. As is known,

$$\bar{x} \in N(a_1, \frac{\sigma^2}{m}),\ \bar{y} \in N(a_2, \frac{\sigma^2}{n}).$$

Then

$$z = \bar{x} - \bar{y} \in N(a_1 - a_2, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}).$$

If the hypothesis under test H_0 is true, that is $a_1 = a_2$, then

$$z \in N(0, \frac{\sigma^2 (m+n)}{mn})$$

and

$$u = \frac{z}{\sqrt{D(z)}} = \frac{(\bar{x} - \bar{y})\sqrt{mn}}{\sigma\sqrt{m+n}} \in N(0,1).$$

Since σ in the given expression is unknown, statistics u is inapplicable for the hypothesis testing, and we will continue searching the appropriate statistics. We know that

$$v_x = \frac{m\bar{s}_x^2}{\sigma^2} \in H_1(m-1), \qquad v_y = \frac{n\bar{s}_y^2}{\sigma^2} \in H_1(n-1).$$

Then statistics

$$v = v_x + v_y = \frac{m\bar{s}_x^2 + n\bar{s}_y^2}{\sigma^2} \in H_1(m+n-2).$$

Let's note, that v does not depend on u, since v_x and v_y are independent from u. Then statistics

$$t = \frac{u}{\sqrt{v}}\sqrt{m+n-2} \in T_1(m+n-2).$$

Substituting here the expressions for u and v, We obtain

$$t = \frac{(\bar{x}-\bar{y})\sqrt{mn(m+n-2)}}{\sqrt{m+n}\sqrt{m\bar{s}_x^2 + n\bar{s}_y^2}} \in T_1(m+n-2). \qquad (13.1).$$

Statistics t does not contain unknown parameters and its distribution law is known. Hence, it is suitable for testing the formulated hypothesis. The hypothesis is tested in the following way. Having set by the significance equation α, by the Table of percentage distribution deviations $T_1(m+n-2)$ we find the quantity $t_{\alpha/2}$, satisfying to equality $p(|t| > t_{\alpha/2}) = \alpha$. Then we find the empirical value of statistics $t_э$ by formula (13.1). If it will appear, that $|t_э| > t_{\alpha/2}$, then the hypothesis H_0 is rejected to the favor of the hypothesis H_1.

For alternatives $H_1 : a_1 > a_2$ and $H_1 : a_1 < a_2$, the right-sided and left-sided significance criteria, respectively, are used.

13.6 Hypothesis test about the dispersions equality

There are two samples of different volumes $x_1,\dots,x_m$ and $y_1,\dots,y_n$ from two normal general populations $N(a_1,\sigma_1^2)$ and, $N(a_2,\sigma_2^2)$, respectively, under conditions, when all parameters are unknown. It is required to test the hypothesis that dispersions of these distributions are equal, that is to check up the two-alternative parametrical hypothesis $\{H_0,H_1\}$, where

$$H_0:\sigma_1^2=\sigma_2^2.$$

For testing this hypothesis, the statistics is used (the big dispersion choose as numerator)

$$F=\frac{s_x^2}{s_y^2},$$

where $s_x^2,\ \ s_y^2$ – are unbiased sample dispersions

$$s_x^2=\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2,$$

$$s_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2.$$

It is easy to show, that if H_0 is true, then $f\in F_1(m-1,n-1)$. Really, since

$$v=\frac{(m-1)s_x^2}{\sigma_1^2}\in H_1(m-1),\quad w=\frac{(n-1)s_y^2}{\sigma_2^2}\in H_1(n-1),$$

and $\sigma_1^2=\sigma_2^2$, then

$$F=\frac{v}{m-1}:\frac{w}{n-1}=\frac{s_x^2}{s_y^2}\in F_1(m-1,n-1).$$

Variable F satisfies to F-distribution with (n_1–1, n_2–1) freedom degrees. The critical area gets out as follows. For a significance value under the α F-distribution table it is defined critical value $F_{\alpha/2;n_1-1,n_2-1}$. If F, calculated on sample, more than this

critical value $F_{\alpha/2;n_1-1,n_2-1}$ hypothesis H_0 should be rejected.

Lecture 14

14.1 Two-dimensional random variables. The two-dimensional distribution law

Two-dimensional random variable (*X, Y*) or system of two random variables – set of two one-dimensional random variables which accept values as a result of carrying out of the same experience.

Two-dimensional random variables are characterised by sets of possible values Ω_X, Ω_Y the a component and the joint (two-dimensional) law of distribution. Depending on the kind of set Ω_X and Ω_Y, two-dimensional random variables can be discrete, continuous and mixed.

The two-dimensional distribution law of probabilities – function (rule, table, etc.) allowing to calculate probabilities of any random events connected by a two-dimensional random variable (*X, Y*):

$$p(\alpha \le X < \beta; \delta \le Y < \gamma), \ \forall \alpha, \beta, \delta, \gamma.$$

14.1.1 Two-dimensional cumulative distribution function

Two-dimensional cumulative distribution function of a two-dimensional random variable (*X, Y*) is equal to probability of joint events $X < x \ Y < y$:

$$F(x, y) = p(X < x; Y < y). \qquad (14.1)$$

Properties of two-dimensional cumulative distribution function:

1. $0 \le F(x, y) \le 1$.
2. $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$, $F(+\infty, +\infty) = 1$.
3. $F(x_1, y) \le F(x_2, y)$, if $x_2 > x_1$; $F(x, y_1) \le F(x, y_2)$, if $y_2 > y_1$.
4. Transition to one-dimensional characteristics:

$$F(x, \infty) = p(X < x, Y < \infty) = p(X < x) = F_X(x); \qquad (14.2)$$

$$F(\infty, y) = p(X < \infty, Y < y) = p(Y < y) = F_Y(y). \qquad (14.3)$$

5. Probability of hit in rectangular area

$$p(\alpha \le X < \beta; \delta \le Y < \gamma) = F(\beta, \gamma) - F(\alpha, \gamma) - F(\beta, \delta) + F(\alpha, \delta).$$

Cumulative distribution function the most universal –form of the law of distribution also can be used for the description both continuous, and discrete two-dimensional random variables.

14.1.3 Two-dimensional probability density function

The two-dimensional random variable (*X, Y*) is continuous if its cumulative distribution function *F (x,)* represents continuous, differentiable function on each of arguments also there is a second mixed derivative $\frac{\partial^2 F(x, y)}{\partial x \partial y}$.

The ***two-dimensional probability density function*** (distribution density) *f(x,y)* characterises the probability density in a vicinity of a point with co-ordinates *(x,y)* and is equal to the second mixed derivative distribution function:

$$f(x,y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{p(\{x\le X<x+\Delta x\}\bigcap\{y\le Y<y+\Delta y\})}{\Delta x\Delta y}=\frac{\partial^2 F(x,y)}{\partial x\partial y}. \qquad (14.4)$$

The probability to get of value of a two-dimensional random variable (*X, Y*) in any area *D* is equal to the sum of all elements of probability for this area:

$$p\{(X,Y)\in D\}=\iint\limits_{(D)} f(x,y)dxdy \qquad (14.5)$$

Properties of two-dimensional density:

1. $f(x,y)\ge 0$
2. A normalizing condition:

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)dxdy=1. \qquad (14.6)$$

Geometrically – the volume of a body limited to a surface of distribution and a plane x0y, is equal to unit.

3. Transition to distribution function:

$$F(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy. \qquad (14.7)$$

4. Transition to one-dimensional characteristics:

$$f_X(x)=\int\limits_{-\infty}^{\infty} f(x,y)dy; \qquad (14.8)$$

$$f_Y(y)=\int\limits_{-\infty}^{\infty} f(x,y)dx. \qquad (14.9)$$

14.2 Dependent and independent random variables

Random variable X ***is independent*** from random variable Y if its distribution law does not depend on what value has accepted random variable Y. Dlja's sizes of independent sizes is carried out following equalites, i.e. criteria of independence:

1) $F(x,y)=p(X<x, Y<y)=p(X<x)\,p(Y<y)=F_X(x)\,F_Y(y)$ x$\forall, y;$ (14.10)

2) $f(x,y)=f_X(x)\,f_Y(y)$ x$\forall, y;$ (13.11)

In the event that criteria are not carried out at least in one point, random variables *X* and *Y* are dependent. For independent random variables two-dimensional forms of the distribution law do not contain any additional information, except that which contains in two one-dimensional laws. Thus, in case of dependence of random variables *X*

and *Y*, transition from two one-dimensional laws to the two-dimensional law it is impossible. For this purpose it is necessary to know conditional distribution laws.

14.3 Conditional distribution laws

As the conditional distribution law is called the distribution of one random variable found provided that other random variable has accepted certain value.

Conditional probability density function for continuous components *X* and *Y* are defined under formulas:

$$f(x/y) = f(x, y)/f_Y(y), \text{ for } f_Y(y) \neq 0; \quad (14.12)$$

$$f(y/x) = f(x, y)/f_X(x), \text{ for } f_X(x) \neq 0. \quad (14.13)$$

Conditional laws of distribution possess all properties of one-dimensional forms of laws of distribution corresponding to them.

If sizes *X* and *Y* are independent, conditional laws of distribution are equal to corresponding unconditional:

$$f(x/y) = f_X(x); \quad (14.14)$$

$$f(y/x) = f_Y(y). \quad (14.15)$$

It is necessary to distinguish functional and statistical dependences between random variables. If X and Y random variables, which are connected among themselves by functional dependence at $Y = (\varphi)$, that, knowing value *X*, it is possible to calculate precisely corresponding value *Y*, and on the contrary.

If between random variables there is a *statistical dependence* (random variables *X* and *Y* are dependent - see (13.10 – 13.11)) on value of one of them it is possible to establish only conditional distribution of probabilities another, i.e. to define, with what probability there will be this or that value of other random variable.

Example. *Y* - a grain yield, *X* - quantity of fertilizers on some ground. It is obvious that between *X* and *Y* there is a statistical dependence as value *Y* (productivity on a site) depends and on many other factors.

Lecture 15

15.1 Numerical characteristics of two-dimensional random variable

Let's consider the basic numerical characteristics of a two-dimensional random variable (*X*, *Y*).

15.1.1 Mixed ordinary moments

The mixed ordinary moment order *k*+*s* it is equal to a mathematical expectation of product and $X^k Y^s$:

$$\alpha_{k,s}(x,y) = \mathrm{M}[X^k Y^s] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s \cdot f(x,y)dxdy, \qquad (15.1)$$

The most often used initial moments - mathematical expectation random variables *X* and *Y*:

$$m_X = \alpha_{1,0}(x,y),\ m_Y = \alpha_{0,1}(x,y); \qquad (15.2)$$

15.1.2 Mixed central moments

The mixed central moment order *k*+*s* it is equal to a mathematical expectation of product of the centered random variables $\overset{\circ}{X}^k\ \overset{\circ}{Y}^s$:

$$\mu_{k,s}(x,y) = \mathrm{M}[(X-m_X)^k \cdot (Y-m_Y)^s] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_X)^k (y-m_Y)^s \cdot f(x,y)dxdy, \qquad (15.3)$$

$f(x,y)$ - two-dimensional probability density function of a continuous random variable (*X*, *Y*).

Let's consider the most often used central moments - - dispersion random variables *X* and *Y*:

$$D_X = \mu_{2,0}(x,y) = \alpha_{2,0}(x,y) - m_X^2,\ D_Y = \mu_{0,2}(x,y) = \alpha_{0,2}(x,y) - m_Y^2. \qquad (15.4)$$

Special role as the characteristic of system of random variables, plays the second mixed central moment of an order 1+1 which $\mu_{1,1}(x,y)$ is called *as the correlation moment* or *covariation* random variables *X*, *Y*.

15.1.3 Correlation moment

The ***correlation moment*** K_{XY} characterises degree of closeness of linear dependence of random variables *X* and *Y* and dispersion of their values around a point (m_X, m_Y):

$$K_{XY} = \mu_{1,1}(x,y) = \alpha_{1,1}(x,y) - m_X m_Y. \qquad (15.5)$$

Properties covariation K_{XY}:

1. $K_{XY} = K_{YX}$.

2. The correlation moment of two independent random variables *X* and *Y* is equal to zero.

3. The absolute size of the correlation moment of two random variables does not exceed a geometrical average of their dispersions

$$\left|K_{xy}\right| \le \sqrt{D_x \cdot D_y} \ \text{Or} \left|K_{xy}\right| \le \sigma_x \cdot \sigma_y . \qquad (15.6)$$

If $K_{XY} < 0$, between random variables *X* and *Y* there is a negative *covariation*, i.e. than it is bigger value of one random variable, are more probable smaller values of another random variable (see statistical dependence in lecture. 13). An *example. X* – number of missed classes of the student, *Y* – an examination mark.

If $K_{XY} > 0$, between sizes *X* and *Y* there is a positive *covariation*, i.e. than it is bigger value of one random variable, are more probable bigger values of another random variable. An *example. X* and *Y* - growth and weight of at random taken student.

If $K_{XY} = 0$, random variables *X* and *Y* are *not correlated*, i.e. between them there is no *linear* dependence.

If $K_{XY} \neq 0$, random variables *X* and *Y* are called *as correlated.*

15.1.4 Correlation coefficient

Correlation coefficient R_{XY} characterises degree of linear dependence of sizes and it is equal:

$$R_{XY} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{K_{XY}}{\sigma_X \sigma_Y} . \qquad (15.7)$$

Properties of correlation coefficient*:*

1. The absolute size of correlation coefficient of two random variables does not exceed unit: $\left|R_{XY}\right| \le 1$.

2. $\left|R_{XY}\right| = 1$, if random variables *X, Y* are connected by linear functional dependence *Y=aX+b*.

The more absolute size of factor of correlation, the more close statistical dependence random variables *X, Y* to linear functional dependence.

3. If random variables *X* and *Y* are independent then $R_{XY} = 0$

15.2 Conditional numerical characteristics

For dependent two-dimensional random variables conditional laws of distribution (see (14.13, 14.14)) can be defined. These laws of distribution possess all properties of unconditional laws, and on their basis under known formulas (after replacement in them of unconditional laws on conditional) numerical characteristics which are called as conditional can be calculated. Conditional mathematical expectation have the greatest practical value.

15.2.1 Regression

As conditional mathematical expectation of random variable *X* is called its mathematical expectation calculated when (to the condition) that random variable *Y* has accepted certain value *Y* = *y*:

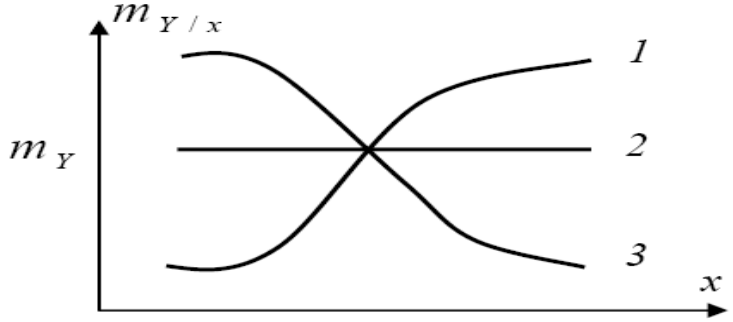
$$M[X / y] = m_{X/y} = \int_{-\infty}^{\infty} x \cdot f(x/y)dx, \qquad (15.8)$$

Similarly and for

$$M[Y / x] = m_{Y/x} = \int_{-\infty}^{\infty} y \cdot f(y/x)dx, \qquad (15.9)$$

The conditional mathematical expectation $m_{X/y}$ is called as *X* on *y*, and a conditional mathematical expectation $m_{Y/x}$ – regression *Y* on *x*. It is obvious that conditional mathematical expectation s represent some functions which depend on the value taken in a condition, i.e. $m_{X/y} = \psi(y)$, and $m_{Y/x} = \varphi(x)$.

Schedules of these dependences are called as regression lines (drawing see).



The regression line 1 specifies that between random variables *X, Y* there is a positive *covariation*, i.e. $K_{XY} > 0$. The regression line 2 specifies that random variables *X, Y* are not dependent, and the regression line 3 - that between random variables *X, Y* exists negative *covariation*, i.e. $K_{XY} < 0$.

The regression analysis allows to reveal character of dependence between random variables *X, Y*. Random variables *X, Y* are called *as linearly correlated* if regression lines are straight lines. The equations of linear regression look like:

$$m_{Y/x} = m_Y + R_{XY}\frac{\sigma_Y}{\sigma_X}(x - m_X), \qquad m_{X/y} = m_X + R_{XY}\frac{\sigma_X}{\sigma_Y}(y - m_Y), \qquad (15.10)$$

Both straight lines pass through a point (m_X, m_Y) which name the centre of joint distribution of sizes *X* and *Y*.

Lecture 16

16.1. Statistical processing of two-dimensional random variables

Let the two-dimensional random variable is spent n independent experiments, in each of which (X, Y) assumes certain values and results of experiments represent two-dimensional sample:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}.$$

Statistical processing of two-dimensional data includes:

- processing and the analysis of components X and Y as one-dimensional variables (lectures 1–3),

- calculation of estimations and the analysis of the parameters inherent only in two-dimensional (multidimensional) random variables.

As a rule, following estimations of numerical characteristics of a two-dimensional random variable (X, Y) are defined:

– Estimations of mathematical expectations:

$$m_X^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \tag{16.1}$$

$$m_Y^* = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \; ; \tag{16.2}$$

– Dispersion estimations:

$$S_0^2(x) = \frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2; \tag{16.3}$$

$$S_0^2(y) = \frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n}{n-1}\overline{y}^2. \tag{16.4}$$

– Standard deviation estimations:

$$S_0(x) = \sqrt{S_0^2(x)} \tag{16.5}$$

$$S_0(y) = \sqrt{S_0^2(y)}. \tag{16.6}$$

16.1.1. An estimation of the correlation moment

The *consistent unbiased* estimation *of the correlation moment* is equal

$$K_{XY}^{*} = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \overline{x})(y_k - \overline{y}), \qquad (16.7)$$

Where x_i, y_i – values random variables *X*, *Y* in *i*-м experiment;

$\overline{x}, \overline{y}$ – Average values of random variables *X* and *Y*, respectively.

16.1.2. The correlation coefficient estimate

The consistent *estimate of the correlation coefficient* is equal to

$$R_{XY}^{*} = \frac{K_{XY}^{*}}{S_0(x) \cdot S_0(y)} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}}, \qquad (16.8)$$

where $S_0(x), S_0(y)$ – are standard deviation estimates of random variables *X* and *Y*, respectively.

16.1.3. The confidence interval for correlation coefficient

The confidence interval for correlation coefficient with reliability γ for the case of two-dimensional normal distribution looks like:

$$I_{\gamma}(R_{XY}) = \left(\frac{e^{2a} - 1}{e^{2a} + 1}; \frac{e^{2b} - 1}{e^{2b} + 1} \right), \qquad (16.9)$$

where $a = 0,5 \cdot \ln\left(\frac{1 + R_{XY}^{*}}{1 - R_{XY}^{*}} \right) - \frac{z_{\gamma}}{\sqrt{n-3}}; b = 0,5 \cdot \ln\left(\frac{1 + R_{XY}^{*}}{1 - R_{XY}^{*}} \right) + \frac{z_{\gamma}}{\sqrt{n-3}};$

z_{γ} – is the argument value of Laplace function, $z_{\gamma} = \arg \Phi(\frac{\gamma}{2})$, i.e. $\Phi(z_{\gamma}) = \frac{\gamma}{2}$.

If the sample size is large enough ($n > 20$), then the distribution law of the correlation coefficient estimate R_{XY}^{*} can be considered asymptotically normal, and confidence interval can be constructed by equation:

$$I_\gamma(R_{XY}) = \left(R_{XY}{}^* - z_\gamma \cdot \frac{1-\left(R_{XY}{}^*\right)^2}{\sqrt{n}}; R_{XY}{}^* + z_\gamma \cdot \frac{1-\left(R_{XY}{}^*\right)^2}{\sqrt{n}} \right). \qquad (16.10)$$

16.2. Statistical criteria of two-dimensional random variables

16.2.1. A hypothesis about absence of correlation dependence

It is supposed that the two-dimensional random variable (*X*, *Y*) is distributed under the normal law. If the module of a dot estimation of factor of the correlation, calculated on initial two-dimensional sample of a random variable (*X*, *Y*), is not great ($\left|R_{XY}^*\right| \le 0,1...0,2$) the hypothesis about absence of correlation dependence between variables *X* and *Y* can be checked up following in the image.

1. The hypothesis is formulated:

$$H_0: R_{XY} = 0;$$

$$H_1: R_{XY} \neq 0.$$

Here – R_{XY} theoretical factor of correlation.

2. The estimation of factor of correlation R_{XY}^* by equation (14.8) is calculated R_{XY}^*.

3. If the sample volume is not great (n <20) items 4 … 6 are carried out, differently – items 7 … 9 are carried out

4. Value of criterion is defined

$$t = \frac{R_{XY}^* \sqrt{n-2}}{\sqrt{1-\left(R_{XY}^*\right)^2}}, \qquad (16.11)$$

Which is distributed under the law of Student with (n–2) freedom degrees if hypothesis H_0 is true.

5. On the set significance value the αconfidential probability is calculated and γ=1$-\alpha$ critical value gets out of the table of Student distribution $t_{\gamma,n-2}$.

6. If $|t| > |t_{\gamma,n-2}|$, hypothesis H_0 deviates, and consequently, variables *X*, *Y* are correlated. Otherwise hypothesis H_0 is accepted.

7. Value of criterion is defined

$$Z = \frac{R^*_{XY}\sqrt{n}}{1-\left(R^*_{XY}\right)^2}, \qquad (16.12)$$

Which is distributed practically under the normal law if hypothesis H_0 is true.

8. On the set significance value from theα?table of function of Laplace critical value is defined $Z_\alpha = \arg\Phi\left(\frac{1-\alpha}{2}\right)$, i.e $\Phi(Z_\alpha) = \frac{1-\alpha}{2}$.

9. If $|Z| > |Z_\alpha|$, hypothesis H_0 deviates, and consequently, variables *X*, *Y* are correlated. Otherwise hypothesis H_0 is accepted.

16.3. A hypothesis about equality of distribution laws

Let $\{x_1, x_2, ..., x_n\}, \{y_1, y_2, ..., y_m\}$ – independent random samples of random variables *X* and *Y*.

The two-alternative hypothesis is formulated:

H_0: $F_X(x) \equiv F_Y(y)$ i.e. two samples belong to the same general population;

H_1: $F_X(x) \neq F_Y(y)$.

Concerning the distribution laws of random variables *X* and *Y* no assumptions made.

For check of the given hypothesis the criterion Wilcoxon (Mann-Witny) of is used. Values $\{x_1, x_2, ..., x_{n_1}\}$ $\{y_1, y_2, ..., y_{n_2}\}$ of both samples are ranked together in increasing order. The criterion is based on the statistics

$$U = \sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{i,j}, \qquad (16.13)$$

where

$$\delta_{i,j} = \begin{cases} 1, & x_i < y_j, \\ 0, & else. \end{cases}$$

This statistics U represents the total number of inversions, when the sample units $x_1,\ldots,x_n$ precede to the sample units $y_1,\ldots,y_m$ in the general variational series. The pair of values (x_i y_j) forms *inversion*, if $y_j < x_i$.

Let, for example, for $n = 4$ and $m = 5$ such sequence has turned out: y_5 x_3 x_4 y_1 y_2 x_2 y_4 y_3 x_1. In our example x_3 and x_4 form on one inversion (with y_5), x_2 forms three inversions (with y_5 y_1 y_2), and x_1 forms five inversions (with all).

It is proved that if H_0 is true, then

$$E(U) = \frac{mn}{2}, \quad D(U) = \frac{mn(m+n+1)}{12}. \qquad (16.14)$$

As a criterion, the variable U– the complete number of inversions - is used. The given variable is distributed under Wilcoxon law and hypothesis H_0 is rejected, if U is larger, than the critical value U_α taken from the Wilcoxon table for the specified significance level α.

For large sample sizes (n and m exceed 25), the critical value U_α is determined by equation

$$U_\alpha = Z_\alpha \sqrt{\frac{nm(n+m+1)}{12}}, \qquad (16.15)$$

where Z_α – is the argument value of Laplace function $Z_\alpha = \arg\Phi\left(\frac{1-\alpha}{2}\right)$.

Lecture 17

17.1. The regression characteristics estimate

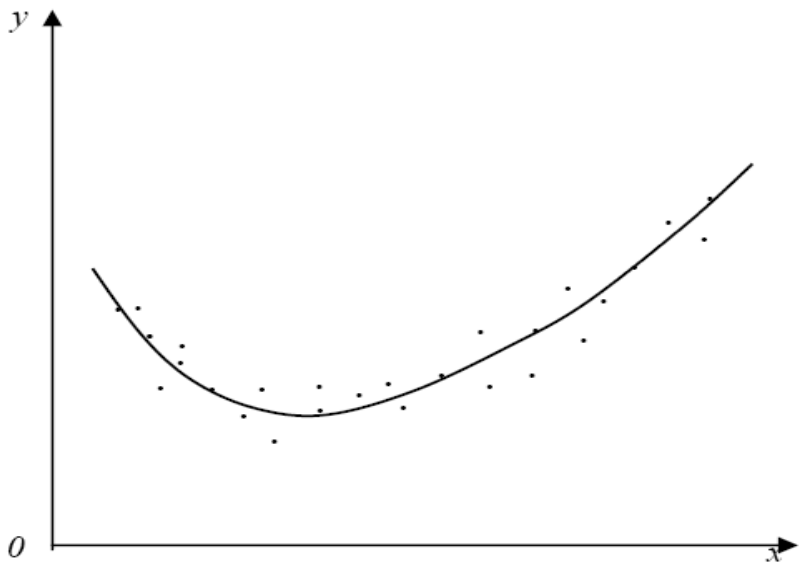
In the regression analysis, the interrelation between random and not random variables based on experimental data is studied. Unlike the correlation analysis, in the regression analysis not all studied interrelated variables are random, and the conditions imposed on the studied variables, are less burdensome. Therefore, it is considered that problems regression analysis are more often in practice in comparison with problems of the correlation analysis.

Let *n* independent experiments are performed , in each of which two-dimensional random variable (*X, Y*) assumes certain values, and experimental results represent two-dimensional sample of the kind $\{(x_1, y_1),(x_2, y_2),\ldots,(x_n, y_n)\}$. It is necessary to reveal the kind of relationship between variables *X*, *Y* based on the available sample, i.e. to obtain the estimate of the conditional mathematical expectation $m^*_{Y/x}$ – regression estimate *Y* for *x*. The given estimate represents a certain function:

$$m^*_{Y/x} = \bar{y}(x) = \varphi(x, a_0, a_1, ..., a_m),$$

where $a_0, a_1, ..., a_m$ - are unknown parameters.

Thus, first, it is necessary to establish the dependence type of $\varphi(x, a_0, a_1, ..., a_m)$, i.e. whether it is linear, quadratic, exponential and etc., second, to determine the values of unknown parameters $a_0, a_1, ..., a_m$. To determine the dependence type, the *scatter diagram* (or *correlation field*) is constructed, which can be obtained if to represent experimental results in the form of points on the plane in the Cartesian coordinates system (see Figure). Based on the analysis of the correlation field, we choose the type *of the empirical regression line* $\bar{y}(x) = \varphi(x, a_0, a_1, ..., a_m)$, which should pass through points $(x_1, y_1)..., (x_n, y_n)$ so, that its graph would correspond to the unknown regression line in the best way, i.e. its values should be approximately equal to *Y* arithmetic mean values for each value of *X=x*.



In many cases, the dependence type can be chosen based on theoretical or other reasons.

To determine the parameter values, at which the best matching of the curve $y = \varphi(x, a_0, a_1, ..., a_m)$ and experimental points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is provided, the least squares method is used.

17.1.1. The least squares method

The essence of the given method is that parameters values $a_0, a_1, ..., a_m$ should be chosen so, that the *sum of squares of the* experimental *points deviations from the fitting curve turned to the minimum:*

$$\sum_{i=1}^{n} \left[y_i - \varphi(x_i, a_0, ..., a_m) \right]^2 = \min . \qquad (17.1)$$

Let's find the values a_j, $j = 1, ..., m$ turning the left part of expression (15.1) into minimum. For this purpose, we differentiate it with respect to a_j, $j = 1, ..., m$ and equate the derivatives to zero (at the point of extremum the derivative is equal to zero):

$$\sum_{i=1}^{n} \left[y_i - \varphi(x_i, a_0, ..., a_m) \right] \frac{\partial \varphi(x_i)}{\partial a_j} = 0, \; j = 0, 1, ..., m , \qquad (17.2)$$

where $\frac{\partial\varphi(x_i)}{\partial a_j}$ – is the value of the private derivative function φ with respect to parameter a_j at point x_i.

The system of the equations (17.2) contains the same number of equations, as the number of unknown parameters, i.e. $m+1$.

It is impossible to solve system (17.2) in a general form, for this purpose, it is necessary to set a particular kind of function φ.

Let y represent the power series:

$$y = \varphi(x, a_0, ..., a_m) = \sum_{j=0}^{m} a_j x^j \ . \qquad (17.3)$$

Then (17.2) will become the linear equations system (LES):

$$\sum_{j=0}^{m} a_j \sum_{i=1}^{n} \left(x_i\right)^{j+k} = \sum_{i=1}^{n} y_i \left(x_i\right)^k, k = 0,1,...,m \qquad (17.4)$$

We divide both parts of equations by the sample size n, the system will become

$$\sum_{j=0}^{m} a_j \alpha^*_{j+k}(x_i) = \alpha^*_{k,1}(x_i, y_i), k = 0,1,...,m \qquad (17.5)$$

where $\alpha^*_k(x)$ – is the ordinary moment estimate of k-order of variable *X:*

$$\alpha^*_k(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i\right)^k ;$$

$\alpha^*_{k,1}(x, y)$ – is the estimate of the mixed ordinary moment of $k+1$-order of variables *X* and *Y:*

$$\alpha^*_{k,1}(x, y) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k y_i \ .$$

In system (17.5), a_j, $j = 1,...,m$ are variables, and the estimates of ordinary moments $\alpha^*_{k,1}(x, y)$ are system of linear equations factors. Having solved the given system, we determine the estimates of the parameters $a^*_0, a^*_1, ..., a^*_m$, providing the best matching of the curve $y = \varphi(x, a_0, a_1, ..., a_m)$ and experimental points $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$.

Example. We determine the linear regression estimate $m_{Y/x} = a_0 + a_1 x$. System (16.5) for *m=1* looks like

$$\begin{cases} \alpha_0^*(x)a_0 + \alpha_1^*(x)a_1 = \alpha_{0,1}^*(x,y) \\ \alpha_1^*(x)a_0 + \alpha_2^*(x)a_1 = \alpha_{1,1}^*(x,y) \end{cases}.$$

Taking into account that $\alpha_0^*(x) = 1, \alpha_1^*(x) = \bar{x}, \alpha_{0,1}^*(x,y) = \bar{y}$, we obtain:

$$\begin{cases} a_0 + \bar{x}a_1 = \bar{y} \\ \bar{x}a_0 + \alpha_2^*(x)a_1 = \alpha_{1,1}^*(x,y) \end{cases}.$$

Whence

$$a_1^* = \frac{\alpha_{1,1}^*(x,y) - \bar{x} \cdot \bar{y}}{\alpha_2^*(x) - \bar{x}^2} = \frac{K_{XY}^*}{S_0^2(x)}, \qquad (17.6)$$

$$a_0^* = \bar{y} - a_1^* \cdot \bar{x}, \qquad (17.7)$$

what corresponds to linear regression equations (15.10).