Lecture 2. Mathematical statistics. Introduction.

§1 Mathematical statistics (MS)

Experiment – any activity or procedure that "may give" a set of outcomes (w) with properties:

a) an experiment is partially or completely unknown;
b) an experiment has a random nature;
c) repeatable (reproducible).

Random variable (r.v.) Given Ω sample space and probability distribution P. a Random variable ξ is a function defined on Ω, taking values in R:

$$\xi : \Omega \to R \qquad (1)$$

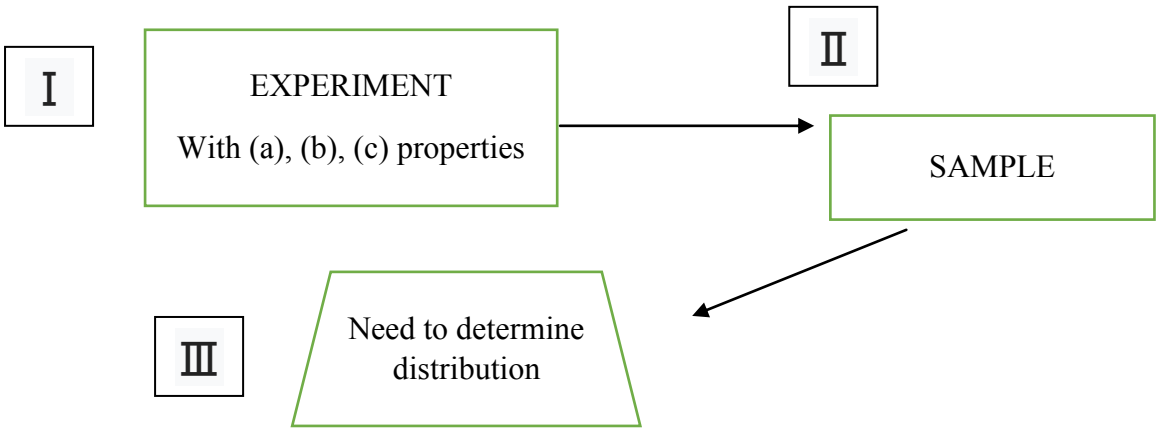
Remark. *Note*! When we consider the experiment in *Mathematical statistics*, we assume that the probabilistic space is given, but we are not interested in it.

Clear, that n times realization of (1) – gives us *experiment* with $x_1, \dots, x_n$ – outcomes (results). ***F*** distribution of ξ satisfies (a)

Let consider set $(x_1, \dots, x_n)$ before experiment and suppose r.v. x_i– identically distributed with ξ and after experiment x_i – value – one of the possible values r.v. x_i.

Definition.

Sample x = $(x_1, \dots, x_n)$ volume n from distribution ***F*** – is a set identically distributed r.v. with ***F*** distribution.



(III) Distribution is characterized by distribution function, density function or table and set of numbers: expectation $E\xi$, variance $D \equiv \mathrm{var}\xi = \mathrm{D}\xi$, and $E\xi^k$.

§2 *Sample characteristics. Descriptive statistics*

Let's recall Definition. The elementary events (outcomes) of a random experiment are such events (atoms of the events) which exactly one of them occurs during the experiment.

Let's consider the sample realization on single elementary outcome

$$u_0: x_1 = x_1(w_0), \dots, x_n = x_n(w_0)$$

and introduce new random variable ξ^*, wich takes values $x_1, x_2, \dots, x_n$ with the same probability (as n-surfaces dice). For ξ^* we obtain the table

ξ^*	x_1	x_2	…	x_n
$\tilde{p}$	1/n	1/n	…	1/n

and the distribution function:

$$F_n^*(y) = \sum_{x_i<y} \frac{1}{n} = \frac{\text{number } x_i \in (-\infty, y)}{n};$$

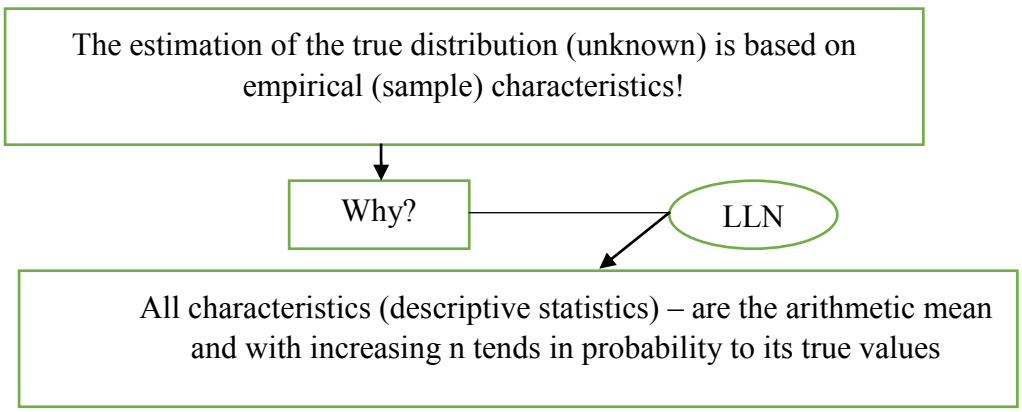
distribution of ξ^* – empirical (or sample) distribution.

$$\tilde{E}\xi^* = \sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n} \sum x_i = \bar{x};$$

$$\tilde{D}\xi^* = \sum_{i=1}^{n} \frac{1}{n} (x_i - E\xi^*)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = S^2; \qquad (2)$$

$$\tilde{E}(\xi^*)^k = \bar{x}^k; \; \tilde{E}g(\xi^*) = \frac{1}{n} \sum_{i=1}^{n} g(x_i) = \overline{g(x)};$$

If we allow an elementary outcome to change then we obtain random variables $F_n^*(y)$, $\bar{x}$, S^2, $\bar{x}^k$, $\overline{g(x)}$ as functions of r. v. $x_1, \dots, x_n$ instead (2).



§3 Sample characteristics (properties). Distribution and density function

I. Empirical distribution function (EDF)

Let F – the unknown true distribution; we can describe one as $\mathrm{F}(\mathrm{y}) = \mathrm{P}(x_1 < y)$.

Definition.

Empirical distribution function determined on the sample $\boldsymbol{x} = (x_1, \dots, x_n)$ volume n – is random function $F_n^*(y): R \times \Omega \to [0,1],\ \forall\, \mathrm{y} \in \mathrm{R},$ such that

$$F_n^*(y) = \frac{number\ X_i \in (-\infty, y)}{n} = \frac{1}{n}\sum_{i=1}^{n} I(x_i < y);\ here\ \mathrm{I}-\text{ indicator function.}$$

$$I(x_i < y) = \begin{cases} 1, & if\ x_i < y \\ 0, & else \end{cases}$$

Remark.

For any y indicator of event $x_i < y$, $I(x_i < y)$ – random variable from Bernouli [bɛ'nju:li] distribution with p – parameter: $p = P(x_i < y) = F(y)$.

Definition. Empirical distribution function:

Let's reorder sample:

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)};\ x_{(1)} = \min_i\{x_i\}, x_{(n)} = \max_{(n)}\{x_i\}$$

$$x_{(k)} - k - th\ element, or\ k - th\ order\ statistics.$$

$F_n^*(y)$ has jumps in each point of the sample; height of jump in x_i – equal to $\frac{m}{n}$, m – number of elements are equal to x_i and

$$F_n^*(y) = \begin{cases} 0, & y \le x_{(1)} \\ \frac{m}{n}, & y \in (x_{(m)}, x_{(m+1)}) \\ 1, & y > x_{(n)}. \end{cases}$$

II. Empirical probability function (EPF) or density function

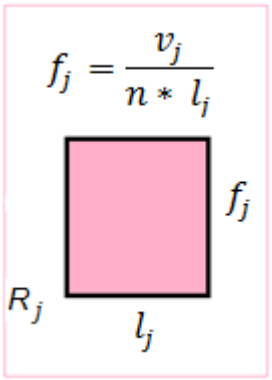
Histogram (of relative frequency).

Table exists for discrete distributions.

Density function – for absolutely continuous distributions.

The histogram – empirical analogue of the density function.

Histogram is based on grouped data; the range of data values $x_{max} - x_{min}$ is devided into several bins $A_1, \dots, A_k$ (consecutive, non-overlapping intervals, often equal size- length): v_j – number elements of the sample got into j – th bin A_j. Let construct rectangle R_j with f_j – rectangle height and l_j – rectangle length (see pic.), where



$$v_j = \{\text{number } x_i \in A_j\} = \sum_{i=1}^{n} I(x_i \in A_j),$$

$$n = \sum_{j=1}^{k} v_j \; for \; \forall A_j$$

$$f_j = \frac{v_j}{n * l_j} - \text{density frequency};$$

$$\frac{v_j}{n} - \text{relative frequency}$$

Definition. Figure, consisting of union of k such rectangles, is called histogram.

Find the square S_j of the painted figure R_j:

$$S_j = f_j * l_j \rightarrow S = \sum_{j=1}^{k} S_j = \sum_{j=1}^{k} f_j l_j = \sum_{j=1}^{k} \frac{v_j}{n} \frac{l_j}{l_j} = \frac{1}{n} \sum_{j=1}^{k} v_j = \frac{n}{n} = 1;$$

Remark: ***On normalized histogram***.

A histogram may also be normalized to display relative frequencies. It then shows the proportion of cases that fall into each of several categories (bins, intervals), with the sum of the heights equaling 1 and $l_j = 1, \; for \; \forall j$.

Remark: ***On relation between a histogram and a density function***

Let density function of elements (sample) – continuous function. If number of intervals tends to infinity $\frac{k(n)}{n} \rightarrow 0$, *here k=k(n), then histogram tends to density function for* $\forall y$.

III. Sample and true moments

Let ξ – random variable belongs to General Population with true known characteristics; x_1(or any x_i) – we can consider as outcome (result of experiment) of ξ and random variable too.

True moments	Estimation for true moments
$E\xi = \mathrm{E}x_1 = a$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;$
	$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2;$
$D\xi = \mathrm{D}x_1 = \sigma^2$	$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2;$ – unbiased
$E\xi^k = \mathrm{E}x_1^k = m_k$	$S_0^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^k;$
$Eg(\xi)$	$\overline{g(x)} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$

§4 *Consistency of the sample characteristics*

- Empirical distribution function (EDF)
- Histogram
- Moments (sample)

Theorem 1.

Let $x_1, \dots, x_n$ – sample from distribution family F with distribution function F and let F_n^* – empirical distribution function, built in accordance with the sample, then

$$F_n^*(y) \xrightarrow{\mathrm{P}} \mathrm{F}(y)$$

when n→ ∞ for ∀$y \in$ R.

Proof:

Using definition of EDF $F_n^*(y) = \frac{number\ X_i \in (-\infty, y)}{n} = \frac{1}{n}\sum_{i=1}^{n} I(x_i < y).$

Random variables $I(x_1 < y)$, $I(x_2 < y)$, … independent and identically distributed =>

$$EI(x_1 < y) = 1 * P(x_1 < y) + 0 * P(x_1 \geq y) = P(x_1 < y) = F(y) < \infty;$$

Repetition LLN: (according to LLN)

$$\text{r.v. } \xi_1, \dots, \xi_n\colon\ E\xi_1 < \infty,\ \ E\xi_1 = a$$

$$\frac{S_n}{n} = \frac{\xi_1 + \cdots + \xi_n}{n} \to a;$$

$$F_n^*(y) = \frac{\sum_{i=1}^{n} I(x_i < y)}{n} \xrightarrow{P} \mathrm{EI}(x_1 < y) = \mathrm{F(y)}$$

Conclusion:

With increasing sample volume empirical distribution function tends in probability to unknown the same distribution for $\forall y \in R$.

Theorem 2. (Glivenko–Cantelli)

In the condition of the theorem.1 we have:

$$\sup_{y \in R} |F_n^*(y) - F(y)| \xrightarrow[n \to \infty]{P} 0$$

Theorem 3. Kolmogorov (about the convergence)

Let $x_1, \dots, x_n$ sample of F with continuous distribution function F, $F_n^*(\mathrm{y})$ – empirical distribution function for $\forall \mathrm{y} \in \mathrm{R}$, then

$$\sqrt{n} \sup_{y \in R} |F_n^*(y) - F(y)| \to \eta, \qquad \mathrm{n} \to \infty$$

and η have Kolmogorov distribution with continuous function

$$F_\xi(\mathrm{y}) = K(y) = \begin{cases} \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 y^2}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Theorem 4. (*on empirical distribution function*)

(based on, average of n independent summand, having Bernoulli distribution)

For any y ∈ R :

1) $EF_n^*(y)$= F(y), $F_n^* - unbiased\ estimation$
2) $DF_n^*(y) = \frac{F(y)(1-F(y))}{n}$;
3) $\sqrt{n}(F_n^*(y) - F(y)) \to N(0, F(y)(1-F(y)))$ tends asymptotically F(y)≠0,1
4) $n\,F_n^*(y)\ has$ Binomial distribution $B(n, F(y))$

Proof:

Known: $I(x_i < y) \in B_{F(y)}, =>$

$$EI(x_1 < y) = F(y), \qquad DI(x_1 - y) = F(y)(1-F(y))$$

1) $I(x_i < y)$ identically distributed, so $EF_n^*(y) = E\frac{\sum_{i=1}^n I(x_i<y)}{n} = \frac{nEI(x_1<y)}{n} = = F(y)$
2) $I(x_i < y)$ – independent and identically distributed
   $DF_n^*(y) = D\frac{\sum_{i=1}^n I(x_i<y)}{n} = \frac{\sum_{i=1}^n DI(x_i<y)}{n^2} = \frac{nDI(x_1<y)}{n^2} = \frac{F(y)(1-F(y))}{\boldsymbol{n}}$;
3) Based on central limit theorem
   $\sqrt{n}\big(F_n^*(y) - F(y)\big) = \sqrt{n}\left(\frac{\sum_{i=1}^n I(x_i<y)}{n} - F(y)\right) = \left(\frac{\sum_{i=1}^n I(x_i<y) - nF(y)}{\sqrt{n}}\right) =$
   $= \frac{\sum_{i=1}^n I(x_i<y) - nEI(x_1<y)}{\sqrt{\boldsymbol{n}}} => N\big(0, DI(x_1 < y)\big) = N(0, F(y)(1-F(y))$;
4) $nF_n^*(y) = I(x_1 < y) + \cdots + I(x_n < y) \in B(n, F(y))$; (according to property: the stability of the summation)

Histogram properties

Let F – absolutely continuous distribution, f – true density, k – intervals of histogram is not depends on n – sample volume. With increasing n square of histogram column tends to square under density function.

Theorem 5:

$$n \to \infty, \qquad \text{for } \forall j = 1, \dots, k;\ l_j * f_j = \frac{v_j}{n} \to \mathrm{P}\left(x_1 \in A_j\right) = \int_{A_j} f(x)dx$$

(without proof)