Lecture 4.

I. Interval estimation

Sample is given $X = (X_1, \dots, X_n) \in \mathcal{F}_\theta, \theta \in \Theta \subseteq R$.

Definition.

Let $0 < \varepsilon < 1$. Interval (θ^-, θ^+) with random ends

$$(\theta^-, \theta^+) = \left(\theta^-(x, \varepsilon), \theta^+(x, \varepsilon)\right)$$

is called *confidence interval* for parameter θ with level of trust $1 - \varepsilon$,

if for $\forall \theta \in \Theta \quad P(\theta^- < \theta < \theta^+) \geq 1 - \varepsilon$,

here $\theta^- < \theta < \theta^+$ means the interval *covers* θ.

Example:

$X = (X_1, \dots, X_n)$ – sample $\in N(a, \sigma^2)$ $a \in R$ – unknown, $\sigma > 0$ – known. Need to construct confidence interval (for a) with level of trust $1 - \varepsilon$.

Based on the statement

if r. v. $\xi_i (i = 1,2)$ with normal distribution $N(a_i, \sigma_i^2)$ and independent, then r. v.

$$\eta = b\xi_1 + c\xi_2 + d$$

has normal distribution with parameters

$$E\eta = ba_1 + ca_2 + d, \qquad D\eta = b^2\sigma_1^2 + c^2\sigma_2^2$$

we have a normal distribution for the sum of the sample elements of any volume n too:

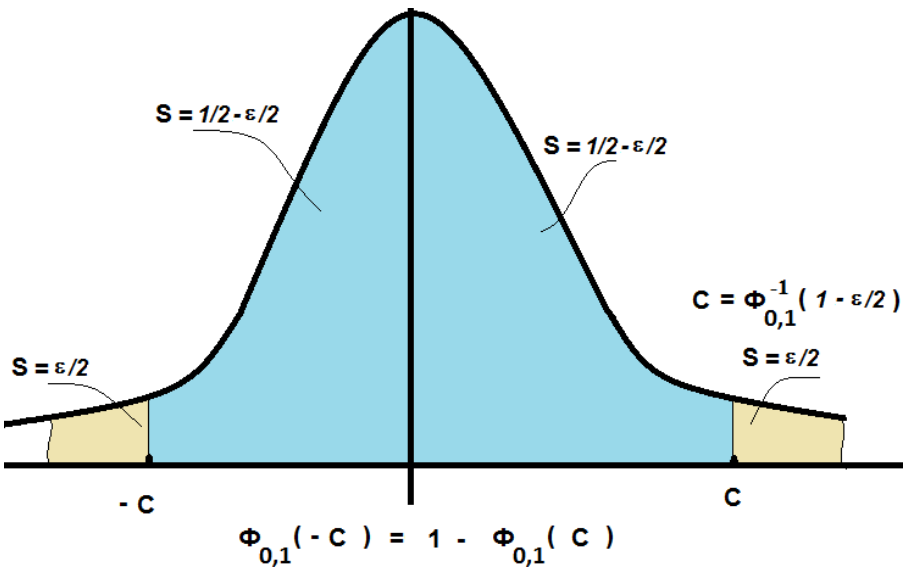
$$n\overline{X} = X_1 + \cdots + X_n \in N(na, n\sigma^2)$$

and a standardized value is equal

$$\eta = \frac{n\overline{X} - na}{\sigma\sqrt{n}} = \sqrt{n}\frac{\overline{X} - a}{\sigma} \in N(0,1).$$

For given $\varepsilon \in (0,1)$ we find c > 0 such that

$P(-c < \eta < c) = 1 - \varepsilon.$



Number c – quantile of $1-\frac{\varepsilon}{2}$ level of standard normal distribution:

$$P(-c<\eta<c)=\Phi_{0,1}(c)-\Phi_{0,1}(-c)=2\Phi_{0,1}(c)-1=1-\varepsilon;$$
$$\rightarrow \qquad \Phi_{0,1}(c)=1-\frac{\varepsilon}{2}.$$

Recall the definition: $\mathcal{F}$ – is a distribution family (with absolutely continuous distribution functions).

Number τ_δ – is a distribution quantile of δ-level ($F\in\mathcal{F}$), if $F(\tau_\delta)=\delta$.

For F – monotonic (in our case – Φ), the quantile is determined in the unique way. Using table (or any statistical package) for it

$$\Phi_{0,1}(x) \text{ can define } \quad c=\tau_{1-\frac{\varepsilon}{2}} \quad or \quad -c=\tau_{\frac{\varepsilon}{2}}.$$

Then solving inequality $-c<\eta<c$ with respect to a, we obtain the exact confidence interval:

$$1-\varepsilon=P(-c<\eta<c)=P\left(-c<\sqrt{n}\frac{\overline{X}-a}{\sigma}<c\right)$$
$$=P\left(\overline{X}-\frac{c\sigma}{\sqrt{n}}<a<\overline{X}+\frac{c\sigma}{\sqrt{n}}\right).$$

You can represent $c=\tau_{1-\frac{\varepsilon}{2}}$ and

$$P\left(\overline{X}-\frac{\sigma\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}}<a<\overline{X}+\frac{\sigma\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}}\right)=1-\varepsilon. \qquad (1)$$

So confidence interval for a is equal

$$\left(\overline{X}-\frac{\sigma\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}},\overline{X}+\frac{\sigma\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}}\right). \qquad (2)$$

Confidence interval construction principles

We need to follow these steps:

1. To find function $G(x,\theta)$ (with distribution function $\mathcal{G}$ does not depend on θ), invertible for any fixed x.
2. To find value g_1, g_2 – quantiles of distribution $\mathcal{G}$ for which $1-\varepsilon=P(g_1<G(x,\theta)<g_2)$.
3. Solving the inequality $g_1<G(x,\theta)<g_2$ with respect to θ and obtain the exact confidence interval.

II. Distributions connected with the Normal one

I. Gamma distribution.

Lemma 1.

The property of stability by summation:

Let $X_1, \dots, X_n$ are independent and ξ_i have Gamma distribution $\Gamma_{\alpha,\lambda_i}$,

$i = 1, \dots, n$, then their *sum* has Gamma distribution with parameters α, $\lambda = \lambda_1 + \cdots + \lambda_n$.

Remark.

The square of r. v. with standard normal distribution has Gamma distribution.

Lemma 2

If ξ has standard normal distribution (SND), then ξ^2 has Gamma distribution $\Gamma_{1/2,1/2}$.

Proof:

Find the derivative of distribution function of r.v. ξ^2 , let's show that this resulting function – is the density function:

$$y \le 0: F_{\xi^2}(y) = P(\xi^2 < y) = 0 \ \rightarrow f_{\xi^2}(y) = 0$$

$$y > 0: F_{\xi^2}(y) = P\left(-\sqrt{y} < \xi < \sqrt{y}\right) = F_\xi\left(\sqrt{y}\right) - F_\xi\left(-\sqrt{y}\right),$$

$$f_{\xi^2}(y) = \left(F_{\xi^2}(y)\right)' = F_\xi'\left(\sqrt{y}\right)\frac{1}{2\sqrt{y}} + F_\xi'\left(-\sqrt{y}\right)\frac{1}{2\sqrt{y}} =$$

$$\left[f_\xi\left(\sqrt{y}\right) + f_\xi\left(-\sqrt{y}\right)\right]\frac{1}{2\sqrt{y}} = (\text{symm } f_\xi) = \frac{f_\xi(\sqrt{y})}{\sqrt{y}} = f_\xi^N = \frac{1}{\sqrt{2\pi y}} e^{\frac{-y}{2}}$$

Remarks. Recall the basic properties of the gamma function:

1) $\Gamma(\alpha, \lambda) = \int_0^\infty x^{\lambda-1} e^{-\alpha x}\, dx = \Gamma(\lambda)/\alpha^\lambda$, parameters α, λ – scale, shape.

2) $\Gamma(\alpha, \lambda) = \int_0^\infty x^{\lambda-1} e^{-x/\alpha}\, dx = \alpha^\lambda \Gamma(\lambda)$ – *other form (1)*

3) $\Gamma(1) = 1, \ (\alpha = 1)$

4) $\Gamma(\lambda + 1) = \lambda\Gamma(\lambda), \ (\alpha = 1)$

5) $\Gamma(1/2) = \sqrt{\pi}$

Continue proof, taking into account both semi-intervals for distribution function

$$\int_{-\infty}^{\infty} f_{\xi^2}(y)\,dy = \int_{-\infty}^{\infty} \left(1/2\right)^{1/2} \frac{y^{1/2-1}e^{-y/2}}{\Gamma\left(1/2\right)}\,dy = \left(1/2\right)^{1/2} \frac{\Gamma\left(1/2, 1/2\right)}{\Gamma\left(1/2\right)} =$$

$$= [\text{ based on } remarks\ (1,4)] = \frac{\left(1/2\right)^{1/2}\Gamma\left(1/2\right)}{\Gamma\left(1/2\right)\left(1/2\right)^{1/2}} = 1$$

so $f_{\xi^2}(y)$ – density function!

II. Distribution χ^2 (Pearson)

Lemma 3. (Follows from lemmas 1,2)

If $\xi_1, \dots, \xi_k$ – are independent and have standard normal distribution, then r. v. $\chi^2 = \xi_1^2 + \cdots + \xi_k^2$ has $\Gamma\left(1/2, k/2\right)$ distribution (and specific for statistics χ^2or Pearson).

Definition.

Distribution of k squares of independent r. v. with SND is called chi-square (χ_k^2) with k-degrees of freedom and denoted H_k.

According to lemma 3 H_k is the same as $\Gamma\left(1/2, k/2\right)$, so distribution function for family distributions H_k depends on k and equal

$$f(y) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(k/2\right)} y^{\frac{k}{2}-1} e^{-\frac{y}{2}}, y > 0 \\ 0, \qquad\qquad y \le 0 \end{cases}$$

Note that $H_2 = \Gamma_{1/2,1} = EXP(1/2)$, Exponential distribution.

Remark:

Stability of χ^2 with respect to summation follows from the stability of Γ-distribution.

A number of properties (P):

P1

If r. v. $\chi^2 \in H_k$ and $\psi^2 \in H_m$, then the sum $\chi^2 + \psi^2 \in H_{m+k}$.

Proof:

Let $\xi_1, \xi_2, \ldots$ – are independent and have NSD N(0,1), then $\xi_1^2 + \cdots + \xi_k^2$ has the same distribution as χ^2, analogically ψ^2 and $\xi_k^2 + \cdots + \xi_{k+m}^2$, so the sum

$$\xi_1^2 + \cdots + \xi_{k+m}^2 \in H_{m+k}, \text{ is proven.}$$

P2

If $\chi^2 \in H_k \Rightarrow E\chi^2 = k,\ D\chi^2 = 2k$

Proof:

$\xi_1, \xi_2, \ldots$ – are independent with NSD, then

$E\xi_1^2 = 1, D\xi_1^2 = E\xi_1^4 - (E\xi_1^2)^2 =$

based on theory prob. property:

$[\, E\xi^{2k} = (2k-1)!! = (2k-1)(2k-3) * \ldots * 3 * 1, \qquad E\xi_1^4 = 3]$

$= 3 - 1 = 2$, so

$E\chi^2 = E(\xi_1^2 + \cdots + \xi_k^2) = k; D\chi^2 = D(\xi_1^2 + \cdots + \xi_k^2) = kD(\xi_1^2) = 2k,$

is proven.

P3

Let $\chi_n^2 \in H_n$, then if $n \to \infty$ $\frac{\chi_n^2}{n} \xrightarrow{P} 1$, $\frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow N(0,1)$.

Proof:

For any n, χ_n^2 has the same distribution as $\xi_1^2 + \cdots + \xi_n^2$, $\xi_i \in \mathrm{N}(0,1)$ and independent.

According to LLN and CLT (central limit theorem), have

$$(\xi_1^2 + \cdots + \xi_n^2)/\mathrm{n} \xrightarrow{\mathrm{P}} \mathrm{E}\xi_1^2 = 1,$$

and

$$\frac{\xi_1^2+\cdots+\xi_n^2-n}{\sqrt{2n}} = \frac{\xi_1^2+\cdots+\xi_n^2-nE\,\xi_1^2}{\sqrt{nD\,\xi_1^2}} \Rightarrow N(0,1),$$

is proven.

P4

Let $\chi_n^2 \in H_n$, then if $n \to \infty$, the following weak convergence takes place

$$\sqrt{2\chi_n^2} - \sqrt{2n-1} \Rightarrow N(0,1)$$

Therefore for large n you can use the approximation for distribution function

$$H_n(x) = P(\chi_n^2 < x) \text{ and } H_n(x) \approx \Phi_{0,1}(\sqrt{2x} - \sqrt{2n-1})$$

Proof:

- $\sqrt{2n} - \sqrt{2n-1} \to 0$, $n \to \infty$ (clear);

$$\sqrt{2\chi_n^2} - \sqrt{2n} = \frac{2}{1+\sqrt{\chi_n^2/n}} \frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow N(0,1)$$

because based on P3 $\frac{2}{1+\sqrt{\chi_n^2/n}} \xrightarrow{p} 1$ and $\frac{\chi_n^2-n}{\sqrt{2n}} \Rightarrow N(0,1)$

- $P(\chi_n^2 < x) = P(\sqrt{2\chi_n^2} - \sqrt{2n-1} < \sqrt{2x} - \sqrt{2n-1})$

P5

If r. v. $\xi_1, \ldots, \xi_k$ – Independent and belong to $N(a, \sigma^2)$, then

$$\chi_k^2 = \sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma}\right)^2 \in H_k$$

III. Student Distribution (StD)

Definition.

Let $\xi_0, \xi_1, \ldots, \xi_k \in N(0,1)$ and independent, the distribution of the random variable

$$t_k = \frac{\xi_0}{\sqrt{\frac{\xi_1^2+\cdots+\xi_k^2}{k}}}$$

is called Student distribution with k degrees of freedom and denoted T_k or the same r. v.:

$$t_k = \frac{\xi}{\sqrt{\chi_n^2/k}}, \xi \in N(0,1), \chi_k^2 \in H_k$$

with density function:

$$f_k(y) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi k}\,\Gamma(k/2)}\left(1+\frac{y^2}{k}\right)^{-(k+1)/2} \quad (3)$$

P6

Student distribution $f_k(y)$ is symmetric: if random variable $t_k \in T_k$, then $-t_k \in T_k$.

P7

Student distribution slightly (weakly) converges to $N(0,1)$ if $n \to \infty$.

Proof:

According to P3, $\frac{\chi_n^2}{n} \xrightarrow{P} 1$, if $n \to \infty$ and $t_k = \frac{\xi}{\sqrt{\chi_k^2/k}} \Rightarrow \xi \in N(0,1)$, is proven.

Remark:

For large k one can use normal density function for approximation StD!

Remark:

For k= 1 Student distribution is Cauchy distribution with density function:

$$f_1(y) = \frac{1}{\pi}(1+y^2)^{-1}$$

follows from $\Gamma\left(1/2\right) = \sqrt{\pi}$ and $\Gamma(1) = 1$.

P8

There are moments of order m<k for Student distribution and there are no moments of order $m \geq k$. All existing moments of odd order are equal to zero. (Why?)

IV. Fisher distribution

Fisher distribution (also closely related with normal) is often called distribution of dispersion relation.

Definition.

Let $\chi_k^2 \in H_k$, $\psi_n^2 \in H_n$ both r. v. independent. Distribution of r. v.

$$f_{k,n} = \frac{\chi_k^2/k}{\psi_n^2/n} = \frac{n}{k}\frac{\chi_k^2}{\psi_n^2}$$

is called Fisher's distribution.

P9

If r. v. $f_{k,n} \in F_{k,n} \rightarrow \frac{1}{f_{k,n}} \in F_{n,k}$.

Proof:

For any x : $F_{k,n}(x) = P\left(f_{k,n} < x\right) = P\left(\frac{1}{f_{k,n}} > \frac{1}{x}\right) = 1 - F_{n,k}\left(\frac{1}{x}\right)$,

is proven.

P10

The Fisher distribution $F_{k,n}$ weakly converges to the degenerate distribution at the point c = 1 at any tendency k,$n \rightarrow \infty$.

P11

Let r.v. $t_k \in T_k$, then $t_k^2 \in F_{1,k}$. Proof follows from definition of t_k , and $f_{1,k}$.