Probability, and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Lecture 11

Statistics II., Interval Estimation

1. Necessary Required Mathematical Background

1.1 The Γ Function

The Γ Function

Definition

The $\Gamma : D = \{z \in \mathbb{C} | \mathrm{Re}(z) > 0\} \to \mathbb{C}$

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt$$

function defined in this way is called the Γ **function**.

The Γ function is attributed to the Swiss mathematician Daniel Bernoulli (1700–1782).

Theorem (Properties)

1. $\Gamma(n) = (n-1)!$ *for* $n \in \mathbb{Z}_+$, *that is, the* Γ *function is a continuous extension of the factorial function.*
2. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

1.2 The $\mathcal{B}$ function

The $\mathcal{B}$ function

Definition

The Beta function $\mathcal{B} : \mathbb{R}_+^2 \to \mathbb{R}$ is defined by

$$\mathcal{B}(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx \qquad (\alpha, \beta \in \mathbb{R}_+)$$

Theorem (Properties)

$$\mathcal{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

2. Distributions derived from the normal distribution

2.1. Chi-squared distribution

Chi-squared distribution

Definition

If $\xi_1, \xi_2, \ldots, \xi_n$ are independent, standard normally distributed random variables, then the distribution χ_n^2 defined by

$$\chi_n^2 \sim \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$$

is called the **chi-squared distribution with n degrees of freedom**.

Theorem (Properties)

The density function of the chi-squared distribution with n degrees of freedom is:

$$f(x,n) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

2.2 Student distribution

Student distribution

Definition

If $\xi_0, \xi_1, \ldots, \xi_n$ are independent standard normally distributed random variables, then the random variable t_n defined by

$$t_n \sim \frac{\xi_0}{\sqrt{\frac{\chi_n^2}{n}}}$$

is called a **Student's t-distributed random variable** (where $\chi_n^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$).

The Student's or t-distribution is named after William Sealy Gosset (1876–1937), who was an English statistician, chemist, and brewer. He worked for the Guinness Brewery and published his results under the pen name "Student".

Properties of the Student's t-distribution

Theorem (Properties)

The probability density function of the t_n distribution is:

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad (x \in \mathbb{R}),$$

- *It can be easily seen that for $n = 1$ we obtain the Cauchy distribution with density function*

$$f_1(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \qquad (x \in \mathbb{R}).$$

- *Moreover, if $n \to \infty$, the t_n distribution approximates the standard normal distribution.*

2.3 Fisher-Snedecor distribution

Definition

χ_n^2 and χ_m^2 are independent chi-squared distributed random variables, then the distribution

$$F_{n,m} \sim \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$$

is called a **Fisher-Snedecor random variable with** n **and** m **degrees of freedom**.

Sir Ronald Aylmer Fisher (1890-1962), British polymath. George Waddel Snedecor (1881-1974), American mathematician and statistician.

Theorem (Properties)

The probability density function of the Fisher-Snedecor distribution is

$$f_{n,m}(x) = \frac{n\Gamma\left(\frac{n+m}{2}\right)}{m\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}-1}}{\left(1+\frac{n}{m}x\right)^{\frac{-n+m}{2}}} \qquad (x \geqslant 0)$$

2.4 The $\Gamma_{n,\lambda}$ distribution

The $\Gamma_{n,\lambda}$ disztribution

Definition

If $\xi_1, \xi_2, \ldots, \xi_n$ are independent exponential random variables with parameter λ, then the

$$\Gamma_{n,\lambda} \sim \xi_1 + \xi_2 + \cdots + \xi_n$$

distribution is called the Γ **distribution with** n **degrees of freedom and parameter** λ.

The density function of $\Gamma_{n,\lambda}$ distribution

Theorem

Density function:

$$f_{n,\lambda}(x) = \begin{cases} \dfrac{1}{\Gamma(n)}\lambda^n e^{-\lambda x}, & ha\ x > 0; \\ 0, & otherwise. \end{cases}$$

Remark

In the density functions of the distributions seen above the degrees of freedom n (due to definitions of them) were positive integers. However, the resulting density functions allow us to take the degrees of freedom as positive real numbers.

3. $1 - \alpha$ Confidence Interval

$1-\alpha$ Confidence Interval

Definition

Let $\xi_1, \xi_2, \ldots, \xi_n$ be an independent sample with unknown parameter ϑ. We look for statistics $\underline{g}$ and $\overline{g}$ such that

$$\mathbb{P}\left(\underline{g}(\xi_1, \xi_2, \ldots, \xi_n) \leqslant \vartheta \leqslant \overline{g}(\xi_1, \xi_2, \ldots, \xi_n)\right) \geqslant 1-\alpha.$$

The obtained interval $(\underline{g}(\xi_1, \xi_2, \ldots, \xi_n), \overline{g}(\xi_1, \xi_2, \ldots, \xi_n))$ is called the $1-\alpha$ **confidence interval (with a given confidence level $1-\alpha$)** for the unknown parameter ϑ.

4. Important Confidence Intervals

Important Confidence Intervals

In the sequence the sample $\xi_1, \xi_2, \ldots, \xi_n$ will be always an independent sample from distribution $\mathcal{N}(\mu, \sigma^2)$. We will investigate the following cases:

- Estimating μ when
    1. σ^2 is known: the applied statistic is: $\frac{\bar{\xi}-\mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$;
    2. σ^2 is unknown: the applied statistic is: $\frac{\bar{\xi}-\mu}{s_n^*}\sqrt{n} \sim t_{n-1}$.
- Estimating σ^2

  the applied statistic is: $\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}$

Estimating μ, and σ^2 is known

Theorem

Preserving our previous notations

$$\frac{\overline{\xi} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0, 1).$$

Proof of the case: estimating μ with known σ^2

Proof.

It is easy to see, that

$$\mathbb{E}(\bar{\xi}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n}n\sum_{i=1}^{n}\mathbb{E}(\xi_i) = \mu,$$

$$\mathbb{D}^2(\bar{\xi}) = \mathbb{D}^2\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}n\mathbb{D}^2(\xi_1) = \frac{1}{n}\sigma^2,$$

thus, $\mathbb{D}(\bar{\xi}) = \sqrt{\frac{1}{n}\sigma^2} = \frac{\sigma}{\sqrt{n}}$, from which, by a known theorem, we get $\bar{\xi} \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$, and the standardization gives the statement. □

The procedure for constructing the confidence interval

1. Compute the value of $\overline{\xi}$ from the sample, with n and σ^2 are given.
2. Compute the value of $u_{\frac{\alpha}{2}}$. Since $\frac{\overline{\xi}-\mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$ we have

$$1-\alpha = \mathbb{P}\left(-u_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi}-\mu}{\sigma}\sqrt{n} \leqslant u_{\frac{\alpha}{2}}\right) =$$
$$= \Phi\left(u_{\frac{\alpha}{2}}\right) - \left(1-\Phi\left(u_{\frac{\alpha}{2}}\right)\right) = 2\Phi\left(u_{\frac{\alpha}{2}}\right) - 1.$$

3. Since the value of α is given, the corresponding $u_{\frac{\alpha}{2}} = \Phi\left(1-\frac{\alpha}{2}\right)$ can be easily found.
4. Based on point 2, the confidence interval is

$$\underline{g}(\xi_1,\xi_2,\ldots,\xi_n) = \overline{\xi} - \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}, \qquad \overline{g}(\xi_1,\xi_2,\ldots,\xi_n) = \overline{\xi} + \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}.$$

Eugene Lukacs's Theorem

Theorem

Preserving our previous notation

1. $\overline{\xi} \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$;
2. $\frac{(n-1)s_n^{*2}}{\sigma^2} \sim \chi^2_{n-1}$;
3. $\overline{\xi}$ *and* s_n^{*2} *are independent.*

Proof.

1. This is clear based on our earlier theorems.

2., 3. Not simple; linear algebraic preliminaries are needed, for example the use of an orthogonal matrix. □

Estimating μ, and σ^2 is unknown

Theorem

If $\xi_0, \xi_1, \xi_2, \ldots, \xi_n$ are independent normally distributed random variables, then

$$\frac{\overline{\xi} - \mu}{s_n^*}\sqrt{n} \sim t_{n-1}.$$

Proof.

This is the care of Lukács Theorem. □

Jenő Lukács (Lukacs, Eugene (1906-1987) Hungarian-born American mathematician).

Remark

The method is easy to remember, since in the

$$\frac{\bar{\xi} - \mu}{\sigma}\sqrt{n}$$

statistic the unknown parameter σ must be replaced by its unbiased estimator s_n^*. Then we obtain that

$$\frac{\bar{\xi} - \mu}{s_n^*}\sqrt{n} \sim t_{n-1}$$

Since the density function f of a t_{n-1} distributed random variable is symmetric with respect to the y-axis. We can use the relation $\mathbb{F}(-x) = 1 - \mathbb{F}(x)$ $(x \in \mathbb{R})$, therefore the construction of the confidence interval is analogous to the case where we estimate μ and σ^2 is known.

Steps for Determining the Confidence Interval

1. Determining $t_{\frac{\alpha}{2}}$

$$\mathbb{P}\left(-t_{\frac{\alpha}{2}} \leqslant \frac{\overline{\xi} - \mu}{s_n^*}\sqrt{n} < t_{\frac{\alpha}{2}}\right) = 2\mathbb{F}\left(t_{\frac{\alpha}{2}}\right) - 1 = 1 - \alpha.$$

   The value of $t_{\frac{\alpha}{2}}$ can be find out using the relevant table by

$$\mathbb{F}\left(t_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

2. Thus we obtain that the confidence interval is

$$\overline{\xi} - \frac{t_{\frac{\alpha}{2}} s_n^*}{\sqrt{n}} < \mu < \overline{\xi} - \frac{t_{\frac{\alpha}{2}} s_n^*}{\sqrt{n}}.$$

Constructing the Confidence Interval for the Unknown σ^2 Parameter

Theorem

Preserving our previous notation

$$\frac{ns_n}{\sigma^2} \sim \chi_{n-1}.$$

Proof.

This is clear based on Eugen Lukács's Theorem. □

Steps for Determining the Confidence Interval

1. Since the density function of the distribution χ^2_{n-1} is not symmetric with respect to the y-axis we can not construct a confidence interval that is symmetric about the origin, and consequently we cannot obtain the shortest confidence interval.
2. We know that $\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}$, thus from

$$\mathbb{P}\left(x_l \leqslant \frac{ns_n^2}{\sigma^2} \leqslant x_u\right) = 1 - \alpha$$

   we can find an appropriate pair of values $x_l = x_{\text{lower}}$ and $x_u = x_{\text{upper}}$ using the corresponding table.

Steps for Determining the Confidence Interval, continuation

3. Since

$$\mathbb{P}\left(x_l \leqslant \frac{ns_n^2}{\sigma^2} \leqslant x_u\right) = \mathbb{F}(x_u) - \mathbb{F}(x_l) = 1 - \alpha,$$

we can find the pair x_a, x_f, for example,

$$\mathbb{F}(x_u) = 1 - \frac{\alpha}{2}, \qquad \mathbb{F}(x_l) = \frac{\alpha}{2}.$$

With these numbers x_l and x_u, the confidence interval is:

$$\frac{ns_n^2}{x_u} \leqslant \sigma^2 \leqslant \frac{ns_n^2}{x_l}.$$

End of Lecture 11