

Valószínűségszámítás és Matematikai Statisztika

Miskolc, 2025.

Dr. Glavosits Tamás

11. előadás

Statisztika II. Intervallumbecslések

1. Szükséges matematikai előismeretek

1.1 A Γ függvény

A Γ függvény

Definíció

A $\Gamma : D = \{z \in \mathbb{C} | \operatorname{Re}(z) > 0\} \rightarrow \mathbb{C}$

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

módon definiált függvényt Γ **függvénynek** nevezzük.

A Γ függvény Daniel Bernoulli svájci matematikus (1700-1782) nevéhez fűződik.

Tétel (Tulajdonságai)

1. $\Gamma(n) = (n-1)!$, ha $n \in \mathbb{Z}_+$, azaz a Γ függvény a faktoriális függvény folytonos kiterjesztésének tekinthető.
2. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

1.2 $A \mathcal{B}$ függvény

A \mathcal{B} függvény

Definíció

A Béta függvényt $\mathcal{B} : \mathbb{R}_+^2 \rightarrow \mathbb{R}$

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

módon definiáljuk.

Tétel (Tulajdonságai)

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

2. A normális eloszlásból származtatott eloszlások

2.1 Khi-négyzet eloszlás

Khi-négyzet eloszlás

Definíció

Ha $\xi_1, \xi_2, \dots, \xi_n$ független, standard normális eloszlású valószínűségi változók, akkor a

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

eloszlást n szabadságfokú khi-négyzet eloszlásnak nevezünk.

Tétel (Tulajdonságai)

Az n szabadságfokú khi-négyzet eloszlás sűrűségfüggvénye:

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

2.2 Student eloszlás

Student eloszlás

Definíció

Ha $\xi_0, \xi_1, \dots, \xi_n$ független standard normális eloszlású valószínűségi változók, akkor a

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

valószínűségi változót **student eloszlású valószínűségi változónak** nevezzük (ahol $\chi_n^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2$).

A Student eloszlás William Seally Gosset (1876-1782) angol statisztikus, vegyész és sörfőző nevéhez fűződik. A Guinness sörgyárnak dolgozott, eredményeit Student álnéven publikálta.

A Student eloszlás tulajdonságai

Tétel (Tulajdonságai)

A t_n eloszlás sűrűségfüggvénye:

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

- $n = 1$ esetben Cauchy eloszlást kapunk

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

sűrűségfüggvénnyel.

- Illetve ha $n \rightarrow \infty$ -ben, akkor a t_n közelít a standard normális eloszláshoz.

2.3 Fisher-Snedcecor eloszlás

Definíció

χ_n^2 és χ_m^2 független, kszí-négyzet eloszlású valószínűségi változók, akkor az

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

eloszlást n, m szabadságfokú **Fisher-Snedecor eloszlású valószínűségi változónak** nevezzük.

Sir Ronald Aylmer Fisher (1890-1962) brit polihisztor. George Waddel Snedecor (1881-1974) amerikai matematikus és statisztikus.

Tétel (Tulajdonságai)

A Fisher-Snedecor eloszlás sűrűségfüggvénye

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

2.4 $\Gamma_{n,\lambda}$ eloszlás

$\Gamma_{n,\lambda}$ eloszlás

Definíció

Ha $\xi_1, \xi_2, \dots, \xi_n$ független λ paraméterű exponenciális eloszlású valószínűségi változók, akkor a

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

eloszlást n szabadságfokú λ paraméterű Γ eloszlásnak nevezzük.

A $\Gamma_{n,\lambda}$ eloszlás sűrűségfüggvénye

Tétel

Sűrűségfüggvénye:

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

Megjegyzés

A fent látott eloszlások sűrűségfüggvényeiben az n szabadságfok (a definíciójuk miatt) pozitív egész volt. Azonban a kapott sűrűségfüggvények lehetővé teszik a számunkra, hogy a szabadságfokok tetszőleges pozitív valós számok legyenek.

3. $1 - \alpha$ megbízhatósági konfidencia intervallum

$1 - \alpha$ megbízhatósági konfidencia intervallum

Definíció

Legyen $\xi_1, \xi_2, \dots, \xi_n$ egy független minta ϑ ismeretlen paraméterrel. Keressük azokat a \underline{g} , \bar{g} statisztikákat, amelyre

$$\mathbb{P}(\underline{g}(\xi_1, \xi_2, \dots, \xi_n) \leq \vartheta \leq \bar{g}(\xi_1, \xi_2, \dots, \xi_n)) \geq 1 - \alpha.$$

Ekkor a $(\underline{g}(\xi_1, \xi_2, \dots, \xi_n), \bar{g}(\xi_1, \xi_2, \dots, \xi_n))$ intervallumot a ϑ ismeretlen paraméter $1 - \alpha$ **konfidencia szintű intervallumbecslésének** nevezzük.

4. Fontosabb intervallumbecslések

Fontosabb intervallumbecslések

A továbbiakban a $\xi_1, \xi_2, \dots, \xi_n$ minden esetben $\mathcal{N}(\mu, \sigma^2)$ eloszlású független minta lesz. Az alábbi esetekkel fogunk foglalkozni:

- μ -t becsüljük

1. σ^2 ismeretlen: az alkalmazott statisztika: $\frac{\bar{\xi} - \mu}{\sigma} \sqrt{n} \sim \mathcal{N}(0, 1)$;

2. σ^2 ismeretlen: az alkalmazott statisztika: $\frac{\bar{\xi} - \mu}{s_n^*} \sqrt{n} \sim t_{n-1}$.

- σ^2 -et becsüljük

az alkalmazott statisztika: $\frac{ns_n^2}{\sigma^2} \sim \chi_{n-1}^2$

μ -t becsüljük, σ^2 ismert

Tétel

Megtartva az előző jelöléseket

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

μ -t becsüljük, σ^2 ismert eset bizonyítása

Bizonyítás.

Rendkívül egyszerű, mivel

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

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

így $\mathbb{D}^2(\bar{\xi}) = \sqrt{\frac{1}{n} \sigma^2} = \frac{\sigma}{\sqrt{n}}$, amiből egy ismert tétel alapján kapjuk, hogy $\bar{\xi} \sim \mathcal{N}\left(\mu, \frac{1}{n} \sigma^2\right)$, amiből standardizálással kapjuk az állítást. □

A konfidencia szerkesztésének a menete

1. A mintából kiszámoljuk a $\bar{\xi}$ értékét, n és σ^2 adott.
2. Kiszámoljuk az $u_{\frac{\alpha}{2}}$ értéket. Mivel $\frac{\bar{\xi} - \mu}{\sigma} \sqrt{n} \sim \mathcal{N}(0, 1)$, így

$$\begin{aligned} 1 - \alpha &= \mathbb{P} \left(-u_{\frac{\alpha}{2}} \leq \frac{\bar{\xi} - \mu}{\sigma} \sqrt{n} \leq u_{\frac{\alpha}{2}} \right) = \\ &= \Phi \left(u_{\frac{\alpha}{2}} \right) - \left(1 - \Phi \left(u_{\frac{\alpha}{2}} \right) \right) = 2\Phi \left(u_{\frac{\alpha}{2}} \right) - 1. \end{aligned}$$

3. Mivel az α értéke is adott, így könnyen visszakereshető a táblázat alapján az $u_{\frac{\alpha}{2}} = \Phi \left(1 - \frac{\alpha}{2} \right)$ értéke.
4. A 2. pont alapján kapjuk, hogy a konfidencia intervallum

$$\underline{g}(\xi_1, \xi_2, \dots, \xi_n) = \bar{\xi} - \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}, \quad \bar{g}(\xi_1, \xi_2, \dots, \xi_n) = \bar{\xi} + \frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}},$$

Lukács Jenő tétele

Tétel

Ha $\xi_1, \xi_2, \dots, \xi_n \mathcal{N}(\mu, \sigma^2)$ eloszlású független valószínűségi változók, akkor

1. $\bar{\xi} \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$;
2. $\frac{(n-1)s_n^{*2}}{\sigma^2} \sim \chi_{n-1}^2$;
3. $\bar{\xi}$ és s_n^{*2} függetlenek.

Bizonyítás.

1. A korábbi tételeink alapján nyilvánvaló.
- 2., 3. Nem egyszerű, lineáris algebrai előismeretek kellenek hozzá, például ortogonális mátrixot használ.



μ -t becsüljük, σ^2 ismeretlen

Tétel

Ha $\xi_0, \xi_1, \xi_2, \dots, \xi_n$ független normális eloszlású valószínűségi változók, akkor

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

Bizonyítás.

Ezt mondja ki a Lukács Tétel. □

Lukács Jenő (Lukacs, Eugene (1906-1987) magyar születésű amerikai matematikus).

Megjegyzés

A módszer könnyen megjegyezhető, ugyanis a

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

statisztikában az ismeretlen σ paramétert ki kell cserélni annak s_n^* torzítatlan becslésére. Így mi kapjuk, hogy

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

Mivel egy t_{n-1} eloszlású valószínűségi változó f sűrűségfüggvénye egy y tengelyre szimmetrikus függvény. Tehát tudjuk alkalmazni az $\mathbb{F}(-x) = 1 - \mathbb{F}(x)$ ($x \in \mathbb{R}$ összefüggést, így a konfidencia intervallum szerkesztése analóg a μ -t keressük, σ^2 ismert esettel.

A konfidencia intervallum meghatározásának lépései

1. A $t_{\frac{\alpha}{2}}$ meghatározása

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

A megfelelő táblázatból a $t_{\frac{\alpha}{2}}$ visszakereshető az

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

összefüggés alapján.

2. Így kapjuk, hogy a konfidencia intervallum

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

Az ismeretlen σ^2 paraméterre vonatkozó konfidencia intervallum szerkesztése.

Tétel

Megőrizve korábbi jelöléseinket

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

Bizonyítás.

Lukács Jenő Tétele alapján nyilvánvaló. □

A konfidencia intervallum szerkesztésének a lépései

1. Mivel a χ_{n-1}^2 eloszlás sűrűségfüggvénye nem szimmetrikus az y tengelyre, így nem tudunk olyan konfidencia intervallumot szerkeszteni, amely az origóra szimmetrikus, (és így) a legrövidebb konfidencia intervallumot sem tudjuk előállítani.
2. Tudjuk, hogy $\frac{ns_n^2}{\sigma^2} \sim \chi_{n-1}^2$, így az

$$\mathbb{P} \left(x_a \leq \frac{ns_n^2}{\sigma^2} \leq x_f \right) = 1 - \alpha$$

összefüggés alapján vissza kell keresnünk a megfelelő táblázatból egy megfelelő $x_a = x_{\text{alsó}}$ és $x_f = x_{\text{felső}}$ értékpár.

A konfidencia intervallum szerkesztésének a lépései, folytatás

3. Mivel

$$\mathbb{P} \left(x_a \leq \frac{ns_n^2}{\sigma^2} \leq x_f \right) = \mathbb{F}(x_f) - \mathbb{F}(x_a) = 1 - \alpha,$$

így elegendő azt az x_a , x_f párt megtalálni, amelyre

$$\mathbb{F}(x_f) = 1 - \frac{\alpha}{2}, \quad \mathbb{F}(x_a) = \frac{\alpha}{2}.$$

Ezekkel az x_a és x_f számokkal a konfidencia intervallum:

$$\frac{ns_n^2}{x_f} \leq \sigma^2 \leq \frac{ns_n^2}{x_a}.$$

Vége az 11. előadásnak