

# A sztochasztika alapjai

## MBNXK262

12. előadás: Konfidenciaintervallumok, próbák

2024/25 tavasz

# Konfidenciaintervallumok

## Definíció

A  $(T_1(\xi), T_2(\xi))$  statisztikapárral definiált intervallum *legalább*  $1 - \varepsilon$  megbízhatósági szintű konfidenciaintervallum a  $\psi(\theta)$  paraméterre, ha

$$\mathbf{P}_\theta(T_1(\xi) < \psi(\theta) < T_2(\xi)) \geq 1 - \varepsilon, \quad \forall \theta \in \Theta,$$

ahol  $\varepsilon > 0$  előre adott kicsi szám.

## Konfidenciaintervallum normális eloszlás várható értékére ismert szórás esetén

Legyenek  $\xi_1, \dots, \xi_n \sim N(\mu, \sigma_0^2)$  független véletlen változók, ahol  $\mu$  az ismeretlen paraméter,  $\sigma_0^2$  ismert. Megadunk  $\mu$ -re pontosan  $1 - \varepsilon$  megbízhatósági szintű konfidenciaintervallumot.

# Konfidenciaintervallum

## Tétel

*Ha  $\xi, \eta$  független normális eloszlású véletlen változók akkor  $\xi + \eta$  is normális eloszlású.*



# Konfidenciaintervallum – ismeretlen szórás

$\sigma_0$  ismert:  $\left(\bar{\xi} - \frac{u_{\varepsilon/2}\sigma_0}{\sqrt{n}}, \bar{\xi} + \frac{u_{\varepsilon/2}\sigma_0}{\sqrt{n}}\right)$

$\sigma$  ismeretlen, becsüljük

$$V_n^* = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2$$

## Definíció

$\eta, \eta_1, \dots, \eta_n$  független standard normálisok

$$T_n = \frac{\eta}{\sqrt{(\eta_1^2 + \dots + \eta_n^2)/n}}$$

változó  $n$  szabadsági fokú Student-eloszlású (vagy  $t$ -eloszlású) véletlen változó. Eloszlásfüggvénye  $\Phi_n(x) = \mathbf{P}(T_n < x)$ .

## Tétel

$\xi_1, \dots, \xi_n$  független normális eloszlású véletlen változók  $\mu$  várható értékkel és  $\sigma$  szórással. Ekkor az  $\bar{\xi}_n$  mintaátlag és a  $V_n$  empirikus szórásnégyzet függetlenek. Továbbá a

$$\sqrt{n} \frac{\bar{\xi}_n - \mu}{\sqrt{V_n^*}}$$

hányados  $n - 1$  szabadsági fokú Student-eloszlású.

Keressük a konfidenciaintervallumot  $[\bar{\xi}_n - r_\varepsilon, \bar{\xi}_n + r_\varepsilon]$  alakban.  
Ekkor

$$\begin{aligned}
 & \mathbf{P}_\mu(\bar{\xi} - r_\varepsilon < \mu < \bar{\xi} + r_\varepsilon) \\
 &= \mathbf{P}_\mu \left( -\frac{r_\varepsilon}{\sqrt{V_n^*}/\sqrt{n}} < \frac{\bar{\xi} - \mu}{\sqrt{V_n^*}/\sqrt{n}} < \frac{r_\varepsilon}{\sqrt{V_n^*}/\sqrt{n}} \right) \\
 &= \Phi_{n-1} \left( \frac{\sqrt{n}r_\varepsilon}{\sqrt{V_n^*}} \right) - \Phi_{n-1} \left( -\frac{\sqrt{n}r_\varepsilon}{\sqrt{V_n^*}} \right) \\
 &= 2\Phi_{n-1} \left( \frac{\sqrt{n}r_\varepsilon}{\sqrt{V_n^*}} \right) - 1 = 1 - \varepsilon,
 \end{aligned}$$

azaz  $\Phi_{n-1} \left( \frac{\sqrt{n}r_\varepsilon}{\sqrt{V_n^*}} \right) = 1 - \frac{\varepsilon}{2}.$

A keresett  $1 - \varepsilon$  megbízhatósági szintű konfidenciaintervallum

$$\left( \bar{\xi} - \frac{t_{\varepsilon/2} \sqrt{V_n^*}}{\sqrt{n}}, \bar{\xi} + \frac{t_{\varepsilon/2} \sqrt{V_n^*}}{\sqrt{n}} \right).$$

# Hipotézisvizsgálat

$\xi_1, \dots, \xi_n$  független, normálisok,  $\sigma_0$  ismert szórással, és *ismeretlen*  $\mu$  várható értékkel. Azt szeretnénk eldönteni, hogy igaz-e hogy a várható érték  $\mu_0$ .

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

Az

$$u = \frac{\bar{\xi}_n - \mu_0}{\sigma_0} \sqrt{n}$$

standard normális a  $H_0$  fennállása esetén. Ha  $H_1$  áll fenn, akkor  $u$  *nem* standard normális. Legyen  $\varepsilon > 0$  rögzített szignifikanciaszint. Az  $u_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2)$  jelöléssel,

$$\begin{aligned} & \mathbf{P}_{\mu_0} \left( \mu_0 \in \left( \bar{\xi}_n - \frac{u_{\varepsilon/2} \sigma_0}{\sqrt{n}}, \bar{\xi}_n + \frac{u_{\varepsilon/2} \sigma_0}{\sqrt{n}} \right) \right) \\ &= \mathbf{P}_{\mu_0} (|u| \leq u_{\varepsilon/2}) = 1 - \varepsilon. \end{aligned}$$

## u-próba

1. A minta alapján kiszámoljuk az  $u$  próbastatisztikát.
2. Rögzített  $\varepsilon > 0$  szignifikanciaszinthez meghatározzuk  $u_{\varepsilon/2}$  értéket.
3. Ha  $|u| \leq u_{\varepsilon/2}$  akkor elfogadjuk  $H_0$ -t, különben elvetjük.

## Példa

Azt szeretnénk tesztelni, hogy az 1kg-os cukor tényleg 1kg-e?  
Tfh: egy csomag tömege normális eloszlást követ  $\mu$  várható értékkel (ezt nem ismerjük), és  $\sigma_0 = 0,05$  ismert szórással.

$$H_0 : \mu = 1 \quad \text{vs.} \quad H_1 : \mu \neq 1. \quad (1)$$

25 elemű minta alapján  $\bar{x}_{25} = 0,98$  adódik. A próbastatisztika értéke

$$u = \frac{\bar{x}_{25} - 1}{0,05} \sqrt{25} = -2.$$

## Példa

$$u = \frac{\bar{x}_{25} - 1}{0,05} \sqrt{25} = -2.$$

Legyen  $\varepsilon = 0,05$ . Ekkor  $u_{\varepsilon/2} = \Phi^{-1}(0,975) = 1,96$ . Mivel  $|u| = 2 > 1,96$ , ezért elvetjük a nullhipotézist, azaz úgy döntünk, hogy nem 1 a várható érték.

## Példa

$$u = \frac{\bar{x}_{25} - 1}{0,05} \sqrt{25} = -2.$$

Legyen  $\varepsilon = 0,05$ . Ekkor  $u_{\varepsilon/2} = \Phi^{-1}(0,975) = 1,96$ . Mivel  $|u| = 2 > 1,96$ , ezért elvetjük a nullhipotézist, azaz úgy döntünk, hogy nem 1 a várható érték.

Ugyanakkor, ha  $\varepsilon = 0,01$ . Ekkor  $u_{\varepsilon/2} = \Phi^{-1}(0,995) = 2,57$ . Mivel  $|u| = 2 < 2,57$ , ezért a nullhipotézist *ezen a szignifikanciaszinten* elfogadjuk.

# Hibák

*Elsőfajú hiba:*  $H_0$  fennáll, mégis elutasítjuk,

*Másodfajú hiba:*  $H_0$  nem áll fenn, mégis elfogadjuk.

## Elsőfajú hiba

Ha a valódi paraméter  $\mu_0$ , és  $|u| \geq u_{\varepsilon/2}$ , ennek valószínűsége  $\mathbf{P}(|u| \geq u_{\varepsilon/2}) = \varepsilon$ , azaz ez éppen az előre megadott  $\varepsilon$  szignifikanciaszint.

## Másodfajú hiba

Ha a valódi paraméter  $\mu \neq \mu_0$ , azaz  $H_1$  áll fönn, és  $|u| \leq u_{\varepsilon/2}$ . Ennek valószínűsége  $\mathbf{P}_\mu(|u| \leq u_{\varepsilon/2})$  függ a valódi paraméter értékétől. A próba *erőfüggvénye*

$$\beta_n(\mu, \varepsilon) = 1 - \mathbf{P}_\mu(|u| \leq u_{\varepsilon/2}) = \mathbf{P}_\mu(|u| \geq u_{\varepsilon/2}).$$

Legyen  $\Delta_n = \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}$ . Ha  $H_1$  áll fenn, és a várható érték valódi értéke  $\mu$ , akkor  $\sqrt{n}(\bar{\xi}_n - \mu)/\sigma_0$  változó lesz standard normális, ezért

$$\begin{aligned}\beta_n(\mu, \varepsilon) &= \mathbf{P}_\mu(|u| \geq u_{\varepsilon/2}) \\ &= \mathbf{P}_\mu \left( \left| \frac{\bar{\xi}_n - \mu}{\sigma_0} \sqrt{n} + \Delta_n \right| \geq u_{\varepsilon/2} \right) \\ &= 1 - \Phi(u_{\varepsilon/2} - \Delta_n) + \Phi(-u_{\varepsilon/2} - \Delta_n) \\ &= 2 - [\Phi(u_{\varepsilon/2} - \Delta_n) + \Phi(u_{\varepsilon/2} + \Delta_n)] .\end{aligned}$$

Innen látjuk, hogy

$$\lim_{n \rightarrow \infty} \beta_n(\mu, \varepsilon) = 1, \quad \text{tetszőleges } \mu \neq \mu_0 \text{ esetén.}$$

Ez utóbbi állítás próba konzisztenciáját jelenti.

## t-próba

A feladat ugyanaz, mint az u-próbánál, csak  $\sigma$  ismeretlen.

Legyen  $\xi \sim N(\mu, \sigma^2)$  *ismeretlen*  $\sigma$ , ismeretlen  $\mu$ . Igaz-e hogy a várható érték  $\mu_0$ .

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

$\sigma$  ismeretlen, becsüljük  $\sqrt{V_n^*}$  korrigált empirikus szórással.  
Különben minden ugyanaz. A tesztstatisztika

$$t = \frac{\bar{\xi}_n - \mu_0}{\sqrt{V_n^*}} \sqrt{n}$$

$H_0$  fennállása esetén Student-eloszlású  $n - 1$  szabadsági fokkal.

**t-próba.** Ezek alapján a következőképpen járunk el:

1. A minta alapján kiszámoljuk az  $t$  próbastatisztikát.
2. Rögzített  $\varepsilon > 0$  szignifikanciaszinthez meghatározzuk  $t_{\varepsilon/2}$  értéket.
3. Ha  $|t| \leq t_{\varepsilon/2}$  akkor elfogadjuk  $H_0$ -t, különben elvetjük.

# Teszt

Rögzített megbízhatósági szint esetén, a mintaelemszám növelésével a konfidenciaintervallum hossza

- ▶ csökken
- ▶ nő
- ▶ nem változik
- ▶ az adott mintától függ

# Teszt

Az u-próbánál a nullhipotézist akkor fogadjuk el, ha a próbastatisztika

- ▶ pozitív
- ▶ 0
- ▶ nagyobb, mint 1
- ▶ beleesik egy megadott 0 középpontú intervallumba

# Teszt

Melyik próba esetén standard exponenciális eloszlású a minta?

- ▶ u
- ▶ t
- ▶ mindkettő
- ▶ egyik sem

# Teszt

Százszor dobunk egy szabályos érmével. Annak a valószínűsége, hogy a fejek száma 45 és 55 közé esik közelítőleg

- ▶ 1
- ▶ 0.5
- ▶  $\Phi(1)$
- ▶  $2\Phi(1) - 1 = 0.68$