A sztochasztika alapjai
MBNXK262

12. előadás: Konfidenciaintervallumok, próbák

2022/23 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 $\xi_1, \ldots, \xi_n \sim \mathrm{N}(\mu, \sigma_0^2)$ független véletlen változók, ahol μ az ismeretlen paraméter, σ_0^2 ismert. Megadunk μ-re pontosan $1-\varepsilon$ megbízhatósági szintű konfidenciaintervallumot.

Konfidenciaintervallum

Tétel

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

Hipotézisvizsgálat

$\xi_1, \ldots, \xi_n$ független, normálisok, ismert σ_0 szórással, és *ismeretlen* μ várható értékkel. Azt szeretnénk eldönti, hogy igaz-e hogy a várható érték μ_0.

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

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

Az

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

statisztika standard normális eloszlású 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,

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

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 μ várható értékkel (ezt nem ismerjük), és $\sigma_0 = 0,05$ ismert szórással.

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

Egy 25 elemű minta alapján a mintaátlagra $\overline{x}_{25} = 0,98$ adódik.
A próbastatisztika értéke

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

Példa

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

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

Példa

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

Legyen $\varepsilon = 0,05$. Ekkor a standard normális eloszlástáblázatából kiolvassuk, hogy $u_{\varepsilon/2} = \Phi^{-1}(0,975) = 1,96$. 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 μ_0, és $|u| \geq u_{\varepsilon/2}$, ennek valószínűsége $\mathbf{P}(|u| \geq u_{\varepsilon/2}) = \varepsilon$, azaz ez éppen az előre megadott ε 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 μ, akkor $\sqrt{n}(\bar{\xi}_n - \mu)/\sigma_0$ változó lesz standard normális, ezért

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

Innen látjuk, hogy

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

Ez utóbbi állítás próba konzisztenciáját jelenti.