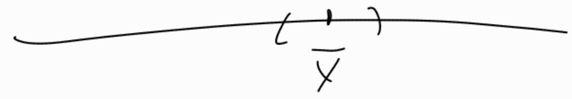
VRSPFMF2026-04-02 zamudil 10 minut

manjkal na prejšnjih predavanjih, zbolel.

Let $X \sim N(\mu, \sigma)$ na populaciji. Potem je $\bar{X} = \frac{X_1+\dots+X_n}{n}$ porazdeljena $N(\mu, \frac{\sigma}{\sqrt{n}})$ oziroma vzorčna statistika

$$Z := \frac{\bar{X}-\mu}{\sigma} \sim N(0,1).$$

Če poznamo σ, potem bomo znali povedati, kako dobra ocena za μ je vzorčno povprečje $\bar{X}$ $\Longrightarrow$ intervali zaupanja.



„Z veliko verjetnostjo je neznani parameter μ na tem intervalu"

Pogosto σ ne poznamo. V tem primeru σ ocenimo s $S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2}$. Toda tedaj ima vzorčna statistika

$T := \frac{\bar{X}-\mu}{S}\cdot\sqrt{n}$ drugačno porazdelitev (ne $N(0,1)$).

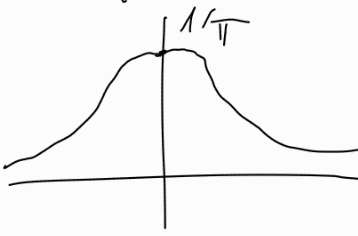
Kako pa je T porazdeljena? Ker je $\chi^2 = \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k-\bar{X})^2 = \frac{(n-1)S^2}{\sigma^2}$, ki je porazdeljena po $\chi^2(n-1)$, je

$Z/T = \frac{S}{\sigma} = \sqrt{\frac{\chi^2}{n-1}}$, torej $T = \frac{Z}{\sqrt{\frac{\chi^2}{n-1}}}$. Pokazati se da (NBD), da sta Z in χ^2 neodvisni sl. spr. Vemo, da je $Z \sim N(0,1)$ in $\chi^2 \sim \chi^2(n-1)$. Od tod lahko izračunamo, da ima T gostoto

$$p_T(t) = \frac{1}{\sqrt{n-1}\cdot B\left(\frac{n-1}{2}, \frac{1}{2}\right)\cdot\left(1+\frac{t^2}{n-1}\right)^{n/2}}, \quad t \in \mathbb{R}$$

Temu rečemo Studentova T-porazdelitev z $n-1$ prostostnimi stopnjami.

Pri $n=2$: $p_T(t) = \frac{1}{B(\frac{1}{2}, \frac{1}{2})}\cdot\frac{1}{1+t^2} = \frac{1}{\pi(1+t^2)}$ ... to je Cauchyjeva porazd.:



NBD: Ko gre $n \to \infty$, gre $\sqrt{n-1}\, B\left(\frac{n-1}{2}, \frac{1}{2}\right) \to \sqrt{2\pi}$, $\left(1+\frac{t^2}{n-1}\right)^{n/2} \to e^{t^2/2}$ in torej $p_T(t) \to \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$, kar je std normalna gostota $N(0,1)$.

To porazdelitev je 1908 odkril William S. Gosset, statistik v dublinski pivovarni Guiness. Objavil pod psevdonimom. Kvaliteto izdelkov je preverjal na majhnih vzorcih (malih).

Za $n \geq 3$ je $E(T) = 0$. Za $n \geq 4$ je $D(T) = \frac{n-1}{n-3}$. (NBD)

$T = \frac{\bar{X}-\mu}{S}\cdot\sqrt{n} \sim \text{Student}(n-1)$.

[3. Metode za pridobivanje cenilk].

1. Metoda momentov

Začetni moment reda $k \in \mathbb{N}$ je $z_k = E(X^k)$.

Let $(X_1, X_2, \dots, X_n)$ vzorec.

Vzorčni začetni moment reda k je $Z_k := \frac{X_1^k + \dots + X_n^k}{n}$.

To je nepristranska cenilka za z_k:

$$E(Z_k) = \frac{1}{n} \sum_{j=1}^{n} \underbrace{E(X_j^k)}_{= z_k} = z_k$$

Po šibkem zakonu velikih števil (izrek Čebiševa) je to tudi dosledna cenilka.

Vzemimo sedaj, da ima X gostoto, ki je odvisna od m parametrov $\xi_1, \xi_2, \dots, \xi_m$, torej $p_X(x, \xi_1, \xi_2, \dots, \xi_m)$.

Potem je $z_k = E(X^k) = \int_{-\infty}^{\infty} x^k p_X(x, \xi_1, \dots, \xi_m)\, dx$. Za k vzemimo $1, 2, \dots, m$.

$$z_1 = f_1(\xi_1, \dots, \xi_m)$$

$$\dots$$

$$z_m = f_m(\xi_1, \dots, \xi_m)$$

Denimo, da iz teh enačb lahko izrazimo parametre kot f-je $z_1, \dots, z_m$:

$$\xi_1 = \varphi_1(z_1, \dots, z_m)$$

$$\dots$$

$$\xi_m = \varphi_m(z_1, \dots, z_m)$$

Potem je $C_k := \varphi_k(\overset{\text{velik } z}{Z_1}, \dots, Z_m)$ cenilka za parameter ξ_k $\forall k \in [m]$ (Z_i je cenilka za z_i).

Zgled 1.) $X \sim N(\mu, \sigma)$. $p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

Potem je $z_1 = E(X) = \mu$

$$z_2 = E(X^2) = D(X) + (E(X))^2 = \sigma^2 + \mu^2$$

Od tod sledi $\mu = z_1$, $\sigma^2 = z_2 - \mu^2 = z_2 - z_1^2$. Cenilka za μ je $Z_1 = \frac{X_1 + \dots + X_n}{n} = \overline{X}$. Cenilka za $\sigma^2 = Z_2 - Z_1^2 =$

$$= \frac{1}{2}\sum_{i=1}^{n} X_i^2 - \Big(\overbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}^{=\overline{X}}\Big)^2.$$

Dokažimo, da je to cenilka enaka cenilki $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 + 2X_i\overline{X} + \overline{X}^2) =$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\overline{X} \cdot n\overline{X} + \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

~~Zgled~~ Zgled 2): Let X enakomerno porazd. na $[a,b]$. Kjer sta a in b neznana parametra, torej

$$p(x; a,b) = \begin{cases} 1/(b-a) & ; \ x \in [a,b] \\ 0 & ; \ \text{sicer} \end{cases}$$

$$z_1 = E(X) = \frac{a+b}{2} \qquad z_2 = E(X^2) = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\,\frac{x^3}{3}\Big|_a^b =$$

$$= \frac{1}{b-a}\cdot\frac{b^3-a^3}{3} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

izrazimo b iz prve: $b = 2z_1 - a$

in vstavimo v drugo:

$$3z_2 = a^2 + a(2z_1-a) + (2z_1-a)^2 = a^2 + 2z_1a - a - 4z_1^2 - 4z_1a + a^2 =$$

$$= a^2 - 2z_1a + 4z_1^2.$$ dobimo kvadratno enačbo

za a: $a^2 - 2z_1 a + (4z_1^2 - 3z_2) = 0$. rešimo jo:

$$D = 4z_1^2 - 4(4z_1^2 - 3z_2) = 12(z_2 - z_1^2) \Rightarrow a_{1,2} = \frac{1}{2}\left(2z_1 \pm \sqrt{12(z_2-z_1^2)}\right) =$$

$$= z_1 \pm \sqrt{3(z_2-z_1^2)}$$

ker je $a<b$, je $a = z_1 - \sqrt{3(z_2-z_1^2)}$; $b = z_1 + \sqrt{3(z_2-z_1^2)}$

cenilka za a je $A := Z_1 - \sqrt{3(Z_2-Z_1^2)} = \overline{X} - S_0\sqrt{3}$

cenilka za b je $B := \overline{X} + S_0\sqrt{3}$

Denimo, da imamo konkreten vzorec: $-2, 0, 1, 2, 4$

$$\overline{x} = \frac{5}{5} = 1, \qquad s_0^2 = \frac{4+0+1+4+16}{5} - 1^2 = 4$$

$\Rightarrow s_0 = 2 \Rightarrow$ vzorčna ~~mednost~~ za A je torej

$1 - 2\sqrt{3} \doteq -2{,}46$

vzorčna vrednost za B je ~~torej~~

$1 + 2\sqrt{3} \doteq 4{,}46$

2. metoda maksimalne zanesljivosti / največjega verjetja

Naj bo ~~gostota~~ za X odvisna od parametra ξ, torej $p_X(x, \xi)$. Funkcija zanesljivosti (likelyhood function):
$$L(x_1, \dots, x_n; \xi) = p(x_1; \xi)\cdot p(x_2;\xi)\cdots p(x_n;\xi).$$
Določimo tak ξ, da bo L največja za dane $x_i \in \mathbb{R}$.
Seveda je ξ_{max} odvisen od $x_1, \dots, x_n$; $\xi_{max} = \varphi(x_1, \dots, x_n)$ za nek φ. Cenilka za ξ je sedaj $C := \varphi(X_1, \dots, X_n)$.