2. Vzorčne statistike in cenilke

Naj bo $(X_1, \dots, X_n)$ vzorec. Vzorčna statistika Y je simetrična funkcija sl. vzorca:

$$Y = g(X_1, \dots, X_n)$$

Pravilom Y ocenjuje vrednost nekega parametra ζ. Tedaj je Y cenilka za ζ.

Y je odvisna od n, zato pišemo tudi $Y_n = g(X_1, \dots, X_n)$.

Y je nepristranska cenilka za ζ, če je $E(Y) = \zeta$.

Cenilka $Y = Y_n$ je **dosledna**, če $Y_n \xrightarrow[n\to\infty]{} \xi$ verjetnostno, tj. za vsak $\varepsilon > 0$ je

$$P\left(|Y_n - \xi| < \varepsilon\right) \xrightarrow[n\to\infty]{} 1.$$

Standardno deviacijo vzorčne statistike Y imenujemo **standardna napaka**: $SE(Y) = \sigma(Y)$

Naj ima X na populaciji mat. upanje $E(X) = \mu$ in $\sigma(X) = \sigma$. **Vzorčno povprečje**

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

je nepristranska cenilka za μ:

$$E(\bar{X}) = \frac{E(X_1) + \dots + E(X_n)}{n} = \frac{n \cdot \mu}{n} = \mu.$$

Po ŠZVŠ (izrek Čebiševa) je $\overline{X}$ tudi dosledna cenilka za μ. Ker je $D(\overline{X}) = \frac{1}{n^2} D(X_1 + \ldots + X_n)$
$= \frac{1}{n^2} \cdot \left(D(X_1) + \ldots + D(X_n)\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$, je

$SE(\overline{X}) = \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$. Čim večji je n, bolje $\overline{X}$ oceni μ. Po CLI ima pri velikih n sl. spremen.

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} = \frac{\overline{X} - \mu}{\sigma} \sqrt{n}$$

pribl. porazd. $N(0,1)$ oziroma $\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Če je $X \sim N(\mu, \sigma)$, je $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ za vsak n.

Trditev Naj bo Y_n cenilka za ξ. Če $E(Y_n) \xrightarrow[n\to\infty]{} \xi$ in $D(Y_n) \xrightarrow[n\to\infty]{} 0$, je Y_n dosledna cenilka za ξ.

Dokaz Fiksirajmo $\varepsilon > 0$. Potem obstaja n_0, da za vse $n \geq n_0$ velja $\left|E(Y_n) - \xi\right| < \frac{\varepsilon}{2}$. Sedaj za vse $n \geq n_0$ velja $\left(|Y_n - \xi| \geq \varepsilon\right) \subseteq$

$$\left(|Y_n - E(Y_n)| + |E(Y_n) - \xi| \geq \varepsilon\right) \subseteq$$

$$\left(|Y_n - E(Y_n)| \geq \frac{\varepsilon}{2}\right).$$ Torej imamo neenakost

$$P(|Y_n - \xi| \geq \varepsilon) \leq P\left(|Y_n - E(Y_n)| \geq \frac{\varepsilon}{2}\right) \leq$$

$$\leq \frac{D(Y_n)}{\varepsilon^2} \cdot 4 \xrightarrow[n\to\infty]{} 0$$ po neenakosti

Čebiševa. Torej $Y_n \xrightarrow[n\to\infty]{} \xi$ verjetnostno.

Porazdelitev $\chi^2(n)$ z n prostostnimi stopnjami ima gostoto

$$p(x) = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad \text{za } x > 0,$$

sicer je $p(x) = 0$.

$E(X) = n$, $D(X) = 2n$, modus $= n-2$, če $n \geq 2$

mediana $\approx n\left(1 - \frac{2}{3n}\right)^3$

Vzorčna disperzija je
$$S_0^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2,$$
popravljena vzorčna disperzija je $S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2$

Kako sta porazdeljeni, če je $X \sim N(\mu, \sigma)$?

Raje vzemimo vzorčno statistiko
$$\chi^2 = \frac{1}{\sigma^2} \sum_{k=1}^{n} (X_k - \overline{X})^2 = \frac{n}{\sigma^2} \cdot S_0^2 = \frac{n-1}{\sigma^2} \cdot S^2$$

Izkaže se, da je $\chi^2 \sim \chi^2(n-1)$:

χ^2 se zapiše kot $Z_1^2 + \dots + Z_{n-1}^2$, kjer je $Z_k \sim N(0,1)$ in so neodvisni; potem se uporabi trditev iz verjetnosti.

Ker je $E(\chi^2) = n-1$, je $E(S_0^2) = \frac{\sigma^2}{n} \cdot E(\chi^2) = \frac{n-1}{n} \cdot \sigma^2$, torej S_0^2 ni nepristranska cenilka za σ^2; ker je $E(S_0^2) \xrightarrow[n \to \infty]{} \sigma^2$, je <u>asimptotično nepristranska</u>. Ker je pa $E(S^2) = \frac{\sigma^2}{n-1} E(\chi^2) = \sigma^2$, je S^2 nepristranska cenilka za σ^2.

Ker je $D(S_0^2) = \frac{\sigma^4}{n^2} D(\chi^2) = \sigma^4 \cdot \frac{2(n-1)}{n^2} \xrightarrow[n\to\infty]{} 0$,
je S_0^2 doslední cenilka za σ^2. Podobno vidimo, da je S^2 tudi dosledna.

Studentova t-porazdelitev z n prost. stop.

ima gostoto: $p(x) = \frac{1}{\sqrt{n} \cdot B\left(\frac{n}{2}, \frac{1}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} =$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n \cdot \pi} \cdot \Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

Leta 1908 jo je odkril W. S. Gosset, statistik v pivovarni Guinness v Dublinu, Student je njegov psevdonim.

Ker je $\lim\limits_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}} = \lim\limits_{n\to\infty}\left(\left(1+\frac{x^2}{n}\right)^n\right)^{\frac{n+1}{2n}} = e^{\frac{x^2}{2}}$ in $\lim\limits_{n\to\infty}\sqrt{n}\cdot B\left(\frac{n}{2},\frac{1}{2}\right) = \sqrt{2\pi}$, je

$$p(x) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$