A sztochasztika alapjai

12. feladatsor: statisztikai alapok

1. Egy játék fizikai terhelhetőségére elvégzett tesztek kg-ban a következő eredményeket adták: 40, 45, 40, 42, 36. A minta alapján adjuk meg a terhelhetőség empirikus eloszlásfüggvényét, mintaátlagát, korrigált/korrigálatlan empirikus szórásnégyzetét és a mediánt!

$n=5 \quad x_1=40 \quad x_2=45 \quad x_3=40 \quad x_4=42 \quad x_5=36$

egy konkrét realizáció

véletlen — determ.

tapasztalati | elméleti

mintaátlag $\overline{x}_5 = \frac{1}{5}\cdot(x_1+x_2+\dots+x_5) = \dots$

$\frac{X_1+\dots+X_n}{n} = \overline{X}_n \qquad E(X)$

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2 = S_n^2 \qquad D^2(X)$

$\frac{1}{n-1}\sum(\ )^2 = S_n^{*2}$

empirikus szórásnégyzet

$s_5^2 = \frac{1}{5}\cdot\sum_{i=1}^{5}(x_i-\overline{x}_5)^2 = \dots \leftarrow$ korrigálatlan

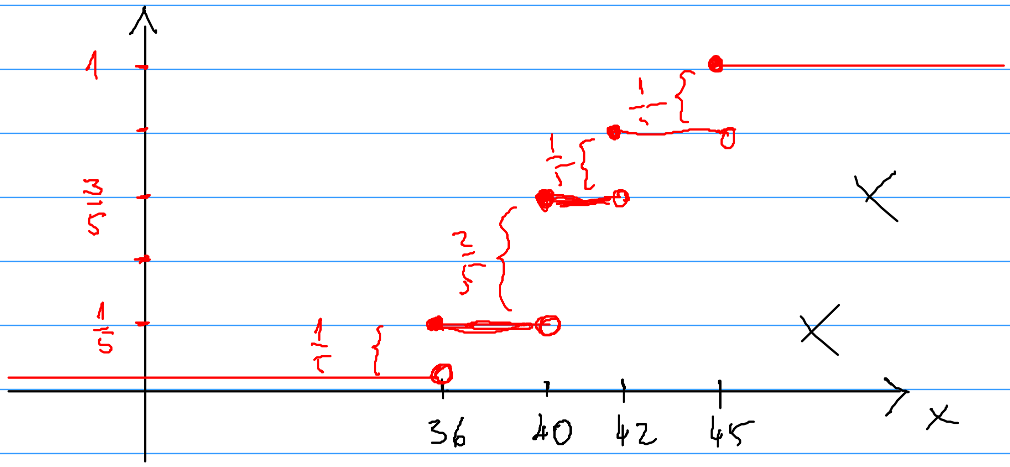
$F_n(x) \qquad F(x)$

$s_5^{*2} = \frac{1}{4}\sum_{i=1}^{5}(x_i-\overline{x}_5)^2 \leftarrow$ korrigált

$\leftarrow (n-1)$-el osztunk

40, 45, 40, 42, 36

Empirikus eloszlásfüggvény: $F_n(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} I(X_i \leq x)$ ideális eloszlásfv.



- 36, 40, 40, 42, 45

medián

-

2. Legyenek $X_1, \ldots, X_n$ független azonos eloszlású véletlen változók véges μ várható értékkel és $\sigma > 0$ szórással. Igazoljuk, hogy $\hat{\mu}_1 = X_1$ torzítatlan becslése a várható értéknek! Igazoljuk, hogy $\hat{\mu}_2 = (X_1 + X_2)/2$ is torzítatlan becslése a várható értéknek. Melyik a hatásosabb?

ϑ-paraméter

$T(X_1, \ldots, X_n)$ torzítatlan becslése $\psi(\vartheta)$-nak $\quad \psi$: függvény

ha $\quad E\big(T(X_1, \ldots, X_n)\big) = \psi(\vartheta)$.

$\hat{\mu}_1 = X_1$ torzítatlan becslése μ-nek

$\overset{def}{\Longleftrightarrow} \quad E(\hat{\mu}_1) = \mu$.

$X_1, X_2, \ldots, X_n \qquad \hat{\mu}_1 = X_1$

$E(X_1) = \mu$ ✓ hát persze.

$$\hat{\mu}_2 = \frac{X_1 + X_2}{2} \qquad E(\hat{\mu}_2) = E\left(\frac{X_1+X_2}{2}\right)$$

$$= \frac{1}{2}\big(E(X_1) + E(X_2)\big) = \mu. ✓$$

Tehát mindkét becslés torzítatlan becslés.

becslés akkor jó ha ~~ezt~~ van a becsült paraméterre

$E(\hat{\mu}_1) = E(\hat{\mu}_2) = \mu$ mindkettő torzítatlan

$\hat{\mu}_2$ jobb becslés, mint $\hat{\mu}_1$ ha

$$D^2(\hat{\mu}_2) < D^2(\hat{\mu}_1)$$

$$D^2(\hat{\mu}_1) = D^2(X_1) = \sigma^2$$

$$D^2(\hat{\mu}_2) = D^2\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}\left(D^2(X_1)+D^2(X_2)\right) = \frac{\sigma^2}{2}$$

$$\frac{\sigma^2}{2} < \sigma^2 \Rightarrow \hat{\mu}_2 \text{ hatásosabb.}$$

$\ulcorner$ $\hat{\mu}_2$ hatásosabb, mint $\hat{\mu}_1$, ha mindkettő torzítatlan és

$$D^2(\hat{\mu}_2) \leq D^2(\mu_1)$$

minden lehetséges háttéreloszlásra, és legalább egy háttéreloszlásnál $<$ van. $\lrcorner$

Várható érték lineáris becslései:

$$\hat{\mu} = \sum_{i=1}^{n} c_i \cdot X_i \qquad c_1, c_2, \dots, c_n \text{ konstansok}$$

Tétel: Ezek közül a legjobb becslés: $\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$.

Biz.: $E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$ torzítatlan.

$$D^2(\overline{X}_n) = \frac{1}{n^2} \cdot D^2\left(\sum X_i\right) \underset{\text{fgtl.}}{=} \frac{1}{n^2} \cdot n \sigma^2 = \frac{\sigma^2}{n}.$$

$\hat{\mu} = \sum_{i=1}^{n} c_i X_i$ torzítatlan, ha

$$E(\hat{\mu}) = \mu$$

$$E(\hat{\mu}) = E\left(\sum c_i X_i\right) = \sum c_i \cdot \mu = \mu \cdot \sum c_i$$

$$\Leftrightarrow \boxed{\sum c_i = 1.}$$

$$D^2(\hat{\mu}) = D^2\left(\sum c_i X_i\right) = \sum_{i=1}^{n} c_i^2 D^2(X_i)$$

$$= \sigma^2 \sum_{i=1}^{n} c_i^2$$

Azt kell belátni, hogy

$$\frac{b^2}{n} \leq b^2 \cdot \sum_{i=1}^{n} c_i^2$$

ha $\sum c_i = 1$.

$$\Leftrightarrow \quad \frac{1}{n} \leq \sum_{i=1}^{n} c_i^2 \quad \text{és} \quad \text{„}=\text{"} \Leftrightarrow c_i = \frac{1}{n}$$

$$\left[\begin{array}{l} a+b=1 \\ a^2+b^2 \to \min \end{array}\right.$$

$$1 = (a+b)^2 = a^2+b^2+2ab \overset{a,b \geq 0}{\geq} a^2+b^2$$

$$\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2} = \frac{1}{2} \Leftrightarrow a^2+b^2 \geq \frac{1}{2}$$

négyzetes

számtani

harmonikus közép $\frac{\frac{1}{a}+\frac{1}{b}}{2}$

harmonikus sor: $1+\frac{1}{2}+\frac{1}{3}+\ldots$

Általános n-re:

$$\sqrt{\frac{c_1^2+\dots+c_n^2}{n}} \geq \frac{c_1+\dots+c_n}{n} = \frac{1}{n}$$

$$c_1^2+\dots+c_n^2 \geq \frac{1}{n}.$$

és $\underset{\shortparallel}{=}$ $\iff$ $c_1=\dots=c_n=\frac{1}{n}.$

vagy Cauchy–Schwarz:

$$\boxed{1 = \sum c_i \cdot 1 \leq \sqrt{\sum c_i^2}\sqrt{\sum 1^2} = \sqrt{\sum c_i^2}\,\sqrt{n}}$$

$X \sim \text{Egyenletes}(0, \theta)$ $E(X) = \int_0^\theta x \cdot \frac{1}{\theta}\, dx$

$= \frac{\theta}{2}$.

$\boxed{\begin{array}{l} \text{Ált: } \frac{b-a}{2} \\ X \sim \text{Egy}(a, b) \end{array}}$

3. Legyenek $X_1, \ldots, X_n$ független, Egyenletes$(0, \theta)$ eloszlású véletlen változók, ahol $\theta > 0$. Határozzuk meg $X_{n,n} = \max\{X_1, \ldots, X_n\}$ eloszlás-, és sűrűségfüggvényét, várható értékét és szórását!

Mutassuk meg, hogy

$$T_1(\mathbf{X}) = \frac{n+1}{2n} X_{n,n}$$

$\mathbf{X} = (X_1, \ldots, X_n)$

torzítatlan becslése $(\theta/2)$-nek, ami hatásosabb, mint a $\overline{X}_n$ mintaátlag! Igazoljuk, hogy T_1 gyengén konzisztens!

$$P(X_{n,n} \leq x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ ? & \theta \geq x \geq 0 \\ 1 & \text{ha } x \geq \theta \end{cases}$$

$x \in [0, \theta]$.

$P(X_{n,n} \leq x) = P(\max(X_1, \ldots, X_n) \leq x)$

$= P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x)$

és, és, és

$\overset{\text{fgtl}}{=} P(X_1 \leq x) \cdot P(X_2 \leq x) \cdots P(X_n \leq x)$

$= \left(\frac{x}{\theta}\right)^n$

$P(X_{n,n} \leq x) = \left(\frac{x}{\theta}\right)^n$

$f(x) = n \cdot x^{n-1} \cdot \theta^{-n}$, ha $x \in (0, \theta)$

Cadastrar? $\frac{n}{n+1}\theta \overset{?}{=} E(X_{n,n}) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{\theta} x\cdot n x^{n-1}\theta^{-n}\,dx$

$$= \frac{n}{\theta^n}\cdot\int_0^\theta x^n\,dx = \frac{n}{\theta^n}\cdot\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{\theta^n}\cdot\frac{\theta^{n+1}}{n+1} = \frac{n\theta}{n+1}$$

$$D^2(X_{n,n}) = E(X_{n,n}^2) - \left(E(X_{n,n})\right)^2$$

$$E(X_{n,n}^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^\theta x^2\cdot n x^{n-1}\cdot\theta^{-n}\,dx$$

$$= \frac{n}{\theta^n}\cdot\int_0^\theta x^{n+1}\,dx = \frac{n}{\theta^n}\cdot\frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2$$

$$D^2(X_{n,n}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2 = \theta^2\cdot\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)$$

$$= \theta^2\cdot\frac{n\cdot(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} = \theta^2\cdot\frac{n}{(n+2)(n+1)^2}$$

$$n(n+1)^2 - n^2(n+2) = n\left[(n^2+2n+1) - n^2 - 2n\right] = n$$

$$T(X) = \frac{n+1}{2n} \cdot X_{n,n}$$ torzítatlan becslése $\frac{\theta}{2}$-nek.

azaz: $$E(T(X)) = \frac{\theta}{2}$$

$$E\left(\frac{n+1}{2n} X_{n,n}\right) = \frac{n+1}{2n} E(X_{n,n}) = \frac{n+1}{2n} \theta \cdot \frac{n}{n+1} = \frac{\theta}{2}$$

azaz $T(X)$ tényleg torzítatlan becslése $\frac{\theta}{2}$-nek (a várható érték)

$$D^2(T(X)) = D^2\left(\frac{n+1}{2n} X_{n,n}\right)$$

$$= \left(\frac{n+1}{2n}\right)^2 D^2(X_{n,n}) = \frac{\cancel{(n+1)^2}}{4n^{\cancel{2}}} \cdot \theta^2 \cdot \frac{\cancel{n}}{\cancel{(n+1)^2}(n+2)}$$

$$= \theta^2 \cdot \frac{1}{4 \cdot n \cdot (n+2)}$$

Tehát $$D^2(T(X)) = \frac{\theta^2}{4n(n+2)}$$

$$D^2(\overline{X}_n) = \frac{1}{n} D^2(X) = \frac{1}{n} \cdot \frac{\vartheta^2}{12}$$

$\overline{X}_n$ torzítatlan becslése $\frac{\vartheta}{2}$-nek (várható értéknek)

$$D^2(\overline{X}_n) > D^2(T(X)) = \frac{\vartheta^2}{4n(n+2)}$$

$$\shortparallel$$

$$\frac{\vartheta^2}{12n}$$

$\Downarrow$

$T(X)$ hatásosabb becslés, mint a mintaátlag.

$$\frac{1}{12n} > \frac{1}{4n(n+2)}$$

$$n+2 > 3 \qquad \underline{\underline{n \geq 2}}$$

$X \sim \text{Egyenletes}(a, b)$

$$E(X) = \frac{b+a}{2}$$

$$D^2(X) = \frac{(b-a)^2}{12}$$

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3}(b^2+ab+a^2)$$

$$E(X^2) - E(X)^2 = \frac{1}{3}(b^2+ab+a^2) - \frac{1}{4}(b^2+2ab+a^2)$$

$$= \frac{1}{12}(4b^2+4ab+4a^2-3b^2-6ab-3a^2)$$

$$= \frac{1}{12}(b^2-2ab+a^2) = \frac{(b-a)^2}{12}$$

véletlen $\quad$ n mintaelemből

$$T(\cancel{X}_n) \text{ gyengén konzisztens becslése } \psi(\theta)\text{-nak} \quad \text{def. (ismertem)}$$

ha $$P\left(\left|T(X_n) - \psi(\theta)\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

$$T(X) = \frac{n}{2(n+1)} X_{n,n} \quad \text{konzisztens becslése } \frac{\theta}{2}\text{-nek.}$$

Legyen $\varepsilon > 0$ tetsz. rögzített.

$$P\left(\left|\frac{n}{2(n+1)} X_{n,n} - \frac{\theta}{2}\right| > \varepsilon\right)$$

$$= P\left(\left|X_{n,n} - \frac{n+1}{n}\theta\right| > \frac{2\varepsilon(n+1)}{n}\right)$$

$$= P\left(X_{n,n} > \frac{n+1}{n}\theta + \frac{2\varepsilon(n+1)}{n} \text{ vagy } X_{n,n} < \frac{n+1}{n}\theta - \frac{2\varepsilon(n+1)}{n}\right)$$

$$= \underbrace{\underbrace{P\left(X_{n,n} > \frac{n+1}{n}\theta + \frac{2\varepsilon(n+1)}{n}\right)}_{\to 0} + P\left(X_{n,n} < \frac{n+1}{n}\theta - \frac{2\varepsilon(n+1)}{n}\right)}_{= 0}$$

$$= P\left(X_{n,n} < \frac{n+1}{n}\theta - \frac{2(n+1)\varepsilon}{n}\right)$$

lie $\quad \frac{n+1}{n}\vartheta - \frac{2(n+1)\varepsilon}{n} < \vartheta$

$$\frac{1}{\cancel{n}}\vartheta < \frac{2(n+1)}{\cancel{n}}\varepsilon$$

$$\frac{\vartheta}{2(n+1)} < \varepsilon \iff \boxed{n > \frac{\vartheta}{2\varepsilon} - 1}$$

$$= \left(\frac{\frac{n+1}{n}\vartheta - \frac{2(n+1)}{n}\varepsilon}{\vartheta} \right)^n =$$

$$= \Big(\underbrace{1 + \frac{1}{n} - \frac{2(n+1)}{\vartheta n}\varepsilon}_{} \Big)^n \longrightarrow 0 .$$

$$\Big\downarrow n \to \infty$$

$$1 - \frac{2\varepsilon}{\vartheta} < 1$$

$\lightning$

4. Egy dobozban két pénzérme van. Az egyik szabályos, a másik cinkelt, 0,7 valószínűséggel ad fejet. Az egyik érmével 4-szer dobunk. Az eredmény 3 fej és 1 írás. Mennyi a valószínűsége, hogy a szabályos érmével dobtunk?

5. Egy almáskertben véletlenszerűen, egymástól függetlenül találhatók fertőzött fák. Tíz egyforma nagy, egyenként három sorból álló ültetvényben rendre 0, 3, 0, 1, 0, 0, 2, 1, 1, 2 beteg fát találtak.

(a) Adjuk meg az empirikus eloszlásfüggvényt, a mintaátlagot és az empirikus szórásnégyzetet!

(b) Tegyük fel, hogy a beteg fák száma Poisson-eloszlást követ. Adjunk maximum likelihood becslést és momentumbecslést az egy sorban található fák számának várható értékére!

6. Egy adatszerverre a lekérdezések exponenciális időközönként érkeznek, ahol ismeretlen paraméterrel. Az időközökre percben mérve a következő adatokat kaptuk: 1,94, 0,33, 2,51, 5,27, 1,73, és 0,61. Adjunk becslést a paraméterre a maximum likelihood és a momentumbecslés alkalmazásával!

7. Legyenek $X_1, \ldots, X_n$ független véletlen változók, $f(x) = \frac{2x}{3\theta^2}$, $\theta \leq x \leq 2\theta$, sűrűségfüggvénnyel. Adjunk becslést θ-ra momentum módszerrel és ML módszerrel is!