10. STATISTIKA

10.1 TOČKASTE PROCJENE

Ω populacija $\longrightarrow$ obilježje X

$\downarrow$ $\longrightarrow$ distribucija $F \to$ nepoznata

uzeti uzorak $\longrightarrow$ $m, \sigma^2 \to$ nepoznato

X sluč. varijabla, populacija

$\leadsto F, f, m, \sigma^2$

$\leadsto F$ može ovisiti o nekim parametrima

$\vartheta_1, \dots, \vartheta_k$

def. Za $X_1, \dots, X_n$ slučajne varijable kažemo da su nezavisne KOPIJE slučajnih varijabli X ako vrijedi:

1) $X_1, \dots, X_n$ nezavisne

2) $X_i \sim F$

$(X_1, \dots, X_n)$ UZORAK

$(x_1, \dots, x_n)$ REALIZACIJA

⇝ pretpostavimo da razdioba F ovisi samo o ϑ:

$$F(x) = F_{\vartheta}(x)$$
$$f(x) = f_{\vartheta}(x)$$

⇝ ϑ procjeniti pomoću $x_1, \ldots, x_n$

$\hat{u} = g(x_1, \ldots, x_n)$ procjena za ϑ

$\Theta = g(X_1, \ldots, X_n)$ STATISTIKA

$g(x_1, \ldots, x_n)$ ⟶ statistika

Θ ...... procjenitelj parametra ϑ

$\hat{u}$ ...... procjena parametra ϑ

Pitanja: 1) Koje statistike su dobre za m i σ^2?

2) Što su to dobre statistike?

⇝ statistika za procjenu očekivanja $a = E(X)$:

$\overline{X} = \frac{1}{n} \cdot (X_1 + \ldots + X_n)$ SREDINA UZORKA

↳ sluč. varijable, statistika

$$E[\overline{X}] = \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{\cancel{n}} \cdot \cancel{n} \cdot a = a$$

$$D(\overline{X}) = \frac{1}{n^2}\left[D(X_1) + \ldots + D(X_n)\right] = \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \sigma^2 = \frac{\sigma^2}{n}$$

⇝ DOBRE STATISTIKE:

1) Θ = NEPRISTRANA STATISTIKA (procjenitelj) za ϑ ako je

$$\boxed{E(\Theta) = \vartheta}$$!

⇝ $\bar{X}$ nepristrani procjenitelj za očekivanje

2) Ako niz statistika $\Theta_n = \Theta_n(X_1, \dots, X_n)$ konvergira po vjerojatnosti prema parametru ϑ tada Θ_n zovemo **VALJANOM STATISTIKOM**

$$\forall \varepsilon > 0 \quad P(|\Theta_n - \vartheta| \geq \varepsilon) \longrightarrow 0 \qquad n \to \infty$$

3) Θ_1 i Θ_2 dvije nepristrane statistike za ϑ

⇝ Θ_1 bolja od Θ_2 ako $D(\Theta_1) < D(\Theta_2)$

$\boxed{\text{TM}}$ NEPRISTRANA STATISTIKA je VALJANA ako joj disperzija teži k 0.

DOKAZ:

$$E[\Theta_n] = \vartheta \quad D(\Theta_n) \to 0, \quad n \to \infty \quad \text{želimo: } \Theta_n \xrightarrow{P} \vartheta$$

$$0 \leq P(|\Theta_n - \vartheta| \geq \varepsilon) \underset{\text{Čebiš.}}{\leq} \frac{D(\Theta_n)}{\varepsilon^2} \longrightarrow 0 \quad \Rightarrow$$

→ $\bar{X}$ valjana za a zbog:

$$D(\bar{X}) = \frac{\sigma^2}{n} \longrightarrow 0 \quad ✓$$

→ STATISTIKA ZA PROCJENU DISPERZIJE:

→ a poznat: $D^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2$

→ a nepoznat: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$

→ D^2 i S^2 su nepristrani i valjani procjenitelji za σ^2

$E(D^2) = \sigma^2$

$D(D^2) = \frac{1}{n}(\mu_4 - \sigma^4) \longrightarrow 0$ ✓

$\theta = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \Rightarrow E(\theta) = \frac{n-1}{n} \sigma^2$

$E\left[\frac{n}{n-1} \theta\right] = \sigma^2$

$E(S^2) = \sigma^2$

Primjer 10.3.

→ 6 rezultata: 3540, 3582, 3559, 3572, 3564, 3584 m

a) $a = 3560$ b) a nepoznat

a) $D^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2 = \frac{1}{6} \sum_{i=1}^{6} (X_i - 3560)^2 = 232,17 \text{ m}^2$

b) $S^2 = \frac{1}{6-1} \sum_{i=1}^{6} (X_i - \overline{X})^2 = \frac{1}{5}\left[(3540 - 3561.17)^2 + \dots + ()^2\right] = 276,97 \text{ m}$

→ podaci mogu biti podijeljeni unaprijed u razrede i dani u tablici:

VRIJEDNOSTI	x_1	x_2	...	x_r
FREKVENCIJE	n_1	n_2	...	n_r

$n = n_1 + \dots + n_r$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r} x_i n_i \qquad s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{r} n_i \cdot x_i^2 - n \cdot \bar{x}^2\right)$$

Primjer 10.4.

$\bar{x},\ s^2 = ?$

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

* PODSJETNIK:

$E(x-c) = E(x) - c$

$D(x-c) = D(x)$

$c = 2620$

$$\bar{x} = \frac{1}{20}\sum_{i=1}^{5}(x_i - 2620)\,n_i + 2620 \qquad \boxed{\bar{x} = 2621}$$

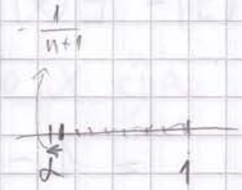
$$s^2 = \frac{1}{19}\left(\sum_{i=1}^{5} n_i \cdot (x_i - 2620)^2 - n\bar{x}^2\right)$$

$$\boxed{s^2 = 967{,}4}$$

Zl-12-4.) Iz intervala $[\alpha, 1]$ odabrano je na sreću n brojeva

$X_1, \dots, X_n \in [\alpha, 1]$. Da bismo procjenili α odabrali smo

statistiku $Y = \text{minimum}\{X_1, \dots, X_n\} - \frac{1}{n+1}$.



a) provjerite nepristranost statistike

$$\boxed{E(Y) = \alpha}$$

→ AKO JE OČEKIVANJE JEDNAKO PRAVOJ VRIJEDNOSTI !

$$Y = \underbrace{\min\{X_1, \dots, X_n\}}_{X} - \frac{1}{n+1}$$

TRIK

$$F_X(x) = P(X < x) = P(\min\{X_1, \dots, X_n\} < x) =$$

MINIMUM UVIJEK >
MAX UVIJEK <

$$= 1 - P(\min\{X_1, \dots, X_n\} > x) =$$

$$= 1 - P(X_1 > x, X_2 > x, \dots, X_n > x) =$$

$$= 1 - P(X_1 > x)^n =$$

$$P(X > a) = \int_a^\infty f(x)\,dx$$

$$= 1 - \left(\int_x^1 \frac{1}{1-\alpha} dx \right)^n$$ JEDNOLIKA RAZDIOBA (pretp.) !

$$= 1 - \left(\frac{1-x}{1-\alpha} \right)^n$$

$$f(x) = F_X'(x) = -n\left(\frac{1-x}{1-\alpha}\right)^{n-1} \cdot \frac{-1}{1-\alpha} = \frac{n}{(1-\alpha)^n} \cdot (1-x)^{n-1}$$

$$E(X) = \int_\alpha^1 x \cdot \frac{n}{(1-\alpha)^n} \underbrace{(1-x)}_{\text{supstit.} = t}{}^{n-1} dx = \dots = \frac{n}{(1-\alpha)^n} \left[\frac{t^n}{n} - \frac{t^{n+1}}{n+1} \right] \Bigg|_0^{1-\alpha} =$$

OKRENI →

$$= \cdots = 1 - \frac{n}{n+1}(1-\alpha)$$

$$E(y) = E(x) - \frac{1}{n+1} = 1 - \frac{n}{n+1}(1-\alpha) - \frac{1}{n+1} =$$

$$= \underbrace{1 - \frac{n}{n+1}}_{\frac{1}{n+1}} + \frac{\alpha \cdot n}{n+1} - \frac{1}{n+1} = \frac{n}{n+1} \cdot \alpha$$

$\Rightarrow$ Ova statistika NIJE NEPRISTRANA. $(E(y) \neq \alpha)$

b) S kojim faktorom treba pomnožiti da bude nepristrana?

$\rightarrow$ treba pomnožiti s $\frac{n+1}{n}$

NIKAD NIJE BILO „PROVJERITI DA JE VALJANA"

$\rightsquigarrow$ MOŽDA PROGURA OVE GODINE DA BUDE

$\boxed{TM}$ $\lim_{n \to \infty} D(x) = 0$

10.2. KRITERIJ NAJVEĆE IZGLEDNOSTI (KNI)

(MILF) — MAXIMUM INTENSITY LIVELIHOOD FUNCTION

def.

Neka je $x_1, \dots, x_n$ realizacija uzorka iz neke populacije X čija funkcija gustoće f ovisi o parametru ϑ.

FUNKCIJA IZGLEDNOSTI se definira kao

$$\boxed{L(\vartheta, x_1, \dots, x_n) = f(\vartheta, x_1) \cdot \dots \cdot f(\vartheta, x_n)}$$

⟹ za procjenu parametra ϑ se uzima vrijednost za koju L poprima **GLOBALNI MAKSIMUM** → derivaciju izjednačiti s nulom!

A ŠTO DRUGO KAD JE MILF U PITANJU

Z1-11-2.) Uzorak $x_1, \dots, x_n$ je izvučen iz populacije s funkcijom gustoće $f(x) = \lambda x^{\lambda-1}$. Pomoću KNI odredite λ.

TRIK!

ZBOG KOMPLIKACIJE SA DERIVACIJOM n UMNOŽA. PRELAZIMO NA ln!

$$L(\lambda, x_1, \dots, x_n) = \lambda \cdot x_1^{\lambda-1} \cdot \lambda \cdot x_2^{\lambda-1} \cdots \lambda \cdot x_n^{\lambda-1}$$

$$= \lambda^n (x_1 \cdots x_n)^{\lambda-1}$$

$$\ln L = n \ln\lambda + (\lambda-1)\ln(x_1 \cdots x_n) \quad \Big/ \frac{\partial}{\partial\lambda}$$

$$\frac{n}{\lambda} + \ln(x_1 \cdots x_n) = 0$$

$$\lambda = -\frac{n}{\ln(x_1 \cdots x_n)} = -\frac{n}{\ln \prod_{i=1}^{n} x_i}$$

mabsimum likelihood intensity function

MILF

21-09-1.)

$$f(x)=\frac{\lambda}{2\sqrt{x}}e^{-\lambda\sqrt{x}}$$, pomoću KNI odredi λ.

$$L(\lambda, x_1,\dots,x_n)=\frac{\lambda}{2\sqrt{x_1}}e^{-\lambda\sqrt{x_1}}\cdot\dots\cdot\frac{\lambda}{2\sqrt{x_n}}e^{-\lambda\sqrt{x_n}}=$$

$$=\frac{\lambda^n}{2^n\sqrt{x_1\cdots x_n}}e^{-\lambda(\sqrt{x_1}+\sqrt{x_2}\cdots+\sqrt{x_n})}$$

$\ln(e)=1$

$$\ln L = n\ln\lambda - \underbrace{\ln 2^n\sqrt{x_1\cdots x_n}}_{\text{konstanta}} - \lambda\sum_{i=1}^{n}\sqrt{x_i} \quad \Big|\frac{\partial}{\partial\lambda}$$

$$\frac{n}{\lambda}-\sum_{i=1}^{n}\sqrt{x_i}=0$$

$$\Rightarrow \boxed{\lambda=\frac{n}{\sum_{i=1}^{n}\sqrt{x_i}}}$$

21-10-1.) Bacamo neispravan novčić. Bacili ga 5 puta i pismo se pojavilo 2 puta, zatim se u 8 puta pojavilo 3 puta. Pomoću KNI (MILF) odredite procjenu za pojavljivanja (vjerojatnost) pisma.

⇝ bez KNI, zdrava logika → $\frac{5}{13}$

$$L(\lambda, x_1,\dots,x_n)=P(X=x_1)\cdots P(X=x_n)$$ → DISKRETAN SLUČAJ

$\binom{n}{k}p^k(1-p)^{n-k}$ ← BINOMNA

BINOMNA RAZDIOBA

$X_1 \sim B(5,p)$, $X_2\sim B(8,p)$

$$L(p)=\binom{5}{2}p^2(1-p)^3\cdot\binom{8}{3}p^3(1-p)^5=\underbrace{\binom{5}{2}\binom{8}{3}}_{K}p^5(1-p)^8$$

$$\ln L = \ln \overset{\text{konstante}}{K} + 5\ln p + 8\ln(1-p) \;\Big/\; \frac{d}{dp} \longrightarrow$$ OBIČNA DERIVAC. SAMO JEDNA VARJABLA!

$$\frac{5}{p} + \frac{-8}{1-p} = 0$$

$$p = \frac{5}{13}$$

HINT:

„NIKAD NIJE BILO S POISSONOVOM RAZDIOBOM, MOGLO BI DOĆI OVE GODINE!"

→ DISKRETNE VAR.: MNOŽIMO VJEROJATNOSTI

KONTINUIRANE VAR.: MNOŽIMO GUSTOĆE

→ KAD-TAD NA ISPITU ĆE DOĆI 10.6. (str. 13) (BILO NA ROKOVIMA...)

→ rješeni boldani zadaci u knjizi

→ 6., 7., 8. zad ne treba, ostalo DA!