Tačkaste ocene parametara raspodele

- Na osnovu uzorka treba da se odredi kakva je raspodela obeležja na populaciji
- Ako je tip raspodele poznat, treba da se odrede parametri raspodele
- Pošto je realizovana vrednost statistike realan broj, to ocene parametara nazivamo **tačkaste ocene**.
- Neka je $X_1, X_2, ..., X_n$ prost slučajan uzorak obima n za posmatrano obeležje X. Statistika $Y = f(X_1, X_2, ..., X_n)$ je slučajna promenljiva koja implicitno zavisi od parametara u raspodeli obeležja X.
- Ako se sa Y ocenjuje parametar θ, tada se statistika Y naziva ocena parametra θ i označava sa $\hat{\theta}$.

Nepristrasnost

- Statistike koje koristimo za ocenu parametara moraju da imaju određene osobine.
- Kvalitet tačkastih ocena se izražava pomoću matematičkog očekivanja i disperzije.
- **Definicija. Neka je $X_1, X_2, \ldots, X_n$ prost slučajan uzorak obima n za obeležje X i neka je θ nepoznati parametar u raspodeli obeležja X. Statistika $Y = f(X_1, \ldots, X_n)$ je *nepristrasna ocena* parametra θ ako je:**

$$E\,(f\,(X_1, \ldots, X_n)) = \theta$$

- Koristi se i termini: ***centrirana*** ili ***nepomerena*** ocena.

Nepristrasnost, nastavak

- Ocena $\hat{\theta}$ parametra θ koja nema osobinu nepristrasnosti se zove ***pristrasna ocena***. Veličina njene pristrasnosti se meri razlikom

$$E(\hat{\theta}) - \theta$$

- Ako nepristrasnost statistike $Y = f(X_1, ..., X_n)$ kojom ocenjujemo neki parametar θ, važi u slučaju kada obim uzorka teži beskonačnosti, tj.

$$\lim_{n\to\infty} E(f(X_1,...,X_n)) = \theta$$

kažemo da statistika Y predstavlja ***asimptotski nepristrasnu*** ocenu posmatranog parametra θ.

Stabilnost

- Za realizovani uzorak $x_1, \ldots, x_n$ računamo realizaciju statistike Y, tj. $y = f(x_1, \ldots, x_n)$. Y je slučajna promenljiva, pa će realizovana vrednost odstupati od $E(Y)$. Poželjno je da se to odstupanje smanjuje ukoliko povećavamo obim uzorka. Navedeno svojstvo je **stabilnost** (**postojanost**).
- **Def. Neka je $X_1, \ldots, X_n$ prost slučajan uzorak obima n za obeležje X i neka je θ nepoznati parametar u raspodeli obeležja X. Statistika $Y = f(X_1, \ldots, X_n)$ je *stabilna ocena* parametra θ ako je nepristrasna i ako**

$$Y \xrightarrow{P} \theta, \qquad n \to \infty$$

- Stabilnost ocene označava da će se sa povećanjem obima uzorka razlika između realizovane vrednosti statistike Y i stvarne vrednosti parametra θ smanjivati.

Efikasnost

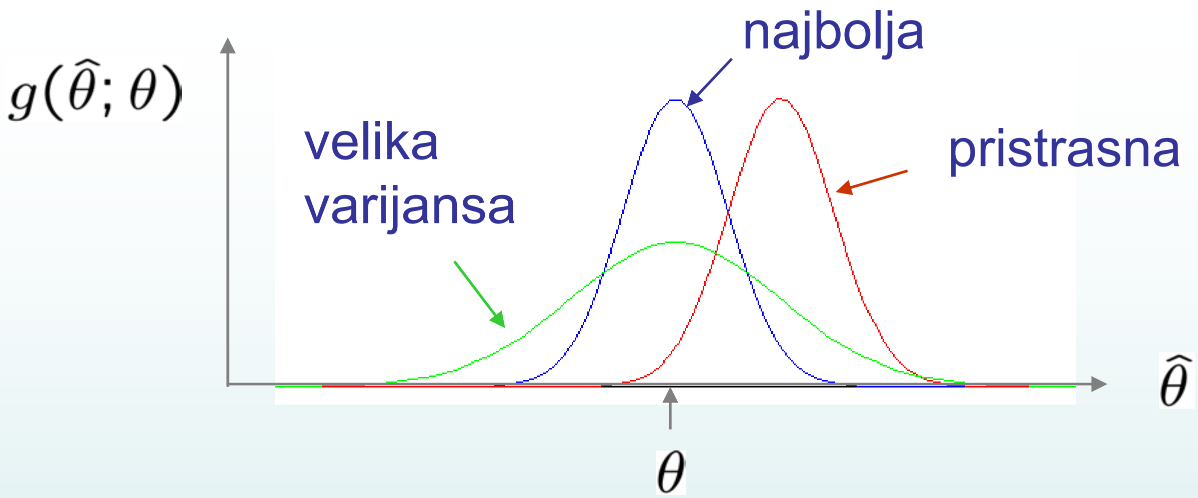
- **Def. Neka je $X_1, ..., X_n$ prost slučajan uzorak obima n za obeležje X i neka je θ nepoznati parametar u raspodeli obeležja X. Neka su statistike $Y = f(X_1, ..., X_n)$ i $Z = g(X_1, ..., X_n)$ nepristrasne ocene parametra θ. Statistika Y je *efikasnija ocena* parametra θ od statistike Z ako je**

$$D(Y) \leq D(Z)$$

- Statistika Y je ***najefikasnija ocena*** parametra θ, ako njena disperzija nije veća od disperzije bilo koje druge nepristrasne ocene istog parametra

$$D(Y) \leq D(Z) \qquad E(Y) = E(Z) = \theta$$

Ocena parametra



- Želimo malu (ili da je nula) pristrasnost $E(\hat{\theta})-\theta$
  - Očekivana vrednost treba da teži tačnoj vrednosti
- Želimo malu varijansu
  - Mala pristrasnost i varijansa su u konfliktu

Nejednakost Rao-Kramera

- Daje donju granicu disperzije nepristrasnih ocena jednog parametra. Neka je $X_1, \ldots, X_n$ prost slučajan uzorak obima n za obeležje X u čijoj raspodeli figuriše nepoznati parametar θ. Neka je gustina raspodele obeležja X

neprekidno obeležje

$$g(x;\theta),\ x\in R,\ \theta\in\Theta$$

- Odnosno, neka je zakon raspodele obeležja X

diskretno obeležje

$$p(x;\theta),\ x\in R,\ \theta\in\Theta$$

Θ označava skup dopustivih vrednosti parametra, tzv. ***parametarski prostor***.

- ***Funkcija verodostojnosti*** je u neprekidnom slučaju:

$$L(X;\theta) = \prod_{j=1}^{n} g(X_j;\theta)$$

Funkcija verodostojnosti

- Funkcija verodostojnosti u diskretnom slučaju je

$$L(X;\theta)=\prod_{j=1}^{n} p(X_j;\theta)$$

- Pod ***uslovima regularnosti***

(1) skup $\{X: L(X;\theta)>0\}$ ne zavisi od θ

(2) L je dvaput diferencijabilna po θ

važi nejednakost Rao-Kramera

$$D(\hat{\theta}) \geq -\frac{1}{E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)}=\frac{1}{E\left(\frac{\partial \ln L}{\partial \theta}\right)^2}$$

$\hat{\theta}$ je nepristrasna ocena parametra θ

Primer nejednakosti Rao-Kramera

- Oceniti matematičko očekivanje uzorka koji ima normalnu raspodelu.

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(X-m)^2}{2\sigma^2}}$$

- Funkcija verodostojnosti dobijena na osnovu prostog slučajnog uzorka X_1, ..., X_n obima n je

$$L(X;m) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(X_j-m)^2}{2\sigma^2}}$$

$$\frac{\partial^2 \ln L}{\partial m^2} = -\frac{n}{\sigma^2} \qquad D(\hat{m}) \geq \frac{\sigma^2}{n}$$

Opravdano korišćenje statistike $\overline{X}_n$ za ocenjivanje parametra m

Asimptotski efikasna ocena

- Ako disperzija statistike $Y = f(X_1, ..., X_n)$ koja je nepristrasna ocena nekog parametra θ, konvergira ka donjoj granici disperzije iz nejednakosti Rao-Kramera, kada obim uzorka teži beskonačnosti, kažemo da statistika Y predstavlja ***asimptotski efikasnu ocenu*** parametra θ.

Metod momenata

- Neka je X_1, X_2, ..., X_n prost slučajan uzorak obima n za obeležje X. Pretpostavimo da postoje konačni momenti reda $1, 2, ..., s$: $E(X^j)$, ($E(|X^j|<\infty$) i da u raspodeli obeležja X ima s nepoznatih parametara θ_1, θ_2, ..., θ_s.
- Izjednačavanjem teorijskih momenata reda j i uzoračkih momenata

$$M_j = \frac{1}{n}\sum_{k=1}^{n} x_k^j$$

istog reda dobijamo sistem jednačina:

$$E(X^j)=M_j, \quad j=1, 2, ..., s \qquad (*)$$

Rešavanjem prethodnog sistema jednačina se dobijaju ocene za θ_1, θ_2, ..., θ_s.

Metoda maksimalne verodostojnosti

- Ocene koje se dobijaju po metodi maksimalne verodostojnosti su često asimptotski nepristrasne i asimptotski efikasne ocene.
- Neka je θ nepoznati parametar u raspodeli obeležja X i neka je Θ skup mogućih vrednosti parametra θ, tzv. parametarski prostor. Neka je $X_1, \ldots, X_n$ prost slučajan uzorak za obeležje X i neka $g(x;\theta)$ označava gustinu raspodele obeležja X. ***Funkcija verodostojnosti*** je funkcija

$$L(X;\theta) = g(X_1;\theta) \cdot \ldots \cdot g(X_n;\theta)$$

- Parametar θ ćemo ocenjivati onom statistikom za koju funkcija verodostojnosti postiže supremum.

Metoda maksimalne verodostojnosti

- Ako je funkcija verodostojnosti diferencijabilna, najpre se određuju (ako postoje) one vrednosti parametra koja su rešenja jednačine

$$\frac{\mathrm{d}}{\mathrm{d}\theta} L(X;\theta) = 0$$

ili

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L(X;\theta)) = 0$$

- Među svim tim vrednostima nalazimo (ako postoje) one vrednosti za koje funkcija L, odnosno $\ln L$ postiže svoj supremum.

FV u diskretnom slučaju

- U diskretnom slučaju funkciju verodostojnosti dobijamo kao proizvod verovatnoća postizanja pojedinih vrednosti iz uzorka:

$$L(X;\theta) = p(X_1;\theta) \cdot \ldots \cdot p(X_n;\theta)$$

  gde je $p(t;\theta) = P(X = t)$

- Zatim se određuje ekstremna vrednost te funkcije ili njenog logaritma.

Višedimenzionalni parametar θ

- U slučaju višedimenzionalnog parametra $\theta=(\theta_1,\ldots,\theta_s)$ i diferencijabilne funkcije verodostojnosti najpre se odrede (ako postoje) one vrednosti parametara iz parametarskog prostora za koje važi:

$$\frac{\partial}{\partial\theta_j} L(X;\theta) = 0$$

ili $\qquad j = 1,\ldots,s$

$$\frac{\partial}{\partial\theta_j} \ln L(X;\theta) = 0$$

- Najčešće su ovi sistemi jednačina nelinearni, pa se rešavaju koristeći numeričke metode analize. Kada se odrede rešenja sistema, uzimaju se ona rešenja za koja se dobija ekstremna vrednost funkcije verodostojnosti.

Primer metoda maks. verodostojnosti

- Neka obeležje X ima Puasonovu $\mathcal{P}(\lambda)$ raspodelu, pri čemu je parametar λ nepoznat. Metodom maksimalne verodostojnosti odrediti statistiku kojom treba ocenjivati parametar λ na osnovu uzorka obima n.
- Funkcija verodostojnosti je

$$L(X;\lambda)=\prod_{k=1}^{n}\frac{\lambda^{X_k}}{X_k!}\mathrm{e}^{-\lambda}=\frac{\lambda^{\sum\limits_{k=1}^{n}X_k}}{\prod\limits_{k=1}^{n}X_k!}\mathrm{e}^{-n\lambda}$$

a njen logaritam

$$\ln L(X;\lambda)=\left(\sum_{k=1}^{n}X_k\right)\ln\lambda-n\lambda-\prod_{k=1}^{n}\ln X_k!$$

Primer metoda maks. verodostojnosti

- Iz uslova $\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln(L(X;\lambda)) = 0$

  dobijamo da je ocena nepoznatog parametra statistika

$$\hat{\lambda} = \overline{X}_n$$

- U toj tački se postiže maksimum funkcije verodostojnosti. Očekivanje Puasonove raspodele je

$$E(X) = \lambda$$

  pa dobijeni rezultat potvrđuje da je upravo uzoračka sredina ocena za očekivanu vrednost obeležja.

Primer

- Neka obeležje X ima $\mathcal{N}(m,\sigma^2)$ raspodelu, pri čemu je parametar m poznat, a parametar σ^2 nije. Metodom maksimalne verodostojnosti (MMV) odrediti ocenu nepoznatog parametra σ^2 na osnovu uzorka obima n.
- Funkcija verodostojnosti je

$$L(X;\sigma^2)=\prod_{k=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(X_k-m)^2}{2\sigma^2}}=\frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(X_k-m)^2}$$

a njen logaritam

$$\ln L(X;\sigma^2)=-n\ln\sigma-n\ln\sqrt{2\pi}-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(X_k-m)^2$$

Primer MMV, nastavak

- Kako je σ^2 nepoznati parametar, za rešenje jednačine

$$\frac{\mathrm{d}}{\mathrm{d}(\sigma^2)} \ln L(X;\sigma^2) = 0$$

dobija se

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2$$

U toj tački je drugi izvod logaritma funkcije verodostojnosti negativan, pa ovo potvrđuje izbor uzoračke disperzije $\widetilde{S}_n^2$ za ocenu disperzije obeležja kad je očekivanje poznato.

Osobine ocena metodom momenata

- Ocene dobijene metodom momenata i metodom maksimalne verodostojnosti se u nekim slučajevima poklapaju,
- Ocene dobijene metodom momenata nisu, u opštem slučaju efikasne, u smislu prethodno datih definicija,
- Ako sistem (*) ima jedinstveno rešenje u obliku neprekidnih funkcija, tada ocena za nepoznati parametar konvergira u verovatnoći ka tom parametru.