Statistika

Tačkasto ocjenjivanje nepoznatog parametra

1. Neka je $(X_1, \dots, X_n)$ (prost slučajni) uzorak obilježja X s konačnim očekivanjem m i disperzijom σ^2.

(a) Dokazati da je $\overline{X}_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ nepristrasna (centrirana) i postojana ocjena za m.

(b) Ispitati nepristrasnost ocjene $\overline{S}_n^2 = \dfrac{1}{n}\sum\limits_{k=1}^{n}(X_k - \overline{X}_n)^2$ za disperziju σ^2.

(c) Odrediti konstantu c tako da $T_n = c\sum\limits_{k=1}^{n-1}(X_{k+1} - X_k)^2$ bude centrirana ocjena za σ^2.

◀

(a)

$$\mathsf{E}\,\overline{X}_n = \mathsf{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\,X_k = \frac{1}{n}\cdot nm = m.$$

Slijedi da je $\overline{X}_n$ centrirana ocjena parametra m.

$$\mathsf{D}\,\overline{X}_n = \mathsf{D}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathsf{D}\,X_k = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Postojanost slijedi na osnovu Hinčinovog zakona velikih brojeva.

Dokažimo po definiciji. Neka je $\varepsilon > 0$ proizvoljan broj. Tada, koristeći Čebišovljevu nejednakost, dobijamo:

$$\mathsf{P}\left\{|\overline{X}_n - m| > \varepsilon\right\} = \mathsf{P}\left\{|\overline{X}_n - \mathsf{E}\overline{X}_n| > \varepsilon\right\} \le \frac{\mathsf{D}\overline{X}_n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, \text{ pri } n \to \infty,$$

pa slijedi da $\overline{X}_n \xrightarrow{\mathsf{P}} m$, pri $n \to \infty$, odnosno $\overline{X}_n$ je postojana ocjena parametra m.

(b) Važi

$$\overline{S}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k^2 - 2X_k\,\overline{X}_n + \overline{X}_n^2)$$
$$= \frac{1}{n}\sum_{k=1}^{n}X_k^2 - 2\overline{X}_n \cdot \frac{1}{n}\sum_{k=1}^{n}X_k + \overline{X}_n^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}X_k^2 - 2\overline{X}_n^2 + \overline{X}_n^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}X_k^2 - \overline{X}_n^2,$$

pa je

$$\mathsf{E}\,\overline{S}_n^2 = \mathsf{E}\left(\frac{1}{n}\sum_{k=1}^{n}X_k^2 - \overline{X}_n^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}X_k^2 - ((\mathsf{E}\overline{X}_n)^2 + \mathsf{D}\overline{X}_n)$$
$$= \frac{1}{n}\cdot n \cdot \mathsf{E}X^2 - (\mathsf{E}X)^2 - \frac{\sigma^2}{n}$$
$$= \sigma^2 - \frac{\sigma^2}{n}$$
$$= \sigma^2\left(1 - \frac{1}{n}\right) < \sigma^2 \text{ za } n > 1.$$

Slijedi da $\overline{S}_n^2$ nije centrirana ocjena za σ^2, ali jeste asimptotski centrirana, jer važi

$$\mathsf{E}\,\overline{S}_n^2 = \sigma^2\left(1 - \frac{1}{n}\right) \to \sigma^2, \text{ pri } n \to \infty.$$

Primijetimo da iz $\mathsf{E}\overline{S}_n^2 = \sigma^2 \cdot \frac{n-1}{n}$ slijedi

$$\mathsf{E}\left(\frac{n}{n-1}\cdot \overline{S}_n^2\right) = \sigma^2,$$

pa je

$$\widehat{S}_n^2 := \frac{n}{n-1}\cdot \overline{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$$

centrirana ocjena za σ^2.

(c)

$$\mathsf{E}T_n = c\sum_{k=1}^{n-1}\mathsf{E}(X_{k+1} - X_k)^2 = c(n-1)\mathsf{E}(X_2 - X_1)^2$$
$$= c(n-1)(\mathsf{E}X_1^2 - 2\mathsf{E}(X_1)\mathsf{E}(X_2) + \mathsf{E}X_2^2)$$
$$= c(n-1)(2\mathsf{E}X^2 - 2(\mathsf{E}X)^2)$$
$$= 2c(n-1)\sigma^2.$$

Da bi ocjena T_n bila centrirana treba da važi $2c(n-1) = 1$, odnosno $c = \dfrac{1}{2(n-1)}$.

▶

2. U kutiji se nalazi N kuglica numerisanih brojevima od 1 do N, N je nepoznato. Iz kutije se po modelu sa vraćanjem izvlače dvije kuglice. Neka je X_1 broj na prvoizvučenoj, a X_2 broj na drugoizvučenoj kuglici. Nepoznato N ocjenjuje se statistikom

$$V = \frac{3}{2}\max\{X_1, X_2\} - \frac{3}{4}.$$

Naći $\mathsf{E}V$.

◀

Neka je $Y_2 = \max\{X_1, X_2\}$ uzorački maksimum.

Raspodjela slučajne veličine Y_2 je:

$$\mathsf{P}\{Y_2 = k\} = \frac{k^2 - (k-1)^2}{N^2} = \frac{2k-1}{N^2}, \quad k = 1, \ldots, N.$$

Očekivanje za Y_2:

$$\begin{aligned} \mathsf{E}Y_2 = \sum_{k=1}^{N} k \cdot \frac{2k-1}{N^2} &= \frac{2}{N^2}\sum_{k=1}^{N} k^2 - \frac{1}{N^2}\sum_{k=1}^{N} k \\ &= \frac{2}{N^2} \cdot \frac{N(N+1)(2N+1)}{6} - \frac{1}{N^2} \cdot \frac{N(N+1)}{2} \\ &= \frac{2}{3}N + \frac{1}{2} - \frac{1}{6N}. \end{aligned}$$

Slijedi

$$\begin{aligned} \mathsf{E}V &= \frac{3}{2}\,\mathsf{E}Y_2 - \frac{3}{4} \\ &= \frac{3}{2}\left(\frac{2}{3}N + \frac{1}{2} - \frac{1}{6N}\right) - \frac{3}{4} \\ &= N - \frac{1}{4N}. \end{aligned}$$

▶

3. U kutiji se nalaze bijele i crne kuglice i znamo da ih ukupno ima $N > 1$. Iz kutije se (a) po modelu sa vraćanjem (b) po modelu bez vraćanja izvlači n kuglica (u slučaju (b) pretpostaviti $n \leqslant N$). Neka je μ_n broj bijelih kuglica u uzorku od n izvučenih, a M nepoznati broj bijelih kuglica u kutiji (prije izvlačenja). Vjerovatnoća $p = \frac{M}{N}$ se ocjenjuje statistikom $p_n^* = \frac{\mu_n}{n}$. Izračunati $\mathsf{E}\,p_n^*$ i $\mathsf{D}\,p_n^*$.

◀

(a) Očigledno $\mu_n : \mathcal{B}(n, p)$, pa je

$$\mathsf{E}\,p_n^* = \frac{1}{n}\,\mathsf{E}\mu_n = \frac{1}{n} \cdot np = p$$

i

$$\mathsf{D}\,p_n^* = \frac{1}{n^2}\,\mathsf{D}\mu_n = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n}.$$

(b) U ovom slučaju je $\mu_n : \mathcal{H}(M, N-M, n)$. Koristeći indikatore (uraditi za vježbu!) dobija se $\mathsf{E}\mu_n = \frac{nM}{N}$ i $\mathsf{D}\mu_n = \frac{nM(N-M)(N-n)}{N^2(N-1)}$, pa je

$$\mathsf{E}\,p_n^* = \frac{1}{n}\,\mathsf{E}\mu_n = \frac{M}{N} = p$$

i

$$\mathsf{D}\,p_n^* = \frac{1}{n^2}\,\mathsf{D}\mu_n = \frac{1}{n^2} \cdot \frac{nM(N-M)(N-n)}{N^2(N-1)} = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}.$$

►

4. Obilježje X ima $\mathcal{U}(0, 2\theta)$, $\theta > 0$, raspodjelu, θ je nepoznati parametar. Parametar θ se na osnovu uzorka $(X_1, \ldots, X_{2n+1})$ ocjenjuje statistikom Y_{n+1} ($n+1$-va po veličini u datom uzorku). Dokazati da je ocjena Y_{n+1} postojana.

◄

Gustinu k-te statistike poretka Y_k smo izvodili na predmetu Teorija vjerovatnoće (pogledati Primjer 2.5.42.).

Potražimo k-tu statistiku poretka uzorka $(X_1, \ldots, X_n)$ apsolutno-neprekidnog obilježja X čija je funkcija raspodjele $F(\cdot)$ i funkcija gustine $f(\cdot)$, na nešto drugačiji način.

Neka je $W_x = \{j : X_j < x\}$. Tada važi

$$F_{Y_k}(x) = \mathsf{P}\{Y_k < x\} = \mathsf{P}\{|W_x| \geqslant k\} = \sum_{j=k}^{n} \binom{n}{j} (F(x))^j (1-F(x))^{n-j}.$$

Diferenciranjem dobijamo

$$\begin{aligned} f_{Y_k}(x) &= \sum_{j=k}^{n} \binom{n}{j} \cdot \big(j \cdot (F(x))^{j-1}(1-F(x))^{n-j} f(x) - (F(x))^j (n-j)(1-F(x))^{n-j-1} f(x)\big) \\ &= f(x) \left(\sum_{j=k}^{n} j \binom{n}{j} (F(x))^{j-1}(1-F(x))^{n-j} - \sum_{j=k}^{n-1} (n-j) \binom{n}{j} (F(x))^j (1-F(x))^{n-j-1} \right). \end{aligned}$$

Korsiteći poznate binomne identitete

$$j\binom{n}{j} = n\binom{n-1}{j-1} \quad \text{i} \quad (n-j)\binom{n}{j} = n\binom{n-1}{j},$$

dobijamo

$$\begin{aligned} f_{Y_k}(x) &= nf(x) \left(\sum_{j=k}^{n} \binom{n-1}{j-1} (F(x))^{j-1}(1-F(x))^{n-1-(j-1)} - \sum_{j=k}^{n-1} \binom{n-1}{j} (F(x))^j (1-F(x))^{n-1-j} \right) \\ &= nf(x) \binom{n-1}{k-1} (F(x))^{k-1} (1-F(x))^{n-k}. \end{aligned}$$

Za dobijanje posljednje jednakosti smo uočili da je

$$\sum_{j=k}^{n-1}\binom{n-1}{j}(F(x))^j(1-F(x))^{n-1-j}=\sum_{j=k+1}^{n}\binom{n-1}{j-1}(F(x))^{j-1}(1-F(x))^{n-1-(j-1)}.$$

Dakle,

$$f_{Y_k}(x)=\frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}f(x).$$

Gustina $(n+1)$-ve statistike poretka uzorka $(X_1,\ldots,X_{2n+1})$ obilježja $X:\mathcal{U}(0,2\theta)$ je:

$$\begin{aligned}
f_{Y_{n+1}}(x;\theta) &= \frac{(2n+1)!}{n!n!}(F(x;\theta))^n(1-F(x;\theta))^n f(x;\theta)\\
&= \begin{cases}\dfrac{(2n+1)!}{n!n!}\left(\dfrac{x}{2\theta}\right)^n\left(1-\dfrac{x}{2\theta}\right)^n\dfrac{1}{2\theta}, & x\in[0,2\theta],\\ 0, & \text{inače.}\end{cases}
\end{aligned}$$

Podsjetnik iz Analize.
Beta funkcija $B(r,s)$, $r,s>0$, definiše se kao

$$B(r,s):=\int_0^1 t^{r-1}(1-t)^{s-1}dt.$$

Pokazuje se da važi

$$B(r,s)=\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$

gdje je sa Γ označena gama funkcija $\Gamma(a):=\int_0^{+\infty}t^{a-1}e^{-t}dt$, $a>0$.

Potražimo očekivanje i disperziju za Y_{n+1}:

$$\begin{aligned}
\mathsf{E}Y_{n+1} &= \frac{(2n+1)!}{n!n!}\int_0^{2\theta}\left(\frac{x}{2\theta}\right)^{n+1}\left(1-\frac{x}{2\theta}\right)^n dx\\
&= \frac{(2n+1)!}{n!n!}\cdot 2\theta\int_0^1 t^{n+1}(1-t)^n dt\\
&= \frac{(2n+1)!}{n!n!}\cdot 2\theta\cdot B(n+2,n+1)\\
&= \frac{(2n+1)!}{n!n!}\cdot 2\theta\cdot\frac{\Gamma(n+2)\Gamma(n+1)}{\Gamma(2n+3)}\\
&= \frac{(2n+1)!}{n!n!}\cdot 2\theta\cdot\frac{(n+1)!n!}{(2n+2)!}\\
&= \theta;
\end{aligned}$$

$$
\begin{aligned}
\mathsf{E}Y_{n+1}^2 &= \frac{(2n+1)!}{n!n!} \cdot 2\theta \int_0^{2\theta} \left(\frac{x}{2\theta}\right)^{n+2} \left(1-\frac{x}{2\theta}\right)^n dx \\
&= \frac{(2n+1)!}{n!n!} \cdot 4\theta^2 \int_0^1 t^{n+2}(1-t)^n dt \\
&= \frac{(2n+1)!}{n!n!} \cdot 4\theta^2 \cdot B(n+3, n+1) \\
&= \frac{(2n+1)!}{n!n!} \cdot 4\theta^2 \cdot \frac{\Gamma(n+3)\Gamma(n+1)}{\Gamma(2n+4)} \\
&= \frac{(2n+1)!}{n!n!} \cdot 4\theta^2 \cdot \frac{(n+2)!n!}{(2n+3)!} \\
&= \frac{2n+4}{2n+3}\,\theta^2,
\end{aligned}
$$

pa je

$$
\mathsf{D}Y_{n+1} = \frac{2n+4}{2n+3}\,\theta^2 - \theta^2 = \frac{\theta^2}{2n+3}.
$$

Neka je $\varepsilon > 0$ proizvoljan broj. Tada, koristeći Čebišovljevu nejednakost, dobijamo:

$$
\mathsf{P}\left\{|Y_{n+1}-\theta| > \varepsilon\right\} = \mathsf{P}\left\{|Y_{n+1}-\mathsf{E}Y_{n+1}| > \varepsilon\right\} \leq \frac{\mathsf{D}Y_{n+1}}{\varepsilon^2} = \frac{\theta^2}{(2n+3)\varepsilon^2} \to 0, \text{ pri } n \to \infty,
$$

pa važi $Y_{n+1} \xrightarrow{\mathsf{P}} \theta$, pri $n \to \infty$, odnosno Y_{n+1} je postojana ocjena parametra θ.

▶