

Оцењивање параметара

Непристрасност: $\hat{\theta}_n$ је непристрасна оцена параметра θ , ако је $E(\hat{\theta}_n) = \theta$.

★ Ако оцена $\hat{\theta}_n$ није непристрасна, али $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$, онда је оцена асимптотски непристрасна.

Постојаност: Оцена $\hat{\theta}_n$ је постојана оцена параметра θ , ако $\hat{\theta}_n \xrightarrow{P} \theta$, односно, ако за

$$(\forall \varepsilon > 0) \lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0.$$

★ Ако је $\hat{\theta}_n$ непристрасна оцена параметра θ и $D(\hat{\theta}_n) \rightarrow 0, n \rightarrow \infty$, онда је $\hat{\theta}_n$ постојана оцена параметра θ , јер је

$$P\{|\hat{\theta}_n - \theta| > \varepsilon\} \leq \frac{E(\hat{\theta}_n - \theta)^2}{\varepsilon^2} = \frac{D(\hat{\theta}_n)}{\varepsilon^2}$$

Постоје различите методе добијања оцена, а као меру квалитета оцена често се користи средње квадратна грешка. Односно

- Ако су $\hat{\theta}_n$ и $\tilde{\theta}_n$ две оцене параметра θ , кажемо да је $\hat{\theta}_n$ боља од $\tilde{\theta}_n$ у средње квадратном смислу, ако је

$$MSE(\hat{\theta}_n) = E(\hat{\theta}_n - \theta)^2 < E(\tilde{\theta}_n - \theta)^2 = MSE(\tilde{\theta}_n).$$

- Ако су $\hat{\theta}_n$ и $\tilde{\theta}_n$ непристрасне оцене, довољно је да упоређујемо дисперзије, јер је онда

$$MSE(\hat{\theta}_n) = E(\hat{\theta}_n - \theta)^2 = E(\hat{\theta}_n - E(\hat{\theta}_n))^2 = D(\hat{\theta}_n).$$

Дакле, у том случају боља је она оцена која има мању дисперзију.

Задатак 1. Обележје X посматране популације има нормалну $\mathcal{N}(m, m)$ расподелу. За оцену непознатог параметра m предложене су две статистике: узорачка средина $\bar{X}_n$ и поправљена узорачка дисперзија $\tilde{S}_n^2$. Испитати непристрасност и постојаност тих оцена. Испитати која је оцена боља у средње квадратном смислу.

Решење. Испитујемо непристрасност обе оцене:

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} n E(X_1) = m,$$

$$E(\tilde{S}_n^2) = E\left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = E\left(\frac{n}{n-1} \bar{S}_n^2\right) = E\left(\frac{n \bar{S}_n^2}{m}\right) \frac{m}{n-1} = m,$$

при чему је коришћено да случајна величина $\frac{n \bar{S}_n^2}{m} \in \chi_{n-1}^2$, па је њено очекивање једнако $n-1$, а дисперзија $2(n-1)$.

Обе оцене су непристрасне, па су њихове средње квадратне грешке једнаке њиховим дисперзијама. Дакле, довољно је да упоредимо дисперзије ових оцена да бисмо испитали која је оцена боља у средње квадратном смислу. Како је

$$D(\bar{X}_n) = D\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} n D(X_1) = \frac{m}{n},$$

$$D(\tilde{S}_n^2) = D\left(\frac{n}{n-1}\bar{S}_n^2\right) = D\left(\frac{m}{n-1}\frac{n\bar{S}_n^2}{m}\right) = \frac{m^2}{(n-1)^2}D\left(\frac{n\bar{S}_n^2}{m}\right) = \frac{2(n-1)m^2}{(n-1)^2} = \frac{2m^2}{n-1},$$

то је

$$D(\bar{X}_n) > D(\tilde{S}_n^2) \iff \frac{m}{n} > \frac{2m^2}{n-1} \iff \frac{n-1}{n} > 2m$$

Када је $\frac{n-1}{n} > 2m$, онда је $\tilde{S}_n^2$ боља у средње квадратном смислу, а када је $\frac{n-1}{n} < 2m$, онда је боља $\bar{X}_n$. Ако важи једнакост подједнако су добре. С обзиром да не постоји универзално правило, које важи за свако $m > 0$, не можемо закључити да је нека универзално боља.

Остало је још да испитамо постојаност ових оцена. Постојаност подразумева да оцена конвергира у вероватноћи ка параметру m . Нека је $\varepsilon > 0$ произвољно:

$$P\{|\bar{X}_n - m| > \varepsilon\} \leq \frac{E(\bar{X}_n - m)^2}{\varepsilon^2} = \frac{D(\bar{X}_n)}{\varepsilon^2} = \frac{m}{n\varepsilon^2},$$

при чему је коришћена Чебишовљева неједнакост. Дакле,

$$0 \leq P\{|\bar{X}_n - m| > \varepsilon\} \leq \frac{m}{n\varepsilon^2} \rightarrow 0, n \rightarrow \infty$$

па по теореме о два полицајца $P\{|\bar{X}_n - m| > \varepsilon\} \rightarrow 0, n \rightarrow \infty$ и то важи за $\forall \varepsilon > 0$. Закључујемо да $\bar{X}_n \xrightarrow{P} m$, односно да је $\bar{X}_n$ постојана оцена за m . Даље,

$$0 \leq P\{|\tilde{S}_n^2 - m| > \varepsilon\} \leq \frac{D(\tilde{S}_n^2)}{\varepsilon^2} = \frac{2m^2}{(n-1)\varepsilon^2} \rightarrow 0, n \rightarrow \infty$$

Теорема о два полицајца даје да $P\{|\tilde{S}_n^2 - m| > \varepsilon\} \rightarrow 0, n \rightarrow \infty$, за ε одабрано произвољно, па ово важи за свако $\varepsilon > 0$, те $\tilde{S}_n^2 \xrightarrow{P} m$. Дакле, оцена $\tilde{S}_n^2$ је такође постојана оцена за m . ✓

Задатак 2. Обележје X има нормалну $\mathcal{N}(m, 1), m \in (0, 1)$, расподелу. За оцену непознатог параметра m , на основу узорка обима n , предложене су две статистике: узорачка средина $\bar{X}_n$ и $T_n = \bar{X}_n I\{0 \leq \bar{X}_n \leq 1\} + I\{\bar{X}_n > 1\}$. Испитати која оцена је боља у средње квадратном смислу.

Решење. Испитајмо која од предложених оцена је боља у средње квадратном смислу, односно желимо да упоредимо $MSE(\bar{X}_n) = E(\bar{X}_n - m)^2$ и $MSE(T_n) = E(T_n - m)^2$. Приметимо да

$$T_n = \begin{cases} 0, & \text{ако } \bar{X}_n < 0 \\ \bar{X}_n, & \text{ако } 0 \leq \bar{X}_n \leq 1 \\ 1, & \text{ако } \bar{X}_n > 1 \end{cases} \implies (T_n - m)^2 = \begin{cases} m^2, & \text{ако } \bar{X}_n < 0 \\ (\bar{X}_n - m)^2, & \text{ако } 0 \leq \bar{X}_n \leq 1 \\ (1 - m)^2, & \text{ако } \bar{X}_n > 1 \end{cases}$$

Дакле,

$$(T_n - m)^2 - (\bar{X}_n - m)^2 = \begin{cases} m^2 - (\bar{X}_n - m)^2, & \text{ако } \bar{X}_n < 0 \\ 0, & \text{ако } 0 \leq \bar{X}_n \leq 1 \\ (1 - m)^2 - (\bar{X}_n - m)^2, & \text{ако } \bar{X}_n > 1 \end{cases}$$

- Ако $\bar{X}_n < 0$: $m \in (0, 1)$, па је $(\bar{X}_n - m)^2 > m^2$, тј. $m^2 - (\bar{X}_n - m)^2 < 0$.
- Ако $\bar{X}_n > 1$: $m \in (0, 1)$, па је $(\bar{X}_n - m)^2 > (1 - m)^2$, тј. $(1 - m)^2 - (\bar{X}_n - m)^2 < 0$.

Стога, $(T_n - m)^2 - (\bar{X}_n - m)^2 \leq 0$, па је и $E((T_n - m)^2 - (\bar{X}_n - m)^2) \leq 0$, односно $E(T_n - m)^2 \leq E(\bar{X}_n - m)^2$, па је T_n ефикаснија од $\bar{X}_n$. ✓

Задатак 3. Обележје X има униформну $\mathcal{U}[0, \theta]$, $\theta > 0$, расподелу. За оцену непознатог параметра θ , на основу узорка обима n , предложене су две статистике: $2\bar{X}_n$ и $\frac{n+1}{n}X_{(n)}$. Испитати која оцена је боља у средње квадратном смислу.

Решење. Испитаћемо непристрасност за почетак:

$$E(2\bar{X}_n) = 2E(X_1) = 2 \cdot \frac{\theta}{2} = \theta,$$

$$E\left(\frac{n+1}{n}X_{(n)}\right) = \frac{n+1}{n}E(X_{(n)}).$$

Да бисмо израчунали очекивање случајне величине $X_{(n)}$ одредимо прво њену расподелу:

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

$$= (P\{X_1 \leq x\})^n = \left(\frac{x}{\theta}\right)^n, \quad x \in [0, \theta],$$

односно

$$f_{(n)}(x) = \frac{d}{dx}F_{(n)}(x) = n \frac{x^{n-1}}{\theta^n}.$$

Дакле,

$$E\left(\frac{n+1}{n}X_{(n)}\right) = \frac{n+1}{n} \int_0^\theta x n \frac{x^{n-1}}{\theta^n} dx = \frac{n+1}{\theta^n} \frac{x^{n+1}}{n+1} \Big|_0^\theta = \theta.$$

Обе оцене су непристрасне, па можемо упоређивати дисперзије. Како је

$$D(2\bar{X}_n) = 4D(\bar{X}_n) = \frac{4}{n^2}nD(X_1) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n},$$

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

$$= \frac{(n+1)^2}{n^2}(E(X_{(n)}^2) - (E(X_{(n)}))^2)$$

$$= \frac{(n+1)^2}{n^2} \left(\int_0^\theta x^2 n \frac{x^{n-1}}{\theta^n} dx - \left(\int_0^\theta x n \frac{x^{n-1}}{\theta^n} dx \right)^2 \right)$$

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

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

то је

$$D\left(\frac{n+1}{n}X_{(n)}\right) < D(2\bar{X}_n) \iff \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{3n} \iff 3 < n+2$$

Ово важи кад год је $n > 1$, па закључујемо да је оцена $\frac{n+1}{n}X_{(n)}$ боља у средње квадратном смислу. ✓

Ефикасност (непристрасних) оцена

За регуларне фамилије дефинишемо **информациону функцију Фишера**

$$I(\theta) = E \left(\frac{\partial}{\partial \theta} \ln p(X; \theta) \right)^2 = E \left(-\frac{\partial^2}{\partial \theta^2} \ln p(X; \theta) \right)$$

и **Рао-Крамерову доњу границу дисперзије свих непристрасних оцена параметра θ**

$$G = \frac{1}{nI(\theta)}.$$

За сваку непристрасну оцену $\hat{\theta}_n$ важи да је $G \leq D(\hat{\theta}_n)$, а ако је $G = D(\hat{\theta}_n)$, онда кажемо да је оцена $\hat{\theta}_n$ **ефикасна**.

У складу са тиме, дефинишемо **ефикасност** оцене $\hat{\theta}_n$

$$ef(\hat{\theta}_n) = \frac{G}{D(\hat{\theta}_n)} \begin{cases} < 1, & \text{оцена није ефикасна} \\ = 1, & \text{оцена јесте ефикасна.} \end{cases}$$

Задатак 4. Обележје X има густину расподеле $f(x; \theta) = \theta x^{\theta-1}, x \in (0, 1], \theta > 0$. За оцену непознатог параметра θ предложена је статистика $V = \frac{n-1}{-\sum_{k=1}^n \ln X_k}$. Испитати да ли је та оцена ефикасна.

Решење. Почињемо испитивањем регуларности дате фамилије расподела. Узорачки простор је $(0, 1]$, па не зависи од непознатог параметра θ . Желимо да испитамо да ли је $\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x; \theta) dx = 0$. Како је

$$\frac{\partial}{\partial \theta} f(x; \theta) = \frac{\partial}{\partial \theta} (\theta x^{\theta-1}) = x^{\theta-1} + \theta x^{\theta-1} \ln x,$$

па је

$$\begin{aligned} \int_0^1 \frac{\partial}{\partial \theta} f(x; \theta) dx &= \int_0^1 x^{\theta-1} dx + \int_0^1 \theta x^{\theta-1} \ln x dx = \left| \ln x = t \implies x = e^t \right| \\ &= \frac{x^\theta}{\theta} \Big|_0^1 + \theta \int_{-\infty}^0 e^{\theta t} t dt = \left| u = t \implies du = dt \right| \\ &= \frac{1}{\theta} + \theta \left(\frac{te^{\theta t}}{\theta} \Big|_{-\infty}^0 - \int_{-\infty}^0 \frac{e^{\theta t}}{\theta} dt \right) = \frac{1}{\theta} - \frac{1}{\theta} = 0. \end{aligned}$$

Фамилија јесте регуларна. Рачунамо информациону функцију Фишера, да бисмо добили доњу границу дисперзије свих непристрасних оцена:

$$I(\theta) = E \left(-\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta) \right).$$

Дакле,

$$\begin{aligned} \ln f(x; \theta) &= \ln \theta + (\theta - 1) \ln x, \\ \frac{\partial}{\partial \theta} \ln f(x; \theta) &= \frac{1}{\theta} + \ln x, \\ \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) &= -\frac{1}{\theta^2} \end{aligned}$$

па је

$$I(\theta) = E\left(\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}.$$

Према томе, доња граница дисперзије свих непристрасних оцена једнака је

$$G = \frac{1}{nI(\theta)} = \frac{\theta^2}{n}.$$

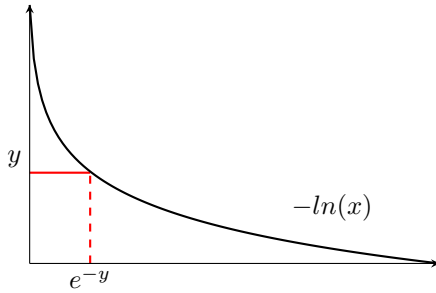
Да би имало смисла да испитујемо ефикасност оцене, морамо прво да проверимо да ли је она непристрасна. Проверавамо да ли је

$$E(V) = E\left(\frac{n-1}{-\sum_{k=1}^n \ln X_k}\right) = \theta.$$

Нека је $U = -\sum_{k=1}^n \ln X_k$, а $Y_k = -\ln X_k$. Пошто је V реална функција од U , довољно је да одредимо расподелу случајне величине U и онда је $E(V) = E\left(\frac{n-1}{U}\right)$ што ћемо лако израчунати.

За почетак, испитујемо расподелу случајних величина $Y_k = -\ln X_k, k = 1, \dots, n$. Приметимо да

$$F_{Y_k}(y) = P\{Y_k \leq y\} = P\{-\ln X_k \leq y\} = (\star)$$



$$\begin{aligned} y < 0: & (\star) = 0; \\ y \geq 0: & \end{aligned}$$

$$\begin{aligned} (\star) &= P\{e^{-y} \leq X_k \leq 1\} \\ &= \int_{e^{-y}}^1 \theta x^{\theta-1} dx \\ &= x^{\theta} \Big|_{e^{-y}}^1 = 1 - e^{-\theta y} \end{aligned}$$

Дакле, случајне величине $Y_k, k = 1, \dots, n$ имају $\mathcal{E}(\theta)$ расподелу и независне су јер су случајне величине $X_1, \dots, X_n$ независне. Закључујемо да

$$U = \sum_{k=1}^n Y_k = -\sum_{k=1}^n \ln X_k \in \gamma(n, \theta).$$

Враћамо се на очекивање случајне величине V :

$$\begin{aligned} E(V) &= E\left(\frac{n-1}{U}\right) = (n-1)E\left(\frac{1}{U}\right) = (n-1) \int_0^{\infty} \frac{1}{u} \frac{u^{n-1} \theta^n e^{-\theta u}}{\Gamma(n)} du \\ &= (n-1) \int_0^{\infty} \frac{u^{n-2} \theta^n e^{-\theta u}}{\Gamma(n-1)} du \cdot \frac{\Gamma(n-1)}{\Gamma(n)} \theta = \frac{(n-1)(n-2)!}{(n-1)!} \theta = \theta, \end{aligned}$$

Према томе, V јесте непристрасна оцена, па сада рачунамо њену дисперзију, да бисмо је упоредили са доњом границом G и видели да ли је ефикасна. Можемо приметити

да

$$\begin{aligned}
 D(V) &= E(V^2) - (E(V))^2 = E\left(\left(\frac{n-1}{U}\right)^2\right) - \theta^2 = (n-1)^2 E\left(\left(\frac{1}{U}\right)^2\right) - \theta^2 \\
 &= (n-1)^2 \int_0^\infty \frac{1}{u^2} \frac{u^{n-1} \theta^n e^{-\theta u}}{\Gamma(n)} du - \theta^2 \\
 &= (n-1)^2 \int_0^\infty \frac{u^{n-3} \theta^{n-2} e^{-\theta u}}{\Gamma(n-2)} du \cdot \frac{\Gamma(n-2)}{\Gamma(n)} \theta^2 - \theta^2 \\
 &= (n-1)^2 \frac{(n-3)!}{(n-1)!} \theta^2 - \theta^2 = \frac{n-1}{n-2} \theta^2 - \theta^2 = \frac{\theta^2}{n-2}.
 \end{aligned}$$

Дакле,

$$ef(V) = \frac{G}{DV} = \frac{\frac{\theta^2}{n}}{\frac{\theta^2}{n-2}} = \frac{n-2}{n} < 1,$$

па ова оцена није ефикасна (али јесте асимптотски ефикасна). $\checkmark$

Задатак 5. Нека је $X_1, X_2, \dots, X_n$ узорак из $\mathcal{N}(0, \sigma^2)$ расподеле. Одредити константу c тако да $c \sum_{i=1}^n |X_i|$ буде непристрасна оцена σ и одредити њену ефикасност.

Решење. Одређујемо c тако да $Y = c \sum_{i=1}^n |X_i|$ буде непристрасна оцена за σ :

$$E(Y) = c \sum_{i=1}^n E(|X_i|) = c \cdot n E(|X_1|).$$

Приметимо да

$$\begin{aligned}
 E(|X_1|) &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = \int_{-\infty}^0 (-x) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_0^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
 &= 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = \left| \frac{\frac{x^2}{2\sigma^2} = t}{x dx = \sigma^2 dt} \right| \\
 &= 2\sigma^2 \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-t} dt = \sqrt{\frac{2}{\pi}} \sigma \left(-e^{-t} \Big|_0^{\infty} \right) = \sqrt{\frac{2}{\pi}} \sigma.
 \end{aligned}$$

Хоћемо да $E(Y) = \sigma$, тј. $c \cdot n \sqrt{\frac{2}{\pi}} \sigma = \sigma$, односно $c = \frac{1}{n} \sqrt{\frac{\pi}{2}}$.

Сада желимо да испитамо ефикасност оцене $\frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |X_i|$. Испитујемо регуларност дате фамилије расподела (узорачки простор је $\mathbb{R}$, па не зависи од σ) :

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \sigma} f(x; \sigma) = 0 \quad (?)$$

Приметимо да

$$\begin{aligned}
 \frac{\partial}{\partial \sigma} f(x; \sigma) &= \frac{\partial}{\partial \sigma} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \right) = \frac{1}{\sqrt{2\pi}} \frac{\partial}{\partial \sigma} \left(\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma} \right) = \frac{1}{\sqrt{2\pi}} \left(\frac{e^{-\frac{x^2}{2\sigma^2}} \frac{2x^2}{2\sigma^3} \sigma - e^{-\frac{x^2}{2\sigma^2}}}{\sigma^2} \right) \\
 &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \left(\frac{x^2}{\sigma^2} - 1 \right)
 \end{aligned}$$

Дакле,

$$\begin{aligned}
 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} \frac{x^2}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx &= \frac{1}{\sigma} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} x^2 e^{-\frac{x^2}{2\sigma^2}} dx - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \right) \\
 &= \frac{1}{\sigma} \left(\frac{1}{\sigma^2} E(X^2) - 1 \right) \\
 &= \frac{1}{\sigma} \left(\frac{1}{\sigma^2} ((D(X) + (E(X))^2) - 1) \right) \\
 &= \frac{1}{\sigma} \left(\frac{1}{\sigma^2} (\sigma^2 + 0) - 1 \right) = 0.
 \end{aligned}$$

Фамилија јесте регуларна, па можемо рачунати информациону функцију Фишера и доњу границу дисперзије непристрасних оцена. Рачунамо информациону функцију Фишера:

$$I(\sigma) = E \left(-\frac{\partial^2}{\partial \sigma^2} \ln f(X; \sigma) \right).$$

Дакле,

$$\begin{aligned}
 \ln f(x; \sigma) &= \ln \frac{1}{\sqrt{2\pi}\sigma^2} - \frac{x^2}{2\sigma^2} = -\frac{1}{2} \ln(2\pi) - \ln \sigma - \frac{x^2}{2\sigma^2}, \\
 \frac{\partial}{\partial \sigma} \ln f(x; \sigma) &= -\frac{1}{\sigma} + \frac{x^2}{\sigma^3}, \\
 \frac{\partial^2}{\partial \sigma^2} \ln f(x; \sigma) &= \frac{1}{\sigma^2} - \frac{3x^2}{\sigma^4},
 \end{aligned}$$

па је

$$I(\sigma) = E \left(-\frac{1}{\sigma^2} + \frac{3X^2}{\sigma^4} \right) = -\frac{1}{\sigma^2} + \frac{3E(X^2)}{\sigma^4} = -\frac{1}{\sigma^2} + \frac{3\sigma^2}{\sigma^4} = -\frac{1}{\sigma^2} + \frac{3}{\sigma^2} = \frac{2}{\sigma^2}.$$

Одавде је $G = \frac{1}{nI(\sigma)} = \frac{\sigma^2}{2n}$.

Пошто смо Y бирали тако да буде непристрасна оцена за σ , још треба само да израчунамо $D(Y)$:

$$\begin{aligned}
 D(Y) &= D \left(c \sum_{i=1}^n |X_i| \right) = c^2 \cdot n D(|X_1|) = c^2 \cdot n (E(X_1^2) - (E(|X_1|))^2) \\
 &= c^2 \cdot n \left(\sigma^2 - \frac{2}{\pi} \sigma^2 \right) = c^2 \cdot n \sigma^2 \left(1 - \frac{2}{\pi} \right) = \frac{\pi}{2} \frac{1}{n^2} n \sigma^2 \left(1 - \frac{2}{\pi} \right) = \frac{\sigma^2(\pi - 2)}{2n}.
 \end{aligned}$$

Како је

$$efY = \frac{G}{D(Y)} = \frac{\frac{\sigma^2}{2n}}{\frac{\sigma^2(\pi-2)}{2n}} = \frac{1}{\pi - 2} < 1,$$

закључујемо да ова оцена није ефикасна. ✓

Метод максималне веродостојности

Задатак 6. Обележје X има Пуасонову $\mathcal{P}(\lambda)$ расподелу. На основу узорка обима n методом максималне веродостојности одредити оцену непознатог параметра λ , а затим испитати непристрасност и ефикасност тако добијене оцене.

Решење. Закон расподеле обележја X је дат са:

$$P\{X = x\} = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0, 1, \dots\}.$$

Претпоставимо да имамо реализовани узорак $(x_1, \dots, x_n)$. Функција веродостојности је тада

$$L(\lambda) = \prod_{k=1}^n P\{X = x_k\} = \prod_{k=1}^n \frac{\lambda^{x_k} e^{-\lambda}}{x_k!} = \frac{e^{-n\lambda} \lambda^{\sum_{k=1}^n x_k}}{\prod_{k=1}^n x_k!},$$

а њен логаритам

$$\ln L(\lambda) = -n\lambda + \sum_{k=1}^n x_k \ln \lambda - \ln \left(\prod_{k=1}^n x_k! \right).$$

Желимо да пронађемо вредност λ која максимизује ову функцију. Тражимо стационарну тачку

$$\frac{\partial}{\partial \lambda} \ln L(\lambda) = -n + \frac{\sum_{k=1}^n x_k}{\lambda} = 0 \implies \lambda = \frac{\sum_{k=1}^n x_k}{n} = \bar{x}_n.$$

Треба проверити да ли је добијено решење заиста максимум: $\frac{\partial}{\partial \lambda} \ln L(\lambda)$ је позитивна за $\lambda < \bar{x}_n$, а негативна за $\lambda > \bar{x}_n$, па функција $\ln L(\lambda)$ расте за $\lambda < \bar{x}_n$, а опада за $\lambda > \bar{x}_n$. Према томе, $\bar{x}_n$ је тачка у којој $\ln L(\lambda)$ достиже максимум. Дакле, оцена параметра λ методом максималне веродостојности на основу реализованог узорка $(x_1, \dots, x_n)$ је

$$\hat{\lambda} = \bar{x}_n.$$

Ово је број, конкретна вредност која се рачуна на основу вредности изабраних у узорак. Као случајна величина, оцена методом максималне веродостојности је

$$\hat{\lambda} = \bar{X}_n.$$

Испитујемо непристрасност добијене оцене:

$$E(\hat{\lambda}) = E(\bar{X}_n) = E(X_1) = \lambda.$$

Оцена је непристрасна, па има смисла да испитујемо њену ефикасност. Испитујемо регуларност дате фамилије расподела:

$$\sum_{x=0}^{\infty} \frac{\partial}{\partial \lambda} p(x; \lambda) = 0 \quad (?)$$

Приметимо да

$$\frac{\partial}{\partial \lambda} p(x; \lambda) = \frac{\partial}{\partial \lambda} \left(\frac{\lambda^x e^{-\lambda}}{x!} \right) = \frac{1}{x!} \left(x \lambda^{x-1} e^{-\lambda} - \lambda^x e^{-\lambda} \right).$$

Односно

$$\sum_{x=0}^{\infty} \frac{\partial}{\partial \lambda} p(x; \lambda) = e^{-\lambda} \left(\sum_{x=1}^{\infty} \frac{x \lambda^{x-1}}{x!} - \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \right) = e^{-\lambda} \left(\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} - \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \right) = e^{-\lambda} (e^{\lambda} - e^{\lambda}) = 0.$$

Дакле, фамилија јесте регуларна. Рачунамо информациону функцију Фишера

$$I(\lambda) = E \left(-\frac{\partial^2}{\partial \lambda^2} \ln p(X; \lambda) \right).$$

$$\ln p(x; \lambda) = \ln \left(\frac{\lambda^x e^{-\lambda}}{x!} \right) = x \ln \lambda - \lambda - \ln x!,$$

$$\frac{\partial}{\partial \lambda} \ln p(x; \lambda) = \frac{x}{\lambda} - 1,$$

$$\frac{\partial^2}{\partial \lambda^2} \ln p(x; \lambda) = -\frac{x}{\lambda^2}.$$

Дакле,

$$I(\lambda) = E \left(\frac{X}{\lambda^2} \right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda},$$

па је

$$G = \frac{1}{nI(\lambda)} = \frac{\lambda}{n}.$$

Рачунамо дисперзију оцено:

$$D(\hat{\lambda}) = D(\bar{X}_n) = \frac{D(X_1)}{n} = \frac{\lambda}{n}.$$

Дакле,

$$ef(\lambda) = \frac{G}{D(\lambda)} = \frac{\frac{\lambda}{n}}{\frac{\lambda}{n}} = 1,$$

па закључујемо да је ова оцена ефикасна. ✓

Задатак 7. Обележје X има густину расподеле $f(x; \theta) = \theta^2 x e^{-\theta x}$, $x \geq 0, \theta > 0$. На основу узорка обима n методом максималне веродостојности одредити оцену непознатог параметра θ .

Решење. Претпоставимо да имамо реализовани узорак $(x_1, \dots, x_n)$. Функција веродостојности је

$$L(\theta) = \prod_{k=1}^n f(x_k; \theta) = \theta^{2n} e^{-\theta \sum_{k=1}^n x_k} \prod_{k=1}^n x_k,$$

па је

$$\ln L(\theta) = 2n \ln \theta - \theta \sum_{k=1}^n x_k + \ln \left(\prod_{k=1}^n x_k \right).$$

Желимо да пронађемо $\theta > 0$ које максимизује $\ln L(\theta)$. Одређујемо стационрну тачку:

$$\frac{\partial}{\partial \theta} \ln L(\theta) = \frac{2n}{\theta} - \sum_{k=1}^n x_k = 0 \implies \frac{2n}{\theta} = \sum_{k=1}^n x_k \implies \theta = \frac{2n}{\sum x_k} = \frac{2}{\bar{x}_n}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је $\frac{\partial}{\partial \theta} \ln L(\theta)$ позитивна за $\theta < \frac{2}{\bar{x}_n}$, а негативна за $\theta > \frac{2}{\bar{x}_n}$, па је функција $\ln L(\theta)$ растућа за $\theta < \frac{2}{\bar{x}_n}$, а опадајућа за $\theta > \frac{2}{\bar{x}_n}$. Према томе, $\frac{2}{\bar{x}_n}$ је тачка у којој $\ln L(\theta)$ достиже максимум. Дакле, оцена параметра θ методом максималне веродостојности на основу реализованог узорка $(x_1, \dots, x_n)$ је

$$\hat{\theta} = \frac{2}{\bar{x}_n}.$$

Када оцену запишемо као случајну величину добија се:

$$\hat{\theta} = \frac{2}{\bar{X}_n}.$$

✓

Метод момената

Нека је $\mathbf{X}_n$ узорак са расподелом F_θ , где је $\theta = (\theta_1, \dots, \theta_p)$ p -димензиони параметар. Како су теоријски моменти обично функције од θ , оцене методом момента добијају се решавањем система од p једначина облика

$$E(X^j) = \frac{1}{n} \sum_{i=1}^n X_i^j, j = 1, \dots, p.$$

по $\theta_1, \dots, \theta_p$. Уместо момената, могу се користити и центрирани моменти, односно решавати систем

$$E(X - EX)^j = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^j, j = 1, \dots, p$$

Оцене добијене овим методом нису јединствене, али за довољно велико n се све оцене параметра понашају слично.

Задатак 8. На основу узорка обима n , наћи оцене методом момената за параметре следећих расподела:

- а) биномне $\mathcal{B}(10, p)$;
- б) униформне $\mathcal{U}[-\theta, \theta]$;
- в) гама $\gamma(\theta + 1, \theta + 2)$.

Решење. а) Изједначавамо моменте првог реда

$$EX = 10p = \bar{X}_n.$$

Према томе, $\hat{p} = \frac{\bar{X}_n}{10}$ је оцена методом момената за p .

б) Пошто је $EX = 0$, изједначавамо моменте другог реда

$$EX^2 = DX = \frac{\theta^2}{3} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Према томе, $\hat{\theta} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}$ је оцена методом момената за θ . Други начин је да изједначимо

$$DX = \frac{\theta^3}{n} = \bar{S}_n^2,$$

па на овај начин добијамо да је $\hat{\theta} = \sqrt{3\bar{S}_n^2}$ оцена методом момената за θ .

в) Изједначавамо моменте првог реда

$$\begin{aligned} EX = \frac{\theta + 1}{\theta + 2} = \bar{X}_n &\implies \theta + 1 = \theta \bar{X}_n + 2\bar{X}_n \\ &\implies \theta(1 - \bar{X}_n) = 2\bar{X}_n - 1. \end{aligned}$$

Према томе, $\hat{\theta} = \frac{2\bar{X}_n - 1}{1 - \bar{X}_n}$ је оцена методом момената за θ .

✓

Задатак 9. Нека X_n има фамилију расподела са густином $p(x; \lambda, \theta) = \lambda e^{-\lambda(x-\theta)}, x \geq \theta, \theta > 0, \lambda > 0$. Наћи оцене методом момената за параметре λ и θ .

Решење. Како је у питању заправо померена експоненцијална расподела (а може и експлицитно да се израчуна), то је:

$$EX = \theta + \frac{1}{\lambda},$$

$$DX = \frac{1}{\lambda^2}.$$

Када изједначимо момента првог и другог реда, решавањем система једначина добијамо $\hat{\lambda} = \frac{1}{\bar{S}_n}$, и $\hat{\theta} = \bar{X}_n - \bar{S}_n$.

✓

Задатак 10. Нека обелезје X има закон расподеле

$$\left(\begin{array}{ccc} -1 & 0 & 2 \\ \frac{\theta(1-\theta)}{4} & \frac{\theta(1-\theta)}{4} & 1 - \frac{\theta(1-\theta)}{2} \end{array} \right),$$

где је $\theta \in (0, 1)$. За оцену непознатог параметра $\theta(1 - \theta)$ на основу узорка обима n предложене су две оцене T_1 и T_2 , где је T_1 оцена добијена методом максималне веродостојности и T_2 оцена добијена методом момената. Одредити која од предложених оцена је боља у средње квадратном смислу, а затим испитати да ли је нека од њих ефикасна.

Решење. Закон расподеле се може записати као

$$p(x; \theta) = \left(\frac{\theta(1-\theta)}{4} \right)^{I\{x=-1\}} \left(\frac{\theta(1-\theta)}{4} \right)^{I\{x=0\}} \left(1 - \frac{\theta(1-\theta)}{2} \right)^{I\{x=2\}}$$

$$= \left(\frac{\theta(1-\theta)}{4} \right)^{1-I\{x=2\}} \left(1 - \frac{\theta(1-\theta)}{2} \right)^{I\{x=2\}},$$

или

$$p(x; \alpha) = \left(\frac{\alpha}{4} \right)^{1-I\{x=2\}} \left(1 - \frac{\alpha}{2} \right)^{I\{x=2\}},$$

где је $\alpha = \theta(1 - \theta)$.

Претпоставимо да имамо реализован узорак $(x_1, x_2, \dots, x_n)$. Функција веродостојности је

$$L(\alpha) = \prod_{k=1}^n p(x_k; \alpha) = \prod_{k=1}^n \left(\frac{\alpha}{4} \right)^{1-I\{x_k=2\}} \left(1 - \frac{\alpha}{2} \right)^{I\{x_k=2\}}$$

$$= \left(\frac{\alpha}{4} \right)^{n - \sum_{k=1}^n I\{x_k=2\}} \left(1 - \frac{\alpha}{2} \right)^{\sum_{k=1}^n I\{x_k=2\}}.$$

па је логаритам веродостојности

$$\ln L(\alpha) = \left(n - \sum_{k=1}^n I\{x_k = 2\} \right) \ln \frac{\alpha}{4} + \sum_{k=1}^n I\{x_k = 2\} \ln \left(1 - \frac{\alpha}{2} \right).$$

Желимо да пронађемо $\alpha > 0$ које максимизује $\ln L(\alpha)$, те стога одређујемо стационарну тачку:

$$\frac{\partial}{\partial \alpha} \ln L(\alpha) = \frac{n - \sum_{k=1}^n I\{x_k = 2\}}{\frac{\alpha}{4}} \cdot \frac{1}{4} + \frac{\sum_{k=1}^n I\{x_k = 2\}}{1 - \frac{\alpha}{2}} \cdot \left(-\frac{1}{2}\right) = 0,$$

одакле је

$$\alpha = \frac{2n - 2 \sum_{k=1}^n I\{x_k = 2\}}{n}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је

$$\frac{\partial^2}{\partial \alpha^2} \ln L(\alpha) = -\frac{n - \sum_{k=1}^n I\{x_k = 2\}}{\alpha^2} - \frac{\sum_{k=1}^n I\{x_k = 2\}}{(2 - \alpha)^2} < 0,$$

ово заиста јесте тачка у којој функција $\ln L(\alpha)$ достиже максимум. Према томе, на основу овог узорка, оцена методом максималне веродостојности је

$$T_1 = \frac{2n - 2 \sum_{k=1}^n I\{X_k = 2\}}{n}.$$

Са друге стране, важи да је

$$EX = -\frac{\theta(1-\theta)}{4} + 2 \left(1 - \frac{\theta(1-\theta)}{2}\right) = 2 - \frac{5}{4}\theta(1-\theta).$$

Сада из $EX = \bar{X}_n$ следи да је

$$T_2 = \frac{4}{5}(2 - \bar{X}_n).$$

Очекивања оцена су:

$$ET_1 = E \left(\frac{2n - 2 \sum_{k=1}^n I\{X_k = 2\}}{n} \right) = 2 - \frac{2}{n} \sum_{k=1}^n P\{X_k = 2\} = 2 - 2P\{X_k = 2\} = 2 - 2 \left(1 - \frac{\alpha}{2}\right) = \alpha,$$

$$ET_2 = E \left(\frac{4}{5}(2 - \bar{X}_n) \right) = \frac{8}{5} - \frac{4}{5}E\bar{X}_n = \frac{8}{5} - \frac{4}{5}EX_1 = \frac{8}{5} - \frac{4}{5} \left(2 - \frac{5}{4}\alpha\right) = \alpha,$$

односно, обе оцене су непристрасне оцене параметра α па можемо поредити њихове дисперзије. Прво нам треба дисперзија од X :

$$DX = EX^2 - (EX)^2 = \frac{\alpha}{4} + 4 - 2\alpha - \left(2 - \frac{5}{4}\alpha\right)^2 = \frac{\alpha(52 - 25\alpha)}{16}.$$

Сада лако рачунамо дисперзије оцена:

$$\begin{aligned} DT_1 &= D \left(2 - \frac{2}{n} \sum_{k=1}^n I\{X_k = 2\} \right) = \frac{4}{n^2} D \left(\sum_{k=1}^n I\{X_k = 2\} \right) \\ &= \frac{4}{n^2} n D(I\{X_k = 2\}) = \frac{4}{n} P\{X_1 = 2\}(1 - P\{X_1 = 2\}) \\ &= \frac{4}{n} \left(1 - \frac{\alpha}{2}\right) \frac{\alpha}{2} = \frac{\alpha(2 - \alpha)}{n}. \end{aligned}$$

$$DT_2 = D \left(\frac{4}{5}(2 - \bar{X}_n) \right) = \frac{16}{25} D\bar{X}_n = \frac{16}{25n} DX_1 = \frac{16}{25n} \frac{\alpha(52 - 25\alpha)}{16} = \frac{\alpha(52 - 25\alpha)}{25n}.$$

Како је за свако $n \in \mathbb{N}$ испуњено $\frac{\alpha(2-\alpha)}{n} < \frac{\alpha(52-25\alpha)}{25n} \iff 50 < 52$, закључујемо да је $DT_1 < DT_2$, па је оцена T_1 боља у средње квадратном смислу од оцено T_2 , па има смисла испитати њену ефикасност. Испитујемо регуларност дате фамилије расподела:

$$\sum_{x \in \{-1, 0, 2\}} \frac{\partial}{\partial \alpha} p(x; \alpha) = \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0.$$

Дата фамилија расподела је регуларна, па можемо рачунати Рао-Крамерову доњу границу дисперзија. Имамо да је

$$\ln p(x; \alpha) = (1 - I\{x = 2\}) \ln \frac{\alpha}{4} + I\{x = 2\} \ln \left(1 - \frac{\alpha}{2}\right),$$

$$\frac{\partial}{\partial \alpha} \ln p(x; \alpha) = \frac{1 - I\{x = 2\}}{\alpha} - \frac{I\{x = 2\}}{2 - \alpha},$$

$$\frac{\partial^2}{\partial \alpha^2} \ln p(x; \alpha) = -\frac{1 - I\{x = 2\}}{\alpha^2} - \frac{I\{x = 2\}}{(2 - \alpha)^2},$$

па је

$$\begin{aligned} I(\alpha) &= E \left(-\frac{\partial^2}{\partial \alpha^2} \ln p(X; \alpha) \right) = E \left(\frac{1 - I\{x = 2\}}{\alpha^2} + \frac{I\{x = 2\}}{(2 - \alpha)^2} \right) \\ &= \frac{1 - P\{x = 2\}}{\alpha^2} + \frac{P\{x = 2\}}{(2 - \alpha)^2} = \frac{1 - (1 - \frac{\alpha}{2})}{\alpha^2} + \frac{1 - \frac{\alpha}{2}}{(2 - \alpha)^2} \\ &= \frac{1}{2\alpha} + \frac{1}{2(2 - \alpha)} = \frac{1}{\alpha(2 - \alpha)}. \end{aligned}$$

Одавде добијамо да је доња граница Рао-Крамера једнака

$$G = \frac{1}{nI(\alpha)} = \frac{\alpha(2 - \alpha)}{n}.$$

Према томе, како је $ef(T_1) = \frac{G}{DT_1} = 1$, закључујемо да је ова оцена ефикасна. ✓