Tema 4

Exercițiul 1

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n de lege

$$\mathbb{P}(X = k) = \frac{\theta^k}{(1+\theta)^{k+1}}$$

unde θ este un parametru pozitiv. Determinați estimatorii obținuți prin metoda momentelor și prin metoda verosimilității maxime și studiați proprietățile acestora.

Exercițiul 2

Să se estimeze, prin metoda verosimilității maxime și prin metoda momentelor, parametrul p al unei repartiții Bernoulli $\mathcal{B}(p)$ plecând de la eșantionul de talie 25:

1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0

Exercițiul 3

Fie X o v.a. de lege Pareto de densitate:

$$f(x) = \begin{cases} \frac{\alpha-1}{x^\alpha}, & x \geq 1 \\ 0, & \text{altfel} \end{cases}$$

unde $\alpha > 2$ este un parametru necunoscut. Determinați estimatorul de verosimilitate maximă plecând de la un eșantion de talie n și arătați că acesta este consistent.

Exercițiul 4

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n dintr-o populație cu densitatea $f_\theta(x) = e^{-(x-\theta)}$, $x \geq \theta$.

a) Determinați estimatorul $\tilde{\theta}$ obținut prin metoda momentelor.

b) Dterminați estimatorul $\hat{\theta}$ obținut prin metoda verosimilității maxime.

c) Determinați legea variabilei $n(\hat{\theta} - \theta)$.

d) Verificați dacă estimatorul $\hat{\theta}$ este nedeplasat.

e) Calculați eroarea medie pătratică a lui $\hat{\theta}$.

f) În cazul în care $\theta = 2$ dorim să generăm 3 valori aleatoare din repartiția lui $X \sim f_\theta(x)$. Pentru aceasta dispunem de trei valori rezultate din repartiția uniformă pe $[0,\ 1]$: $u_1 = 0.25$, $u_2 = 0.4$ și $u_3 = 0.5$. Descrieți procedura.

Exercițiul 5

Fie $(X_i)_{1\leq i\leq n}$ un eșantion de talie n dintr-o populație de densitate f_θ dată de

$$f_\theta = \theta e^{-\theta x}\mathbf{1}_{[0,+\infty)}(x),$$

cu $\theta > 0$ un parametru necunoscut.

1. Ne propunem să estimăm θ prin $Y_n = \frac{n}{\sum_{i=1}^n X_i}$
   a) Arătați că estimatorul Y_n este bine definit.
   b) Explicați de ce este logic să alegem Y_n ca estimator pentru θ.
   c) Determinați legea limită a $\sqrt{n}(Y_n - \theta)$.
   d) Determinați legea sumei $\sum_{i=1}^n X_i$ și calculați $\mathbb{E}_\theta[(Y_n - \theta)^2]$.
2. Pentru estimarea lui θ propunem estimatorul $Z_n = \frac{n-1}{n}Y_n$
   a) Verifică Z_n proprietăți similare de convergență cu Y_n?
   b) Pe care dintre cei doi estimatori, Y_n sau Z_n, îl preferați?

Exercițiul 6

Fie $(X_i)_{1\leq i\leq n}$ un eșantion de talie n dintr-o populație de densitate

$$f_\theta(x) = \frac{2}{\theta^2}x\mathbf{1}_{[0,\theta]}(x),$$

cu $\theta > 0$ un parametru necunoscut.

1. Considerăm $X_{(n)}$ statistica de ordine de rang n
   a) Determinați densitatea lui $X_{(n)}$.
   b) Calculați media și varianța lui $X_{(n)}$.
   c) Arătați că $X_{(n)}$ converge în probabilitate la θ.
2. Studiați convergența lui $\bar{X}_n$ și determinați un estimator consistent pentru θ.
3. Pe care dintre estimatorii $X_{(n)}$ și $\frac{2}{3}\bar{X}_n$ îl preferați pentru a estima pe θ?

Tema 4

Soluții

Exercițiul 1

Fie $X_1, X_2, \dots, X_n$ un eșantion de talie n de lege

$$\mathbb{P}(X = k) = \frac{\theta^k}{(1+\theta)^{k+1}}$$

unde θ este un parametru pozitiv. Determinați estimatorii obținuți prin metoda momentelor și prin metoda verosimilității maxime și studiați proprietățile acestora.

Vom începe prin a determina estimatorul obținut prin metoda momentelor. Pentru aceasta trebuie să determinăm $\mathbb{E}[X_1]$. Avem

$$\begin{aligned}
\mathbb{E}[X_1] &= \sum_{k\geq 0} k\mathbb{P}(X_1 = k) = \sum_{k\geq 0} k\frac{\theta^k}{(1+\theta)^{k+1}} = \sum_{k\geq 1} k\frac{\theta^k}{(1+\theta)^{k+1}} \\
&= \frac{\theta}{(1+\theta)^2}\sum_{k\geq 1} k\left(\frac{\theta}{1+\theta}\right)^{k-1} = \frac{\theta}{(1+\theta)^2}\sum_{k\geq 0}(k+1)\left(\frac{\theta}{1+\theta}\right)^{k} \overset{q=\frac{\theta}{1+\theta}}{=} \frac{\theta}{(1+\theta)^2}\sum_{k\geq 0}(k+1)q^k \\
&= \frac{\theta}{(1+\theta)^2}\sum_{k\geq 0}\frac{d}{dq}q^{k+1} = \frac{\theta}{(1+\theta)^2}\frac{d}{dq}\left(\sum_{k\geq 0}q^{k+1}\right) = \frac{\theta}{(1+\theta)^2}\frac{d}{dq}\left(\frac{1}{1-q} - 1\right) \\
&= \frac{\theta}{(1+\theta)^2}\frac{1}{(1-q)^2} = \frac{\theta}{(1+\theta)^2}\frac{1}{\left(1-\frac{\theta}{1+\theta}\right)^2} = \theta.
\end{aligned}$$

Astfel estimatorul obținut prin metoda momentelor este $\tilde{\theta}_n = \bar{X}_n$.

Pentru a determina estimatorul de verosimilitate maximă trebuie să calculăm funcția de verosimilitate:

$$L(\theta|x_1,\dots,x_n) = f_\theta(x_1)\cdots f_\theta(x_n) = \frac{\theta^{x_1}}{(1+\theta)^{x_1+1}}\cdots\frac{\theta^{x_n}}{(1+\theta)^{x_n+1}} = \frac{\theta^{\sum_{i=1}^n x_i}}{(1+\theta)^{n+\sum_{i=1}^n x_i}}.$$

Vom găsi maximul considerând logaritmul funcției de verosimilitate

$$l(\theta|x_1,\dots,x_n) = \sum_{i=1}^n x_i\log(\theta) - \left(n + \sum_{i=1}^n x_i\right)\log(1+\theta)$$

care prin derivare devine $\frac{d}{d\theta}l(\theta|x_1,\dots,x_n) = \frac{1}{\theta}\sum_{i=1}^n x_i - \left(n+\sum_{i=1}^n x_i\right)\frac{1}{1+\theta}$. Rezolvând ecuația $\frac{d}{d\theta}l(\theta|x_1,\dots,x_n) = 0$ obținem estimatorul de verosimilitate maximă $\hat{\theta}_n = \bar{X}_n$ care putem observa că este același cu cel rezultat prin metoda momentelor.

Din metoda momentelor am văzut că estimatorul $\hat{\theta}_n (= \tilde{\theta}_n)$ este nedeplasat. Aplicând *Legea Numerelor Mari* deducem că $\bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X_1] = \theta$ de unde obținem că $\hat{\theta}_n$ este consistent. De asemenea, din *Teorema Limită Centrală* avem că

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \sqrt{n}\left(\bar{X}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, Var(X_1))$$

Pentru $Var(X_1)$ avem nevoie de momentul de ordin 2:

$$\begin{aligned}
\mathbb{E}[X_1^2] &= \sum_{k\geq 0} k^2 \mathbb{P}(X_1 = k) = \sum_{k\geq 0} k^2 \frac{\theta^k}{(1+\theta)^{k+1}} = \frac{\theta}{(1+\theta)^2}\sum_{k\geq 0}(k+1)^2\left(\frac{\theta}{1+\theta}\right)^k \\
&= \frac{\theta}{(1+\theta)^2}\sum_{k\geq 0}(k^2+2k+1)\left(\frac{\theta}{1+\theta}\right)^k \\
&= \frac{\theta}{(1+\theta)^2}\left[\sum_{k\geq 0} k^2\left(\frac{\theta}{1+\theta}\right)^k + \sum_{k\geq 0} 2(k+1)\left(\frac{\theta}{1+\theta}\right)^k - \sum_{k\geq 0}\left(\frac{\theta}{1+\theta}\right)^k\right] \\
&= \frac{\theta}{(1+\theta)^2}\left[(\theta+1)\mathbb{E}[X_1^2] + 2(\theta+1)^2 - \frac{1}{1-\frac{\theta}{1+\theta}}\right] = \frac{\theta}{(1+\theta)^2}\left[(\theta+1)\mathbb{E}[X_1^2] + 2(\theta+1)^2 - (\theta+1)\right]
\end{aligned}$$

de unde obținem $\mathbb{E}[X_1^2] = 2\theta^2 + \theta$ și $Var(X_1) = \theta^2 + \theta$. Prin urmare

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, \theta(\theta+1)).$$

De asemenea eroarea pătratică medie devine

$$MSE_\theta(\hat{\theta}_n) = Var_\theta(\hat{\theta}_n) + b_\theta(\hat{\theta}_n)^2 = Var_\theta(\hat{\theta}_n) = \frac{\theta(1+\theta)}{n}.$$

Exercițiul 2

Să se estimeze, prin metoda verosimilității maxime și prin metoda momentelor, parametrul p al unei repartiții Bernoulli $\mathcal{B}(p)$ plecând de la eșantionul de talie 25: 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 1 1 1 0

Dacă $X \sim \mathcal{B}(p)$ atunci $\mathbb{P}(X = x) = f_p(x) = p^x(1-p)^{1-x}$, media este $\mathbb{E}[X] = p$ iar varianța $Var(X) = p(1-p)$. Astfel, egalând media aritmetică cu cea empirică, metoda momentelor conduce la estimatorul $\tilde{p}_n = \bar{X}_n$.

Funcția de verosimilitate este:

$$L(p|x_1,\dots,x_n) = f_p(x_1)\cdots f_p(x_n) = p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

iar logaritmul funcției de verosimilitate devine

$$l(p|x_1,\dots,x_n) = \sum_{i=1}^n x_i \log(p) + \left(n - \sum_{i=1}^n x_i\right)\log(1-p).$$

Pentru a determina estimatorul de verosimilitate maximă trebuie să rezolvăm ecuația $\frac{d}{dp}l(p|x_1,\dots,x_n) = 0$ care implică

$$\frac{\sum_{i=1}^{n} x_i}{p} = \left(n - \sum_{i=1}^{n} x_i\right)\frac{1}{1-p}$$

de unde obținem estimatorul $\hat{p}_n = \bar{X}_n$ (observăm că $\hat{p}_n = \tilde{p}_n$). Folosind eșantionul din enunțul problemei găsim $\hat{p}_n = \tilde{p}_n = 0.28$.

Exercițiul 3

Fie X o v.a. de lege Pareto de densitate:

$$f(x) = \begin{cases} \frac{\alpha-1}{x^{\alpha}}, & x \geq 1 \\ 0, & \text{altfel} \end{cases}$$

unde $\alpha > 2$ este un parametru necunoscut. Determinați estimatorul de verosimilitate maximă plecând de la un eșantion de talie n și arătați că acesta este consistent.

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n dintr-o populație Pareto de parametru α. Se poate observa că funcția de verosimilitate este

$$L(\alpha|x_1,\ldots,x_n) = f_{\alpha}(x_1)\cdots f_{\alpha}(x_n) = \frac{\alpha-1}{x_1^{\alpha}}\cdots\frac{\alpha-1}{x_n^{\alpha}} = \frac{(\alpha-1)^n}{(x_1\cdots x_n)^{\alpha}}$$

iar logaritmul funcției de verosimilitate devine

$$l(\alpha|x_1,\ldots,x_n) = n\log(\alpha-1) - \alpha\log(x_1\cdots x_n).$$

Pentru a determina maximul funcției de verosimilitate trebuie să rezolvăm ecuația $\frac{d}{d\alpha}l(\alpha|x_1,\ldots,x_n) = 0$, care conduce la

$$\frac{n}{\alpha-1} = \log(x_1\cdots x_n)$$

de unde estimatorul de verosimilitate maximă este $\hat{\alpha}_n = 1 + \frac{1}{\frac{\sum_{i=1}^{n}\log(X_i)}{n}}$.

Cum variabilele aleatoare $X_1, X_2, \ldots, X_n$ sunt i.i.d. și cum $g(x) = \log(x)$ este continuă avem că $\log(X_1), \log(X_2), \ldots, \log(X_n)$ sunt i.i.d. iar din *Legea Numerelor Mari* obținem că

$$\frac{\sum_{i=1}^{n}\log(X_i)}{n} \overset{a.s.}{\to} \mathbb{E}[\log(X)].$$

Cum funcția $h(x) = 1 + \frac{1}{x}$ este continuă, din *Teorema aplicațiilor continue* găsim că

$$\hat{\alpha}_n = 1 + \frac{1}{\frac{\sum_{i=1}^{n}\log(X_i)}{n}} \overset{a.s.}{\to} 1 + \frac{1}{\mathbb{E}[\log(X)]}.$$

Pentru a verifica proprietatea de consistență trebuie să determinăm $\mathbb{E}[\log(X)]$,

$$\begin{aligned}
\mathbb{E}[\log(X)] &= \int_1^\infty \log(x)\frac{\alpha-1}{x^\alpha}\,dx = \int_1^\infty \log(x)\left(-x^{-\alpha+1}\right)'\,dx\\
&= -x^{1-\alpha}\log(x)|_1^\infty + \int_1^\infty \frac{1}{x}x^{1-\alpha}\,dx = \int_1^\infty x^{-\alpha}\,dx\\
&= \left.\frac{x^{1-\alpha}}{1-\alpha}\right|_1^\infty = \frac{1}{\alpha-1}.
\end{aligned}$$

Prin urmare găsim

$$\hat{\alpha}_n = 1 + \frac{1}{\frac{\sum_{i=1}^n \log(X_i)}{n}} \xrightarrow{a.s.} 1 + \frac{1}{\frac{1}{\alpha-1}} = \alpha,$$

ceea ce implică faptul că $\hat{\alpha}_n$ este un estimator consistent pentru α.

Exercițiul 4

Fie $X_1, X_2, \dots, X_n$ un eșantion de talie n dintr-o populație cu densitatea $f_\theta(x) = e^{-(x-\theta)}$, $x \geq \theta$.

a) Determinați estimatorul $\tilde{\theta}$ obținut prin metoda momentelor.

b) Dterminați estimatorul $\hat{\theta}$ obținut prin metoda verosimilității maxime.

c) Determinați legea variabilei $n(\hat{\theta}-\theta)$.

d) Verificați dacă estimatorul $\hat{\theta}$ este nedeplasat.

e) Calculați eroarea medie pătratică a lui $\hat{\theta}$.

f) În cazul în care $\theta = 2$ dorim să generăm 3 valori aleatoare din repartiția lui $X \sim f_\theta(x)$. Pentru aceasta dispunem de trei valori rezultate din repartiția uniformă pe $[0, 1]$: $u_1 = 0.25$, $u_2 = 0.4$ și $u_3 = 0.5$. Descrieți procedura.

a) Pentru metoda momentelor este necesar să calculăm $\mathbb{E}[X]$. Avem

$$\begin{aligned}
\mathbb{E}[X] &= \int_\theta^\infty xe^{-(x-\theta)}\,dx = e^\theta\int_\theta^\infty x\left(-e^{-x}\right)'\,dx\\
&= e^\theta\left[-xe^{-x}|_\theta^\infty + \int_\theta^\infty e^{-x}\,dx\right] = e^\theta\left[\theta e^{-\theta} + e^{-\theta}\right]\\
&= \theta + 1
\end{aligned}$$

și egalând cu media eșantionului, $\bar{X}_n$, găsim că estimatorul rezultat din metoda momentelor este $\tilde{\theta}_n = \bar{X}_n - 1$.

b) Funcția de verosimilitate este

$$\begin{aligned}
L(\theta|x_1,\dots,x_n) &= f_\theta(x_1)\cdots f_\theta(x_n) = e^{-(x_1-\theta)}\mathbf{1}_{[\theta,\infty)}(x_1)\cdots e^{-(x_n-\theta)}\mathbf{1}_{[\theta,\infty)}(x_n)\\
&= e^{-\left(\sum_{i=1}^n x_i - n\theta\right)}\mathbf{1}_{[\theta,\infty)}(x_1)\cdots\mathbf{1}_{[\theta,\infty)}(x_n) = e^{-\left(\sum_{i=1}^n x_i - n\theta\right)}\mathbf{1}_{(0,x_1]}(\theta)\cdots\mathbf{1}_{(0,x_n]}(\theta)\\
&= e^{-\left(\sum_{i=1}^n x_i - n\theta\right)}\mathbf{1}_{(0,x_1]\cap\cdots\cap(0,x_n]}(\theta) = e^{n\theta - \sum_{i=1}^n x_i}\mathbf{1}_{(0,x_{(1)}]}(\theta).
\end{aligned}$$

Observăm că funcția de verosimilitate $L(\theta|x_1,\dots,x_n)$ este crescătoare (dar nu derivabilă) iar estimatorul de verosimilitate maximă este

$$\hat{\theta}_n = \arg max_{\theta} L(\theta|X_1,\dots,X_n) = X_{(1)}.$$

c) Pentru a determina legea variabilei $n(\hat{\theta}_n - \theta)$ plecăm de la repartiția estimatorului $\hat{\theta}_n$. Folosind independența variabilelor $X_1, X_2, \dots, X_n$ avem

$$\mathbb{P}_\theta(X_{(1)} > x) = \mathbb{P}_\theta(X_1 > x, X_2 > x, \dots, X_n > x) = \mathbb{P}_\theta(X_1 > x)^n = e^{-n(x-\theta)}, \quad x \geq \theta.$$

Găsim că funcția de repartiție a variabilei $n(\hat{\theta}_n - \theta)$ este

$$\mathbb{P}_\theta\left(n(\hat{\theta}_n - \theta) \leq x\right) = \mathbb{P}_\theta\left(\hat{\theta}_n \leq \theta + \frac{x}{n}\right) = 1 - e^{-n\left(\theta + \frac{x}{n} - \theta\right)} = 1 - e^{-x},$$

care coincide funcția de repartiție a exponențialei $\mathcal{E}(1)$. Astfel deducem că $n(\hat{\theta}_n - \theta) \sim \mathcal{E}(1)$.

d) Plecând de la rezultatul obținut la punctul precendent avem că

$$1 = \mathbb{E}_\theta[n(\hat{\theta}_n - \theta)] = n\mathbb{E}_\theta[\hat{\theta}_n] - n\theta$$

ceea ce conduce la $\mathbb{E}_\theta[\hat{\theta}_n] = \theta + \frac{1}{n}$ de unde conclusionăm că estimatorul $\hat{\theta}_n$ este deplasat.

e) Pentru a calcula eroarea pătratică medie folosim descompunerea $MSE_\theta(\hat{\theta}_n) = Var(\hat{\theta}_n) + b_\theta(\hat{\theta}_n)^2$. Observăm, din punctul d) că abaterea este $b_\theta(\hat{\theta}_n) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta = \frac{1}{n}$. Pentru varianța estimatorului $Var(\hat{\theta}_n)$ folosim tot rezultatul din punctul c) ($Var(\mathcal{E}(1)) = 1$) și avem

$$1 = Var\left(n(\hat{\theta}_n - \theta)\right) = n^2 Var(\hat{\theta}_n)$$

de unde $Var(\hat{\theta}_n) = \frac{1}{n^2}$ și $MSE_\theta(\hat{\theta}_n) = \frac{2}{n^2}$.

f) Pentru a genera trei observații din repartiția $f_\theta(x)$ cu $\theta = 2$ vom folosi *metoda inversă* bazată pe *Teorema de universalitate a repartiției uniforme*. Pentru aceasta trebuie să determinăm funcția cuantilă. Funcția de repartiție este

$$F_\theta(x) = \int_{-\infty}^{x} e^{-(t-\theta)} \mathbf{1}_{[\theta,\infty)}(t)\,dt = e^{\theta}\int_{\theta}^{x} e^{-t}\,dt = e^{\theta}(e^{-\theta} - e^{-x}) = 1 - e^{-\theta x}, \quad x \geq \theta,$$

iar pentru $\theta = 2$ avem $F_{\theta=2}(x) = 1 - e^{-2x}, \quad x \geq 2$. Pentru a determina funcția cunatilă avem $F_2(x) = u$ de unde $F_2^{-1}(u) = -\frac{\log(1-u)}{2}$. Înlocuind cu valorile $u_1 = 0.25$, $u_2 = 0.4$ și $u_3 = 0.5$ repartizate uniform pe $[0,1]$ găsim valorile $x_1 = 0.143841$, $x_2 = 0.2554128$ și $x_3 = 0.3465736$.

Exercițiul 5

Fie $(X_i)_{1\leq i\leq n}$ un eșantion de talie n dintr-o populație de densitate f_θ dată de

$$f_\theta = \theta e^{-\theta x}\mathbf{1}_{[0,+\infty)}(x),$$

cu $\theta > 0$ un parametru necunoscut.

1. Ne propunem să estimăm θ prin $Y_n = \frac{n}{\sum_{i=1}^{n} X_i}$

a) Arătați că estimatorul Y_n este bine definit.
b) Explicați de ce este logic să alegem Y_n ca estimator pentru θ.
c) Determinați legea limită a $\sqrt{n}(Y_n - \theta)$.
d) Determinați legea sumei $\sum_{i=1}^n X_i$ și calculați $\mathbb{E}_\theta[(Y_n - \theta)^2]$.

2. Pentru estimarea lui θ propunem estimatorul $Z_n = \frac{n-1}{n} Y_n$
   a) Verifică Z_n proprietăți similare de convergență cu Y_n?
   b) Pe care dintre cei doi estimatori, Y_n sau Z_n, îl preferați?

1. a) Pentru a arăta că variabila aleatoare Y_n este bine definită trebuie să verificăm că numitorul, $\sum_{i=1}^n X_i$, nu se anulează cu probabilitate 1. Cum X_i sunt absolut continue în raport cu măsura Lebesgue (admit densitatea f_θ) rezultă că $\mathbb{P}(X_i = 0) = 0$ prin urmare $\sum_{i=1}^n X_i > 0$ a.s. de unde reiese concluzia.

b) Observăm că f_θ este densitatea unei exponențiale, prin urmare $X_i \sim \mathcal{E}(\theta)$ și $\mathbb{E}[X_i] = \frac{1}{\theta}$. Din *Legea Numerelor Mari*, aplicată variabilelor X_i (care sunt i.i.d. și integrabile) obținem că $\bar{X}_n \overset{a.s.}{\to} \frac{1}{\theta} > 0$ și aplicând teorema de continuitate găsim că $Y_n = \frac{1}{\bar{X}_n} \overset{a.s.}{\to} \theta$. Această relație arată că Y_n este un estimator consistent pentru θ deci un estimator rezonabil.

c) Cum $Var(X_1) = \frac{1}{\theta^2} < \infty$ din *Teorema Limită Centrală* avem

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\theta}\right) \overset{d}{\to} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

Aplicând metoda delta cu $g(x) = \frac{1}{x}$, care este derivabilă pe $\mathbb{R}_+^*$ cu derivata în punctul $\frac{1}{\theta}$ egală cu $g'(1/\theta) = -\theta^2$, găsim

$$\sqrt{n}(Y_n - \theta) \overset{d}{\to} \mathcal{N}\left(0, \frac{(-\theta^2)^2}{\theta^2}\right) = \mathcal{N}(0, \theta^2).$$

d) Cum variabilele aleatoare X_i sunt repartizate exponențial de parametru θ și $\mathcal{E}(\theta) = \Gamma(1, \theta)$ deducem că[1] $\sum_{i=1}^n X_i \sim \Gamma(n, \theta)$. Pentru a calcula eroarea medie pătratică vom folosi relația:

$$MSE_\theta(Y_n) = \mathbb{E}[(Y_n - \theta)^2] = Var(Y_n) + b_\theta(Y_n)^2.$$

Pentru calculul $Var(Y_n)$ avem nevoie de $\mathbb{E}[Y_n]$ și $\mathbb{E}[Y_n^2]$. Pentru $\mathbb{E}[Y_n]$ avem

$$\begin{aligned}\mathbb{E}[Y_n] &= \mathbb{E}\left[\frac{n}{\sum_{i=1}^n X_i}\right] = n\mathbb{E}\left[\frac{1}{\sum_{i=1}^n X_i}\right] = n\int_0^\infty \frac{1}{x}\frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}\, dx \\ &\overset{u=\theta x}{=} \frac{n\theta}{\Gamma(n)}\int_0^\infty u^{n-2} e^{-u}\, du = \frac{n\theta}{\Gamma(n)}\Gamma(n-1) = \frac{n\theta}{n-1}\end{aligned}$$

iar pentru $\mathbb{E}[Y_n^2]$ și $Var(Y_n)$ obținem

$$\begin{aligned}\mathbb{E}[Y_n^2] &= \frac{n^2\theta^2}{\Gamma(n)}\Gamma(n-2) = \frac{n^2\theta^2}{(n-1)(n-2)}, \\ Var(Y_n) &= \mathbb{E}[Y_n^2] = \mathbb{E}[Y_n] = \frac{n^2\theta^2}{(n-1)^2(n-2)}.\end{aligned}$$

[1] A se vedea Exercițiul 5 din Tema 2 pentru o demonstrație a acestui rezultat.

Combinând rezultatele de mai sus deducem că

$$\begin{aligned} MSE_\theta(Y_n) &= Var(Y_n) + (\mathbb{E}[Y_n] - \theta)^2 = \frac{n^2\theta^2}{(n-1)^2(n-2)} + \left(\frac{n\theta}{n-1} - \theta\right)^2 \\ &= \frac{\theta^2(n^2+n-2)}{(n-1)^2(n-2)}. \end{aligned}$$

2. a) Proprietățile asimptotice ale lui Z_n sunt similare cu cele ale lui Y_n: cum $\frac{n-1}{n} \to 1$ și $Y_n \overset{a.s.}{\to} \theta$ deducem că $Z_n \overset{a.s.}{\to} \theta$. În plus, cum $\sqrt{n}(Y_n - \theta) \overset{d}{\to} \mathcal{N}(0, \theta^2)$ avem că

$$\sqrt{n}(Z_n - \theta) = \sqrt{n}(Y_n - \theta) - \sqrt{n}(Y_n - Z_n) = \sqrt{n}(Y_n - \theta) - \frac{Y_n}{\sqrt{n}} \overset{d}{\to} \mathcal{N}(0, \theta^2)$$

unde în ultima relație am folosit *Teorema lui Slutsky* și faptul că $\frac{Y_n}{\sqrt{n}}$ converge în probabilitate la 0.

b) Pentru a compara cei doi estimatori calculăm eroarea medie pătratică a lui Z_n și avem

$$\begin{aligned} \mathbb{E}[Z_n] &= \frac{n-1}{n}\mathbb{E}[Y_n] = \theta \quad \text{(estimator nedeplasat)} \\ Var(Z_n) &= \left(\frac{n-1}{n}\right)^2 Var(Y_n) = \frac{\theta^2}{n-2} \\ MSE_\theta(Z_n) &= Var(Z_n) + 0 = \frac{\theta^2}{n-2}. \end{aligned}$$

Se poate arăta cu ușurință că eroarea medie pătratică a estimatorului Z_n este inferioară erorii medii pătratice a lui Y_n prin urmare estimatorul Z_n este de preferat.

Exercițiul 6

Fie $(X_i)_{1\leq i\leq n}$ un eșantion de talie n dintr-o populație de densitate

$$f_\theta(x) = \frac{2}{\theta^2} x \mathbf{1}_{[0,\theta]}(x),$$

cu $\theta > 0$ un parametru necunoscut.

1. Considerăm $X_{(n)}$ statistica de ordine de rang n
   a) Determinați densitatea lui $X_{(n)}$.
   b) Calculați media și varianța lui $X_{(n)}$.
   c) Arătați că $X_{(n)}$ converge în probabilitate la θ.
2. Studiați convergența lui $\bar{X}_n$ și determinați un estimator consistent pentru θ.
3. Pe care dintre estimatorii $X_{(n)}$ și $\frac{2}{3}\bar{X}_n$ îl preferați pentru a estima pe θ?

1. a) Observăm că pentru toate valorile $x \in \mathbb{R}$ avem

$$\mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_1 \leq x)^n = \begin{cases} 0, & x < 0 \\ \left(\frac{x}{\theta}\right)^{2n}, & 0 \leq x \leq \theta \\ 1 & x > \theta \end{cases}$$

Astfel funcția de repartiție a lui $X_{(n)}$ este continuă și de clasă C^1 pe porțiuni. Densitatea variabilei aleatoare $X_{(n)}$ se poate calcula derivând funcția de repartiție de unde

$$f_{X_{(n)}}(x) = \frac{2nx^{2n-1}}{\theta^{2n}}\mathbb{1}_{[0,\theta]}(x).$$

b) Avem

$$\mathbb{E}[X_{(n)}] = \int_0^\theta x\frac{2nx^{2n-1}}{\theta^{2n}}\,dx = \frac{2n}{2n+1}\theta,$$
$$\mathbb{E}[X_{(n)}^2] = \int_0^\theta x^2\frac{2nx^{2n-1}}{\theta^{2n}}\,dx = \frac{2n}{2n+2}\theta^2$$
$$Var(X_{(n)}) = \frac{2n}{2n+2}\theta^2 - \left(\frac{2n}{2n+1}\theta\right)^2 = \frac{n}{(n+1)(2n+1)^2}\theta^2.$$

c) Cum variabila aleatoare $X_{(n)}$ ia valori în inervalul $[0,\theta]$ avem că $X_{(n)} \leq \theta$ a.s.. Prin urmare pentru $\epsilon > 0$ avem

$$\mathbb{P}(|X_{(n)} - \theta| > \epsilon) = \mathbb{P}(\theta - X_{(n)} > \epsilon) = \mathbb{P}(X_{(n)} < \theta - \epsilon) = \begin{cases} 0, & \epsilon > \theta \\ \frac{(\theta-\epsilon)^{2n}}{\theta^{2n}}, & \epsilon \leq \theta \end{cases}$$

deci $\mathbb{P}(|X_{(n)} - \theta| > \epsilon) \to 0$ pentru orice $\epsilon > 0$, ceea ce arată că $X_{(n)} \xrightarrow{\mathbb{P}} \theta$.

2. Cum variabilele aleatoare X_i sunt mărginite ele sunt și integrabile iar media lor comună este

$$\mathbb{E}[X_1] = \frac{2}{\theta^2}\int_0^\theta x^2\,dx = \frac{2\theta}{3}.$$

Aplicând *Legea Numerelor Mari* deducem că $\bar{X}_n \xrightarrow{a.s.} \frac{2\theta}{3}$, prin urmare $\frac{3\bar{X}_n}{2} \xrightarrow{a.s.} \theta$ deci $\frac{3\bar{X}_n}{2}$ este un estimator consistent pentru θ.

3. Pentru a compara cei doi estimatori, $X_{(n)}$ și $\frac{3\bar{X}_n}{2}$, trebuie să calculăm erorile lor medii pătratice. Pentru estimatorul $\frac{3\bar{X}_n}{2}$ avem

$$\mathbb{E}\left[\frac{3\bar{X}_n}{2}\right] = \frac{3}{2}\mathbb{E}[X_1] = \theta$$
$$Var\left(\frac{3\bar{X}_n}{2}\right) = \frac{9}{4}Var(\bar{X}_n) = \frac{9}{4n}Var(X_1) = \frac{9}{4n}\left(\frac{2}{\theta^2}\int_0^\theta x^3\,dx - \left(\frac{2\theta}{3}\right)^2\right)$$
$$= \frac{9}{4n}\left(\frac{2\theta^4}{4\theta^2} - \frac{4\theta^2}{9}\right) = \frac{\theta^2}{8n},$$

prin urmare eroarea medie pătratică este $MSE_\theta(\frac{3\bar{X}_n}{2}) = \frac{\theta^2}{8n}$.

În mod similar, eroarea medie pătratică pentru estimatorul $X_{(n)}$ este

$$MSE_\theta(X_{(n)}) = (\mathbb{E}[X_{(n)}] - \theta)^2 + Var(X_{(n)}) = \left(\frac{2n}{2n+1}\theta - \theta\right)^2 + \frac{n\theta^2}{(n+1)(2n+1)^2}$$
$$= \frac{\theta^2}{(n+1)(2n+1)}.$$

Pentru $\theta > 0$ avem că

$$\frac{MSE_{\theta}(X_{(n)})}{MSE_{\theta}\left(\frac{3\bar{X}_n}{2}\right)} = \frac{8n}{(n+1)(2n+1)}$$

ceea ce implică faptul că estimatorul $X_{(n)}$ este preferabil estimatorului $\frac{3\bar{X}_n}{2}$ în raport cu eroarea pătratică medie pentru $n \geq 3$.