

Tema 3

Soluții

Exercițiul 1



- a) Nivelul de zgomot al unei mașini de spălat este o v.a. de medie 44 dB și de abatere standard 5 dB. Admițând aproximarea normală care este probabilitatea să găsim o medie a zgomotului superioară la 48 dB într-un eșantion de talie 10 mașini de spălat ?
- b) O telecabină are o capacitate de 100 de persoane. Știind că greutatea populației (țării) este o v.a. de medie 66.3 Kg și o abatere standard de 15.6 Kg și presupunând că persoanele care au urcat în telecabină au fost alese în mod aleator din populație, care este probabilitatea ca greutatea totală acestora să depășească 7000 Kg ?

- a) Fie X nivelul de zgomot produs de o mașină de spălat luată la intamplare și \bar{X}_{10} media unui eșantion de talie 10. Presupunem că aproximarea gaussiană are loc pentru $n = 10$. Avem

$$\bar{X}_{10} \sim \mathcal{N}\left(44, \frac{5^2}{10}\right),$$

de unde $\mathbb{P}(\bar{X}_{10} > 48) \simeq \mathbb{P}\left(Z > \frac{48-44}{5/\sqrt{10}}\right) = \mathbb{P}(Z > 2.53) = 1 - 0.9943 = 0.0057$, unde $Z \sim \mathcal{N}(0, 1)$. Observăm că această probabilitate este foarte mică.

- b) Fie X greutatea unei persoane luate la intamplare și \bar{X}_{100} greutatea media a unui eșantion de 100 de persoane. Aplicând aproximarea gaussiană (Teorema Limită Centrală) avem

$$\bar{X}_{100} \simeq \mathcal{N}\left(66.3, \frac{15.6^2}{100}\right),$$

de unde $\mathbb{P}(\bar{X}_{100} > \frac{7000}{100}) \simeq \mathbb{P}\left(Z > \frac{70-66.3}{15.6/\sqrt{100}}\right) = \mathbb{P}(Z > 2.37) = 0.0089$, unde $Z \sim \mathcal{N}(0, 1)$.

Exercițiul 2



Fie X_1, \dots, X_n un eșantion de talie n dintr-o populație de medie μ și varianță σ^2 . Arătați că varianța varianței eșantionului este:

$$\mathbb{V}(S^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right)$$

unde $\mu_4 = \mathbb{E}[(X_i - \mu)^4]$ este momentul centrat de ordin 4. Ce revine această formulă în cazul Gaussian (normal) ?

Am văzut la curs că

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - \frac{1}{n} \left[\sum_{i=1}^n (X_i - \mu) \right]^2.$$

Dacă notăm cu $Z_i = X_i - \mu$ atunci observăm că v.a. Z_i sunt i.i.d. iar $\mathbb{E}[Z_i] = 0$, $\mathbb{E}[Z_i^2] = \sigma^2$ și $\mathbb{E}[Z_i^4] = \mu_4$. Avem că

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^n (Z_i)^2 - \frac{1}{n} \left(\sum_{i=1}^n Z_i \right)^2 = \sum_{i=1}^n (Z_i)^2 - \frac{1}{n} \left(\sum_{i=1}^n Z_i^2 + 2 \sum_{i < j} Z_i Z_j \right)^2 \\ &= \frac{n-1}{n} \sum_{i=1}^n (Z_i)^2 - \frac{2}{n} \sum_{i < j} Z_i Z_j\end{aligned}$$

de unde obținem

$$\begin{aligned}(n-1)^2 \mathbb{E}[(S^2)^2] &= \mathbb{E} \left[\left(\frac{n-1}{n} \sum_{i=1}^n (Z_i)^2 - \frac{2}{n} \sum_{i < j} Z_i Z_j \right)^2 \right] \\ &= \left(\frac{n-1}{n} \right)^2 \mathbb{E} \left[\sum_{i=1}^n Z_i^4 + 2 \sum_{i < j} Z_i^2 Z_j^2 \right] - \frac{4(n-1)}{n^2} \mathbb{E} \left[\left(\sum_{k=1}^n Z_k^2 \right) \left(\sum_{i < j} Z_i Z_j \right) \right] \\ &\quad + \frac{4}{n^2} \mathbb{E} \left[\left(\sum_{i < j} Z_i Z_j \right)^2 \right] \tag{*}\end{aligned}$$

Pentru primul termen din suma de mai sus avem

$$\left(\frac{n-1}{n} \right)^2 \mathbb{E} \left[\sum_{i=1}^n Z_i^4 + 2 \sum_{i < j} Z_i^2 Z_j^2 \right] = \left(\frac{n-1}{n} \right)^2 (n\mu_4 + n(n-1)\sigma^4).$$

Termenul al doilea din ecuația (*) este 0 deoarece conține sau termeni de forma $\mathbb{E}[Z_i Z_j Z_k^2]$, cu $i \neq j \neq k$, sau termeni de forma $\mathbb{E}[Z_j Z_k^3]$ cu $j \neq k$.

Pentru ultimul termen avem din ecuația (*) avem

$$\frac{4}{n^2} \mathbb{E} \left[\left(\sum_{i < j} Z_i Z_j \right)^2 \right] = \frac{4}{n^2} \mathbb{E} \left[\sum_{i < j} Z_i^2 Z_j^2 \right] = \frac{2(n-1)}{n} \sigma^4,$$

restul termenilor fiind zero deoarece sunt de forma $\mathbb{E}[Z_i^2 Z_j Z_k]$ sau $\mathbb{E}[Z_i Z_j Z_k Z_l]$ cu $i \neq j \neq k \neq l$.

Combinand rezultatele obținem că

$$\begin{aligned}(n-1)^2 \mathbb{V}[S^2] &= \frac{(n-1)^2}{n} \mu_4 + \frac{(n-1)^3}{n} \sigma^4 + 2 \frac{n-1}{n} \sigma^4 - (n-1)^2 \mathbb{E}[S^2]^2 \\ &= \frac{(n-1)^2}{n} \mu_4 + \frac{(n-1)(3-n)}{n} \sigma^4\end{aligned}$$

prin urmare $\mathbb{V}[S^2] = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right)$.

În cazul normal avem că $\mu_4 = 3\sigma^4$ (de ce ?) deci $\mathbb{V}[S^2] = \frac{2\sigma^4}{n-1}$ (vedeți legea χ^2).

Exercițiul 3



Fie X_1, \dots, X_n un eșantion de talie n dintr-o populație de medie μ și varianță σ^2 . Arătați că

$$\text{Cov}(\bar{X}, S^2) = \frac{\mu_3}{n}$$

unde $\mu_3 = \mathbb{E}[(X_i - \mu)^3]$ este momentul centrat de ordin 3. Acest rezultat ne arată că cele două statistici sunt asimptotic *necorelate*.

Dacă notăm cu $Z_i = X_i - \mu$, atunci $\bar{X} - \mu = \bar{Z}$ și $\mathbb{E}[\bar{Z}] = 0$. Mai mult,

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2$$

prin urmare

$$\begin{aligned} \text{Cov}(\bar{X}, S^2) &= \text{Cov}(\bar{X} - \mu, S^2) = \text{Cov}\left(\bar{Z}, \frac{1}{n-1} \left[\sum_{i=1}^n Z_i^2 - n\bar{Z}^2 \right]\right) \\ &= \frac{1}{n-1} \mathbb{E}\left[\bar{Z} \left(\sum_{i=1}^n Z_i^2 - n\bar{Z}^2 \right)\right] = \frac{1}{n-1} \left[\mathbb{E}\left[\bar{Z} \left(\sum_{i=1}^n Z_i^2 \right)\right] - n\mathbb{E}[\bar{Z}^3] \right] \end{aligned}$$

Cum

$$\mathbb{E}\left[\bar{Z} \left(\sum_{i=1}^n Z_i^2 \right)\right] = \frac{1}{n} \mathbb{E}\left[\left(\sum_{j=1}^n Z_j \right) \left(\sum_{i=1}^n Z_i^2 \right)\right] = \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^n Z_i^3 \right] = \mu_3$$

și

$$\mathbb{E}[\bar{Z}^3] = \frac{1}{n^3} \mathbb{E}\left[\left(\sum_{i=1}^n Z_i \right) \left(\sum_{j=1}^n Z_j \right) \left(\sum_{k=1}^n Z_k \right)\right] = \frac{1}{n^3} \mathbb{E}\left[\sum_{i=1}^n Z_i^3 \right] = \frac{\mu_3}{n^2}$$

rezultă că $\text{Cov}(\bar{X}, S^2) = \frac{1}{n-1} \left(\mu_3 - \frac{\mu_3}{n} \right) = \frac{\mu_3}{n}$.

Exercițiul 4



Fie X_1, X_2, \dots, X_n un eșantion de talie n dintr-o populație F cu $\mathbb{E}[X_1^2] < \infty$.

1. a) Arătați că $\mathbb{E}[X_1] = \arg \min_{t \in \mathbb{R}} \mathbb{E}[(X_1 - t)^2]$.

b) Determinați $\arg \min_{t \in \mathbb{R}} \sum_{i=1}^n \frac{(X_i - t)^2}{n}$.

2. Notăm cu $x_{\frac{1}{2}} = F^{-1}\left(\frac{1}{2}\right)$ mediana repartiției lui X_1

a) Arătați că dacă F este continuă pe \mathbb{R} și strict crescătoare pe o vecinătate a lui $x_{\frac{1}{2}}$ atunci

$$x_{\frac{1}{2}} = \arg \min_{t \in \mathbb{R}} \mathbb{E}[|X_1 - t|].$$

b) Determinați, în funcție de paritatea lui n , $\arg \min_{t \in \mathbb{R}} \sum_{i=1}^n \frac{|X_i - t|}{n}$.

- a) Dacă definim $f : t \in \mathbb{R} \rightarrow \mathbb{E}[(X_1 - t)^2]$, atunci $f(t) = t^2 - 2\mathbb{E}[X_1]t + \mathbb{E}[X_1^2]$ este o funcție de gradul doi strict convexă. Astfel are sens să vorbim de minimul (unic) acesteia care se determină rezolvând ecuația $f'(t) = 2t - 2\mathbb{E}[X_1] = 0 \iff t = \mathbb{E}[X_1]$.

b) În mod similar, funcția $g(t) = t^2 - 2\frac{\sum_{i=1}^n X_i}{n}t + \frac{\sum_{i=1}^n X_i^2}{n}$ admite un unic minim în punctul critic dat de $t = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}_n$.
- În acest exercițiu vom folosi o metodă diferită de cea din Exercițiul 4 din Tema 1 de a arăta rezultatul de la punctul a).

Reamintim (a se vedea Exercițiul 6 din Tema 1) că dacă $Z \geq 0$ atunci

$$\mathbb{E}[Z] = \int_0^\infty \mathbb{P}(Z > z) dz. \quad (\star)$$

a) Să observăm pentru început că $|X_1 - t| = (X_1 - t)_+ + (t - X_1)_+$, prin urmare

$$\begin{aligned} \mathbb{P}(|X_1 - t| > z) &= \mathbb{P}((X_1 - t)_+ + (t - X_1)_+ > z) \\ &= \mathbb{P}((X_1 - t)_+ + 0 > z \text{ și } X_1 \geq t) + \mathbb{P}(0 + (t - X_1)_+ > z \text{ și } X_1 < t) \\ &= \mathbb{P}(X_1 - t > z) + \mathbb{P}(t - X_1 > z). \end{aligned}$$

Folosind relația din (\star) avem

$$\begin{aligned} \mathbb{E}[|X_1 - t|] &= \int_0^\infty \mathbb{P}(|X_1 - t| > z) dz \\ &= \int_0^\infty \mathbb{P}(X_1 - t > z) dz + \int_0^\infty \mathbb{P}(t - X_1 > z) dz \\ &= \int_0^\infty (1 - F(z + t)) dz + \int_0^\infty F(t - z) dz \\ &= \int_t^\infty (1 - F(u)) du + \int_{-\infty}^t F(u) du. \end{aligned}$$

Avem că (folosim că F este continuă)

$$\frac{\partial}{\partial t} \mathbb{E}[|X_1 - t|] = -(1 - F(t)) + F(t) = 2F(t) - 1$$

și cum F este o bijecție strict crescătoare într-o vecinătate a lui $x_{\frac{1}{2}}$ deducem că

$$2F(t) - 1 > 0 \iff t > x_{\frac{1}{2}} \text{ și } 2F(t) - 1 < 0 \iff t < x_{\frac{1}{2}}$$

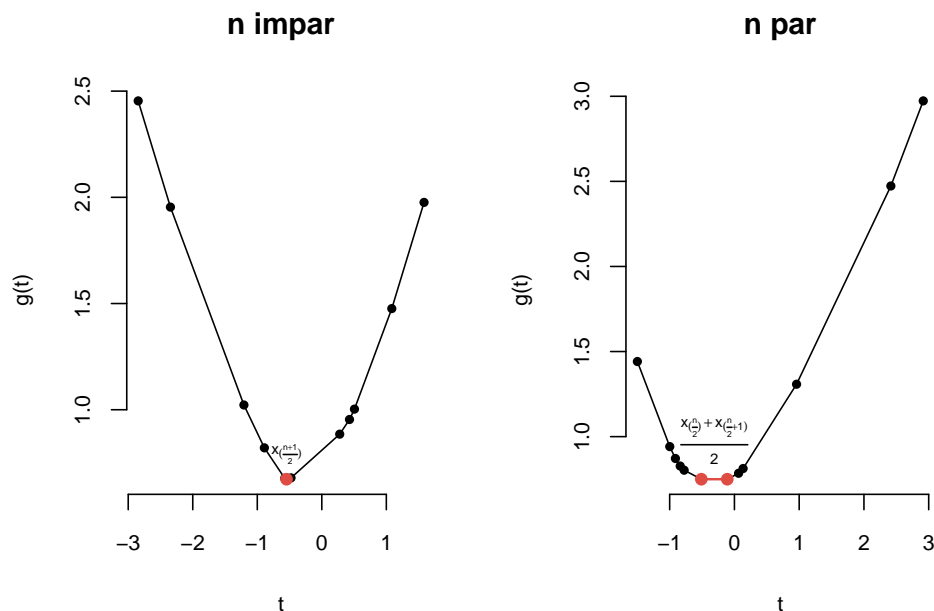
de unde funcția $t \rightarrow \mathbb{E}[|X_1 - t|]$ își atinge minimul în $t = x_{\frac{1}{2}}$.

- b) Funcția $g(t) = \sum_{i=1}^n \frac{|X_i - t|}{n}$ este liniară pe porțiuni iar panta sa se schimbă în fiecare X_i . Fie $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ statisticile de ordine de rang $1, 2, \dots, n$ și să notăm cu $X_{(0)} = -\infty$ și $X_{(n+1)} = \infty$. Putem observa că funcția g are panta $-\frac{n}{n}$ pe intervalul $(-\infty, X_{(1)})$, $-\frac{n-2}{n}$ pe intervalul $[X_{(1)}, X_{(2)})$, etc. și apoi panta $\frac{n-2}{n}$ pe intervalul $[X_{(n-1)}, X_{(n)})$ și $\frac{n}{n}$ pe intervalul $[X_{(n)}, \infty)$. Altfel spus, panta este $\frac{2(k-1)-n}{n}$ pe intervalul $[X_{(k-1)}, X_{(k)})$.

Distingem două cazuri în funcție de paritatea lui n :

- dacă n este par, panta lui g se va anula pe un interval care corespunde la $[X_{(\frac{n}{2})}, X_{(\frac{n}{2}+1)})$ prin urmare g este minimă pe acest interval (în practică se ia mijlocul acestui interval)
- dacă n este impar atunci panta lui g nu se anulează și în acest caz funcția atinge valoarea minimă în $X_{(\frac{n+1}{2})}$

Astfel observăm că valoarea care minimizează funcția $\sum_{i=1}^n \frac{|X_i - t|}{n}$ este chiar mediana empirică.



Exercițiul 5



X_1, \dots, X_n un eșantion de talie n cu funcția de repartiție $F(x)$ și densitatea $f(x)$ și (Y_1, \dots, Y_n) versiunea ordonată crescător a acestuia. Notăm cu $H_k(x)$ și $h_k(x)$ funcția de repartiție și densitatea v.a. Y_k . Fie $Y_1 = \inf X_i$ și $Y_n = \sup X_i$.

- Care este funcția de repartiție și densitatea lui Y_1 și Y_n ?
- Care este probabilitatea ca o observație dintr-o v.a. de lege $\mathcal{N}(\mu, \sigma^2)$ să depășească $\mu + 3\sigma$?
- Dar într-un eșantion de talie 100 cat este această probabilitate (i.e. probabilitatea ca o observație să depășească $\mu + 3\sigma$)?
- Dintr-un eșantion de talie 100 dintr-o populație repartizată $\mathcal{N}(0, 1)$ ce valoare nu poate fi depășită cu o probabilitate de 99% ?

- e) O societate de analiză a calității apei și a mediului efectuează un sondaj în laboratoarele sale (50 la număr, repartizate pe tot teritoriul României) pentru a testa dacă efectuează măsurători corecte. Pentru aceasta, serviciul de calitate trimite la fiecare laborator un eșantion de apă care conține o anumită concentrație de crom și le cere să determine această concentrație de crom. Ținând cont de fluctuațiile care apar în prepararea soluției, precum și de imprecizia aparatelor de măsură, societatea presupune că repartiția concentrației de crom (mg/l) este $\mathcal{N}(10, 1)$.

Printre rezultatele obținute de la laboratoare, două dintre acestea au înregistrat măsurători mai diferite decât celelalte: laboratorul L_1 a înregistrat o concentrație de 6 mg/l (cea mai mică valoare înregistrată) iar laboratorul L_2 a măsurat o concentrație de 13 mg/l (cea mai mare dintre măsurători).

Puteți spune, cu o probabilitate de 99%, că aceste valori sunt coerente sau că valorile obținute sunt aberante (datorită erorilor de măsurare, de calibrare a aparatelor, etc.) ?

- a) Se observă cu ușurință că

$$H_n(x) = \mathbb{P}(Y_n \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \stackrel{\text{indep.}}{=} F(x)^n$$

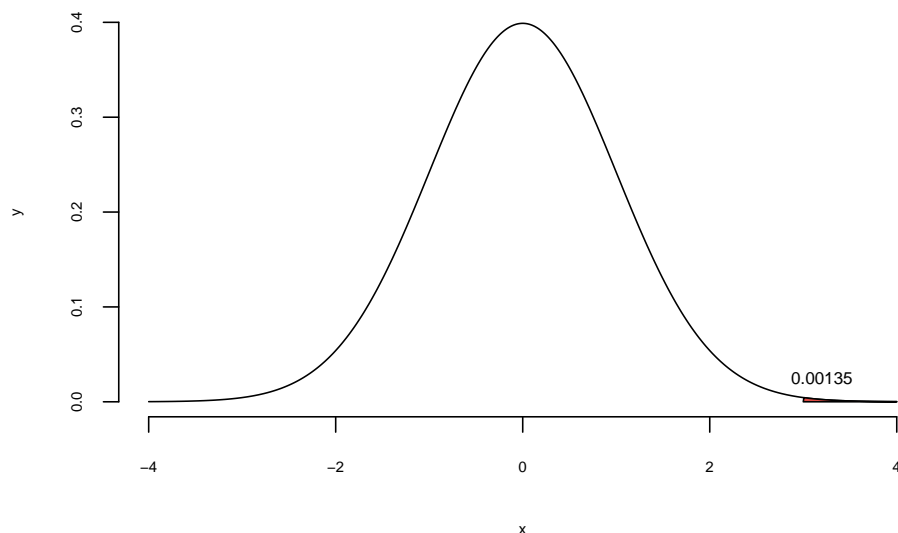
$$h_n(x) = \frac{d}{dx} H_n(x) = n f(x) F(x)^{n-1}$$

$$H_1(x) = \mathbb{P}(Y_1 \leq x) = 1 - \mathbb{P}(Y_1 > x) = 1 - \mathbb{P}(X_1 > x, \dots, X_n > x) \stackrel{\text{indep.}}{=} 1 - (1 - F(x))^n$$

$$h_1(x) = \frac{d}{dx} H_1(x) = n f(x) (1 - F(x))^{n-1}$$

- b) Fie $X \sim \mathcal{N}(\mu, \sigma^2)$. Problema cere să găsim probabilitatea $\mathbb{P}(X > \mu + 3\sigma)$. Avem (vezi porțiunea roșie din figură)

$$\mathbb{P}(X > \mu + 3\sigma) = \mathbb{P}\left(\frac{X - \mu}{\sigma} > 3\right) = 1 - \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq 3\right) = 0.00135$$



- c) Fie X_1, X_2, \dots, X_n un eșantion de talie $n = 100$ dintr-o populație normală $\mathcal{N}(\mu, \sigma^2)$ și fie $Z_i = \mathbf{1}_{\{X_i > \mu + 3\sigma\}}$ variabilele Bernoulli care iau valoarea 1 atunci când $X_i > \mu + 3\sigma$ și 0 în rest. Problema revine la a determina probabilitatea

$$\mathbb{P}(Z_1 + \dots + Z_n = 1) \stackrel{i.i.d.}{=} \binom{n}{1} \mathbb{P}(Z_1 = 1) \mathbb{P}(Z_1 = 0)^{n-1} = n \mathbb{P}(X_1 > \mu + 3\sigma) \mathbb{P}(X_1 < \mu + 3\sigma)^{n-1} \simeq 0.11809$$

- d) Fie X_1, X_2, \dots, X_n un eșantion de talie $n = 100$ dintr-o populație normală $\mathcal{N}(0, 1)$. Problema ne cere să găsim valoarea lui x pentru care probabilitatea $\mathbb{P}(X_1 < x, X_2 < x, \dots, X_n < x) = 0.99$. Prin urmare vrem să găsim pe x așa încât $H_n(x) = 0.99$. Din punctul a) avem $H_n(x) = F(x)^n$ deci $x = F^{-1}(\sqrt[n]{0.99}) = 3.7177$.

- e) Fie X_1, X_2, \dots, X_n un eșantion de talie $n = 50$ dintr-o populație normală $\mathcal{N}(10, 1)$ ($n = 50$ reprezintă numărul de laboratoare iar X_i este concentrația de crom din laboratorul i). Din datele problemei avem că laboratorul 1 a înregistrat cea mai mică valoare (6 mg/l) iar laboratorul 2 a înregistrat cea mai mare valoare (13 mg/l). Problema ne cere să evaluăm probabilitatea

$$\mathbb{P}(Y_1 \leq 6, Y_n \geq 13) = 1 - \mathbb{P}(\{Y_1 > 6\} \cup \{Y_n < 13\}) = 1 - \mathbb{P}(Y_1 > 6) - \mathbb{P}(Y_n < 13) + \mathbb{P}(Y_1 > 6, Y_n < 13).$$

$$\text{Avem că } \mathbb{P}(Y_1 > 6) = \mathbb{P}(X_1 > 6, \dots, X_n > 6) = (1 - F(6))^n \text{ iar } F(6) = \mathbb{P}(X_1 \leq 6) = \mathbb{P}\left(\frac{X_1 - 10}{1} \leq -4\right) \simeq 0.00003 \text{ deci } \mathbb{P}(Y_1 > 6) \simeq 0.99871.$$

De asemenea $\mathbb{P}(Y_n < 13) = F(13)^n$ iar cum $F(13) = \mathbb{P}(X_1 \leq 13) = \mathbb{P}\left(\frac{X_1 - 10}{1} \leq 3\right) \simeq 0.9986$ rezultă că $\mathbb{P}(Y_n < 13) \simeq 0.9346$.

În mod similar, $\mathbb{P}(Y_1 > 6, Y_n < 13) = \mathbb{P}(6 < X_1 < 13, \dots, 6 < X_n < 13) = \mathbb{P}(6 < X_1 < 13)^n$ și cum $\mathbb{P}(6 < X_1 < 13) = \mathbb{P}(X_1 < 13) - \mathbb{P}(X_1 \leq 6) \simeq 0.9986$ obținem că $\mathbb{P}(Y_1 > 6, Y_n < 13) \simeq 0.9332$.

În concluzie avem că $\mathbb{P}(Y_1 \leq 6, Y_n \geq 13) \simeq 0.0001$.

Exercițiul 6



Fie X_1, X_2, \dots, X_n un eșantion de talie n dintr-o populație $\mathcal{U}([0, \theta])$ cu $\theta > 0$ necunoscut.

- Fie $\hat{\theta}_n = \max\{X_1, \dots, X_n\}$. Determinați funcția de repartiție a lui $\hat{\theta}_n$.
- Arătați că $\hat{\theta}_n$ este un estimator consistent pentru θ .
- Arătați că $\hat{\theta}_n$ nu este un estimator nedeplasat pentru θ și construiți un asemenea estimator.

- a) Observăm că funcția de repartiție pentru $X \sim \mathcal{U}(0, \theta)$ este $F_\theta(x) = \frac{x}{\theta}$ dacă $x \in (0, \theta)$ și $F_\theta(x) = 0$ altfel. Cum X_1, X_2, \dots, X_n sunt i.i.d. $\mathcal{U}(0, \theta)$, funcția de repartiție pentru $\hat{\theta}_n = \max\{X_1, \dots, X_n\}$ este

$$F_{\hat{\theta}_n}(x) = \mathbb{P}_\theta(\hat{\theta}_n \leq x) = \mathbb{P}_\theta(X_1 \leq x, \dots, X_n \leq x) = (\mathbb{P}_\theta(X_1 \leq x))^n = \left(\frac{x}{\theta}\right)^n, \quad x \in (0, \theta).$$

- b) Pentru a arăta că $\hat{\theta}_n$ este consistent pentru θ trebuie verificat că $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$. Putem remarca că $\theta \geq \hat{\theta}_n$ deoarece fiecare X_i este strict mai mic decât θ . Pentru $\varepsilon > 0$, avem

$$\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = \mathbb{P}_\theta(\theta - \hat{\theta}_n > \varepsilon) = \mathbb{P}_\theta(\hat{\theta}_n \leq \theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

Dacă $\varepsilon < \theta$ atunci membrul drept converge la 0 pentru $n \rightarrow \infty$ de unde obținem concluzia. În caz că $\varepsilon > \theta$ atunci membrul drept este egal cu 0 de unde și limita.

- c) Pentru a verifica dacă estimatorul $\hat{\theta}_n$ este deplasat trebuie să calculăm $\mathbb{E}_\theta[\hat{\theta}_n]$. Cum funcția de repartiție a lui $\hat{\theta}_n$ este $F_{\hat{\theta}_n}(x) = \left(\frac{x}{\theta}\right)^n$ putem găsi cu ușurință că densitatea este $f_{\hat{\theta}_n}(x) = n\frac{x^{n-1}}{\theta^n}$ pentru $x \in (0, \theta)$ și 0 altfel. Prin urmare

$$\mathbb{E}_\theta[\hat{\theta}_n] = \int_0^\theta x f_{\hat{\theta}_n}(x) dx = n \int_0^\theta \left(\frac{x}{\theta}\right)^n dx \stackrel{y=x/\theta}{=} n\theta \int_0^1 y^n dy = \frac{n\theta}{n+1}.$$

Cum $\mathbb{E}_\theta[\hat{\theta}_n] \neq \theta$ concluzionăm că estimatorul este deplasat. Dacă definim $\tilde{\theta}_n = \frac{n}{n+1}\theta$, atunci se observă că $\tilde{\theta}_n$ este nedepășat și cum $\hat{\theta}_n$ era consistent iar $\frac{n}{n+1}$ converge la 1 deducem că $\tilde{\theta}_n$ este un estimator consistent.

Exercițiul 7



Fie $X \sim B(10, \theta)$ cu $\theta \in (0, 1)$ necunoscut. Fie $\hat{\theta}_1 = \frac{X}{10}$ și $\hat{\theta}_2 = \frac{X+1}{12}$ doi estimatori pentru θ .

- Calculați $\mathbb{E}_\theta[\hat{\theta}_1]$ și $\mathbb{E}_\theta[\hat{\theta}_2]$.
- Calculați erorile medii pătratice: $MSE_\theta(\hat{\theta}_1)$ și $MSE_\theta(\hat{\theta}_2)$.
- Trasați pe același grafic erorile medii pătratice ale celor doi estimatori ca funcții de θ . Pe care dintre cei doi estimatori îl preferați?

a) Cum $\mathbb{E}_\theta[X] = 10\theta$ obținem că $\mathbb{E}_\theta[\hat{\theta}_1] = \theta$ și $\mathbb{E}_\theta[\hat{\theta}_2] = \frac{10\theta+1}{12}$.

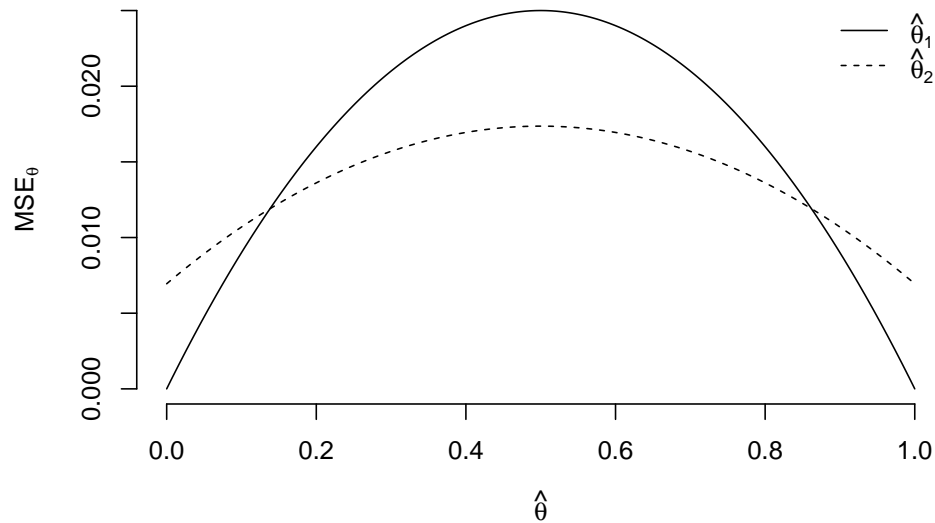
b) Pentru calculul erorii medii pătratice vom folosi următoarea formulă $MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta}) + B_\theta(\hat{\theta})^2$. Cum $\hat{\theta}_1$ este un estimator nedepășat rezultă că $B_\theta(\hat{\theta}_1) = 0$ și

$$MSE_\theta(\hat{\theta}_1) = Var_\theta(\hat{\theta}_1) = 10^{-2}Var_\theta(X) = \frac{\theta(1-\theta)}{10}.$$

Pentru $\hat{\theta}_2$ avem $B_\theta(\hat{\theta}_2) = \frac{10\theta+1}{12} - \theta$ de unde

$$MSE_\theta(\hat{\theta}_2) = \frac{Var_\theta(X)}{12^2} + \left(\frac{10\theta+1}{12} - \theta\right)^2 = \frac{6\theta - 6\theta^2 + 1}{144}.$$

c) Avem următoarea figură:



Chiar dacă $\hat{\theta}_1$ este nedeplasat și $\hat{\theta}_2$ este deplasat, niciunul dintre cei doi estimatori nu are eroarea medie pătratică uniform mai mică. Cu toate acestea, eroarea medie pătratică pentru estimatorul $\hat{\theta}_2$ este mai mică decât cea pentru estimatorul $\hat{\theta}_1$ pe aproape toată plaja de valori a lui θ (mai exact pe intervalul $\theta \in \left[\frac{1-\sqrt{\frac{11}{12}}}{2}, \frac{1+\sqrt{\frac{11}{12}}}{2} \right]$). Cum eroarea medie pătratică este mai importantă decât nedeplasarea, recomand folosirea estimatorului $\hat{\theta}_2$.