

Tema 3

Soluții

Exercițiul 1

- 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. Aplicand 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

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

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

- 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 5

a) Se observă cu ușurință că

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

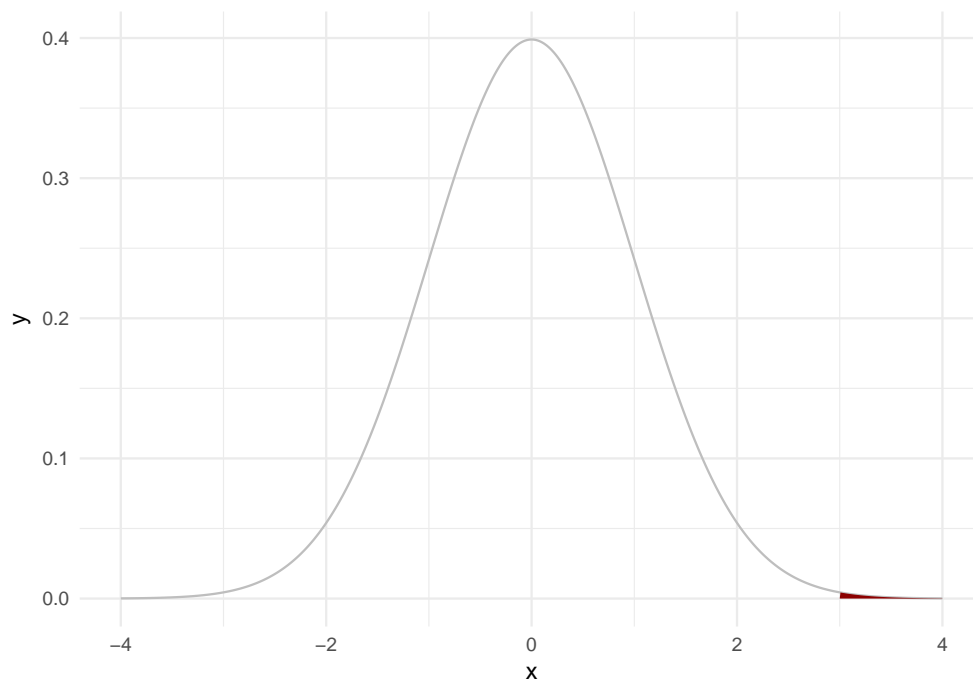
$$h_n(x) = \frac{d}{dx} H_n(x) = n f(x) F(x)^{n-1}$$

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

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

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 e santion 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 cand $X_i > \mu + 3\sigma$ și 0 in 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 e santion 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 incat $H_n(x) = 0.99$. 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 e santion 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

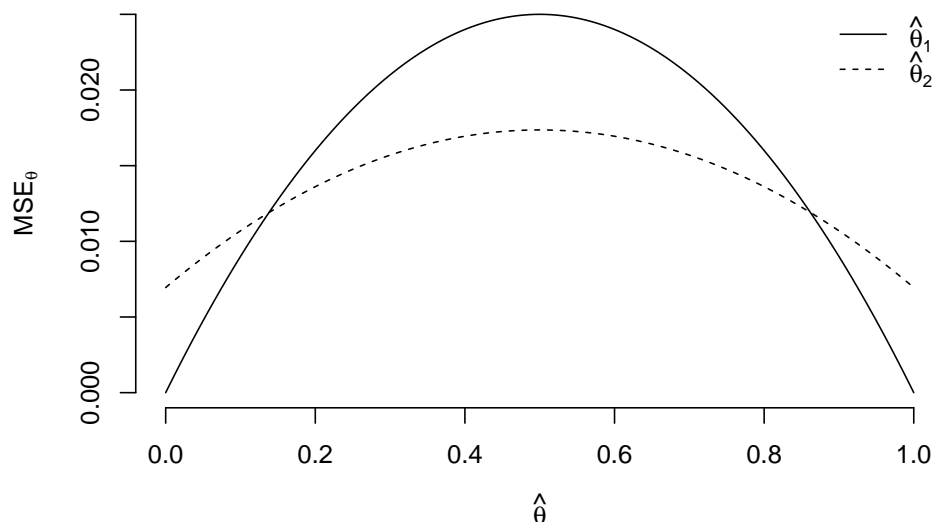
- a) Cum $\mathbb{E}_\theta[X] = 10\theta$ obținem că $\mathbb{E}_\theta[\hat{\theta}_1] = \theta$ și $\mathbb{E}_\theta[\hat{\theta}_2] = \frac{10\theta+1}{12}$.
- b) Pentru calculul erorii medii pătratice vom folosi următoarea formulă $MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta}) + B_\theta(\hat{\theta})^2$. Cum $\hat{\theta}_1$ este un estimator nedeplasat 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 $\text{hat}\theta_2$ este deplasat, niciunul dintre cei doi estimatori nu are eroarea medie pătratică uniform mai mică. Cu toate acestea, eroarea medie pătratică pentru estimatorul $\text{hat}\theta_2$ este mai mică decât cea pentru estimatorul $\text{hat}\theta_1$ 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 $\text{hat}\theta_2$.