# Seminarul 7

**1. i)** Estimaţi parametrul necunoscut $p \in (0,1)$ pentru distribuţia binomială a unei caracteristici cercetate: $X \sim Bino(N,p)$, unde $N \in \mathbb{N}^*$ este cunoscut, cu metoda momentelor, respectiv metoda verosimilităţii maxime. Sunt estimatorii obţinuţi nedeplasaţi, respectiv consistenţi?
**ii)** Într-o urnă sunt bile albe şi negre. Proporţia de bile albe $p \in (0,1)$ este necunoscută. În urma a $n = 6$ serii a câte $N = 5$ extrageri cu returnarea bilei extrase în urnă s-au obţinut: 3, 4, 2, 0, 2, respectiv 1, bile albe. Estimaţi valoarea lui $p$ cu metoda momentelor, respectiv metoda verosimilităţii maxime.

**2. i)** O caracteristică cercetată $X$ are funcţia de densitate

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

unde $\lambda > 0$ este fixat. Estimaţi parametrul necunoscut $\lambda$ cu metoda momentelor, respectiv metoda verosimilităţii maxime. Sunt estimatorii obţinuţi consistenţi?
**ii)** Durata culorii roşii (în minute) $X$ a unui anumit semafor are funcţia de densitate $f_X$ dată mai sus, cu parametrul $\lambda > 0$ necunoscut. Un taximetrist (curios din fire) a observat următoarele durate (în minute) ale culorii roşii pentru acest semafor: $1, \frac{3}{2}, 3, 2, 3, \frac{5}{2}, 1, 2$. Aplicaţi metoda momentelor, respectiv metoda verosimilităţii maxime, pentru a estima valoarea lui $\lambda$, folosind datele furnizate de taximetrist.

Pentru rezolvarea următoarelor probleme de statistică, folosiți următoarele valori numerice calculate în Python: (pt intervale de încredere)

nu se vacue calc. cuantilelor
se vor da la examen

```
[1]: from scipy.stats import norm, t, chi2
```

```
[2]: norm.ppf([0.025,0.05,0.95,0.975])
```

```
[2]: array([-1.95996398, -1.64485363,  1.64485363,  1.95996398])
```

```
[3]: t.ppf([0.025,0.05,0.95,0.975],4), t.ppf([0.025,0.05,0.95,0.975],99)
```

```
[3]: (array([-2.77644511, -2.13184678,  2.13184678,  2.77644511]),
      array([-1.98421695, -1.66039116,  1.66039116,  1.98421695]))
```

```
[4]: chi2.ppf([0.025,0.05,0.95,0.975],2), chi2.ppf([0.025,0.05,0.95,0.975],4)
```

```
[4]: (array([0.05063562, 0.10258659, 5.99146455, 7.37775891]),
      array([ 0.48441856,  0.71072302,  9.48772904, 11.14328678]))
```

**3.** Considerăm următoarele date statistice pentru masa corporală a persoanelor dintr-o anumită populaţie: 71 kg; 68 kg; 77 kg; 69 kg; 65 kg. Presupunem că masa corporală este o caracteristică ce urmează distribuţia normală.
**a)** Ştiind că varianţa/dispersia masei corporale este 20, determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.
**b)** Ştiind că varianţa/dispersia masei corporale este necunoscută, determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.

$\to$ prob. de a avea succes.

**1. i)** Estimaţi parametrul necunoscut $p \in (0,1)$ pentru distribuţia binomială a unei caracteristici cercetate: $X \sim Bino(N, p)$, unde $N \in \mathbb{N}^*$ este cunoscut, cu metoda momentelor, respectiv metoda verosimilităţii maxime. Sunt estimatorii obţinuţi nedeplasaţi, respectiv consistenţi?
**ii)** Într-o urnă sunt bile albe şi negre. Proporţia de bile albe $p \in (0,1)$ este necunoscută. În urma a $n = 6$ serii a câte $N = 5$ extrageri cu returnarea bilei extrase în urnă s-au obţinut: 3, 4, 2, 0, 2, respectiv 1, bile albe. Estimaţi valoarea lui $p$ cu metoda momentelor, respectiv metoda verosimilităţii maxime.

$x_1, \ldots, x_n$ – date statistice pt. caracteristica $X \sim Bino(N, p)$, $p$ necunoscut

$X_1, \ldots, X_n$ – variabile de selecţie

**Metoda momentelor** $\quad E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\quad$ $p$ val. estimată

$$E(X) = N \cdot p \Rightarrow N \cdot p = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow p = \frac{1}{Nn}\sum_{i=1}^{n} x_i \Rightarrow \hat{p}(x_1, \ldots, x_n) = \frac{1}{N \cdot n} \cdot \sum_{i=1}^{n} x_i$$

**Metoda verosimilităţii maxime**

$$L(x_1, \ldots, x_n, p) = P(X = x_1) \cdot P(X = x_2) \cdot \ldots \cdot P(X = x_n)$$

$$P(X = x_i) = C_N^{x_i} \cdot p^{x_i} \cdot (1-p)^{N - x_i}, \quad x_i \in \{0, 1, \ldots, N\} \quad \Big| \Rightarrow L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} C_N^{x_i} p^{x_i} \cdot (1-p)^{N - x_i}$$

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} C_N^{x_i} \cdot p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{nN - \sum_{i=1}^{n} x_i}$$

$$\ln L(x_1, \ldots, x_n, p) = \ln \prod_{i=1}^{n} C_N^{x_i} + \sum_{i=1}^{n} x_i \cdot \ln p + \left(nN - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

$$\frac{d}{dp} L(x_1, \ldots, x_n, p) = \sum_{i=1}^{n} x_i \cdot \frac{1}{p} - \left(nN - \sum_{i=1}^{n} x_i\right) \frac{1}{1-p} = 0$$

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

$$\hat{p}(x_1, \ldots, x_n) = \frac{1}{N \cdot n} \cdot \sum_{i=1}^{n} x_i$$

$$\frac{d^2}{dp^2} \ln L(x_1, \ldots, x_n, p) = -\sum_{i=1}^{n} x_i \frac{1}{p^2} - \left(Nn - \sum_{i=1}^{n} x_i\right) \frac{1}{(1-p)^2} \leq 0$$

$$\hat{p}(X_1, \ldots, X_m) = \frac{1}{Nm} \cdot \sum_{i=1}^{m} X_i$$

$\hat{p}(X_1, \ldots, X_m)$ nedeplasat dacă $E(\hat{p}(X_1, \ldots, X_m)) = p$

$\hat{p}(X_1, \ldots, X_m)$ consistent dacă $\hat{p}(X_1, \ldots, X_m) \xrightarrow{a.s.} p$

$$E(\hat{p}(X_1, \ldots, X_m)) = E\left(\frac{1}{Nm} \cdot \sum_{i=1}^{m} X_i\right) = \frac{1}{Nm} \cdot \sum_{i=1}^{m} \underbrace{E(X)}_{N \cdot p} = \frac{1}{\cancel{N \cdot m}} \cdot \cancel{m N} p = p$$

$\Downarrow$ nedeplasat

$$LTNM \Rightarrow \frac{1}{m}(X_1 + \ldots + X_m) \xrightarrow{a.s.} E(X)$$

$$\hat{p}(X_1, \ldots, X_m) = \frac{1}{N} \cdot \frac{1}{m}(X_1 + \ldots + X_m) \xrightarrow{a.s.} \frac{1}{N} E(X) = \frac{1}{N} \cdot Np = p$$

ii) $p \in (0, 1)$ proporție bile albe necunoscute

$m = 6$ serii $N = 5$ extrageri

$x_1 = 3$, $x_2 = 4$, $x_3 = 2$, $x_4 = 0$, $x_5 = 2$, $x_6 = 1$

$$\hat{p}(x_1, \ldots, x_m) = \frac{1}{N \cdot m} \cdot \sum_{i=1}^{m} x_i = \frac{1}{6 \cdot 5} \cdot (3 + 4 + 2 + 0 + 2 + 1) = \frac{2}{5}$$

**2. i)** O caracteristică cercetată $X$ are funcţia de densitate

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

unde $\lambda > 0$ este fixat. Estimaţi parametrul necunoscut $\lambda$ cu metoda momentelor, respectiv metoda verosimilităţii maxime. Sunt estimatorii obţinuţi consistenţi?

**ii)** Durata culorii roşii (în minute) $X$ a unui anumit semafor are funcţia de densitate $f_X$ dată mai sus, cu parametrul $\lambda > 0$ necunoscut. Un taximetrist (curios din fire) a observat următoarele durate (în minute) ale culorii roşii pentru acest semafor: $1, \frac{3}{2}, 3, 2, 3, \frac{5}{2}, 1, 2$. Aplicaţi metoda momentelor, respectiv metoda verosimilităţii maxime, pentru a estima valoarea lui $\lambda$, folosind datele furnizate de taximetrist.

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \quad \lambda > 0 \text{ fixat necunoscut}$$

$x_1, \ldots, x_n$ date statistice

$X_1, \ldots, X_n$ variabile de selecție coresp.

Metoda momentelor

$$E(X) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \frac{2}{\lambda} \Bigg\} \Rightarrow$$

$$\Rightarrow \frac{2}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i \Rightarrow \hat{\lambda}(x_1, \dots, x_n) = \frac{2n}{\sum_{i=1}^{n} x_i}$$ val. estimată

Met. verosimilității maxime

$$L(x_1, \dots, x_n, \lambda) = f_X(x_1) \cdot \dots \cdot f_X(x_n) = \prod_{i=1}^{n} \lambda^2 x_i e^{-\lambda x_i}$$

$$= \lambda^{2n} \cdot \prod_{i=1}^{n} x_i \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ln L(x_1, \dots, x_n, \lambda) = 2n \ln \lambda + \ln\left(\prod_{i=1}^{n} x_i\right) - \lambda \sum_{i=1}^{n} x_i$$

$$\left.\begin{aligned} &\frac{d}{d\lambda} \ln L(x_1, \dots, x_n, \lambda) = \frac{2n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \\ &\frac{d^2}{d\lambda^2} \ln L(x_1, \dots, x_n, \lambda) = -\frac{2n}{\lambda^2} < 0 \end{aligned}\right\} \Rightarrow \hat{\lambda}(x_1, \dots, x_n) = \frac{2n}{\sum_{i=1}^{n} x_i}$$ val. estimată

$$\hat{\lambda}(X_1, \dots, X_n) = \frac{2n}{\sum_{i=1}^{n} x_i}$$ estimatorul

$$\text{LTNM}, \ \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{a.s.} E(X) = \frac{2}{\lambda}$$

$$\hat{\lambda}(X_1, \dots, X_n) = \frac{2}{\frac{1}{n}\sum_{i=1}^{n} x_i} \xrightarrow{a.s.} \frac{2}{\frac{2}{\lambda}} = \lambda$$ consistent

ii) $x_1 = 1$, $x_2 = \frac{3}{2}$, $x_3 = 3$, $x_4 = 2$, $x_5 = 3$, $x_6 = \frac{5}{2}$, $x_7 = 1$, $x_8 = 2$

$n = 8$

$$\sum_{i=1}^{n} x_i = 1 + \frac{3}{2} + 3 + 2 + 3 + \frac{5}{2} + 1 + 2 = 16$$

$$\lambda = \frac{2 \cdot 8}{16} = 1 \text{ minut}$$

**3.** Considerăm următoarele date statistice pentru masa corporală a persoanelor dintr-o anumită populaţie: 71 kg; 68 kg; 77 kg; 69 kg; 65 kg. Presupunem că masa corporală este o caracteristică ce urmează distribuţia normală.
**a)** Ştiind că varianţa/dispersia masei corporale este 20, determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.
**b)** Ştiind că varianţa/dispersia masei corporale este necunoscută, determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.

a) $V(X) = 20 \Rightarrow \sigma = \sqrt{V(X)}$ $\quad \mu = E(X)$ val. medie a masei corp.

$1-\alpha = 95\%$ nivelul de încredere

$\alpha$ – nivel de semnificație, $\alpha = 5\% = 0,05$

intervalul bilateral de încredere:

$$\left(\overline{x_n} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \overline{x_n} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$$

$$\overline{x_n} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \underbrace{z_{1-\frac{\alpha}{2}}}_{\text{cuantila}} = \text{norm.ppf}(\underbrace{1-0,025}_{0,975}) \simeq 1,96$$

$$\overline{x_5} = \frac{71+68+77+69+65}{5} = 70$$

$$\left(70 - \frac{\sqrt{20}}{\sqrt{5}} \cdot 1,96, 70 + \frac{\sqrt{20}}{\sqrt{5}} \cdot 1,96\right) \Rightarrow (66,08 ; 73,92)$$

$H_0: \mu = 75 \qquad H_1: \mu \neq 75$

$75 \notin (66,08 ; 73,92) \Rightarrow$ ipoteza $H_0$ se respinge, $H_1$ se acceptă

b) $1-\alpha = 95\%$, $\alpha = 5\%$

$$\left(\overline{x_n} - \frac{s_n}{\sqrt{n}} t_{1-\alpha}, \overline{x_n} + \frac{s_n}{\sqrt{n}} t_{1-\alpha}\right)$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x_n})^2 \qquad t_{1-\alpha} \text{ cuantilele legii Student}$$

val. variației de selecție $\qquad T(n-1)$

$$S_m^2 = \frac{1}{4}\left[(71-70)^2+(68-70)^2+(77-70)^2+(69-70)^2+(65-70)^2\right]$$

$$= \frac{1}{4}\, 80 = 20$$

$$t_{1-\alpha} = t.\mathrm{ppf}_4(0{,}975) \approx 2{,}78$$

$$\left(70 - \frac{\sqrt{20}}{\sqrt{5}}\cdot 2{,}78 \,;\; 70 + \frac{\sqrt{20}}{\sqrt{5}}\cdot 2{,}78\right)$$

$$(64{,}44 \,;\; 75{,}56)$$

**c)** Determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru varianţa masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că abaterea standard a masei corporale este 10 kg.

**4.** Într-un sondaj de opinie, suntem interesaţi de proporţia $p$ a persoanelor dintr-un anumit oraş care ar vota candidatul $A$ împotriva candidatului $B$. 576 din participanții la sondaj au declarat că ar vota candidatul $A$, iar 324 din participanți au declarat că ar vota cu candidatul $B$.
**a)** Determinaţi un interval de încredere bilateral cu nivelul de încredere 95% pentru $p$.
**b)** Testați cu probabilitatea de risc 5% ipoteza că $p = 0{,}6$.

**5.** Un provider de internet îşi asigură clienţii că viteza conexiunii la internet este în medie 250 Mbps între orele 20:00 şi 22:00. Pe de altă parte, providerul susţine că în acest interval orar conexiunea nu este stabilă, având o abatere standard de 40 Mbps. În urma unei selecţii de 100 de clienţi s-a constatat că valoarea mediei de selecţie este 242 Mbps pentru viteza conexiunii între orele specificate.
i) Să se construiască un interval de încredere unilateral stâng cu nivelul de încredere 95% pentru media vitezei conexiunii.
ii) Să se testeze, cu nivelul de semnificaţie 5%, dacă media vitezei este cea pretinsă de provider.

**6.** O companie doreşte înlocuirea unui sistem de frânare pentru un anumit tip de maşină cu unul nou, care să reducă semnificativ distanţa de frânare. Media distanţei de frânare pentru vechiul sistem este mai mare sau egală decât 50 m, pentru o viteză de 80 km/h pe ploaie. În urma testării a 100 de maşini cu noul sistem de frânare instalat, pentru o viteză de 80 km/h pe ploaie, s-a constatat că valoarea mediei de selecţie este 49 m şi că valoarea abaterii standard de selecţie este 1 m pentru distanţa de frânare a acestui eşantion.
i) Să se construiască un interval de încredere unilateral drept cu nivelul de încredere 95% pentru media distanței de frânare a noului sistem.
ii) Să se testeze cu probabilitatea de risc 5% dacă noul sistem de frânare este diferit de cel vechi.

**7.** Un sondaj de opinie are în vedere studiul a două caracteristici: $F$, genul de film preferat, și $M$, genul de muzică preferat. S-au obținut următoarele date statistice:

| Film \ Muzică | Rock | Pop |
|---|---|---|
| Acțiune | 25 | 45 |
| Comedie | 26 | 39 |
| Dramă | 39 | 26 |

Testați, cu nivelul de semnificație 5%, dacă $F$ și $M$ sunt independente.

**8.** Într-o urnă sunt 100 de bile. Fiecare bilă este colorată cu una din culorile: roșu, albastru, verde. În urma extragerii a 120 bile, cu returnare, s-au obținut 34 de bile roșii, 55 de bile albastre și 31 de bile verzi. Folosind un test statistic cu nivelul de semnificaţie 5%, se poate afirma că în urnă sunt 25 de bile roșii, 50 de bile albastre și 25 de bile verzi?