PS - Simulare examen scris 2021 (timp de lucru 30 de minute) - soluţii

S1. Fie $N_1, ..., N_5 \sim \begin{pmatrix} 1 & 2 & \dots & 10 \\ \frac{1}{10} & \frac{1}{10} & \dots & \frac{1}{10} \end{pmatrix}$ variabile aleatoare independente, care urmează distribuţia discretă uniformă $Unid(10)$. Fie vectorul de valori $U = [1, 2, 3, 1, 2, 4, 1, 2, 2, 4]$ şi vectorul (de date aleatore alese din U) $X = [U_{N_1}, ..., U_{N_5}]$. Fie Z variabila aleatoare care indică de câte ori apare 1 în vectorul X.

a) Determinaţi $P(Z = 3)$, $P(\{Z < 3\} \cup \{Z > 4\})$ şi $P(Z < 3|Z \geq 1)$.

b) Să se scrie distribuţia de probabilitate a variabilei aleatoare Z.

R: Probabilitatea p ca un element în vectorul X să fie egal cu 1 este de fapt probabilitatea ca o valoare aleasă aleator din vectorul U să fie egală cu 1. Are loc $p = \frac{3}{10}$.

a) $P(Z = 3) = P($"1 apare exact de trei ori în vectorul X"$) = C_5^3 p^3 (1-p)^2 = 10 \cdot 0.3^3 \cdot 0.7^2$;

$P(\{Z < 3\} \cup \{Z > 4\}) = P(Z < 3) + P(Z > 4)$ (cele două evenimente sunt disjuncte)

$\Longrightarrow P(\{Z < 3\} \cup \{Z > 4\}) = P(Z \in \{0, 1, 2\}) + P(Z = 5) = \sum_{k \in \{0,1,2,5\}} C_5^k p^k (1-p)^{5-k} = \sum_{k \in \{0,1,2,5\}} C_5^k \cdot 0.3^k \cdot 0.7^{5-k}$

$P(Z < 3|Z \geq 1) = \dfrac{P(1 \leq Z < 3)}{P(Z \geq 1)} = \dfrac{P(Z = 1) + P(Z = 2)}{1 - P(Z < 1)} = \dfrac{C_5^1 \cdot 0.3 \cdot 0.7^4 + C_5^2 \cdot 0.3^2 \cdot 0.7^3}{1 - 0.7^5}$.

b) Distribuţia de probabilitate a variabilei aleatoare Z este dată prin

$P(Z = k) = C_5^k \cdot 0.3^k \cdot 0.7^{5-k},\ k \in \{0, 1, 2, 3, 4, 5\}$; adică $Z \sim Bino(5, 0.3)$.

S2. Fie $x_1 = 1, x_2 = 0, x_3 = 3, x_4 = 2, x_5 = 0, x_6 = 1, x_7 = 4, x_8 = 5$ date statistice pentru caracteristica X, care are următoarea distribuţie:

$$P(X = k) = p(1-p)^k \text{ pentru } k \in \{0, 1, 2, \dots\},$$

iar $p \in (0, 1)$ este parametru necunoscut. Folosind metoda verosimilităţii maxime, estimaţi valoarea parametrului necunoscut p.

R: $n = 8$; $p \in (0, 1)$; $L(x_1, \dots, x_n; p) = \prod_{i=1}^{n} P(X = x_i) = p^n (1-p)^{\sum_{i=1}^{n} x_i} \Longrightarrow \ln L(x_1, \dots, x_n; p) = n \ln p + \ln(1-p) \sum_{i=1}^{n} x_i$.

$\frac{\partial \ln L}{\partial p}(x_1, \dots, x_n; p) = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i}{1-p} = 0 \Longrightarrow \frac{1-p}{p} = \frac{\sum_{i=1}^{n} x_i}{n} \Longrightarrow \hat{p} = \frac{n}{\sum_{i=1}^{n} x_i + n} = \frac{1}{1+\bar{x}_n}$;

$\frac{\partial^2 \ln L}{\partial p^2}(x_1, \dots, x_n; p) = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} x_i}{(1-p)^2} < 0 \Longrightarrow \hat{p}$ e punct de maxim. $\hat{p}(1, 0, 3, 2, 0, 1, 4, 5) = \frac{8}{\sum_{i=1}^{8} x_i + 8} = \frac{8}{16+8} = \frac{1}{3}$.