

Materia PS pentru examenul oral: întrebări teoretice, scurte probleme, întrebări legate de programe Octave

1) Definiția clasică a probabilității, definiția axiomatică a probabilității

Definiția clasică: Probabilitatea realizării unui eveniment este raportul dintre numărul de cazuri favorabile și numărul de cazuri posibile.

Definiția axiomatică: Definiția clasică a probabilității poate fi utilizată numai în cazul în care numărul cazurilor posibile este finit. Dacă numărul evenimentelor elementare este infinit, atunci există evenimente pentru care probabilitatea, în sensul clasic, nu are niciun înțeles.

2) Cum se folosește probabilitatea geometrică

Alegem un punct aleator $A \in D$ (D - spațiul de selecție (o mulțime)). Probabilitatea ca punctul A să fie inclus în M (M - un spațiu inclus în D) este:

$$P = \text{măsura}(M) / \text{măsura}(D)$$

Măsura unei mulțimi - lungimea (în R), aria (în R^2), volumul (în R^3).

3) Două evenimente independente; n evenimente independente (în totalitate)

Evenimentele A și B sunt independente dacă apariția evenimentului A nu influențează apariția evenimentului B (și invers). Două evenimente independente: două aruncări succesive de zar.

4) Definiția probabilității condiționate

Fie (Ω, K, P) un spațiu de probabilitate și fie $A, B \in K$. Probabilitatea condiționată a evenimentului A de evenimentul B este $P(\cdot | B) : K \rightarrow [0, 1]$, definită prin:

$$P(A | B) = P(A \cap B) / P(B)$$

$P(B) > 0$, $P(A | B)$ - probabilitatea apariției evenimentului A , știind că evenimentul B s-a produs

5) Formula probabilității totale, formula lui Bayes, regula de înmulțire

Formula probabilității totale:

Într-un spațiu de probabilitate (Ω, K, P) considerăm partiția $\{H_1, \dots, H_n\} \in \Omega$ cu H_i din K și $P(H_i) > 0$, $i = 1..n$, și fie $A \in K$. Atunci are loc:

$$P(A) = P(A|H_1)P(H_1) + \dots + P(A|H_n)P(H_n)$$

Formula lui Bayes:

Formula este o metodă de a "corecta" pe baza unor date disponibile o probabilitate determinată. Se pornește cu o estimare pentru probabilitatea unei anumite ipoteze I . Dacă avem noi date D , ce privesc ipoteza I , se poate calcula o probabilitate "corectă" pentru ipoteza I , numită probabilitate posterioară.

$$P(I|D) = \frac{P(D|I) \cdot P(I)}{P(D)} = \frac{P(D|I) \cdot P(I)}{P(D|I) \cdot P(I) + P(D|\bar{I}) \cdot P(\bar{I})}.$$

Regula de înmulțire: Fie (Ω, K, P) un spațiu de probabilitate și Fie $A_1, \dots, A_n \in K$ astfel încât $P(A_1 \cap \dots \cap A_{n-1}) > 0$.

Atunci:

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

6) Distribuția de probabilitate a variabilei aleatoare discrete

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots \\ p_1 & p_2 & \dots & p_i & \dots \end{pmatrix} = \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

7) Distribuții discrete clasice: distribuția uniformă, distribuția Bernoulli, distribuția binomială, distribuția hipergeometrică, distribuția geometrică

Distribuția uniformă: $X \sim Unid(n), n \in \mathbb{N}^*$

$$X \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Distribuția Bernoulli: $X \sim Bernoulli(p), p \in (0, 1)$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Distribuția binomială: $X \sim Bino(n, p), p \in (0, 1), n \in \mathbb{N}^*$

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k \in \{0, \dots, n\}.$$

Distribuția hipergeometrică: $X \sim Hyge(n, n_1, n_2)$

$$\Rightarrow P(X = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n}, \quad k \in \{0, \dots, n^*\}.$$

Distribuția geometrică: $X \sim Geo(p), p \in (0, 1)$

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

8) Distribuția de probabilitate a unui vector aleator discret (X,Y)

$$\mathbb{X} = (X, Y) \sim \begin{pmatrix} (x_i, y_j) \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$$

9) Dacă se cunoaște distribuția vectorului aleator discret (X,Y) , cum se determină distribuțiile lui X și Y ?

$$p_{ij} = P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i)P(Y = y_j) \quad \forall i \in I, j \in J.$$

$$P(X = x_i) = \sum_{j \in J} p_{ij} \quad \forall i \in I$$

$$P(Y = y_j) = \sum_{i \in I} p_{ij} \quad \forall j \in J.$$

10) **Rețeaua Bayes (noduri: părinte, descendent, nondescendent; probabilitățile și probabilitățile condiționate asociate nodurilor din rețea, proprietatea rețelei Bayes)**

Rețele Bayes – grafuri orientate aciclice, în care nodurile sunt variabile aleatoare și există anumite proprietăți de independență între noduri.

Evenimentele A, B din K sunt condițional independente, cunoscând evenimentul C din K , dacă și numai dacă:

$$P(A \cap B|C) = P(A|C)P(B|C)$$

Nodul Y este **părinte** pentru nodul X dacă există o muchie orientată de la Y la X . ($p(X)$)

Nodul Y este **descendent** al nodului X dacă există un drum orientat de la X la Y . ($d(X)$)

Într-o rețea în care există o structură *cauzală*, nodurile din $p(X)$ reprezintă cauzele pentru X , iar nodurile din $d(X)$ sunt efectele nodului X .

Nodul Y este **nondescendent** al nodului X dacă nu este descendent al nodului X . ($nd(X)$)

Proprietatea rețelei Bayes: orice nod X și nondescendenții săi sunt condițional independenți dacă se cunosc valorile părinților. Dacă $p(X) = \emptyset$, atunci X și $nd(X)$ sunt independenți.

11) **Funcția de densitate a unei variabile aleatoare continue (definiția, proprietăți)**

Def. 19. *Funcția de densitate* $f : \mathbb{R} \rightarrow \mathbb{R}$ a unei v.a. continue este funcția pentru care are loc

$$P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad \forall x \in \mathbb{R}.$$

P. 14. Fie f funcția de densitate a unei v.a. continue X . Au loc proprietățile:

(1) $f(t) \geq 0$ pentru orice $t \in \mathbb{R}$;

(2) $\int_{-\infty}^{\infty} f(t) dt = 1$;

(3) $P(a < X \leq b) = \int_a^b f(t)dt \quad \forall a, b \in \mathbb{R}, a < b$;

(4) $P(X = a) = 0 \quad \forall a \in \mathbb{R}$;

(5) pentru $\forall a < b, a, b \in \mathbb{R}$ au loc

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(t)dt.$$

12) Funcția de repartiție (definiția, proprietăți)

Def. 20. *Funcția de repartiție* $F : \mathbb{R} \rightarrow [0, 1]$ a unei variabile aleatoare X (discrete sau continue) este

$$F(x) = P(X \leq x) \quad \forall x \in \mathbb{R}.$$

P. 15. *Funcția de repartiție* F a unei variabile aleatoare X (discrete sau continue) are următoarele proprietăți:

(1) F este monoton crescătoare, adică pentru orice $x_1 < x_2$ rezultă $F(x_1) \leq F(x_2)$.

(2) $\lim_{x \rightarrow \infty} F(x) = 1$ și $\lim_{x \rightarrow -\infty} F(x) = 0$.

(3) F este continuă la dreapta, adică $\lim_{x \searrow x_0} F(x) = F(x_0) \quad \forall x_0 \in \mathbb{R}$.

(4) $P(a < X \leq b) = F(b) - F(a) \quad \forall a, b \in \mathbb{R}, a < b$.

Observație importantă:

▷ Orice funcție $F : \mathbb{R} \rightarrow \mathbb{R}$, care are proprietățile (1), (2), (3) din **P.15** este o funcție de repartiție.
Matlab/Octave:

Distribuția v.a. discrete X	Generare valori aleatoare	Funcția de repartiție $F_X(x)$	Probabilitate $P(X = x)$
$Bino(n, p)$	<code>binornd(n, p)</code>	<code>binocdf(x, n, p)</code>	<code>binopdf(x, n, p)</code>
$Unid(n)$	<code>unidrnd(n)</code>	<code>unidcdf(x, n)</code>	<code>unidpdf(x, n)</code>
$Hyge(n, n_1, n_2)$	<code>hygernd(n_1+n_2, n_1, n)</code>	<code>hygecdf(x, n_1+n_2, n_1, n)</code>	<code>hygepdf(x, n_1+n_2, n_1, n)</code>
$Geo(p)$	<code>geornd(p)</code>	<code>geocdf(x, p)</code>	<code>geopdf(x, p)</code>

Distribuția v.a. continue X	Generare valori aleatoare	Funcția de repartiție $F_X(x)$	Funcția de densitate $f_X(x)$
$Unif[a, b]$	<code>unifrnd(a, b)</code>	<code>unifcdf(x, a, b)</code>	<code>unifpdf(x, a, b)</code>
$N(m, \sigma^2)$	<code>normrnd(m, sigma)</code>	<code>normcdf(x, m, sigma)</code>	<code>normpdf(x, m, sigma)</code>
$Exp(\lambda)$	<code>exprnd(1/lambda)</code>	<code>expcdf(x, 1/lambda)</code>	<code>exppdf(x, 1/lambda)</code>

13) Funcția de densitate a unui vector aleator continuu (X,Y) (definiția, proprietăți)

Def. 21. (X_1, \dots, X_n) este un *vector aleator continuu* dacă fiecare componentă a sa este o variabilă aleatoare continuă.

Def. 22. $f_{(X,Y)} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ este *funcția de densitate a vectorului aleator continuu* (X, Y) , dacă

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(s, t) ds dt \quad \forall x, y \in \mathbb{R}.$$

14) Funcția de repartiție a unui vector aleator (X,Y) (definiția, proprietăți)

Def. 23. $F_{(X,Y)} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ este *funcția de repartiție a vectorului aleator* (X, Y) (discret sau continuu), dacă

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) \quad \forall x, y \in \mathbb{R}.$$

15) n variabile aleatoare independente (discrete, respectiv continue): definiția, proprietăți

Def. 24. X_1, \dots, X_n sunt *n variabilele aleatoare independente* (discrete sau continue), dacă

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

P. 16. Variabilele aleatoare continue X_1 (cu funcția de densitate f_{X_1}) și X_2 (cu funcția de densitate f_{X_2}) sunt independente, dacă și numai dacă

$$f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \quad \forall x_1, x_2 \in \mathbb{R},$$

unde $f_{(X_1, X_2)}$ este funcția de densitate a vectorului aleator (X_1, X_2) .

16) Valoarea medie a unei variabile aleatoare discrete, respectiv variabile aleatoare continue (definiția, proprietăți)

Def. 25. Valoarea medie a unei v.a. continue X , care are funcția de densitate f , este

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt, \text{ dacă } \int_{-\infty}^{\infty} |t|f(t)dt < \infty.$$

▷ Valoarea medie a unei variabile aleatoare caracterizează tendința centrală a valorilor acesteia.

P. 18. Proprietăți ale valorii medii; fie X, Y v.a. continue:

→ $E(aX + b) = aE(X) + b$ pentru orice $a, b \in \mathbb{R}$;

→ $E(X + Y) = E(X) + E(Y)$;

→ Dacă X și Y sunt variabile aleatoare **independente**, atunci $E(X \cdot Y) = E(X)E(Y)$.

→ Dacă $g : \mathbb{R} \rightarrow \mathbb{R}$ e o funcție, astfel încât $g(X)$ este o v.a. continuă, atunci

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx,$$

$$\text{dacă } \int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty.$$

Def. 14. Valoarea medie a unei variabile aleatoare discrete (numerice) X , care ia valorile $\{x_i, i \in I\}$, este

$$E(X) = \sum_{i \in I} x_i P(X = x_i),$$

dacă $\sum_{i \in I} |x_i| P(X = x_i) < \infty$.

P. 9. Fie X și Y v.a. discrete. Au loc proprietățile:

→ $E(aX + b) = aE(X) + b$ pentru orice $a, b \in \mathbb{R}$;

→ $E(X + Y) = E(X) + E(Y)$;

→ Dacă X și Y sunt v.a. independente, atunci $E(X \cdot Y) = E(X)E(Y)$.

→ Dacă $g : \mathbb{R} \rightarrow \mathbb{R}$ e o funcție astfel încât $g(X)$ este v.a., atunci

$$E(g(X)) = \sum_{i \in I} g(x_i) P(X = x_i),$$

dacă $\sum_{i \in I} |g(x_i)| P(X = x_i) < \infty$.

17) Varianța unei variabile aleatoare (definiția, proprietăți). Deviația standard (abaterea standard) a unei variabile aleatoare (definiția).

Def. 26. Varianța (dispersia) unei variabile aleatoare X (discrete sau continue) este

$$V(X) = E\left((X - E(X))^2\right),$$

(dacă valoarea medie $E\left((X - E(X))^2\right)$ există). Valoarea $\sqrt{V(X)}$ se numește **deviația standard** a lui X și o notăm cu $Std(X)$.

▷ Varianța unei variabile aleatoare caracterizează împrăștierea (dispersia) valorilor lui X în jurul valorii medii $E(X)$.

P. 19. Proprietăți ale varianței:

→ $V(X) = E(X^2) - E^2(X)$.

→ $V(aX + b) = a^2 V(X) \forall a, b \in \mathbb{R}$.

→ Dacă X și Y sunt variabile aleatoare **independente**, atunci $V(X + Y) = V(X) + V(Y)$.

18) Legea tare a numerelor mari pentru un șir variabile aleatoare independente, care au aceeași distribuție

Legea numerelor mari se referă la descrierea rezultatelor unui experiment repetat de foarte multe ori. Conform acestei legi, rezultatul mediu obținut se apropie tot mai mult de valoarea așteptată, cu cât experimentul se repetă de mai multe ori. Aceasta se explică prin faptul că abaterile aleatoare se compensează reciproc.

Legea numerelor mari are două formulări: **legea slabă a numerelor mari** (LSNM) și **legea tare a numerelor mari** (LTNM).

Def. 29. Șirul de v.a. $(X_n)_n$ cu $E|X_n| < \infty \forall n \in \mathbb{N}$ verifică **legea tare a numerelor mari (LTNM)** dacă

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (X_k(\omega) - E(X_k)) = 0\right\}\right) = 1,$$

adică

$$\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{a.s.} 0.$$

19) Cum se estimează probabilitatea teoretică $P(A)$ a unui eveniment A , folosind frecvența relativă?

$$f_n(A) = \frac{1}{n}(X_1 + \dots + X_n); \quad f_n(A) \text{ este o v.a.}$$

20) Ce sunt variabilele de selecție ale unei caracteristici statistice?

Variabilele de selecție X_1, \dots, X_n sunt variabile aleatoare independente cu aceeași distribuție ca X .
 X_i este output-ul repetiției i a experimentului X .

21) Estimator nedeplasat, estimator consistent

Def. 30. $g(X_1, \dots, X_n)$ este **estimator nedeplasat** pentru parametrul necunoscut θ , dacă

$$E(g(X_1, \dots, X_n)) = \theta.$$

$g(X_1, \dots, X_n)$ este **estimator consistent** pentru parametrul necunoscut θ , dacă

$$g(X_1, \dots, X_n) \xrightarrow{a.s.} \theta.$$

Fie $g_1 = g_1(X_1, \dots, X_n)$ și $g_2 = g_2(X_1, \dots, X_n)$ estimatori **nedeplasați** pentru parametrul necunoscut θ . $g_1(X_1, \dots, X_n)$ este **mai eficient** decât $g_2(X_1, \dots, X_n)$, dacă $V(g_1) < V(g_2)$.

22) Media de selecție, varianța de selecție, abaterea standard de selecție, funcția de repartiție de selecție (definițiile lor; cum se estimează acestea în Octave)

► **media de selecție (empirică)**

$$\bar{X}_n = \frac{1}{n} (X_1 + \dots + X_n)$$

► **varianța (dispersia) de selecție (empirică)**

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

► abaterea standard de selecție (empirică)

$$S_n = \left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right)^{\frac{1}{2}}$$

► funcția de repartiție empirică $\mathcal{F}_n : \mathbb{R} \times \Omega \rightarrow [0, 1]$

$$\mathcal{F}_n(x, \omega) = \frac{\#\{i \in \{1, \dots, n\} : X_i(\omega) \leq x\}}{n}, x \in \mathbb{R}$$

În Octave:

- media de selecție: `mean(d)`, `d` – vectorul datelor statistice
- varianța de selecție: `var(d)`, `d` – vectorul datelor statistice
- abaterea standard de selecție: `std(d)`
- funcția de repartiție: `empirical_cdf(x, d)`

23) Intervale de încredere pentru medie, varianță, proporție (folosind, dacă este necesar, programul Octave)

▷ distribuția normală $N(0, 1)$

cuantila $z_\alpha = \text{norminv}(\alpha, 0, 1)$; funcția de repartiție $F_{N(0,1)}(x) = \text{normcdf}(x, 0, 1)$

▷ distribuția Student $St(n)$

cuantila $t_\alpha = \text{tinv}(\alpha, n)$; funcția de repartiție $F_{St(n)}(x) = \text{tcdf}(x, n)$

▷ distribuția Chi-pătrat $\chi^2(n)$

cuantila $c_\alpha = \text{chi2inv}(\alpha, n)$; funcția de repartiție $F_{\chi^2(n)}(x) = \text{chi2cdf}(x, n)$

Interval de încredere pentru media teoretică
când dispersia σ^2 este cunoscută:

bilateral

unilateral

Expresia intervalului de încredere,
folosind datele statistice

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} \right)$$

$$\left(-\infty, \bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_\alpha \right)$$

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}, \infty \right)$$

Interval de încredere pentru media teoretică când dispersia este necunoscută	Expresia intervalului de încredere, folosind datele statistice
bilateral	$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} \right)$
unilateral	$\left(-\infty, \bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{\alpha} \right)$ $\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\alpha}, \infty \right)$

Interval de încredere pentru varianța (dispersia) teoretică $V(X)$	Expresia intervalului de încredere, folosind datele statistice
bilateral	$\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot s_n^2, \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot s_n^2 \right)$
unilateral	$\left(0, \frac{n-1}{c_{\alpha}} \cdot s_n^2 \right)$ $\left(\frac{n-1}{c_{1-\alpha}} \cdot s_n^2, \infty \right)$

Interval de încredere pentru abaterea standard teoretică $Std(X)$	Expresia intervalului de încredere, folosind datele statistice
bilateral	$\left(\sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n, \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n \right)$
unilateral	$\left(0, \sqrt{\frac{n-1}{c_{\alpha}}} \cdot s_n \right)$ $\left(\sqrt{\frac{n-1}{c_{1-\alpha}}} \cdot s_n, \infty \right)$

Interval de încredere pentru proporția p	Expresia intervalului de încredere, folosind datele statistice
bilateral	$\left(\bar{x}_n - \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x}_n + \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \cdot z_{1-\frac{\alpha}{2}} \right) \cap (0, 1)$
unilateral	$\left(0, \bar{x}_n - \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \cdot z_{\alpha} \right) \cap (0, 1)$ $\left(\bar{x}_n - \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \cdot z_{1-\alpha}, 1 \right) \cap (0, 1)$

24) Teste statistice (folosind, dacă este necesar, programul Octave)

	I. $H_0: m = m_0$ $H_1: m \neq m_0$	II. $H_0: m \geq m_0$ $H_1: m < m_0$	III. $H_0: m \leq m_0$ $H_1: m > m_0$
Se acceptă H_0 dacă	$ z < z_{1-\frac{\alpha}{2}}$	$z > z_{\alpha}$	$z < z_{1-\alpha}$
Se respinge H_0 în favoarea lui H_1 , dacă	$ z \geq z_{1-\frac{\alpha}{2}}$	$z \leq z_{\alpha}$	$z \geq z_{1-\alpha}$

► în Octave/Matlab: `ztest`

```
x=normrnd(0,1,1,1000);
[a1,~,a2]=ztest(x,0,1,'tail','both','alpha',0.01) % cazul I
[b1,~,b2]= ztest(x,0,1,'tail','left','alpha',0.01) % cazul II
[c1,~,c2]=ztest(x,0,1,'tail','right','alpha',0.01) % cazul III
```