PROBABILITĂŢI

Sem. 1, Funcţiile lui Euler

Funcţia Gamma a lui Euler $\Gamma : (0,\infty) \to (0,\infty)$ $\quad \Gamma(a) = \int\limits_0^\infty x^{a-1}e^{-x}dx$

1. $\Gamma(1) = 1$; **2.** $\Gamma(a+1) = a\Gamma(a)$, $\forall a > 0$;

3. $\Gamma(n+1) = n!$, $\forall n \in \mathbb{N}$; **4.** $\Gamma\left(\frac{1}{2}\right) = \sqrt{2}\int\limits_0^\infty e^{-\frac{t^2}{2}}dt = \int\limits_{\mathbb{R}} e^{-t^2}dt = \sqrt{\pi}$.

Funcţia Beta a lui Euler $\beta : (0,\infty)\times(0,\infty) \to (0,\infty)$ $\quad \beta(a,b) = \int\limits_0^1 x^{a-1}(1-x)^{b-1}dx$

1. $\beta(a,1) = \frac{1}{a}$, $\forall a > 0$; **2.** $\beta(a,b) = \beta(b,a)$, $\forall a,b > 0$; **3.** $\beta(a,b) = \frac{a-1}{b}\beta(a-1,b+1)$, $\forall a > 1, b > 0$;

4. $\beta(a,b) = \frac{b-1}{a+b-1}\beta(a,b-1) = \frac{a-1}{a+b-1}\beta(a-1,b)$, $\forall a,b > 1$; **5.** $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\forall a,b > 0$.

Sem. 2, Prob. clasică, geometrică, condiţionată; Evenimente independente; Formula lui Bayes

Probabilitate clasică: $P(A) = \frac{\text{nr. of favorable outcomes}}{\text{total nr. of possible outcomes}}$.

Probabilitate condiţionată: $P(A|B) = \frac{P(A\cap B)}{P(B)}$, $P(B) \neq 0$.

Evenimente independente: A, B independente $<=> P(A\cap B) = P(A)P(B) <=> P(A|B) = P(A)$

Formula probabilităţii totale: $\{A_i\}_{i\in I}$ o partiţie a lui S, atunci $P(A) = \sum\limits_{i\in I} P(A_i)P(A_i|A)$

Probabilitatea intersecţiei: $P\left(\bigcap\limits_{i=1}^{n}\right)A_i = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1\cap A_2)\;...\;P\left(A_n|\bigcap\limits_{i=1}^{n-1}A_i\right)$

Formula lui Bayes: $\{A_i\}_{i\in I}$ o partiţie a lui S, atunci $P(A_j|A) = \frac{P(A|A_j)\,P(A_j)}{\sum\limits_{i\in I} P(A|A_i)\,P(A_i)}$, $\forall j \in I$

Sem. 3, Scheme probabilistice

Schema binomială: Probabilitatea ca din n repetări independente ale unui experiment, un eveniment care apare cu aceeaşi probabilitate p, la fiecare repetare, să se realizeze de k ori, este $P(n,k) = C_n^k p^k q^{n-k}$, $k = \overline{0,n}$.

Schema multinomială: Probabilitatea ca din $n = n_1 + n_2 + ... + n_r$ repetări ale experimentului, evenimentul E_i să se realizeze de n_i ori, unde $p_i = P(E_i)$, $i = \overline{1,r}$, is $P(n; n_1, ..., n_r) = \frac{n!}{n_1!n_2!...n_r!}p_1^{n_1}p_2^{n_2}...p_r^{n_r}$.

Schema lui Bernoulli fără revenire (hipergeometrică): Probabilitatea ca din n bile extrase din totalul de N bile albe şi negre existente în urnă (din care n_1 albe), k să fie albe şi $n-k$ negre, $(0 \le k \le n_1$, $0 \le n-k \le N-n_1)$, este $P(n;k) = \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}$.

Schema lui Bernoulli fără revenire cu r stări: Probabilitatea ca din $M = m_1 + m_2 + ... + m_r$ extrageri, să avem m_i bile de culoarea i din totalul de n_i, $i = \overline{1,r}$, $(n = n_1 + n_2 + ... + n_r)$, este $P(n; m_1, ..., m_r) = \frac{C_{n_1}^{m_1}C_{n_2}^{m_2}...C_{n_r}^{m_r}}{C_n^M}$.

Schema lui Poisson: Probabilitatea de a avea k succese din n încercări, cu probabilitate de succes p_i la a $i-$a încercare $(q_i = 1-p_i)$, $i = \overline{1,n}$ este $P(n;k) = \sum\limits_{1\le i_1<...<i_k\le n} p_{i_1}...p_{i_k}q_{i_{k+1}}...q_{i_n}$, $\quad i_{k+1}, ..., i_n \in \{1,...,n\} \setminus \{i_1,...,i_k\}$ = coeficientul lui x^k în dezvoltarea $(p_1x+q_1)(p_2x+q_2)\ldots(p_nx+q_n)$.

Schema lui Pascal: Probabilitatea ca al $n-$lea succes să apară după k insuccese, probabilitatea de succes fiind p $(q = 1-p)$, este $P(n;k) = C_{n+k-1}^{n-1}p^nq^k = C_{n+k-1}^k p^nq^k$.

Schema geometrică: Probabilitatea ca primul succes să apară după k insuccese, probabilitatea de succes fiind p $(q = 1-p)$, este $p_k = pq^k$.

Sem. 4, Variabile aleatoare şi vectori aleatori de tip discret

Distribuţia Bernoulli de parametru $p \in (0,1)$: $X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

Distribuţia binomială de parametri $n \in \mathbb{N}, p \in (0,1)$: $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$

Distribuţia hipergeometrică de parametri $a, b, n \in \mathbb{N}, n \le a$: $X \begin{pmatrix} k \\ p_k \end{pmatrix}_{k=\overline{0,n}}$, $p_k = \dfrac{C_a^k C_b^{n-k}}{C_{a+b}^n}$

Distribuţia Poisson de parametru $\lambda > 0$: $X \begin{pmatrix} k \\ p_k \end{pmatrix}_{k\in\mathbb{N}}$, where $p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$

Distribuţia Pascal de parametri $n \in \mathbb{N}, p \in (0,1)$: $X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k\in\mathbb{N}}$

Distribuţia geometrică de parametru $p \in (0,1)$: $X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k\in\mathbb{N}}$

Distribuţia uniformă de tip discret de parametru $m \in \mathbb{N}$: $X \begin{pmatrix} k \\ \dfrac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$

Funcţie de repartiţie $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X < x)$

Vector aleator de tip discret: $(X,Y) : S \to \mathbb{R}^2$,
pdf $p_{ij} = P(X = x_i, Y = y_j)$, $(i,j) \in I \times J$,

cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) = P(X < x, Y < y) = \sum\limits_{x_i<x} \sum\limits_{y_j<y} p_{ij}$, $\forall (x,y) \in \mathbb{R}^2$,

$p_i = P(X = x_i) = \sum\limits_{j\in J} p_{ij}$, $\forall i \in I$, $q_j = P(Y = y_j) = \sum\limits_{i\in I} p_{ij}$, $\forall j \in J$ (densităţi marginale)

Operaţii: $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J}$
X şi Y sunt **independente** $<=>$ $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j) = p_i q_j$.

$X + Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}$, $\alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i\in I}$, $XY \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}$, $X/Y \begin{pmatrix} x_i/y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}$ $(y_j \neq 0)$

Sem. 5, Var. aleatoare şi vectori aleatori de tip continuu; Operaţii cu var. al. de tip cont.

$X : S \to \mathbb{R}$ variabilă aleatoare de tip continuu cu pdf $f : \mathbb{R} \to \mathbb{R}$, cdf $F : \mathbb{R} \to \mathbb{R}$. Proprietăţi:

1. F este absolut continuă şi $F(x) = P(X < x) = \int\limits_{-\infty}^{x} f(t)dt$ **2.** $f(x) \ge 0, \forall x \in \mathbb{R}$, $\int\limits_{\mathbb{R}} f(x) = 1$

3. $P(X = x) = 0, \forall x \in \mathbb{R}$, $P(a < X < b) = \int\limits_{a}^{b} f(t)dt$ **4.** F este continuă la stânga şi crescătoare

5. $F(-\infty) = 0$, $F(\infty) = 1$

Vector aleator de tip cont.: $(X,Y) : S \to \mathbb{R}^2$, pdf $f = f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) = P(X < x, Y < y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u,v)\, dv\, du$, $\forall (x,y) \in \mathbb{R}^2$. Proprietăţi:

1. $P(a_1 \le X < b_1, a_2 \le Y < b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$
2. F este continuă la stânga şi crescătoare în raport cu fiecare variabilă
3. $F(\infty, \infty) = 1$, $F(-\infty, y) = F(x, -\infty) = 0$, $\forall x, y \in \mathbb{R}$
4. X, Y independente $<=>$ $F(x,y) = F_X(x) F_Y(y)$, $<=>$ $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$
5. $F_X(x) = F(x, \infty)$, $F_Y(y) = F(\infty, y)$, $\forall x, y \in \mathbb{R}$ **6.** $P((X,Y) \in D) = \iint\limits_{D} f(x,y)\, dy\, dx$

7. $f_X(x) = \int\limits_{\mathbb{R}} f(x,y)dy$, $\forall x \in \mathbb{R}$, $f_Y(y) = \int\limits_{\mathbb{R}} f(x,y)dx$, $\forall y \in \mathbb{R}$ (densităţi marginale)

8. Operaţii:

Suma: $f_{X+Y}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}(u, z-u)du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(u)f_Y(z-u)du$

Produsul: $f_{XY}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}\left(u, \frac{z}{u}\right)\frac{1}{|u|}du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(u)f_Y\left(\frac{z}{u}\right)\frac{1}{|u|}du$

Câtul: $f_{X/Y}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}(uz, u)|u|du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(uz)f_Y(u)|u|du$

Funcţia $Y = g(X)$: $g : \mathbb{R} \to \mathbb{R}$ derivabilă cu $g' \neq 0$, strict monotonă, $f_Y(y) = \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$, $y \in g(\mathbb{R})$

Sem. 6, Caracteristici numerice ale variabilelor aleatoare

Valoarea medie:

X discr. cu pdf $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $M(X) = \sum\limits_{i\in I} x_i p_i$, X cont. cu pdf $f : \mathbb{R} \to \mathbb{R}$, $M(X) = \int\limits_{\mathbb{R}} xf(x)dx$.

Dispersia: $D(X) = M\left((X - M(X))^2\right) = M(X^2) - (M(X))^2$.

Deviaţia standard: $\sigma(X) = \sqrt{D(X)}$.

Momente de ordinul k:

- **iniţial** $\nu_k = M\left(X^k\right)$,
- **absolut** $\underline{\nu_k} = M\left(|X|^k\right)$,
- **central** $\mu_k = M\left((X - M(X))^k\right)$.

Covarianţa: $\text{cov}(X,Y) = M((X - M(X))(Y - M(Y))) = M(XY) - M(X)M(Y)$

Coeficientul de corelaţie : $\rho(X,Y) = \dfrac{\text{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

Proprietăţi:

1. $M(aX + b) = aM(X) + b$, $D(aX + b) = a^2D(X)$ **2.** $M(X + Y) = M(X) + M(Y)$

3. dacă X şi Y sunt independente, atunci $M(XY) = M(X)M(Y)$ şi $D(X + Y) = D(X) + D(Y)$

4. $h : \mathbb{R} \to \mathbb{R}$, X discr., atunci $M(h(X)) = \sum\limits_{i\in I} h(x_i)p_i$, X cont., atunci $M(h(X)) = \int\limits_{\mathbb{R}} h(x)f(x)dx$

5. $\text{cov}(X,Y) = M(XY) - M(X)M(Y)$ **6.** $D\left(\sum\limits_{i=1}^{n} a_iX_i\right) = \sum\limits_{i=1}^{n} a_i^2D(X_i) + 2\sum\limits_{1\leq i<j\leq n} a_ia_j\text{cov}(X_i, X_j)$

7. X, Y independente $=>$ $\text{cov}(X,Y) = \rho(X,Y) = 0$ (X şi Y sunt *necorelate*)

8. $-1 \leq \rho(X,Y) \leq 1$; $\rho(X,Y) = \pm 1$ $<=>$ $\exists\, a, b \in \mathbb{R}$, $a \neq 0$ a. î. $Y = aX + b$

9. (X,Y) vect. al. de tip cont. cu pdf $f(x,y)$, $h : \mathbb{R}^2 \to \mathbb{R}^2$, atunci $M(h(X,Y)) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$.

Sem. 7, Inegalităţi; Şiruri de variabile aleatoare

Inegalitatea lui Hölder: $M(|XY|) \leq (M(|X|^p))^{\frac{1}{p}} \cdot (M(|Y|^q))^{\frac{1}{q}}$, $\forall p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

Inegalitatea lui Markov: $P(|X| \geq a) \leq \frac{1}{a}M(|X|)$, $\forall a > 0$.

Inegalitatea lui Chebyshev: $P(|X - M(X)| \geq \epsilon) \leq \dfrac{D(X)}{\epsilon^2}$, $\forall \epsilon > 0$.

Convergenţă:

1) **în probabilitate** $X_n \xrightarrow{p} X$, dacă $\lim\limits_{n\to\infty} P\Big(|X_n - X| < \varepsilon\Big) = 1, \quad \forall\varepsilon > 0$;

2) **tare** $X_n \xrightarrow{t} X$, dacă $\lim\limits_{n\to\infty} P\Big(\bigcap\limits_{k\geq n}\{|X_k - X| < \varepsilon\}\Big) = 1, \quad \forall\varepsilon > 0$;

3) **aproape sigură** $X_n \xrightarrow{a.s.} X$, dacă $P\Big(\lim\limits_{n\to\infty} X_n = X\Big) = 1$;

4) **în distribuţie** $X_n \xrightarrow{d} X$, dacă $\lim\limits_{n\to\infty} F_n(x) = F(x)$, $\forall x \in \mathbb{R}$ punct de continuitate pentru F;

5) **în medie de ordin** r, $0 < r < \infty$ $X_n \xrightarrow{L^r} X$, dacă $\lim\limits_{n\to\infty} M(|X_n - X|^r) = 0$.

Proprietăţi 1. 2) $<=>$ 3) $=>$ 1) $=>$ 4) **2.** 5) $=>$ 1).

STATISTICĂ

X o caracteristică de populaţie, $X_1, X_2, ..., X_n$ selecţie de volum n, i.e. variabile independente şi identic distribuite, cu acelaşi pdf ca X; θ parametrul ţintă, $\overline{\theta} = \overline{\theta}(X_1, X_2, ..., X_n)$ estimator punctual.

Medie de selecţie: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

Moment de selecţie: $\overline{\nu_k} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$,

Moment absolut de selecţie: $\overline{\mu_k} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$,

Dispersie de selecţie: $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Funcţie de verosimilitate a unei selecţii: $L(X_1, ..., X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$.

Informaţia lui Fisher: $I_n(\theta) = M\left[\left(\dfrac{\partial \ln L(X_1, ..., X_n|\theta)}{\partial\theta}\right)^2\right]$.

- dacă domeniul valorilor lui X nu depinde de θ, atunci $I_n(\theta) = -M\left[\dfrac{\partial^2 \ln L(X_1, ..., X_n|\theta)}{\partial\theta^2}\right]$ şi $I_n(\theta) = nI_1(\theta)$.

Eficienţa unui estimator absolut corect: $e(\overline{\theta}) = \dfrac{1}{I_n(\theta)D(\overline{\theta})}$.

Estimatorul $\overline{\theta}$ este
- **nedeplasat**: $M(\overline{\theta}) = \theta$;
- **optimal**: $M(\overline{\theta}) = \theta$ şi $D(\overline{\theta}) \leq D(\hat{\theta})$, $\forall\ \hat{\theta}$ estimator nedeplasat;
- **absolut corect**: $M(\overline{\theta}) = \theta$ şi $\lim_{n\to\infty} D(\overline{\theta}) = 0$;
- **eficient**: absolut corect şi $e(\overline{\theta}) = 1$.

Statistica $S = S(X_1, X_2, ..., X_n)$ este
- **suficientă pentru** θ: pdf. cond. $f(X_1, ..., X_n|S)$ nu depinde de $\theta \overset{\text{Crit.Fact.}}{<=>} L(x_1, ..., x_n|\theta) = g(s, \theta)h(x_1, ..., x_n)$;
- **completă pentru familia de distribuţii** $f(x \mid \theta),\ \theta \in A$: $\quad M(\varphi(S)) = 0, \forall\ \theta \in A\ =>\ \varphi \overset{\text{a.s.}}{=} 0$.

Metoda momentelor:
Se rezolvă sistemul $\nu_k = \overline{\nu_k}$ pentru toţi parametrii necunoscuţi .

Metoda verosimilităţii maxime:

Se rezolvă sistemul $\dfrac{\partial L(X_1, ..., X_n|\theta)}{\partial\theta_j} = 0$ sau $\dfrac{\partial \ln L(X_1, ..., X_n|\theta)}{\partial\theta_j} = 0$, $j = \overline{1, m}$ pentru parametrii necunoscuţi $\theta = (\theta_1, ..., \theta_m)$.

Teorema Lehmann-Scheffé: Fie $\hat{\theta}$ un estimator şi S o statistică suficientă şi completă pentru θ. Atunci $\overline{\theta} = M(\hat{\theta}|S)$ este un estimator optimal.

Inegalitatea Rao-Cramer: Fie $\overline{\theta}$ un estimator absolut corect pentru θ. Atunci $D(\overline{\theta}) \geq \dfrac{1}{I_n(\theta)}$.

Testarea ipotezelor statistice: $H_0 : \theta = \theta_0$, cu una din alternativele $H_1 : \begin{cases} \theta < \theta_0 & \text{(test la stânga)}, \\ \theta > \theta_0 & \text{(test la dreapta)}, \\ \theta \neq \theta_0 & \text{(test bilateral)}. \end{cases}$

Nivel de semnificaţie: $\alpha = P(\text{ eroare de tipul I}) = P(\text{ se respinge } H_0 \mid H_0) = P(TS \in RR \mid \theta = \theta_0)$.
Eroare de tipul II: $\beta = P(\text{ eroare de tipul II}) = P(\text{ se acceptă } H_0 \mid H_1) = P(TS \notin RR\ |H_1)$.
Puterea unui test: $\pi(\theta^*) = P(\text{ se respinge } H_0 \mid \theta = \theta^*) = P(TS \in RR \mid \theta = \theta^*)$.
Lema Neyman-Pearson(LNP): Se testează ipotezele simple $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$. Fie $L(\theta^*)$ funcţie de verosimilitate a selecţiei, atunci când $\theta = \theta^*$. Atunci pentru orice $\alpha \in (0, 1)$, un cel mai puternic test (un test care maximizează puterea $\pi(\theta_1)$) este testul cu $RC = \left\{\dfrac{L(\theta_1)}{L(\theta_0)} \geq k_\alpha\right\}$, unde $k_\alpha > 0$ este o constantă.