

Curs 5

Variabile aleatoare discrete

5.1 Funcția generatoare

În problemele în care apar variabile aleatoare nenegative se utilizează funcția generatoare a variabilei aleatoare X , $G_X(z)$ care se definește prin relația

$$G_X(z) = \sum_{k=0}^{\infty} p_k z^k, \quad (5.1)$$

unde $P(\{X = k\}) = p_k, k = 0, 1, 2, \dots$. Deci cunoscând probabilitățile p_k putem determina funcția generatoare. Deoarece $\sum_{k=0}^{\infty} p_k = 1$, seria de puteri (5.1) converge pentru $|z| \leq 1$. Pentru $|z| < 1$ seria de puteri se poate deriva termen cu termen și obținem:

$$G_X^{(j)}(z) = \sum_{k=j}^{\infty} k(k-1)\dots(k-j+1)p_k z^{k-j} = \sum_{k=j}^{\infty} C_{k,j}^j p_k z^{k-j}.$$

Dacă în relația de mai sus facem $z = 0$ obținem

$$G_X^{(j)}(0) = j!p_j \text{ sau } p_j = \frac{1}{j!}G_X^{(j)}(0).$$

Aceasta ne arată că putem determina toate probabilitățile p_k cu ajutorul funcției generatoare. Din această cauză $G_X^{(j)}$ se numește funcție generatoare. În particular, punând $z = 1$ în G_X' și în G_X'' obținem:

$$M[X] = G_X'(1), \quad M[X^2] = G_X'(1) + G_X''(1).$$

5.2 Legi discrete de repartiție

Vom prezenta legi teoretice de repartiție a unor variabile aleatoare discrete. Există multe fenomene întâlnite în inginerie care "ascultă" de aceste legi în sensul că anumite selecții de date experimentale sunt organizate după modele teoretice standard care trebuie cunoscute. Statisticianul apelează la rezultatele teoretice care îi orientează calculele și deciziile.

5.2.1 V.a. repartizată uniform

Cea mai simplă variabilă aleatoare este aceea care are un număr finit de valori, iar fiecare eveniment are aceeași probabilitate. În acest caz, vom avea

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}.$$

Dacă presupunem că X ia valorile întregi $m, m+1, \dots, m+n+1$, notăm $X \in \text{Uniform}[m, n]$ și obținem

$$M[X] = m + \frac{n+1}{2},$$

$$D[X] = \frac{n^2-1}{12}.$$

5.2.2 V. a. repartizată binomial

Evenimentele A și contrarul său, \bar{A} se repetă de mai multe ori și se produc cu aceeași probabilitate, $P(A) = p$ respectiv $P(\bar{A}) = q = 1 - p$. Probabilitatea producerii de k ori a evenimentului A fi $C_n^k p^k q^{n-k}$, iar v. a. care ia ca valori numărul de produceri ale evenimentului A are tabloul de repartiție

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & k & \dots & n \\ C_n^0 p^0 q^n & C_n^1 p^1 q^{n-1} & C_n^2 p^2 q^{n-2} & \dots & C_n^k p^k q^{n-k} & \dots & C_n^n p^n q^0 \end{pmatrix}.$$

În acest caz scriem $X \in \text{Binomial}[n, p]$, $M[X] = np$, $D^2[X] = npq$.

Într-adevăr, variabila aleatoare X poate fi văzută ca suma a n variabile aleatoare independente X_i care ia valoarea 1 cu probabilitatea p , adică la a i -a repetare a experienței evenimentul A s-a produs și valoarea 0 cu probabilitatea q , adică la a i -a repetare a experienței evenimentul A nu s-a produs

$$X_i : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad i = \overline{1, n}, \quad M[X_i] = p.$$

Cu notațiile introduse, putem scrie

$$X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i,$$

$$M[X] = M\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n M[X_i] = np,$$

$$D[X] = D\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n D[X_i],$$

deoarece variabilele aleatoare $X_i, i = \overline{1, n}$, sunt independente. Dar

$$D[X_i] = M[X_i^2] - (M[X_i])^2.$$

Deoarece

$$X_i^2 : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad M[X_i^2] = p,$$

avem

$$D[X_i] = p - p^2 = pq \text{ și } D[X] = npq.$$

Exercițiul 1 Arătați că dacă $X \in \text{Binomial}[n, p]$ atunci $G_X(z) = (pz + q)^n$ și deduceți cu ajutorul ei relațiile $M[X] = np$, $D^2[X] = npq$.

Problemele în care se utilizează repartiția binomială sunt cele în care facem următoarele presupuneri:

- sunt numai două posibilități pentru rezultatul experienței (succes/insucces, realizări/ne-realizări, bit corect/eronat, aparat bun/defect etc.),
- probabilitatea succesului este aceeași pentru fiecare experiență,
- sunt n repetări ale experienței, unde n este constant,
- cele n experiențe sunt independente.

Facem observația că în calculul factorialilor care intervin în combinări, pentru n suficient de mare, se recomandă formula aproximativă a lui Stirling,

$$n! \approx n^n e^{-n} \sqrt{2n\pi}.$$

5.2.3 V. a. repartizată binomial cu exponent negativ

Facem observația că în cazul *variabilei aleatoare binomiale* se numără realizările evenimentului A în cazul în care se efectuează n experiențe. Deci *numărul de repetări ale experienței este fixat iar numărul succeselor este aleator*.

În cazul *v. a. binomiale cu exponent negativ* variabila ia ca valori numărul de experiențe necesare pentru a obține m realizări. În acest caz *numărul realizărilor ale evenimentului A este fixat iar numărul repetărilor ale evenimentului A este aleator*.

Se spune că variabila aleatoare X are repartiție binomială cu exponent negativ, cu parametrul m și p , ($m = 1, 2, \dots, 0 < p < 1$), dacă X poate lua valorile $m, m + 1, m + 2, \dots$ iar

$$P(\{X = k\}) = C_{m+k-1}^{m-1} p^m q^k, q = 1 - p.$$

Vom avea așadar

$$X : \left(\begin{array}{cccccc} m & m+1 & m+2 & \dots & m+k & \dots \\ C_{m-1}^{m-1} p^m & C_m^{m-1} p^m q & C_{m+1}^{m-1} p^m q^2 & \dots & C_{m+k-1}^{m-1} p^m q^k & \dots \end{array} \right)$$

Vom scrie $X \in \text{NegBin}[m, p]$.

Această repartiție apare în modele de tipul următor: să presupunem că s-a efectuat un număr de observații independente, în cursul fiecărei observații evenimentul A (de exemplu nefuncționarea unui subansamblu al unui aparat) poate să se producă cu probabilitatea p . Observațiile continuă până când evenimentul se produce de m ori și se caută probabilitatea pentru ca aceasta să se producă exact în decursul a $m + k$ observații. Evenimentul $\{X = k\}$ se scrie ca intersecție a două evenimente: evenimentul B_1 care constă în faptul că în primele $m + k - 1$ observații evenimentul A se produce de $m - 1$ ori și evenimentul B_2 care constă în faptul că în observația $m + k$ se produce evenimentul A . Din schema lui Bernoulli se deduce că probabilitatea primului eveniment este $P(B_1) = C_{m+k-1}^{m-1} p^{m-1} q^k$ iar probabilitatea celui de-al doilea eveniment este egală cu p , $P(B_2) = p$. Deci

$$P(\{X = k\}) = P(B_1 \cap B_2) = P(B_1)P(B_2) = C_{m+k-1}^{m-1} p^{m-1} q^k p = C_{m+k-1}^{m-1} p^m q^k.$$

Numele acestei repartiții provine din faptul că probabilitatea căutată este coeficient în dezvoltarea în serie a lui

$$\left(\frac{1}{p} - \frac{q}{p} \right)^{-m} = \sum_{k=0}^{\infty} C_{m+k-1}^{m-1} p^m q^k. \quad (5.2)$$

Justificăm această relație:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \text{ pentru } |x| < 1, \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1},$$

$$\frac{2}{(1-x)^3} = \sum_{k=2}^{\infty} k(k-1)x^{k-2} \Rightarrow \frac{1}{(1-x)^3} = \sum_{k=2}^{\infty} C_k^2 x^{k-2}.$$

Se demonstrează prin inducție că $\frac{1}{(1-x)^m} = \sum_{k=m-1}^{\infty} C_k^{m-1} x^{k-m+1}$.

Trecem k în $k-1$ și obținem: $(1-x)^{-m} = \sum_{k=m}^{\infty} C_{k-1}^{m-1} x^{k-m}$ care reprezintă dezvoltarea în serie binomială cu putere negativă. Trecem $x \rightarrow q$ și obținem:

$$\left(\frac{1}{p} - \frac{q}{p}\right)^{-m} = p^m (1-q)^{-m} = p^m \sum_{k=m}^{\infty} C_{k-1}^{m-1} q^{k-m} = \sum_{k=m}^{\infty} C_{k-1}^{m-1} p^m q^{k-m} = \sum_{k=0}^{\infty} C_{m+k-1}^{m-1} p^m q^k.$$

Observăm că definirea acestei variabile aleatoare este corectă deoarece $P(\{X = k\}) > 0$ și $\sum_{k=0}^{\infty} C_{m+k-1}^{m-1} p^m q^k = 1$.

Vom considera v.a.

$$Y : \begin{pmatrix} 0 & 1 & 2 & \dots & k & \dots \\ C_{m-1}^{m-1} p^m & C_m^{m-1} p^m q & C_{m+1}^{m-1} p^m q^2 & \dots & C_{m+k-1}^{m-1} p^m q^k & \dots \end{pmatrix}$$

Atunci

$$X = Y + m.$$

Calculăm funcția generatoare pentru Y .

$$G_Y(z) = \sum_{k=0}^{\infty} C_{m+k-1}^{m-1} p^m q^k z^k = \left(\frac{1}{p} - \frac{zq}{p}\right)^{-m},$$

$$G'_Y(z) = m \left(\frac{1}{p} - \frac{zq}{p}\right)^{-m-1} \frac{q}{p},$$

$$M[Y] = G'_Y(1) = m \frac{q}{p},$$

$$G''_Y(z) = m(m+1) \left(\frac{1}{p} - \frac{zq}{p}\right)^{-m-1} \left(\frac{q}{p}\right)^2,$$

$$G''_Y(1) = m(m+1) \left(\frac{q}{p}\right)^2,$$

$$D[Y] = m(m+1) \left(\frac{q}{p}\right)^2 + m \frac{q}{p} - \left(m \frac{q}{p}\right)^2 = \frac{mq}{p^2}.$$

Atunci

$$M[X] = M[Y] + m = \frac{m}{p},$$

$$D[X] = D[Y] = \frac{mq}{p^2}.$$

Exercițiul 2 Calculatorul A transmite un mesaj calculatorului B . Mesajul este codat astfel încât B detectează când se produc erori în cursul transmiterii mesajului. Dacă B detectează o eroare cere calculatorului A retransmiterea mesajului. Presupunem că probabilitatea ca transmiterea mesajului să fie eronată este $p = 0, 1$. Vrem să definim variabila aleatoare X care ia ca valori numărul de mesaje care trebuie transmise până la obținerea a patru mesaje eronate.

V. a. X are o distribuție binomială cu exponent negativ. În acest caz $m = 4$. Valorile pe care le ia v. a. sunt 4, 5, 6, ... cu probabilitățile din tabel

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & \dots \\ 0.0001 & 0.00036 & 0.00081 & 0.001458 & \dots \end{pmatrix}$$

$$P(\{X = 0\}) = C_3^3 (0.1)^4 = 0.0001,$$

$$P(\{X = 1\}) = C_4^3 (0.1)^4 (0.9) = 0.00009 \cdot 4 = 0.00036$$

$$P(\{X = 2\}) = C_5^3 (0.1)^4 (0.9)^2 = 8.1 \times 10^{-5} \frac{5 \cdot 4}{2} = 0.00081$$

$$P(\{X = 3\}) = C_6^3 (0.1)^4 (0.9)^3 = 7.29 \times 10^{-5} \frac{6 \cdot 5 \cdot 4}{2 \cdot 3} = 0.001458 \text{ etc.}$$

$$M[X] = m \frac{q}{p} = 4 \frac{0.9}{0.1} = 36, \text{ deci cam la } 36 \text{ de mesaje se obțin patru mesaje eronate.}$$

5.2.4 V. a. repartizată geometric

Spre deosebire de variabila aleatoare binomială care apare când fixăm un număr de repetări ale unei experiențe A , în cazul variabilei geometrice numărăm numărul m de efectuări ale experienței până la prima realizare a evenimentului A (experiențe de tip Bernoulli, adică independente și repetate în aceleași condiții). În acest caz variabila aleatoare X ia valorile $X = k$, $k = 1, 2, \dots, m, \dots$ și avem

$$P(\{X = k\}) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots,$$

unde $P(A) = p$ este probabilitatea de realizare a evenimentului A . Vom avea deci

$$X : \begin{pmatrix} 1 & 2 & \dots & k & \dots \\ p & pq & \dots & pq^{k-1} & \dots \end{pmatrix}$$

Calculăm media și dispersia variabilei aleatoare. Se observă faptul că v.a. de tip geometric este un caz particular al v.a. binomială cu exponent negativ, pentru $m = 1$. Atunci

$$M[X] = \frac{1}{p},$$

$$D[X] = \frac{q}{p^2}.$$

Exercițiul 3 Calculatorul A transmite un mesaj calculatorului B . Mesajul este codat astfel încât B detectează când se produc erori în cursul transmiterii mesajului. Dacă B detectează o eroare cere calculatorului A retransmiterea mesajului. Dacă probabilitatea ca transmiterea mesajului să fie eronată este $p = 0,1$, care este probabilitatea ca mesajul să necesite mai mult de două transmisii.

Fiecare transmitere a mesajului urmează o repartiție binomială cu probabilitatea ca mesajul să fie corect $q = 1 - p$. Experiențele se repetă până când prima transmisie este corectă. Probabilitatea ca mesajul să necesite mai mult de două transmisii este

$$P(\{X > 2\}) = qp^2 + qp^3 + \dots = qp^2 \frac{1}{1-p} = p^2 = 10^{-2}.$$

Media variabilei aleatoare din acest exercițiu este $M[X] = \frac{1}{0.1} = 10$ iar dispersia este $D[X] = \frac{0.9}{0.01} = 90$

Interpretarea este următoarea: media aritmetică este 10, deci la un număr mare de transmisii numărul mediu de transmisii până la primul mesaj eronat este 10.

5.2.5 V. a. repartizată hipergeometric

Fie o mulțime care conține N obiecte, $k \leq N$ dintre ele considerate ca succes și $N - k$ considerate ca eșec. Un eșantion de dimensiune $n \leq k$ de obiecte este selectat aleator (fără revenirea obiectului în mulțime). Fie X variabila aleatoare care ia ca valori numărul de succese din eșantion.

$$X : \left(\begin{array}{ccccccc} 0 & 1 & 2 & \dots & i & \dots & n \\ \frac{C_k^0 C_{N-k}^n}{C_N^n} & \frac{C_k^1 C_{N-k}^{n-1}}{C_N^n} & \frac{C_k^2 C_{N-k}^{n-2}}{C_N^n} & \dots & \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} & \dots & \frac{C_k^n C_{N-k}^0}{C_N^n} \end{array} \right).$$

Vom scrie $X \in \text{Hip}[n, k, N]$. Caracteristicile numerice ale variabilei aleatoare X sunt

$$M[X] = \frac{kn}{N} = np, p = \frac{k}{N}.$$

$$D[X] = \frac{np(1-p)(N-n)}{N-1}$$

Media se poate calcula astfel: ca și în cazul binomial, variabila aleatoare hipergeometrică poate fi scrisă sub forma unei sume de variabile aleatoare. Să notăm cu X_1 numărul de succese la prima extragere, cu X_2 numărul de succese obținute la a doua extragere etc. Tabloul de repartiție al variabilei X_1 este

$$X_1 : \left(\begin{array}{cc} 0 & 1 \\ q & p \end{array} \right).$$

Vom arăta că și celelalte variabile aleatoare $X_i, i = \overline{2, n}$, au aceeași repartiție. Pentru aceasta calculăm probabilitatea obținerii unui succese la extracția de ordin i , atunci când se fac extrageri fără revenirea obiectului în mulțime. Numărul cazurilor posibile A_N^i . Numărul cazurilor favorabile coincide cu numărul de moduri în care pot fi obținute i obiecte astfel încât pe locul i să fie succes. Un succes poate fi obținut în k moduri.

Celelalte $i - 1$ locuri pot fi ocupate în A_{N-1}^{i-1} moduri, în total kA_{N-1}^{i-1} cazuri favorabile. Probabilitatea cătată va fi deci

$$\frac{kA_{N-1}^{i-1}}{A_N^i} = \frac{k(N-1)(N-2)\dots(N-i+1)}{N(N-1)\dots(N-i+1)} = \frac{k}{N} = p.$$

Rezultă că X_i are repartiția

$$X_i : \left(\begin{array}{cc} 0 & 1 \\ q & p \end{array} \right).$$

Din relațiile

$$X = X_1 + X_2 + \dots + X_n$$

și

$$M[X_1] = M[X_2] = \dots = M[X_n] = p,$$

rezultă

$$M[X] = np.$$

Ca și în cazul v. a. binomiale, X se scrie ca o sumă de variabile aleatoare având aceeași repartiție

$$\left(\begin{array}{cc} 0 & 1 \\ q & p \end{array} \right).$$

Deosebirea este esențială: în cazul repartiției hipergeometrice, variabilele aleatoare nu sunt independente. Din această cauză nu putem calcula dispersia variabilei aleatoare X ca suma dispersiilor variabilelor aleatoare X_i .

O altă variantă de calcul a mediei și dispersiei va fi prezentată în continuare.

Să observăm că, pentru orice $a, b \in \mathbb{N}$, avem

$$\begin{aligned} & \left(\sum_{i=0}^a C_a^i \cdot x^i \right) \cdot \left(\sum_{j=0}^b C_b^j \cdot x^j \right) = (1+x)^a (1+x)^b \\ & = (1+x)^{a+b} = \sum_{r=0}^{a+b} C_{a+b}^r \cdot x^r, \quad \forall x \in \mathbb{R}. \end{aligned}$$

Rezultă deci, identificând coeficienții lui x^r , egalitatea

$$C_{a+b}^r = \sum_{i=0}^r C_a^i \cdot C_b^{r-i}, \quad \forall r = \overline{0, a+b},$$

numită identitatea hipergeometrică (de aici și denumirea v.a.). Să calculăm media și dispersia:

$$\begin{aligned} M[X] &= \sum_{i=0}^n \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} \cdot i = \sum_{i=0}^n \frac{k!i}{i![(k-1)-(i-1)]!} \cdot \frac{C_{(N-1)-(k-1)}^{(n-1)-(i-1)}}{C_{N-1}^{n-1} \cdot \frac{N}{n}} \\ &= \frac{kn}{N} \cdot \sum_{i=1}^n C_{k-1}^{i-1} \cdot \frac{C_{(N-1)-(k-1)}^{(n-1)-(i-1)}}{C_{N-1}^{n-1}} = np \underbrace{\left(\sum_{i=0}^{n-1} \frac{C_{k-1}^i C_{(N-1)-(k-1)}^{(n-1)-i}}{C_{N-1}^{n-1}} \right)}_{=1} = np. \end{aligned}$$

De asemenea,

$$\begin{aligned} M[X^2] - M[X] &= M[X^2 - X] = \sum_{i=0}^n \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} \cdot i(i-1) \\ &= \sum_{i=2}^n \frac{k(k-1)(k-2)!i(i-1)}{i(i-1)(i-2)![(k-2)-(i-2)]!} \frac{C_{(N-2)-(k-2)}^{(n-2)-(i-2)}}{C_{N-2}^{n-2} \cdot \frac{N(N-1)}{n(n-1)}} \\ &= \frac{k(k-1)n(n-1)}{N(N-1)} \sum_{i=2}^n C_{k-2}^{i-2} \cdot \frac{C_{(N-2)-(k-2)}^{(n-2)-(i-2)}}{C_{N-2}^{n-2}} \\ &= \frac{k(k-1)n(n-1)}{N(N-1)} \underbrace{\left(\sum_{i=0}^{n-2} \frac{C_{k-2}^i C_{(N-2)-(k-2)}^{(n-2)-i}}{C_{N-2}^{n-2}} \right)}_{=1} = \frac{k(k-1)n(n-1)}{N(N-1)}. \end{aligned}$$

De aici,

$$\begin{aligned} D[X] &= M[X^2] - M[X] + M[X] - (M[X])^2 \\ &= np \frac{(k-1)(n-1)}{(N-1)} + np - n^2 p^2 \\ &= np \left(\frac{(k-1)(n-1)}{(N-1)} + 1 - \frac{nk}{N} \right) \\ &= np \frac{(1-p)(N-n)}{N-1}. \end{aligned}$$

Teorema 5.2.1 Dacă X este o variabilă aleatoare repartizată hipergeometric cu parametrii N, k, n , pentru care

$$P(\{X = i\}) = \frac{C_k^i C_{N-k}^{n-i}}{C_N^n},$$

atunci

$$\lim_{N \rightarrow \infty} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} = C_n^i p^i q^{n-i}, \quad (5.3)$$

unde $p = \frac{k}{N}$, $q = 1-p$ și deci pentru valori mari ale lui N , o variabilă aleatoare repartizată hipergeometric poate fi aproximată cu o variabilă repartizată binomială $\text{Binomial}(n, \frac{k}{N})$.

Demonstrație. Avem

$$\begin{aligned} & \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} = \\ &= \frac{k(k-1)\dots(k-i+1)(N-k)(N-k-1)\dots(N-k-n+i-1)}{i!(n-i)!} \frac{n!}{N\dots(N-n+1)} \\ &= \frac{n!}{i!(n-i)!} \frac{k(k-1)\dots(k-i+1)(N-k)(N-k-1)\dots(N-k-n+i-1)}{N\dots(N-n+1)} \\ &= C_n^i \frac{k}{N} \frac{k-1}{N} \dots \frac{k-i+1}{N} \frac{N-k}{N} \frac{N-k-1}{N} \dots \frac{N-k-n+i-1}{N} \frac{N^n}{N\dots(N-n+1)} \\ &= C_n^i \frac{k}{N} \frac{k-1}{N} \dots \frac{k-i+1}{N} \frac{N-k}{N} \frac{N-k-1}{N} \dots \frac{N-k-n+i-1}{N} \frac{N^n}{N\dots(N-n+1)} \\ &= C_n^i p(p - \frac{1}{N}) \dots (p - \frac{i-1}{N}) q(q - \frac{1}{N}) \dots (q - \frac{n-i+1}{N}) \frac{N^n}{N\dots(N-n+1)}. \end{aligned}$$

Trecând la limită obținem (5.3). □

5.2.6 V. a. repartizată Poisson

Distribuția Poisson se folosește ca model pentru evenimente a căror număr de repetare nu are o limită superioară. Spunem că variabila aleatoare X este repartizată Poisson, cu parametru λ , $\lambda \in \mathbb{R}_+$, $X \in \text{Poisson}[\lambda]$, dacă tabloul său de repartiție este

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & k & \dots \\ e^{-\lambda} & \frac{\lambda}{1!} e^{-\lambda} & \frac{\lambda^2}{2!} e^{-\lambda} & \dots & \frac{\lambda^k}{k!} e^{-\lambda} & \dots \end{pmatrix}.$$

Folosim notația $P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$. Evident $P(k; \lambda) > 0$ și $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$.

Proprietăți ale repartiției Poisson.

P1. Media repartiției Poisson este aceeași cu dispersia repartiției și egală cu λ . Aceasta este o particularitate remarcabilă și este o modalitate de a recunoaște că un fenomen urmează o repartiție Poisson. $M[X] = \lambda$, $D[X] = \lambda$. Într-adevăr

$$M[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$$

Pentru calculul dispersiei folosim formula $D[X] = M[X^2] - (M[X])^2$.

$$\begin{aligned} M[X^2] &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=2}^{\infty} (k^2 - k) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \\ &= \lambda^2 \sum_{k=2}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} + \lambda = \lambda^2 + \lambda, \\ D[X] &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

P2. Repartiția Poisson se obține ca un caz limită a repartiției binomiale când $n \rightarrow \infty$ și p scade astfel încât $np = \lambda = \text{const.}$ (în general, $n > 30, p \in (0; 0, 1)$).

Repartiția Poisson a apărut odată cu studiul variabilelor aleatoare cu un număr foarte mare de valori. Exemple clasice de astfel de variabile aleatoare sunt cele legate de **evenimentele care apar rar** într-un interval de timp, ca de exemplu:

- numărul de particule emise de o suprafață radioactivă într-un interval de timp;
- numărul accidentelor dintr-o fabrică în decurs de o săptămână;
- numărul mașinilor care trec printr-un anumit loc într-un interval de timp;
- numărul cererilor de I/O la o rețea de calculatoare într-un interval de timp.

Toate aceste procese se numesc Poisson și ele vor fi studiate în Capitolul Procese Stochastice. În aceste cazuri λ este proporțională cu durata intervalului de timp (adică $\lambda = tw$, w coeficient de proporționalitate). Parametrul $w = \frac{\lambda}{t}$ se mai numește parametru de intensitate, reprezentând media numărului de evenimente care se produc în unitatea de timp.

P3. Deoarece repartiția Poisson se întâlnește în cazul evenimentelor care se întâmplă rar (n mare, $np = \text{constant}$, rezultă p mic), ea mai poartă denumirea de **legea evenimentelor rare**.

Teorema 5.2.2 *Dacă X urmează o repartiție binomială atunci*

$$\lim_{n \rightarrow \infty} C_n^k p^k q^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda},$$

unde $np = \lambda = \text{const.}$

Demonstrație. Avem

$$\begin{aligned} \lim_{n \rightarrow \infty} C_n^k p^k q^{n-k} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square \end{aligned}$$

Observația 5.2.3 *Teorema afirmă că experimentele Bernoulli cu probabilitate de succes din ce în ce mai mică se apropie de legea Poisson de repartiție cu parametrul $\lambda = np$.*

Pentru calculul probabilităților $P(k, \lambda)$ s-au întocmit tabele pentru λ cuprins între 0, 1, ..., 20. Aceste tabele dau probabilitățile ca evenimentele să se realizeze de cel puțin m ori.

$$P(\{X \geq m\}) = \sum_{k=m}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Teorema 5.2.2 ne permite să aproximăm probabilitățile unei repartiții binomiale.

Exercițiul 4 Probabilitatea de eroare la transmiterea unui bit printr-o linie de comunicație este de 10^{-3} . Să se calculeze probabilitatea ca la transmiterea unui bloc de 1000 de biti să avem cinci sau mai mult de cinci erori.

Transmiterea fiecărui bit reprezintă o experiență Bernoulli care se realizează dacă s-a produs o eroare la transmiterea bitului. Probabilitatea ca să se producă k erori în 1000 de transmisii este dată de probabilitatea calculată cu legea binomială cu $n = 1000$ și $p = 10^{-3}$. Legea Poisson cu parametrul $\lambda = np = 1000 \cdot 10^{-3} = 1$ aproximează legea binomială. Astfel $P[N \geq 5] = 1 - P[N < 5] = 1 - \sum_{k=0}^4 \frac{1}{k!} e^{-1} = 1 - e^{-1} \{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\} = 0,00366$.

Exercițiul 5 O fabrică produce becuri și dă 2 % rebut. Să se aproximeze probabilitatea ca într-o cutie de 100 de becuri să obținem cel mult trei becuri defecte?

Presupunând independența, avem de-a face cu o variabilă aleatoare repartizată binomial cu $p = 0,02$ și $n = 100$. Dacă dorim să calculăm probabilitatea cu ajutorul schemei binomiale, după calcule laborioase obținem $\sum_{k=0}^3 C_{100}^k (0,02)^k (0,98)^{100-k} = 0,859$. Dacă aproximăm X cu o variabilă aleatoare repartizată Poisson,

$$\lambda = 100 \times 0,02 = 2, \sum_{k=0}^3 \frac{2^k e^{-2}}{k!} = 0,857.$$

Repartitia	p_i	$M[X]$	$D[X]$	Observații
binomială	$C_n^i p^i q^{n-i}$	np	npq	$\lim_{n \rightarrow \infty} C_n^i p^i q^{n-i} = \frac{\lambda^i}{i!} e^{-\lambda}$ $np = \lambda$
binomială cu exponent negativ	$C_{m+i-1}^{m-1} p^m q^i$	$\frac{mq}{p}$	$\frac{mq}{p^2}$	
hipergeometrică	$\frac{C_k^i C_{N-k}^{n-i}}{C_N^n}$	np	$\frac{np(1-p)(N-n)}{N-1}$	$\lim_{N \rightarrow \infty} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} = C_n^k p^k$
Poisson	$\frac{\lambda^i}{i!} e^{-\lambda}$	λ	λ	
geometrică	pq^{i-1}	$\frac{1}{p}$	$\frac{q}{p^2}$	