Curs 5

Variabile aleatoare

5.1 Funcţia generatoare

În problemele în care apar variabile aleatoare nenegative se utilizează funcţia generatoare a variabilei aleatoare X, $G_X(z)$ care se defineşte prin relaţia

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

unde $P(\{X = k\}) = p_k, k = 0, 1, 2, \ldots$. Deci cunoscând probabilităţile p_k putem determina funcţia 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 şi obţinem:

$$G_X^{(j)}(z) = \sum_{k=j}^{\infty} k(k-1)\ldots(k-j+1)p_k z^{k-j} = \sum_{k=j}^{\infty} C_k^j j! p_j z^{k-j}.$$

Dacă în relaţia de mai sus facem $z = 0$ obţinem

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

Aceasta ne arată că putem determina toate probabilităţile p_k cu ajutorul funcţiei generatoare. Din acestă cauză $G_X^{(j)}$ se numeşte funcţie generatoare. În particular, punând $z = 1$ în G'_X şi în G''_X obţinem:

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

5.2 Legi discrete de repartiţie

Vom prezenta legi teoretice de repartiţie a unor variabile aleatoare discrete. Există multe fenomene întâlnite în inginerie care "ascultă" de aceste legi în sensul că anumite selecţii de date experimentale sunt organizate după modele teoretice standard care trebuie cunoscute. Statisticianul apelează la rezultatele teoretice care îi orientează calculele şi 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 aceeaşi probabilitate. În acest caz, vom avea

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

Dacă presupunem că X ia valorile întregi $m, m+1, ..., m+n+1$, notăm $X \in \text{Uniform}[m,n]$ şi obţinem

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

5.2.2 V. a. repartizată binomial

Evenimentele A şi contrarul său, $\overline{A}$ se repetă de mai multe ori şi se produc cu aceeaşi probabilitate, $P(A) = p$ respectiv $P(\overline{A}) = q = 1 - p$. 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 evenimntului A are tabloul de repartiţie

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

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 experienţei evenimentul A s-a produs şi valoarea 0 cu probabilitatea q, adică la a i-a repetare a experienţei evenimentul A nu s-a produs

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

Cu notaţiile introduse, putem scrie

$$X = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^n X_i,$$
$$M[X] = M[\sum_{i=1}^n X_i] = \sum_{i=1}^n M[X_i] = np,$$
$$D[X] = D[\sum_{i=1}^n X_i] = \sum_{i=1}^n D[X_i],$$

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

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

Deoarece

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

avem

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

Exerciţiul 1 Arătaţi că dacă $X \in \text{Binomial}[n,p]$ atunci $G_X(z) = (pz+q)^n$ şi deduceţi cu ajutorul ei relaţiile $M[X] = np, D^2[X] = npq$.

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

-sunt numai două posibilităţi pentru rezultatul experienţei (succes/insucces, realizări/nerealizări, bit corect/eronat, aparat bun/defect etc.),

-probabilitatea succesului este aceeaşi pentru fiecare experienţă,

-sunt n repetări ale experienţei, unde n este constant,

-cele n experienţe sunt independente.

Facem observaţia că în calculul factorialelor 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. multinomială

O generalizare a distribuţiei binomiale este repartiţia obţinută pentru modelul următor.

În urma unui experiment se pot realiza k evenimente diferite independente $A_1, A_2, ..., A_k$. Aceste evenimente se produc cu probabilităţile $p_1, p_2, ...p_k$, $\sum\limits_{j=1}^{k} p_j = 1,\ 0 \le p_j \le 1$. Se repetă experienţa de n ori şi se constată că evenimentul $A_j, j = \overline{1,k}$ s-a realizat de n_j ori.

Teorema 5.2.1 *Dacă definim v. a. X_j =numărul de apariţii ale evenimentului A_j în cele n probe, atunci:*

$$P\left(X_1 = n_1, X_2 = n_2, ..., X_k = n_k\right) = \frac{n!}{n_1!n_2!...n_k!}\left(p_1\right)^{n_1}\left(p_2\right)^{n_2}...\left(p_k\right)^{n_k}$$

cu $n_1 + n_2 + ... + n_k = n$ şi $0 \le n_j \le n$.

Observaţia 5.2.2 *Repartiţia depinde de $k-1$ parametri.*

Spunem în acest caz că $\left(X_1, X_2, ..., X_k\right) \in \text{Multinomial}\left[n, p_1, p_2, ..., p_k\right]$.

Exerciţiul 2 presupunem că avem o urnă care conţine 10 bile roşii, 20 de bile albe şi 30 bile negre. Extragem 10 bile din urnă cu repunerea bilei extrase în urnă. Să se determine proobabilitatea ca să obţinem 3 bile roşii, 4 bile albe şi 3 bile negre.

Deoarece extragerea se face cu repunerea bilei extrase în urnă evenimentele sunt independente şi deci:

$$P\left(X_1 = 3, X_2 = 4, X_3 = 3\right) = \frac{10!}{3!4!3!}\left(\frac{10}{60}\right)^3\left(\frac{20}{60}\right)^4\left(\frac{30}{60}\right)^3 = 3.000\,7 \times 10^{-2}.$$

5.2.4 V. a. repartizată binomial cu exponent negativ

Facem observaţia că în cazul *variabilei aleatoare binomiale* se numără realizările evenimentului A în cazul în care se efectuează n experienţe. Deci *numărul de repetări ele experienţei este fixat iar numărul succeselor este aleator.*

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

Se spune că variabila aleatoare X are repartiţie binomială cu exponent negativ, cu parametrii m şi p, $(m = 1, 2, \ldots, 0 < p < 1)$, dacă X poate lua valorile $m, m+1, m+2, \ldots$ iar

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

Vom avea aşadar

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

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

Acestă repartiţie apare în modele de tipul următor: să presupunem că s-a efectuat un număr de observaţii independente, în cursul fiecărei observaţii evenimentul A (de exemplu nefuncţionarea unui subansamblu al unui aparat) poate să se producă cu probabilitatea p. Observaţiile continuă până când evenimentul se produce de m ori şi se caută probabilitatea pentru ca aceasta să se producă exact în decursul a $m + k$ observaţii. Evenimentul $\{X = k\}$ se scrie ca intersecţie a două evenimente: evenimenul B_1 care constă în faptul că în primele $m + k - 1$ observaţii evenimentul A se produce de $m - 1$ ori şi evenimentul B_2 care constă în faptul că în observaţia $m + k$ 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 repartiţii provine din faptul că probabilitatea căutată este coeficient în dezvoltarea în serie a lui

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

Justificăm această relaţie:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \text{ pentru } |x| < 1, \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} k x^{k-1},$$
$$\frac{2}{(1-x)^3} = \sum_{k=2}^{\infty} k(k-1) x^{k-2} \Rightarrow \frac{1}{(1-x)^3} = \sum_{k=2}^{\infty} C_k^2 x^{k-2}.$$

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

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

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

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

Calculăm funcţia generatoare.

$$\begin{aligned}
G_X(z) &= \sum_{k=0}^{\infty} C_{m+k-1}^{m-1} p^m q^k z^k = \left(\frac{1}{p} - \frac{zq}{p}\right)^{-m}, \\
G'_X(z) &= m\left(\frac{1}{p} - \frac{zq}{p}\right)^{-m-1} \frac{q}{p}, \\
M[X] &= G'_X(1) = m\frac{q}{p}, \\
G''_X(z) &= m\,(m+1)\left(\frac{1}{p} - \frac{zq}{p}\right)^{-m-1} \left(\frac{q}{p}\right)^2, \\
G''_X(1) &= m\,(m+1)\left(\frac{q}{p}\right)^2, \\
D[X] &= m\,(m+1)\left(\frac{q}{p}\right)^2 + m\frac{q}{p} - \left(m\frac{q}{p}\right)^2 = \frac{mq}{p^2}.
\end{aligned}$$

Exerciţiul 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. Presupunem că probabilitatea ca transmiterea mesajului să fie eronată este $p = 0,1$. Vrem să definim variabila aleatore X care ia ca valori numărul de mesaje care trebuie transmise pâna la obţinerea a patru mesaje eronate.

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

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

$P\left(\{X=0\}\right) = C_3^3\,(0.1)^4 = 0.000\,1,$

$P\left(\{X=1\}\right) = C_4^3\,(0.1)^4\,(0.9) = 0.000\,09 \cdot 4 = 0.000\,36$

$P\left(\{X=2\}\right) = C_5^3\,(0.1)^4\,(0.9)^2 = 8.1 \times 10^{-5}\frac{5\cdot 4}{2} = 0.000\,81$

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

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

5.2.5 V. a. repartizată geometric

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

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

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

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

Calculăm media şi dispersia variabilei aleatoare. Notăm $q = 1 - p$. Calculăm funcţia generatoare.

$$\begin{aligned} G_X(z) &= \sum_{k=1}^{\infty} p\,(qz)^k = \frac{pqz}{1-qz}, \\ G'_X(z) &= \frac{p}{(1-qz)^2}, \quad G'_X(1) = \frac{1}{p}, \\ G''_X(z) &= \frac{2pq}{(1-qz)^3}, \quad G''_X(1) = \frac{2q}{p^2}, \\ D[X] &= \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}. \end{aligned}$$

Exerciţiul 4 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 repartiţie binomială cu probabilitatea ca mesajul să fie corect $q = 1 - p$. Experienţele se repetă până când prima transmisie este corectă. Probabilitatea ca mesajul să necesite mai mult de două transmisii este

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

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

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.

5.2.6 V. a. repartizată hipergeometric

Fie o mulţime care conţine N obiecte, $k \leq N$ dintre ele considerate ca succes şi $N - k$ considerate ca eşec. Un eşantion de dimensiune $n \leq k$ de obiecte este selectat aleator (fără revenirea obiectului în mulţime). Fie X variabila aleatoare care ia ca valori numărul de succese din eşantion.

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

Vom scrie $X \in \mathrm{Hip}[n, k, N]$. 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 şi 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 obţinute la a doua extragere etc. Tabloul de repartiţie al variabilei X_1 este

$$X_1 : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Vom arăta că şi celelalte variabile aleatoare $X_i, i = \overline{2, n}$, au aceeaşi repartiţie. Pentru aceasta calculăm probabilitatea obţinerii unui succese la extracţia de ordin i, atunci când se fac extrageri fără revenirea obiectului în mulţime. Numărul cazurilor posibile A_N^i. Numărul cazurilor favorabile coincide cu numărul de moduri în care pot fi obţinute i obiecte astfel încât pe locul i să fie succes. Un succes poate fi obţinut 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 repartiţia

$$X_i : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Din relaţiile

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

şi

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

rezultă

$$M[X] = np.$$

Ca şi în cazul v. a. binomiale, X se scrie ca o sumă de variabile aleatoare având aceeaşi repartiţie

$$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Deosebirea este esenţială: în cazul repartiţiei 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 şi dispersiei va fi prezentată în continuare.

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

$$\left(\sum_{i=0}^{a} C_a^i \cdot x^i\right) \cdot \left(\sum_{j=0}^{b} C_b^i \cdot x^j\right) = (1+x)^a (1+x)^b$$

$$= (1+x)^{a+b} = \sum_{r=0}^{a+b} C_{a+b}^r \cdot x^r, \quad \forall x \in \mathbb{R}.$$

Rezultă deci, identificând coeficienţii 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 şi denumirea v.a.). Să calculăm media şi dispersia:

$$M[X] = \sum_{i=0}^{n} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} \cdot i = \sum_{i=0}^{n} \frac{k! i}{i![(k-1)-(i-1)]!} \cdot \frac{C_{(N-1)-(k-1)}^{(n-1)-(i-1)}}{C_{N-1}^{n-1} \cdot \frac{N}{n}}$$

$$= \frac{kn}{N} \cdot \sum_{i=1}^{n} C_{k-1}^{i-1} \cdot \frac{C_{(N-1)-(k-1)}^{(n-1)-(i-1)}}{C_{N-1}^{n-1}} = np \underbrace{\left(\sum_{i=0}^{n-1} \frac{C_{k-1}^{i} C_{(N-1)-(k-1)}^{(n-1)-i}}{C_{N-1}^{n-1}} \right)}_{=1} = np.$$

De asemenea,

$$M[X^2] - M[X] = M[X^2 - X] = \sum_{i=0}^{n} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} \cdot i(i-1)$$

$$= \sum_{i=2}^{n} \frac{k(k-1)(k-2)! i(i-1)}{i(i-1)(i-2)![(k-2)-(i-2)]!} \frac{C_{(N-2)-(k-2)}^{(n-2)-(i-2)}}{C_{N-2}^{n-2} \cdot \frac{N(N-1)}{n(n-1)}}$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)} \sum_{i=2}^{n} C_{k-2}^{i-1} \cdot \frac{C_{(N-2)-(k-2)}^{(n-2)-(i-2)}}{C_{N-2}^{n-2}}$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)} \underbrace{\left(\sum_{i=0}^{n-2} \frac{C_{k-2}^{i} C_{(N-2)-(k-2)}^{(n-2)-i}}{C_{N-2}^{n-2}} \right)}_{=1} = \frac{k(k-1)n(n-1)}{N(N-1)}.$$

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 5.2.3 *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

$$\lim_{N \to \infty} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} = C_n^i p^i q^{n-i}, \tag{5.3}$$

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

Demonstraţie. Avem

$$\frac{C_k^i C_{N-k}^{n-i}}{C_N^n} =$$

$$= \frac{k(k-1)\ldots(k-i+1)}{i!} \frac{(N-k)(N-k-1)\ldots(N-k-n+i-1)}{(n-i)!} \frac{n!}{N\ldots(N-n+1)}$$

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

Trecând la limită obţinem (5.3). □

5.2.7 V. a. repartizată Poisson

Distribuţia Poisson se foloseşte ca model pentru evenimente a căror număr de repetare nu are o limită superioară. Spunem că variabila aleatoare X este repartizată Poisson, cu paramatru $\lambda, \lambda \in \mathbb{R}_+$, $X \in \text{Poisson}\left[\lambda\right]$, dacă tabloul său de repartiţie 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 notaţia $P(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$. Evident $P(k;\lambda) > 0$ şi $\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!} = 1$.

Proprietăţi ale repartiţiei Poisson.

P1. Media repartiţiei Poisson este aceeaşi cu dispersia repartiţiei şi egală cu λ. Aceasta este o particularitate remarcabilă şi este o modalitate de a recunoaşte că un fenomen urmează o repartiţie 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 $D[X] = M[X^2] - (M[X])^2$.

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

P2. Repartiţia Poisson se obţine ca un caz limită a repartiţiei binomiale când $n \to \infty$ şi p scade astfel încât $np = \lambda = \text{const.}$(în general, $n > 30, p \in (0; 0,1)$).

Repartiţia 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 suprafaţă radioactivă într-un interval de timp;

- numărul accidentelor dintr-o fabrică în decurs de o săptămână;

- numărul maşinilor care trec printr-un anumit loc într-un interval de timp;

- numărul cererilor de I/O la o reţea de calculatoare într-un interval de timp.

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

P3. Deoarece repartiţia Poisson se întâlneşte în cazul evenimentelor care se întîmplă rar (n mare, $np =$ constant, rezultă p mic), ea mai poartă denumirea de **legea evenimentelor rare**.

Teorema 5.2.4 *Dacă X urmează o repartiţie binomială atunci*

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

unde $np = \lambda = const.$

Demonstraţie. Avem

$$\lim_{n\to\infty} C_n^k p^k q^{n-k} = \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} (\frac{\lambda}{n})^k (1-\frac{\lambda}{n})^{n-k}$$
$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} (1-\frac{\lambda}{n})^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square$$

Observaţia 5.2.5 *Teorema afirmă că experimentele Bernoulli cu probabilitate de succes din ce în ce mai mică se apropie de legea Poisson de repartiţie cu parametul $\lambda = np$.*

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

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

Teorema 5.2.4 ne permite să aproximăm probabilităţile unei repartiţii binomiale.

Exerciţiul 5 Probabilitatea de eroare la transmiterea unui bit printr-o linie de comunicaţie 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 experienţă 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 $n = 1000$ şi $p = 10^{-3}$. Legea Poisson cu parametrul $\lambda = np = 1000 \cdot 10^{-3} = 1$ aproximează legea binomială. Astfel $P[N \geq 5] = 1 - P[N < 5] = 1 - \sum_{k=0}^{4} \frac{1}{k!} e^{-1} = 1 - e^{-1}\{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\} = 0,00366.$

Exerciţiul 6 O fabrică produce becuri şi dă 2 % rebut. Să se aproximeze probabilitatea ca într-o cutie de 100 de becuri să obţinem cel mult trei becuri defecte?

Presupunând indepependenţa, avem de-a face cu o variabilă aleatoare repatizată binomial cu $p = 0,02$ şi $n = 100$. Dacă dorim să calculăm probabilitatea cu ajutorul schemei binomiale, după calcule laborioase obţinem $\sum_{k=0}^{3} C_{100}^k (0,2)^k (0,98)^{100-k} = 0,859$.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	$M[X]$	$D[X]$	Observaţii
binomială	$C_n^i p^i q^{n-i}$	np	npq	$\lim_{n\to\infty} C_n^i p^i q^{n-i} =$ $= \frac{\lambda^i}{i!} e^{-\lambda}$ $np = \lambda$
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	$\frac{np(1-p)(N-n)}{N-1}$	$\lim_{N\to\infty} \frac{C_k^i C_{N-k}^{n-i}}{C_N^n} = C_n^k p^k$
Poisson	$\frac{\lambda^i}{i!} e^{-\lambda}$	λ	λ	
geometrică	pq^{i-1}	$\frac{1}{p}$	$\frac{q}{p^2}$	