1.7 Variabile aleatoare discrete

Numim variabilă aleatoare discretă o variabilă aleatoare ce poate lua un număr cel mult numărabil de valori (adică fie un număr finit de valori, fie o mulţime de valori numărabilă, adică ale cărei elemente formează un şir).

Notând cu $x_1, x_2, \ldots$ valorile posibile ale variabilei aleatoare X şi cu $p_i = P(X = x_i)$, $i = 1, 2, \ldots$, probabilităţile corespunzătoare, reprezentăm variabila aleatoare discretă X sub forma

$$X = \begin{pmatrix} x_1 & x_2 & \ldots \\ p_1 & p_2 & \ldots \end{pmatrix}.$$

Observaţia 1.7.1 *Cum variabila aleatoare X ia numai valorile $x_1, x_2, \ldots$ cu probabilităţile corespunzătoare $p_1, p_2, \ldots$, rezultă că avem*

$$\sum_{i \geq 1} p_i = 1$$

şi

$$P(X = x) = 0, \qquad \textit{oricare ar fi } x \in \mathbb{R} - \{x_1, x_2, \ldots\}.$$

Graficul funcţiei de distribuţie F_X a unei variabile aleatoare discrete X este o funcţie în scară, ce are discontinuităţi la dreapta in punctele x_i (salturi egale cu p_i în aceste puncte).

Câteva caracteristici numerice ale unei variabile aleatoare discrete:

- Media: este suma valorilor luate de variabila aleatoare înmulţite cu probabilităţile corespunzătoare:

  $$M(X) = \sum_{i \geq 1} x_i P(X = x_i) = \sum_{i \geq 1} x_i p_i,$$

  dacă această serie este absolut convergentă (în caz contrar, seria fie nu este convergentă, fie valoarea sumei se poate schimba dacă se rearanjează termenii $x_1, x_2, \ldots$ într-o altă ordine);

- Momentul de ordin r: este media variabilei aleatoare X^r, adică

  $$M_r(X) = M(X^r) = \sum_{i \geq 1} x_i^r p_i;$$

- Momentul centrat de ordin r: este media variabilei aleatoare $(X - M(X))^r$

  $$\mu_r(X) = M\left((X - M(X))^r\right) = \sum_{i \geq 1} (x_i - M(X))^r p_i;$$

- Dispersia: este momentul centrat de ordin doi, şi se mai notează

  $$\sigma^2(X) = \mu_2(X) = \sum_{i \geq 1} (x_i - M(X))^2 p_i;$$

- Media de ordin r:

  $$m_r = (M_r(X))^{1/r}.$$

1.7.1 Variabila aleatoare uniformă

Modelează efectuarea unui experiment ce are ca rezultat un un număr finit de posibilităţi egal probabile, spre exemplu aruncarea unei monede, a unui zar, jocul la ruletă, etc.

O variabilă aleatoare uniformă ce ia valorile $1, \ldots, n$ $(n \geq 1)$ este de forma

$$X = \begin{pmatrix} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{n}{1} \end{pmatrix}.$$

Media şi dispersia variabilei aleatoare uniforme sunt

$$M(X) = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2},$$

respectiv

$$\begin{aligned} \sigma^2(X) &= \sum_{i=1}^{n} \left(i - \frac{n+1}{2}\right)^2 \cdot \frac{1}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} i^2 - \frac{n+1}{n}\sum_{i=1}^{n} i + \frac{1}{n}\sum_{i=1}^{n} \frac{(n+1)^2}{4} \\ &= \frac{1}{n}\frac{n(n+1)(2n+1)}{6} - \frac{n+1}{n}\frac{n(n+1)}{2} + \frac{1}{n}n\frac{(n+1)^2}{4} \\ &= \frac{n^2-1}{12}. \end{aligned}$$

1.7.2 Variabila aleatoare Bernoulli

Numim *experiment Berrnoulli* un experiment ce poate rezulta în unul din două cazuri posibile, pe care le vom numi *succes (notat prin* 1), respectiv *insucces* (notat prin 0).

Notând prin $p = P(X = 1)$ probabilitatea succesului (şi deci $q = 1 - p = P(X = 0)$ este probabilitatea insuccesului), o variabilă aleatoare Bernoulli se poate reprezenta sub forma

$$X = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \qquad p \in (0,1).$$

Media şi dispersia sunt

$$M(X) = 0 \cdot (1-p) + 1 \cdot p = p,$$

şi respectiv

$$\sigma^2(X) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p) = pq.$$

1.7.3 Variabila aleatoare binomială

O variabilă aleatoare X reprezentând numărul de succese obşinute în n repetări independente ale unui experiment Bernoulli în care probabilitatea succesului este p, se numeşte variabilă aleatoare binomială cu parametrii n şi p:

$$X = \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix},$$

unde $p_i = P(X = i) = C_n^i p^i (1-p)^{n-i}$, este probabilitatea obţinerii a i succese (şi a $n - i$ insuccese), $i = 0, 1, \dots, n$.

Observaţia 1.7.2 *Să observăm că are loc relaţia*

$$\sum_{i=0}^{n} p_i = \sum_{i=0}^{n} C_n^i p^i (1-p)^{n-i} = (p + (1-p))^n = 1.$$

Media şi dispersia sunt date de

$$M(X) = \sum_{i=0}^{n} i C_n^i p^i (1-p)^{n-i} = np,$$

şi respectiv

$$\sigma^2(X) = \sum_{i=0}^{n} (i - np)^2 C_n^i p^i (1-p)^{n-i} = np(1-p) = npq.$$

Are loc următoarea:

Propoziţia 1.7.3 *Dacă X este o variabilă aleatoare cu parametrii n şi p, atunci pentru $i = 0, 1, \dots, n$, $P(X = i)$ creşte, apoi descreşte, având un maxim pentru $i = [(n+1)p]$ (cel mai mare număr întreg mai mic sau cel mult egal cu $(n+1)p$).*

Demonstraţie. Avem:

$$\frac{P(X = i)}{P(X = i-1)} = \frac{C_n^i p^i (1-p)^{n-i}}{C_n^{i-1} p^{i-1} (1-p)^{n-i+1}} = \frac{(n-i+1)p}{i(1-p)} \geq 1$$

dacă şi numai dacă

$$i \leq (n+1)p,$$

şi deci

$$P(X = i) \leq P(X = i-1) \quad \Leftrightarrow \quad i \leq (n+1)p,$$

ceea ce arată că $P(X = i)$ creşte, apoi descreşte, având un maxim pentru $i = [(n+1)p]$. ∎

Exemplul 1.7.4 *Se aruncă* 4 *monede, şi presupunem că aruncările sunt independente, în fiecare din ele probabilitatea de apariţie a stemei fiind* $1/4$*. Atunci variabila aleatoare* X *ce reprezintă numărul de steme obţinut este o variabilă aleatoare binomială (ea dă numărul de succese obţinut, apariţia stemei fiind considerată ca fiind un succes) cu parametrii* $n = 4$ *şi* $p = 1/2$*, si deci avem*

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ p_0 & p_1 & p_2 & p_3 & p_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \end{pmatrix},$$

unde $p_k = C_4^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k}$*,* $k = 0, \ldots, 4$.

1.7.4 Variabila aleatoare geometrică

Numim variabil[aleatoare geometriocă cu parametrul $p \in (0, 1)$ o variabilă aleatoare reprezentând numărul de încercări efectuate într+un şir de experimente Bernoulli independente, având fiecare probabilitatea succesului egală cu p, până la obţinerea primului succes.

Variabila aleatoare geometrică este deci

$$X = \begin{pmatrix} 1 & 2 & 3 & \ldots \\ p_1 & p_2 & p_3 & \ldots \end{pmatrix},$$

unde $p_k = P(X = k) = (1 - p)^{k-1} \cdot p$ (deoarece primele $k - 1$ încercări sunt insuccese iar încercarea k este succes), $k \geq 1$.

Observaţia 1.7.5 *Relaţia* $\sum_{k=1}^{\infty} p_k = 1$ *este verificată:*

$$\begin{aligned} \sum_{k=1}^{\infty} p_k &= \sum_{k=1}^{\infty} (1 - p)^{k-1} p \\ &= p\left(1 + (1 - p) + (1 - p)^2 + \ldots\right) \\ &= p\frac{1}{1 - (1 - p)} \\ &= p\frac{1}{p} \\ &= 1. \end{aligned}$$

Exemplul 1.7.6 *Se aruncă în mod repetat un zar. Dacă* X *reprezintă numărul de aruncări până la prima apariţie a feţei* 1*, atunci* X *este o variabilă aleatoare geometrică cu paramaterul* $p = 1/6$ *(probabilitatea de apariţie a succesului), şi spre exemplu probabilitatea ca faţa* 1 *să apară pentru prima dată la a treia încercarea este*

$$P(X = 3) = \left(1 - \frac{1}{6}\right)^{3-1} \cdot \frac{1}{6} = \frac{25}{216} \approx 0.1157.$$

Media şi dispersia variabilei aleatoare geometrice cu parametrul p sunt

$$M(X) = \sum_{k=1}^{\infty} k \cdot p_k = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = \frac{1}{p},$$

respectiv

$$\sigma^2(X) = \sum_{k=1}^{\infty} \left(k - \frac{1}{p}\right)^2 \cdot p_k = \sum_{k=1}^{\infty} \left(k - \frac{1}{p}\right)^2 \cdot (1-p)^{k-1} p = \frac{1-p}{p^2}.$$

Propoziţia 1.7.7 ("Lipsa de memorie" a v.a. geometrice) *Dacă X este o variabilă aleatoare geometrică cu parametrul p, atunci*

$$P(X = n+k \mid X > n) = P(X = k),$$

pentru orice $k, n \geq 1$.

Demonstraţie. Avem:

$$\begin{aligned} P(X = n+k \mid X > n) &= \frac{P(X = n+k,\ X > n)}{P(X > n)} = \frac{P(X = n+k)}{1 - P(X \leq n)} \\ &= \frac{(1-p)^{n+k-1} p}{1 - \sum_{i=1}^{n} P(X = i)} = \frac{(1-p)^{n+k-1} p}{1 - \sum_{i=1}^{n} (1-p)^{i-1} p} \\ &= \frac{(1-p)^{n+k-1} p}{1 - p\frac{(1-p)^n - 1}{1-p-1}} = \frac{(1-p)^{n+k-1} p}{(1-p)^n} = (1-p)^{k-1} p \\ &= P(X = k). \end{aligned}$$

■

Observaţia 1.7.8 *Spre exemplu, propoziţia anterioară arată ca dacă X reprezintă numărul de aruncări ale unui ban până la prima apariţie a stemei, atunci probabilitatea de apariţie a stemei pentru prima dată la încercarea* 106*, condiţionat de faptul că ea nu a apărut în primele* 100 *de aruncări, este aceeaşi cu probabilitatea de apariţie a stemei la încercarea* $106 - 100 = 6$*:*

$$P(X = n + 106 \mid X > 100) = P(X = 6),$$

şi deci variabila aleatoare geometrică nu are "memoria" primelor 100 *de insuccese (adică a feţei "ban").*

Se poate arăta că variabila aleatoare geometrică este singura variabilă aleatoare discretă ce are această proprietate de lipsă a memoriei.

1.7.5 Variabila aleatoare binomială negativă

Numim variabilă aleatoare negativă cu parametrii $r \in \mathbb{N}^*$ şi $p \in (0, 1)$ o variabilă aleatoare reprezentând numărul de experimente Bernoulli independente efectuate (in fiecare din ele probabilitatea succesului fiind p), până la apariţia a r succese.

Variabila aleatoare binomială cu parametrii r şi p este deci:

$$X = \begin{pmatrix} r & r+1 & r+2 & \dots \\ p_r & p_{r+1} & p_{r+2} & \dots \end{pmatrix},$$

unde $p_k = P(X = k) = C_{k-1}^{r-1} (1-p)^{k-r} p^r$, $k = r, r+1, \dots$

Observaţia 1.7.9 *Pentru* $r = 1$ *se obţine variabila aleatoare geometrică, şi deci variabila aleatoare binomială negativă este o generalizare a variabilei aleatoare geometrice.*

Observaţia 1.7.10 *numele de variabilă aleatoare* binomială negativă *provine din faptul că pentru variabila aleatoare binomială numărul de încercări este determinat (fixat) şi cel al succeselor este aleator, pe când pentru variabila aleatoare binomială negativă numărul de succese este determinat iar cel al încercărilor este aleator, în acest sens ea fiind opusă ("negativă") variabilei aleatoare binomiale, fapt ce justifică acest nume (variabila aleatoare binomială negativă).*

Se poate arăta că media şi dispersia unei variabile aleatoare binomiale negative cu parametrii $r \in \mathbb{N}^*$ şi $p \in (0,1)$ sunt date de

$$M(X) = r\frac{1-p}{p},$$

respectiv

$$\sigma^2(X) = r\frac{1-p}{p^2}.$$

1.7.6 Variabila aleatoare Poisson

Numim variabilă aleatoare Poisson o variabilă aleatoare ce ia valorile $0, 1, 2, \dots$ cu probabilităţile $p_0, p_1, p_2, \dots$, unde $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \dots$, adică

$$X = \begin{pmatrix} 0 & 1 & 2 & \dots \\ p_0 & p_1 & p_2 & \dots \end{pmatrix}.$$

Observaţia 1.7.11 *Relaţia* $\sum_{k=1}^{\infty} p_k = 1$ *este verificată, deoarece*

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots\right) = e^{-\lambda}e^{\lambda} = 1.$$

Variabila aleatoare Poisson este foarte utilă în practică, deoarece ea poate fi folosită ca o aproximaţie a variabilei aleatoare binomiale cu parametrii n şi p, unde relaţia de legătură între parametrii este $\lambda = np$.

Astfel, pentru n "mare" şi p "mic" astfel încât $\lambda = np$ are o valoare moderată, atunci dacă Y este o variabilă aleatoare binomială cu parametrii n şi p, avem:

$$\begin{aligned}
P(Y=k) &= C_n^k (1-p)^{n-k} p^k = \frac{n!}{(n-k)!k!}\left(1-\frac{\lambda}{n}\right)^{n-k}\frac{\lambda^k}{n^k} \\
&= \frac{n(n-1)\dots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n}\frac{1}{\left(1-\frac{\lambda}{n}\right)^k} \\
&\approx 1\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1 \\
&= P(X=k),
\end{aligned}$$

şi deci putem aproxima variabila aleatoare binomială Y (cu parametrii n şi p) prin variabila aleatoare Poisson X (cu parametrul $\lambda = np$).

Exemplul 1.7.12 *Exemple de variabile aleatoare binomiale ce por fi aproximate prin variabile aleatoare Poisson: numărul de greşeli tipografice de pe o pagină, numărul de greşeli telefonice dintr-o zi, numărul de cutremure de pământ dintr-o anumită perioadă de timp, numărul de clienţi ce intră într-un magazin (pentru toate aceste exemple, probabilitatea p a succesului este mică, iar n – numărul de încercări efectuate este mare, astfel încât $\lambda = np$ are o valoare moderată).*

Se poate arăta că media şi dispersia unei variabile aleatoare Poisson cu parametrul λ sunt

$$M(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = \lambda,$$

respectiv

$$\sigma^2(X) = \sum_{k=0}^{\infty}(k-\lambda)^2 p_k = \sum_{k=0}^{\infty}(k-\lambda)^2 e^{-\lambda}\frac{\lambda^k}{k!} = \lambda.$$

1.7.7 Exerciţii

1. Într-un şir de încercări Bernoulli cu probabilitatea succesului $p \in (0,1)$, care este probabilitatea apariţiei a n succese înainte de apariţia a m insuccese?

2. Din două cutii de chibrituri, fiecare conţinând N beţe, se scot independent, cu probabilităţi egale, câte un chibrit. Care este probabilitatea ca atunci când una din cutii este goală, cealaltă să conţină k chibrituri? ($k = 0, 1, \dots, N$)

3. Fie X o variabilă aleatoare Poisson cu parametrul $\lambda > 0$. Să se arate că probabilitatea $P(X=n)$ creşte, apoi descreşte, atunci când $n = 0, 1, 2, \dots$. Pentru ce valoare a lui n este probabilitatea $P(X=n)$ maximă?

4. O familie are $n \geq 1$ copii cu probabilitate αp^n, unde $p \in (0,1)$ şi $\alpha \leq \frac{1-p}{p}$.

(a) Ce procent de familii nu au nici un copil?

(b) Dacă probabilitatea unui copil de a fi băiat este 1/2 şi este independentă de sexul celorlalţi copii din familie, ce procent de familii au exact k băieţi (şi orice număr de fete)?

5. O urnă conţine N bile numerotate $1, 2, \ldots, N$. Dacă din urnă se extrag (fără întoarcere în urnă) $n \leq N$ bile, să se determine probabilitatea $P(X = k)$, unde X este variabila aleatoare reprezentând maximul numerelor extrase.

6. Să se compare aproximaţia variabilei aleatoare binomiale (cu parametrii indicaţi) prin variabila aleatoare Poisson (cu parametrul corespunzător), în cazurile:

   (a) $P(X = 2)$, în cazul $n = 8$, $p = 0.1$;

   (b) $P(X = 9)$, în cazul $n = 10$, $p = 0.95$;

   (c) $P(X = 0)$, în cazul $n = 10$, $p = 0.1$;

   (d) $P(X = 4)$, în cazul $n = 9$, $p = 0.2$;

7. O persoană cumpără 50 de lozuri, pentru fiecare probabilitatea de câştig fiind 1/100. Care este probabilitatea obţinerii

   (a) Măcar a unui loz câştigător?

   (b) Exact a unui loz câştigător?

   (c) A nici unui loz câştigător?

8. Numărul de ori de care o persoană răceşte într-un an este o variabilă aleatoare Poisson cu parametrul $\lambda = 5$. Un medicament reduce acest parametru la $\lambda = 3$ pentru 75% din populaţie şi este ineficient pentru restul de 25% al populaţiei.

   O persoană, căreia i s-a administrat medicamentul, a răcit de două ori într-un an. Cât de probabil este că medicamentul a fost eficient pentru ea?

9. Oamenii intră într-un magazin cu rata de 1 persoană / 2 minute. Folosind aproximaţia prin variabilă aleatoare Poisson (cu parametru corspunzător), să se determine:

   (a) Care este probabilitatea ca nimeni să nu intre în magazin în intervalul $12:00 - 12:05$?

   (b) Cel puţin 4 persoane să intre în magazin în acest interval?

10. O ruletă este marcată cu numerele $1, 2, \ldots, 36$. Dacă o persoană pariază pe numerele $1 - 12$, care este probabilitatea de a pierde în primele 5 încercări? Dar de a câştiga pentru prima oară în a 4-a încercare?

11. O monedă se aruncă repeptat, până la cea de-a 10-a apariţie a feţei ban. Dacă X este variabila aleatoare ce reprezintă numărul de feţe stemă obţinut, să se determine probabilitatea $P(X = n)$, $n = 0, 1, 2, \ldots$.