1 Дискретни случайни величини и неравенство на Чебишев

Нека n, M и N са естествени числа, $\max(n, M) \leq N$, и $(\Omega, \mathfrak{A}, \mathbf{P})$ е вероятностното пространство на експеримент $\mathcal{E}$.

Дефиниция 1.1. *Ще казваме, че случайната величина X е хипергеометрично разпределена с параметри n, M и N, което ще записваме чрез $X \in HG(n, M, N)$, ако $X(\Omega) = \{0, 1, \ldots, \min(n, M)\}$ и тегловата функция на X има вида $k \longmapsto \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, за $k = 0, 1, \ldots, \min(n, M)$.*

Една възможна интерпретация на хипергеометрично разпределена случайна величина с параметри n, M и N, дава следната схема: дадени са N обекта, като точно M от тях имат фиксирано свойство P. Избираме произволни n обекта измежду дадените N. Вероятността на събитието "точно k измежду избраните n да имат свойство P" - се дава чрез функцията

$$k \longmapsto \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Дефиниция 1.2. *Ще казваме, че случайната величина X е геометрично разпределена с параметър $p \in (0, 1)$, което ще записваме чрез $X \in G(p)$, ако $X(\Omega) = \{1, 2 \ldots, \}$ или $X(\Omega) = \{0, 1, \ldots, \}$ и тегловата функция на X има съответно вида $k \longmapsto (1-p)^{k-1}p$ и $k \longmapsto (1-p)^k p$, $k = 0, 1, \ldots$*

Интерпретация на геометрично разпределена случайна величина с параметър p, дава биномна схема с параметри (∞, p), като вероятността на събитието "първия успех настъпва на $k-$тия опит" е равна на $(1-p)^{k-1}p$. Аналогично, вероятността на събитието "k неуспеха (брой неуспехи) до първи успех" е равна на $(1-p)^k p$.

Пример 1.3. *Нека $X \in G(p)$ е с теглова функция:*

а) $k \longmapsto (1-p)^k p$, $k = 0, 1, \ldots$;

б) $k \longmapsto (1-p)^{k-1}p$, $k = 1, 2, \ldots$

Да се докаже, че $\mathbf{E}X = \frac{1-p}{p}$, $\mathbf{D}X = \frac{1-p}{p^2}$ за подточка а); $\mathbf{E}X = \frac{1}{p}$, $\mathbf{D}X = \frac{1-p}{p^2}$ за подточка б).

Пример 1.4. *(Многомерно хипергеометрично разпределение): Нека $n, k, m_1, m_2, \ldots, m_k, N$ са естествени числа. Дадени са N обекта, като точно m_1 от тях имат свойство P_1, точно m_2 от тях имат свойство P_2, ... , точно m_k от тях имат свойство P_k, като $m_1 + \cdots + m_k = N$ и всеки обект има точно едно свойство. Избираме произволни n обекта измежду дадените N. Вероятността на събитието - точно s_1 измежду избраните n имат свойство P_1, точно s_2 имат свойство P_2, ... , точно s_k имат свойство P_k, като $s_1 + \cdots + s_k = n$ - е равна на*

$$\frac{\binom{m_1}{s_1}\binom{m_2}{s_2}\cdots\binom{m_k}{s_k}}{\binom{N}{n}}.$$

Дефиниция 1.5. *Ще казваме, че случайната величина X е Поасоново разпределена с параметър $\lambda \in \mathbb{R}^+$, което ще записваме чрез $X \in Po(\lambda)$, ако $X(\Omega) = \{0, 1, \ldots, \}$ и тегловата функция на X има вида $k \longmapsto \frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, \ldots$*

Интерпретация на Поасоново разпределена случайна величина с параметър λ ще дадем, когато разгледаме непрекъснати случайни величини, конкретно - експоненциално разпределена случайна величина с параметър λ, и стохастични процеси - поасонов процес. Поасоново разпределение $\mathrm{Po}(\lambda)$ се използва в случаите, когато за дадено събитие A е известено

- средният брой настъпванията на A, за всеки единичен времеви интервал е константа$(= \lambda)$
- всяко настъпване на A е независимо от предходните, тоест предисторията на процеса не влияе на текущото му състояние

В този случай, вероятността събитието A да настъпи точно k пъти за единица време е $\frac{e^{-\lambda}\lambda^k}{k!}$.

Нека $(\Omega, \mathfrak{A}, \mathbf{P})$ е вероятностно пространство на експеримента $\mathcal{E}$ и $X : \Omega \to \mathbb{R}$, $\;Y : \Omega \to \mathbb{R}$ са дискретни случайни величини.

Теорема 1.6. *Сума на краен брой дискретни случайни величини върху Ω е дискретна случайна величина. Произведените на краен брой дискретни случайни величини е дискретна случайна величина. Произведение на дискретна случайна величина с реално число също е дискретна случайна величина. По-общо, ако $g : \mathbb{R} \to \mathbb{R}$ е "произволна"(Борелова) функция, то $g \circ X$ е дискретна случайна величина.*

Директно следствие от теорема 1.6 е, че множеството от дискретните случайни величини върху Ω снабдено с операциите сума и произведение е алгебра , в частност - линейно пространство над $\mathbb{R}$. Нека $X : \Omega \to \mathbb{R}$ е дискретна случайна величина със средно EX. Изображението $I : \Omega \to \{1\}$, $\omega \longmapsto 1$ е дискретна случайна величина, тогава rI също е дискретна случайна величина за всяко $r \in \mathbb{R}$. Така $X - (EX)I$ е дискретна случайна величина, която ще записваме чрез $X - EX$. Ако съществува числото $E((X - EX)^2)$, то се нарича вариация на X и се означава с DX. Понеже математическото очакване E е линеен функционал (теорема 1.7) върху линейното пространство на случайните величини, то $DX = E((X - EX)^2) = E(X^2 - 2XEX + (EX)^2) = E(X^2) - (EX)^2$.
Да означим със $\mathfrak{S}(\Omega, \mathfrak{A}, P)$ алгебрата на дискретните случайни величини върху $(\Omega, \mathfrak{A}, P)$.

Теорема 1.7. *Средната стойност E е линеен функционал върху $\mathfrak{S}(\Omega, \mathfrak{A}, P)$, тоест $E(X + Y) = EX + EY$ и $E(rX) = rEX$ за произволни $X, Y \in \mathfrak{S}(\Omega, \mathfrak{A}, P)$ и $r \in \mathbb{R}$. Означение: $E \in \mathsf{Hom}(\mathfrak{S}(\Omega, \mathfrak{A}, P),\ \mathbb{R})$.*

Дефиниция 1.8. *Дискретните случайни величини X_k, $\;k = 1, 2, \ldots, n$ са наричат независими, ако за всяко $l = 2, \ldots, n$ и всеки $x_1, \ldots, x_n \in \mathbb{R}$ е в сила $P(X_{k_1} = x_1, \ldots, X_{k_l} = x_l) = P(X_{k_1} = x_1) \times \cdots \times P(X_{k_l} = x_l)$, при всеки $k_1, \ldots, k_l$ такива, че $1 \leq k_1 < \ldots < k_l \leq n$.*

Теорема 1.9. *Нека X_k, $\;k = 1, 2, \ldots, n$ са независими дискретни случайни величини със средни стойности съответно равни на EX_k. Ако съществува средното $E(X_1 X_2 \cdots X_n)$, то $E(X_1 X_2 \cdots X_n) = EX_1 EX_2 \cdots EX_n$.*

Теорема 1.10. *Нека $X_k,\ k = 1, 2, \dots, n$ са независими дискретни случайни величини с вариации съответно равни на DX_k. Ако съществува вариацията $D(X_1 + \dots + X_n)$, то $D(X_1 + \dots + X_n) = DX_1 + \dots + DX_n$.*

Доказателство: Индукция: при $n = 2$ получаваме $D(X_1 + X_2) = E((X_1 + X_2)^2) - (E(X_1 + X_2))^2 = EX_1^2 + EX_2^2 + 2EX_1X_2 - (EX_1)^2 - (EX_2)^2 - 2EX_1EX_2 = DX_1 + DX_2 + 2EX_1X_2 - 2EX_1EX_2 = DX_1 + DX_2$, поради равенството $EX_1X_2 = EX_1EX_2$, което следва от теорема 1.9. Нека предположим, че твърдението е вярно за $n-1$ и да разгледаме случая на n независими величини. Имаме $D(X_1 + \dots + X_{n-1} + X_n) = D((X_1 + \dots + X_{n-1}) + X_n) = D(X_1 + \dots + X_{n-1}) + D(X_n) = DX_1 + \dots + DX_{n-1} + DX_n$. □

Теорема 1.11. *Нека X е дискретна случайна величина и $g : \mathbb{R} \to \mathbb{R}$ е Борелова функция. Ако съществува средното $E(g \circ X)$, то е в сила равенството $E(g \circ X) = \sum_{x \in X(\Omega)} g(x)P(X = x)$.*

Теорема 1.12. *Нека X е дискретна случайна величина със средна стойност и вариация съответно равни на EX и DX. Тогава за всяко $\alpha > 0$ е в сила $P(|X - EX| \geq \alpha) \leq \frac{DX}{\alpha^2}$.*

Доказателство:

$$DX = E((X - EX)^2) = \sum_{x \in X(\Omega)} (x - EX)^2 P(X = x) \geq \sum_{|x - EX| \geq \alpha} (x - EX)^2 P(X = x)$$

$$\geq \sum_{|x - EX| \geq \alpha} \alpha^2 P(X = x) = \alpha^2 \sum_{|x - EX| \geq \alpha} P(X = x) = \alpha^2 P(\cup_{|x - EX| \geq \alpha} \{X = x\})$$

$$= \alpha^2 P(|X - EX| \geq \alpha) \iff P(|X - EX| \geq \alpha) \leq \frac{DX}{\alpha^2}.$$

□

1.1 Условия на задачите от упражнение 8

Задача 1 Хвърлят се два зара. Нека случайната величина X е сумата от падналите се точки. Да се намери разпределението, очакването и дисперсията на X, ако заровете са:
а) правилни;
б) неправилни с $P(1) = P(6) = 1/4,\ P(2) = P(3) = P(4) = P(5) = 1/8$.
Ще бъде ли необичайно, ако при хвърлянето на 1000 зара сумата е била повече от 3700?

Задача 2 От урна съдържаща 5 бели и 3 черни топки се избират последователно, една по една топки докато се появи бяла. Да се намери разпределението на случайната величина - "брой на изтеглените черни топки" и да се пресметне математическото очакване и дисперсията и, при извадка:
а) без връщане;
б) с връщане.
Опитът се повтаря 1000 пъти. Да се оцени вероятността да са извадени повече от 900 черни топки.

Задача 3 Вероятността за улучване на цел при един изстрел е 0,001. За поразяване на са необходими поне две попадения. Каква е вероятността за поразяване на целта, ако са направени

5000 изстрела?

Задача 4 В кутия има 7 лампи от които 3 са дефектни. По случаен начин се избират за проверка 4 лампи. Да се намери разпределението на случайната величина "брой на изпробваните качествени лампи" и да се пресметне нейното очакване.

Задача 5 В Патагония на месец се регистрират средно две слаби земетресения. Каква е вероятността за три месеца да има по-малко от четири слаби земетресения?

1.2 Решения на задачите от упражнение 8

Задача 1 Нека $X_i,\ i = 1, 2$ са случайните величини - брой точки паднали се на $i-$тия зар.
a) Тегловата функция на $X = X_1 + X_2$ е $P(X = k) = \begin{cases} \frac{k-1}{36} & \text{за} \quad 2 \le k \le 7; \\ \frac{13-k}{36} & \text{за} \quad 8 \le k \le 12. \end{cases}$
$EX = \sum_{k=2}^{12} kP(X = k) = \sum_{k=2}^{7} \frac{k(k-1)}{36} + \sum_{k=8}^{13} \frac{k(13-k)}{36} = 7$. Имаме $EX_1 = EX_2 = \frac{\sum_{k=1}^{6} k}{6} = 3.5$ От $X = X_1 + X_2$ и по теорема 1.7 получаваме $EX = E(X_1 + X_2) = EX_1 + EX_2 = 7$. Понеже X_1 и X_2 са независими и $DX_1 = DX_2 = \frac{\sum_{k=1}^{6} k^2}{6} - 3.5^2 = \frac{35}{12}$, то $DX = D(X_1 + X_2) = DX_1 + DX_2 = 2DX_1 = \frac{35}{6}$.
b) Означаваме $p_k = P(X = k),\ k = 2, 3, \dots, 12$. Тогава за разпределението на X получаваме:

$$p_k = p_{14-k}, \quad p_2 = p_3 = \frac{1}{16},\ p_4 = \frac{5}{64},\ p_5 = \frac{3}{32},\ p_6 = \frac{7}{64},\ p_7 = \frac{3}{16}.$$

$EX_1 = EX_2 = \frac{1+6}{4} + \frac{2+3+4+5}{8} = \frac{7}{2}$ и $DX_1 = DX_2 = \frac{1^2+6^2}{4} + \frac{2^2+3^2+4^2+5^2}{8} - 3.5^2 = \frac{15}{4}$, следователно $EX = 7$ и $DX = 7.5$
Нека $X,\ X_i,\ i = 1, 2, \dots 1000$ са случайните величини - сума на падналите се точки при 1 хвърляне на 1000 зара; брой точки паднали се на $i-$тия зар . Тогава $X = \sum_{k=1}^{1000} X_k$ и $EX_1 = \dots = EX_{1000} = 3.5$ откъдето $EX = E\left(\sum_{k=1}^{1000} X_k\right) = \sum_{k=1}^{1000} EX_k = 1000EX_1 = 3500$. Понеже $X_i,\ i = 1, 2, \dots 1000$ са независими и $DX_1 = \dots = DX_{1000} = \frac{35}{12}$ за a) и $DX_1 = \dots = DX_{1000} = \frac{15}{4}$ за b), то $DX = D\left(\sum_{k=1}^{1000} X_k\right) = \sum_{k=1}^{1000} DX_k = 1000DX_1 = \begin{cases} 2916.6 & \text{за a);} \\ 3750 & \text{за b).} \end{cases}$
Прилагаме неравенството на Чебишев $P(|X - EX| \ge k) \le \frac{DX}{k^2}$ за $k = 201 :\ P(X > 3700) \le P(|X - EX| \ge 201) \le \frac{DX}{201^2} = \begin{cases} 0.07 & \text{за a);} \\ 0.09 & \text{за b),} \end{cases}$
което означава, че е събитито $\{X > 3700\}$ е малко вероятно.
Задача 2 Нека X е случайната величина - брой изтеглени черни топки.
a) Тегловата функция $k \longmapsto p_k$ на X е : $p_0 = \frac{5}{8},\ p_1 = \frac{3}{8} \times \frac{5}{7},\ p_2 = \frac{3}{8} \times \frac{2}{7} \times \frac{5}{6},\ p_3 = \frac{3}{8} \times \frac{2}{7} \times \frac{1}{6}$. Тогава $EX = \sum_{k=0}^{3} kp_k = 1 \times \frac{3}{8} \times \frac{5}{7} + 2 \times \frac{3}{8} \times \frac{2}{7} \times \frac{5}{6} + 3 \times \frac{3}{8} \times \frac{2}{7} \times \frac{1}{6} = \frac{1}{2}$ и $DX = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \sum_{k=0}^{3} k^2 p_k - \frac{1}{4} = \frac{15}{28}$.
b) В случая $X \in \mathrm{Ge}(\frac{5}{8})$ с теглова функция $k \longmapsto (1 - \frac{5}{8})^k \frac{5}{8}$. По-общо, ако $X \in \mathrm{Ge}(p)$, то $\mathrm{E}X = \sum_{k=0}^{\infty} kq^k p = pq \sum_{k=1}^{\infty} kq^{k-1} = pq \frac{d}{dx} \left(\sum_{k=1}^{\infty} x^k\right)|_{x=q} = pq \frac{d}{dx}\left(\frac{x}{1-x}\right)|_{x=q} = pq\left(\frac{1}{(1-x)^2}\right)|_{x=q} = \frac{q}{p} = \frac{1-p}{p}$ и $\mathrm{D}X = \sum_{k=0}^{\infty} k^2 q^k p - \sum_{k=0}^{\infty} kq^k p = pq^2\left(\sum_{k=1}^{\infty} k(k-1)q^{k-2} + \frac{1}{q}\sum_{k=1}^{\infty} kq^{k-1}\right) - \frac{q^2}{p^2} =$

$pq^2\left(\frac{d^2}{dx^2}(\sum_{k=1}^{\infty}x^k)|_{x=q}+\frac{1}{q}\frac{d}{dx}(\sum_{k=1}^{\infty}x^k)|_{x=q}\right)-\frac{q^2}{p^2}=\frac{1-p}{p^2}$. Следователно $\mathrm{E}X=\frac{q}{p}=\frac{3}{5}$, $\mathrm{D}X=\frac{1-p}{p^2}=\frac{24}{25}$.

Нека $X,\ X_i,\ i=1,2,\dots 1000$ са случайните величини - брой изтеглени черни топки при 1000 независими опита; брой изтеглени черни топки при $i-$тия опит. Тогава $X=\sum_{i=1}^{1000}X_i$

$$\mathrm{E}X=1000\mathrm{E}X_1=\begin{cases}500 & \text{за a)}\\ 600 & \text{за b)}\end{cases}$$

$$\mathrm{D}X=1000\mathrm{D}X_1=\begin{cases}535.7 & \text{за a)}\\ 960 & \text{за b)}\end{cases}$$

Прилагаме неравенството на Чебишев

$P(X>900)\le P(|X-EX|\ge 400)\le\frac{DX}{400^2}=\frac{535.7}{400^2}=0.003$ за a),

$P(X>900)\le P(|X-EX|\ge 300)\le\frac{DX}{300^2}=\frac{960}{300^2}=0.01$ за b).

Задача 3 Нека $X\in\mathrm{Bi}(5000,\ \frac{1}{1000})$, то търсената вероятност е $P=P(X\ge 2)=1-P(X<2)=1-P(\{X=0\}\cup\{X=1\})=1-P(X=0)-P(X=1)=1-(1-\frac{1}{1000})^{5000}-\binom{5000}{1}\frac{1}{1000}\times(1-\frac{1}{1000})^{4999}\approx 1-e^{-5}-5e^{-5}=1-6e^{-5}\approx 0.959$

Задача 4 Нека X е случайната величина - брой изпробвани качествени лампи. Тогава $X\in HG(n,M,N)$, където $n=M=4,\ N=7$, тоест X е с хипегеометрично разпределение и теглова функция $k\longmapsto p_k=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\ k=0,1,\dots,\min(n,M)$. Пресмятаме $\mathrm{E}X=...=\frac{nM}{N}$ и в частност при $n=M=4,\ N=7$ имаме $\mathrm{E}X=\frac{16}{7}$.

Задача 5 Нека $X_i,\ i=1,2,3$ са случайните величини - брой земетресения за $i-$тия месец. По условие $X_i\in\mathrm{Po}(2)$ са независими. Търсим вероятността $P(X_1+X_2+X_3<4)$. Ще докажем, че ако $Y_j\in Po(\lambda_j),\ j=1,\dots,n$ са независими, то случайната величина $Y=\sum_{j=1}^{n}Y_j$ е поасоново разпределена с параметър $\sum_{j=1}^{n}\lambda_j$. При $n=2$ имаме $P(Y_1+Y_2=k)=P(\cup_{j=0}^{k}\{Y_1=j,\ Y_2=k-j\})=\sum_{j=0}^{k}P(\{Y_1=j,\ Y_2=k-j\})=\sum_{j=0}^{k}P(\{Y_1=j\}\cap\{Y_2=k-j\})=\sum_{j=0}^{k}P(Y_1=j)P(Y_2=k-j)=\sum_{j=0}^{k}\frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^j\lambda_2^{k-j}}{j!(k-j)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\frac{k!}{j!(k-j)!}\times\lambda_1^j\lambda_2^{k-j}=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\binom{n}{j}\lambda_1^j\lambda_2^{k-j}=\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}$ ... (е базата на индукцията)... - ще го докажем чрез пораждащи функции на упражнение 8.

В частност $X=X_1+X_2+X_3\in Po(6)$ и $P(X<4)=\sum_{k=0}^{3}\frac{e^{-6}6^k}{k!}=61e^{-6}\approx 0.1512$