СЕМ, лекция 6

(2020-11-05)

Дефиниция: (**Пораждаща функция / преговор**) Ако $X \in \mathbb{N}_0^+$, то
$g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$, $|s| < 1$ се нарича пораждаща функция.

$\oplus$ $\quad \mathbb{P}(X = 1) = p$ и $\mathbb{P}(X = 0) = 1 - p \Rightarrow g_X(s) = 1 - p + ps$.

Пораждащата функция е като една торбичка, която носи много обекти и може да ги вадим един по един чрез различни операции, като например:

$$\begin{cases} \mathbb{E}X = g_X'(1) \\ DX = g_X''(1) + g_X'(1) - \left(g_X'(1)\right)^2 \\ \mathbb{P}(X = k) = \frac{g_X^{(k)}(0)}{k!} \end{cases}$$

Дефиниция: (**Независимост на (дискретни) случайни величини**) Случайните величини $X_1, X_2, \ldots, X_N$ са независими в съвкупност, ако $\forall\ 1 \leq m \leq N$ и $\{j_1, j_2, \ldots, j_m\} \subseteq \{1, 2, \ldots, N\}$ и възможни стойности $(x_1, x_2, \ldots, x_m)$ е вярно, че
$\mathbb{P}(X_{j_1} = x_1 \cap X_{j_2} = x_2 \cap \ldots \cap X_{j_m} = x_m) = \prod_{i=1}^{m} \mathbb{P}(X_{j_i} = x_i)$.

Вероятността да се сбъдне кой да е подбор от случайни величини се разпада на произведението на индивидуалните вероятности.

Твърдение: Ако $X_1, \ldots, X_n$ са независими в съвкупност, целочислени случайни величини и $Y = \sum_{j=1}^{n} X_j$, то $g_Y(s) = \prod_{j=1}^{n} g_{X_j}(s)$.

Логика на доказателството: $g_Y(s) = \mathbb{E}s^{\sum_{j=1}^{n} X_j} = \mathbb{E}\prod_{j=1}^{n} s^{X_j}$.

Коментари:

$\mathbb{E}\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} \mathbb{E}X_j$, винаги (когато очакванията са добре дефинирани)

$D\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} DX$, ако $X_1, X_2, \ldots, X_n$ са независими в съвкупност.

Някои целочислени случайни величини

$X_1, X_2, \ldots, X_m, \ldots$ - взаимно независими (дискретни) сл. вел.

X_i	0	1
$\mathbb{P}$	q	p

$i \geq 1$
$p + q = 1$

p и q са фиксирани $\forall i$ (не зависят от i). Все едно имаме n експеримента, които са еднородни по своя характер (хвърляне на монета).

Тази постановка се нарича схема на Бернули.

А. Разпределение на Бернули

$X \in Ber(p)$, ако имаме разпределението

X_i	0	1
$\mathbb{P}$	q	p

$p + q = 1$

$$\begin{cases} \mathbb{E}X = 0.q + p.1 = p \\ \mathbb{E}X^2 = 0^2 . q + 1^2 . p = p \end{cases} \Rightarrow DX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1-p) = pq$$

$$g_X(s) = \mathbb{E}s^X = s^0 q + s^1 p = q + ps = 1 - p + ps$$

Б. Биномно разпределение

Взимаме само първите n случайни величини от схемата на Бернули: $X_1, X_2, \ldots, X_n$ и образуваме $X = \sum_{i=1}^{n} X_i$, което ни отразява броя успехи измежду n експеримента.

$X \in Bin(n, p)$ се нарича биномно разпределена (дискретна) случайна величина с параметри n и p.

Твърдение: Нека X е биномно разпределена сл. вел. с параметри n и p. Тогава:

а) $g_X(s) = (q + ps)^n$

б) $\mathbb{E}X = np$, $DX = npq$

в)

X	0	1	$\ldots$	k	$\ldots$	n
$\mathbb{P}$	q^n	npq^{n-1}	$\ldots$	$\binom{n}{k} p^k q^{n-k}$	$\ldots$	p^n

k успеха измежду n експеримента
всеки един измежду които се случва
независимо от останалите с вероятност $\underbrace{p}_{\text{за успех}}$

Доказателство:

а) $X = \sum_{j}^{n} X_j$ (сума на независими бернулиеви случайни величини) $\Rightarrow$

$$g_X(s) \overset{\text{твърдение}}{=} \prod_{j=1}^{n} g_{X_j}(s) = \prod_{j=1}^{n} (q+ps) = (q+ps)^n\ ;$$

б) $$\mathbb{E}X = \mathbb{E}\sum_{j=1}^{n} X_j \overset{\text{лин. функц.}}{=} \sum_{j=1}^{n} \mathbb{E}X_j = \sum_{j=1}^{n} p = np\ ;$$

$$DX = D\sum_{j=1}^{n} X_j \overset{\text{незав.}}{=} \sum_{j=1}^{n} DX_j = \sum_{j=1}^{n} pq = npq\ ;$$

в) Комбинаторно избираме k експеримента от общо n, които да са успешни по $\binom{n}{k}$ начина и умножаваме по вероятността за успех p - k пъти и k пъти по вероятността за неуспех $\Rightarrow \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k}$.

Но може да сведем комбинаторния проблем до чисто аналитичен и да го получим по следния начин:

$k!\mathbb{P}(X=k) = g_X^{(k)}(0),\ 0 \le k \le n.$

Но, $g_X^{(k)}(0) = \dfrac{\partial^k}{\partial s^k}(q+ps)^n\Big|_{s=0} = n(n-1)\dots(n-k+1)q^{n-k}p^k \Rightarrow$

$$\Rightarrow \mathbb{P}(X=k) = \frac{n(n-1)\dots(n-k+1)q^{n-k}p^k}{k!} = \frac{n!}{(n-k)!k!}p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}$$

$\oplus$ Колко заразени биха влезли в България от чужбина, ако вероятността човек да е заразен е $p = 1\,\%$ и са се върнали $100\,000$ души. Тогава, не лош модел би бил случайната величина $X \sim Bin(n = 100\,000,\ p = 0.01)$, която брой колко от върналите се в България са заразени.

B. Геометрично разпределение

Тук може да си мислим за горния пример, но по следния начин: Колко незаразени хора ще влязат в България, преди да влезе първия заразен, върнал се от чужбина?

<u>Дефиниция</u>: $X \in Ge(p)$ и $X = \min\left\{j \ge 1 \,\middle|\, \sum_{i=1}^{j} X_i = 1\right\} - 1$. Т.е. най-малкото j, за което сумата $\sum_{i=1}^{j} X_i$ става единица, като от нея вадим 1-ца, за да извадим последната стъпка (успех/заразен посетител). Това е броя неуспехи до първи успех (в нашия случай с „успех“ контраинтуитивно сме маркирали заразен човек).

$\underbrace{0\,0\,0}_{\text{неуспехи}}\ \underbrace{1}_{\text{успех}} \Rightarrow X = 3$

$\underbrace{0}_{\text{неуспех}}\ \underbrace{1}_{\text{успех}} \Rightarrow X = 1$

$\underbrace{1}_{\text{успех}} \Rightarrow X = 0$

<u>Твърдение</u>: $X \in Ge(p)$. Тогава:

а)

X	0	1	2	$\dots$	k	$\dots$
$\mathbb{P}$	p	pq	pq^2	$\dots$	pq^k	$\dots$

$\Rightarrow \mathbb{P}(X = k) = pq^k$

б) $g_X(s) = \dfrac{p}{1 - qs}, \ |s| < 1$

<u>Доказателство</u>:

а)

$\mathbb{P}(X = 0) = \mathbb{P}(X_1 = 1) = p$ $\qquad 1$

$\mathbb{P}(X = 1) = \mathbb{P}(X_1 = 0;\ X_2 = 1) = pq$ $\qquad 01$

$\dots$ $\qquad \dots$

$\mathbb{P}(X = k) = \mathbb{P}(X_1 = 0;\ X_2 = 0;\ \dots X_k = 0;\ X_{k+1}) = pq^k$ $\qquad \underbrace{00...0}_{k}1$

$\overset{\text{твърдение}}{\Rightarrow}$ а) е доказано.

б) $$g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k pq^k = p \sum_{k=0}^{\infty} s^k q^k \overset{|qs|<1}{=} \frac{p}{1 - qs}.$$

<u>Следствие</u>: $X \sim Ge(p) \Rightarrow \mathbb{E}X = \dfrac{q}{p}$ и $DX = \dfrac{q}{p^2}$

<u>Доказателство</u>: $g_X(s) = \dfrac{p}{1 - qs}$

$$\mathbb{E}X = g_X'(1) = p\frac{\partial}{\partial s}\left(\frac{1}{1 - qs}\right) = p(-1)\frac{\frac{\partial}{\partial s}(1 - qs)}{(1 - qs)^2}\Bigg|_{s=1} = \frac{pq}{(1 - q)^2} = \frac{pq}{p^2} = \frac{q}{p};$$

$$g_X''(1) = \frac{\partial}{\partial s}\left(\frac{pq}{(1 - qs)^2}\right) = pq\frac{\partial}{\partial s}\left(\frac{1}{(1 - qs)^2}\right) = pq(-2)\frac{\frac{\partial}{\partial s}(1 - qs)}{(1 - qs)^3} =$$

$$= \frac{2pq^2}{(1-qs)^3}\bigg|_{s=1} = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2};$$

$$DX = g_X''(1) + g_X'(1) - \left(g_X'(1)\right)^2 = \frac{\cancel{2}q^2}{p^2} + \frac{q}{p} - \cancel{\left(\frac{q}{p}\right)^2} = \frac{q^2}{p^2} + \frac{q}{p} =$$

$$= \frac{q^2 + qp}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2}$$

⊕ Брой „тури“ при хвърляне на монета до първо „ези“ $\sim Ge\left(p = \frac{1}{2}\right)$;

⊕ Брой неуспешни атаки (в баскетбол) до първи кош. Ако грубо казано не се променят играчите и имат вероятност за успех, например $80\,\%$ (за кош), то броя на неуспешните атаки до първия успех може да се моделира добре като се допусне, че е разпределен геометрично с параметър $p = 0.8$.

Твърдение: (**Безпаметност на геометричното разпределение**) Нека $X \in Ge(p)$. Тогава $\forall n \geq 0$ и $k \geq 0$: $\mathbb{P}(X \geq m + k \,|\, X \geq m) = \mathbb{P}(X \geq k) = q^k$.

Тоест, ако сме наблюдавали нашия процесор да работи 20 месеца, то вероятността да работи следващия месец е грубо казано същата, като вероятността да работи в момента, в който сме го стартирали в първия месец до края на първия месец.

Доказателство:

$$\mathbb{P}(X \geq l) = \sum_{j=l}^{\infty} pq^j = pq^l \sum_{j=0}^{\infty} q^j = pq^l \frac{1}{1-q} = q^l.$$ Тогава,

$$\mathbb{P}(X \geq m + k \,|\, X \geq m) = \frac{\mathbb{P}(X \geq m + k \cap X \geq m)}{\mathbb{P}(X \geq m)} = \frac{\mathbb{P}(X \geq m + k)}{\mathbb{P}(X \geq m)} =$$

$$= \frac{q^{m+k}}{q^m} = q^k = \mathbb{P}(X \geq k).$$

Г. Отрицателно биномно разпределение

Това разпределение се конструира от няколко геометрично разпределени случайни величини. Т.е. тук ще моделираме броя на неуспехите, но не до първия успех, а до r-тия успех.

Дефиниция: $X \in NB(r, p)$ и $X = \min\left\{j \geq 1 \,\middle|\, \sum_{i=1}^{j} X_j = r\right\} - r$.

вероятност за успех
в схемата на бернули

Това е броя неуспехи до r-тия успех.

$r=2$ $\quad\underbrace{0\,0\,0}\,1\,\underbrace{0\,0\,0\,0\,0}\,1$ $\quad X=8$

общо 8 неуспеха

$r=4$ $\quad 1\,1\,\underbrace{0}\,1\,1$ $\quad X=1$

$r=1$ $\quad \underbrace{0\,0\,0}\,1$ $\quad X=3$

$\oplus$ За $r=1 \to NB(1,p)=Ge(p)$

<u>Твърдение</u>: Ако $X \in NB(r,p)$, то $X=\sum_{j=1}^{r} Y_j$, където Y_j са геометрични с вероятност p $\left(Y_j \in Ge(p)\right)$, за $1 \le j \le r$ и Y_j са независими в съвкупност.

За $r=2: X \in NB(2,p)$ $\quad \underbrace{0\,0\,0}\,1\,\underbrace{0\,0}\,1$

$X \overset{?}{=} \sum_{j=1}^{2} Y_j$, където $Y_j \in Ge(p)$ и Y_j са независими в съвкупност.

Ще проверим, че $Y_1 \perp\!\!\!\perp Y_2$ $\left(\text{Ясно е, че } X=Y_1+Y_2 \text{ и } Y_1 \in Ge(p) \text{ и } Y_2 \in Ge(p)\right)$

$\mathbb{P}(Y_1=l \cap Y_2=m) \overset{?}{=} \mathbb{P}(Y_1=l)\mathbb{P}(Y_2=m),\ \forall\ l,m \ge 0.$

$\mathbb{P}(Y_1=l \cap Y_2=m)=$

$$=\mathbb{P}(\underbrace{X_1,0,\ldots,X_l=0,X_{l+1}=1,X_{l+2}=0,\ldots,X_{l+m}=0,X_{l+m+1}=1}_{\substack{\text{независими експерименти}\\ \text{от схемата на Бернули}}})=$$

$$=\underbrace{\mathbb{P}(X_1=0)\ldots\mathbb{P}(X_2=0)\mathbb{P}(X_{l+1}=1)}\ \underbrace{\mathbb{P}(X_{l+2}=0)\ldots\mathbb{P}(X_{l+m}=0)\mathbb{P}(X_{l+m+1}=1)}=$$

$=\mathbb{P}(Y_1=l)\mathbb{P}(Y_2=m) \Rightarrow$ независими.

<u>Твърдение</u>: Ако $X \sim NB(r,p)$, то $g_X(s)=\left(\dfrac{p}{1-qs}\right)^r$, $\mathbb{E}X=\dfrac{rq}{p}$ и $DX=\dfrac{rq}{p^2}$.

Доказателство:

$$\mathbb{E}X = \mathbb{E}\sum_{j=1}^{r} \underbrace{Y_j}_{\substack{\text{геом.}\\ \text{незав.}}} = \sum_{j=1}^{r} \mathbb{E}Y_j = r\frac{q}{p};$$

$$DX = D\sum_{j=1}^{r} Y_j \overset{\text{незав.}}{=} \sum_{j=1}^{r} DY_j = r\frac{q}{p^2};$$

$$g_X(s) \overset{\text{твърдение}}{=} \prod_{j=1}^{r} g_{Y_j}(s) = \prod_{j=1}^{r} \frac{p}{1-qs} = \left(\frac{p}{1-qs}\right)^r.$$

Твърдение: $X \sim NB(r, p)$. Тогава:

X	0	1	$\dots$	k	$\dots$
$\mathbb{P}$				$\binom{r+k-1}{k} p^r q^k$	

Аналитичен подход:

От пораждащата функция знаем, че $g_X(s) = \left(\dfrac{p}{1-qs}\right)^r$, но освен това знаем, че

$$k!\mathbb{P}(X=k) = g_X^{(k)}(s)\Big|_{s=0}$$

$$g_X(s) = \frac{p^r}{(1-qs)^r}, \text{ но } \frac{1}{(1-x)^r} \underset{|x|<1}{\overset{\substack{\text{ред на}\\ \text{Тейлър}}}{=}} \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k$$

$$\frac{\partial^n}{\partial x^n}\frac{1}{(1-x)^r}\Bigg|_{x=0} = r(r+1)\dots(r+1-n)\frac{1}{(1-x)^{r+n}}\Bigg|_{x=0} = r(r+1)\dots(r-n+1)$$

$$\Rightarrow g_X(s) = p^r \sum_{k=0}^{\infty} \binom{r+k-1}{k} \underbrace{q^k s^k}_{x^k}$$

$$g_X^{(k)}(0)\Big|_{s=0} = p^r \binom{r+k-1}{k} q^k k! = k!\mathbb{P}(X=k)$$

$$\Rightarrow \mathbb{P}(X=k) = \binom{r+k-1}{k} p^r q^k.$$

Комбинаторен подход:

k нули и $r-1$ единици трябва да се поставят на $r+k-1$ позиции след което да се последват от $1^{\text{-ца}}$:

$$\underbrace{\binom{r+k-1}{k}}_{\text{нули}} \underbrace{q^k}_{\text{вер.}} \; p^{r-1}p$$

Д. Поасоново разпределение

Дефиниция: Нека $\lambda > 0$. Казваме, че X е поасоново разпределена случайна величина с параметър λ и бележим с $X \sim Pois(\lambda)$, ако $\mathbb{P}(X=k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$, $k \geq 0$

$$1 \stackrel{?}{=} \sum_{k=0}^{\infty} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda} \times \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\substack{\text{развитие в ред} \\ \text{на Тейлър} \\ \text{за exp}}} = e^{-\lambda}e^{\lambda} = 1$$

Какво може да се моделира добре с поасоновото разпределение на случайна величина:

- ⊕ Брой постъпили заявки към сървър за единица време ;
- ⊕ Брой катастрофи за $1^{\text{-ца}}$ време на дадена територия/кръстовище ;
- ⊕ Брой насекоми за единица площ ;
- ⊕ Брой голове за $1^{\text{-ца}}$ време ;
- ⊕ Брой звезди в някакъв регион.

Твърдение:

$X \in Pois(\lambda)$. Тогава:

а) $g_X(s) = e^{-\lambda+\lambda s}$, за $|s| \leq 1$

б) $\mathbb{E}X = DX = \lambda$

Доказателство:

а) $$g_X(s) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda}e^{\lambda s} ;$$

б) $\mathbb{E}X = g_X'(1) = \lambda e^{-\lambda} e^{\lambda s}\Big|_{s=1} = \lambda$

$$g_X''(1) = \frac{\partial}{\partial s}\left(\lambda e^{-\lambda} e^{\lambda s}\right) = \lambda^2 e^{-\lambda} e^{\lambda s}\Big|_{s=1} = \lambda^2$$

$$DX = g_X'(1) + g_X''(1) - \left(g_X'(1)\right)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = \mathbb{E}X.$$

Теорема: (**Поасон**) Нека $\forall\ n \geq 1 : X_n \sim Bin(n, p_n)$, където $p_n = \dfrac{\lambda}{n} + \dfrac{V_n}{n}$, където $\lambda > 0$ и $\lim\limits_{n\to\infty} V_n = 0$.

Т.е. $P_N = \dfrac{\lambda}{n} + O\left(\dfrac{1}{n}\right)$. Тогава $\forall\ k \geq 0 : \lim\limits_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$ или казано по друг начин, $X \sim Pois(\lambda)$.

$\oplus$ $p = 0{,}01$ - вероятност на зараза и $n = 1000$.

$$\mathbb{P}(X = 50) = \binom{1000}{50} \times 0{,}01^{50} \times 0{,}99^{950} \underset{\lambda=np=10}{\overset{Pois}{\approx}} \frac{10^{50}}{50!} e^{-10}$$

$$\mathbb{P}(X = 0) \approx \frac{10^0}{0!} e^{-10} = e^{-10}.$$