

СЕМ, лекция 6 (2020-11-05)

Дефиниция (Пораждаща функция / преговор). Ако $X \in \mathbb{N}_0^+$, то

$$g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k), \quad |s| < 1 \text{ се нарича пораждаща функция.}$$

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

Пораждащата функция може да си я представяме като торбичка, в която има много обекти и може да ги изваждаме един по един чрез различни математически операции, като например:

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

Дефиниция (Независимост на (дискретни) случайни величини). Случайните величини X_1, X_2, \dots, X_N са независими в съвкупност, ако $\forall 1 \leq m \leq N$ и $\{j_1, j_2, \dots, j_m\} \subseteq \{1, 2, \dots, N\}$ и възможни стойности (x_1, x_2, \dots, x_m) е вярно, че

$$\mathbb{P}(X_{j_1} = x_1 \cap X_{j_2} = x_2 \cap \dots \cap X_{j_m} = x_m) = \prod_{i=1}^m \mathbb{P}(X_{j_i} = x_i).$$

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

Твърдение. Ако X_1, \dots, 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.

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

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

Някои целочислени случайни величини

$X_1, X_2, \dots, X_m, \dots$ – взаимно независими (дискретни) сл. вел.

X_i	0	1	$i \geq 1$
\mathbb{P}	q	p	$p + q = 1$

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	$p + q = 1$
	\mathbb{P}	q	p	

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

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

Б. Биномно разпределение

Взимаме само първите n случайни величини от схемата на Бернули: X_1, X_2, \dots, X_n и образуваме сл. вел. $X = \sum_{i=1}^n X_i$, която 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:

- $g_X(s) = (q + ps)^n$
- $\mathbb{E}X = np, \mathbb{D}X = npq$
-

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

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

Доказателство.

a) $X = \sum_{j=1}^n X_j$ (сума на независими бернулиеви случайни величини) \Rightarrow

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

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

$$\mathbb{D}X = \mathbb{D} \sum_{j=1}^n X_j \stackrel{\text{незав.}}{=} \sum_{j=1}^n \mathbb{D}X_j = \sum_{j=1}^n pq = npq;$$

с) Комбинаторно избираме k експеримента от общо n , които да са успешни по $\binom{n}{k}$ начина и умножаваме по вероятността за успех p точно k пъти и $n - k$ 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 получим резултата по следния начин:

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

$$\text{Но, } g_X^{(k)}(0) = \left. \frac{\partial^k}{\partial s^k} (q + ps)^n \right|_{s=0} = n \times (n-1) \times \dots \times (n-k+1) \times q^{n-k} \times p^k \Rightarrow$$

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

⊕ 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 $X \sim \text{Bin}(n = 100\,000, p = 0.01)$, която брой колко от върналите се в България са заразени. В случая неинтуитивно приемаме за успех вероятността човек да е заразен.

В. 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?

Дефиниция. $X \in Ge(p)$ и $X = \min \left\{ j \geq 1 \mid \sum_{i=1}^j X_i = 1 \right\} - 1$. 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).

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

неуспехи успех

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

неуспех успех

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

успех

Твърдение. $X \in Ge(p)$. Тогава:

а)

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

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

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

Доказателство.

а)

$\mathbb{P}(X = 0) = \mathbb{P}(X_1 = 1) = p$	1
$\mathbb{P}(X = 1) = \mathbb{P}(X_1 = 0; X_2 = 1) = pq$	01
...	...
$\mathbb{P}(X = k) = \mathbb{P}(X_1 = 0; X_2 = 0; \dots X_k = 0; X_{k+1}) = pq^k$	<u>00...0</u> 1
	k

твърдение
 \Rightarrow а) е доказано.

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

Следствие. $X \sim Ge(p) \Rightarrow \mathbb{E}X = \frac{q}{p}$ и $\mathbb{D}X = \frac{q}{p^2}$

Доказателство. $g_X(s) = \frac{p}{1 - qs}$

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

$$\begin{aligned} g''_X(1) &= \frac{\partial}{\partial s} \left(\frac{pq}{(1 - qs)^2} \right) = pq \frac{\partial}{\partial s} \left(\frac{1}{(1 - qs)^2} \right) = pq(-2) \frac{\frac{\partial}{\partial s}(1 - qs)}{(1 - qs)^3} = \\ &= \frac{2pq^2}{(1 - qs)^3} \Big|_{s=1} = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2}; \end{aligned}$$

$$\begin{aligned} \mathbb{D}X &= g''_X(1) + g'_X(1) - (g'_X(1))^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p} \right)^2 = \frac{q^2}{p^2} + \frac{q}{p} = \\ &= \frac{q^2 + qp}{p^2} = \frac{q(p + q)}{p^2} = \frac{q}{p^2}; \end{aligned}$$

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

\oplus Брой неуспешни атаки (в баскетбол) до първи кош. Ако грубо казано не се променят играчите и имат вероятност за успех, например 80 % за кош при атака, 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 (Безпамятност 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.

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

$$\begin{aligned} \mathbb{P}(X \geq m + k | X \geq m) &= \frac{\mathbb{P}(X \geq m + k \cap X \geq m)}{\mathbb{P}(X \geq m)} = \frac{\mathbb{P}(X \geq m + k)}{\mathbb{P}(X \geq m)} = \\ &= \frac{q^{m+k}}{q^m} = q^k = \mathbb{P}(X \geq k) \end{aligned}$$

.

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 $X = \min \left\{ j \geq 1 \mid \sum_{i=1}^j X_i = r \right\} - r$.
 вероятност за успех
 в схемата на бернули

Това е броя неуспехи до r -ТИЯ успех.

$r = 2$ 0 0 0 1 0 0 0 0 1 $X = 8$
 общо 8 неуспеха

$r = 4$ 1 1 0 1 1 $X = 1$

$r = 1$ 0 0 0 1 $X = 3$

\oplus За $r = 1 \rightarrow NB(1, p) = Ge(p)$

Твърдение. Ако $X \in NB(r, p)$, то $X = \sum_{j=1}^r Y_j$, където Y_j са геометрични с вероятност p ($Y_j \in Ge(p)$), за $1 \leq j \leq r$ и Y_j са независими в съвкупност.

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

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

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

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

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

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

$$= \underbrace{\mathbb{P}(X_1 = 0) \dots \mathbb{P}(X_2 = 0) \mathbb{P}(X_{l+1} = 1)}_{\text{геом. незав.}} \underbrace{\mathbb{P}(X_{l+2} = 0) \dots \mathbb{P}(X_{l+m} = 0) \mathbb{P}(X_{l+m+1} = 1)}_{\text{геом. незав.}} = \mathbb{P}(Y_1 = l) \mathbb{P}(Y_2 = m) \Rightarrow \text{независими.}$$

Твърдение. Ако $X \sim NB(r, p)$, то $g_X(s) = \left(\frac{p}{1 - qs}\right)^r$, $\mathbb{E}X = \frac{rq}{p}$ и $\mathbb{D}X = \frac{rq}{p^2}$.

Доказателство.

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

$$\mathbb{D}X = \mathbb{D} \sum_{j=1}^r Y_j \stackrel{\text{незав.}}{=} \sum_{j=1}^r \mathbb{D}Y_j = r \times \frac{q}{p^2};$$

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

Твърдение. $X \sim NB(r, p)$. Тогава:

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

Аналитичен подход.

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

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

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

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

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

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

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

Комбинаторен подход.

k нули и $r - 1$ единици трябва да се поставят на $r + k - 1$ позиции след което да се последват от 1-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 $X \sim \text{Pois}(\lambda)$, ако $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \geq 0$

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

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

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

Твърдение.

$X \in Pois(\lambda)$. Тогава:

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

b) $\mathbb{E}X = \mathbb{D}X = \lambda$

Доказателство.

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

b) $\mathbb{E}X = g'_X(1) = \lambda e^{-\lambda} e^{\lambda s} \Big|_{s=1} = \lambda, g''_X(1) = \frac{\partial}{\partial s} (\lambda e^{-\lambda} e^{\lambda s}) = \lambda^2 e^{-\lambda} e^{\lambda s} \Big|_{s=1} = \lambda^2$

$$\mathbb{D}X = g'_X(1) + g''_X(1) - (g'_X(1))^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = \mathbb{E}X.$$

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

$\lim_{n \rightarrow \infty} V_n = 0$. Тоест, $p_n = \frac{\lambda}{n} + O\left(\frac{1}{n}\right)$. Тогава $\forall k \geq 0$ е изпълнено, че

$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, където $Y \sim Pois(\lambda)$ или казано по друг начин, $X_n \sim Pois(\lambda)$.

Това на практика означава, че с поасоновото разпределение може много добре да приближаваме биномното разпределение, когато имаме много на брой опити.

\oplus $p = 0,01$ – вероятност на зараза и $n = 1\,000$.

$$\mathbb{P}(X = 50) = \binom{1\,000}{50} \times 0,01^{50} \times 0,99^{950} \underset{\lambda=np=10}{\approx} \frac{10^{50}}{50!} e^{-10}$$

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

Доказателство.

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \simeq \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2) \dots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \frac{n^k + O(n^{k-1})}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

Тъй като $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right) = e^{-\lambda}$ и $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$, то това автоматично води до $\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!}$, което искахме да докажем.

Равенството може да се докаже и по алтернативен начин, като се използва апроксимацията на Стърлинг за факториел.