# СЕМ, лекция 7

(2020-11-12)

**Теорема (Поасон / преговор).** Нека $X_n \in Bin(n, p)$, където $p_n = \dfrac{\lambda}{n} + \dfrac{V_n}{n}$, $\left(\text{т.е. } \dfrac{V_n}{n} \text{ клони по-бързо от } \dfrac{\lambda}{n} \text{ към } 0\right)$. Тогава $\forall k \geq 0$ е вярно, че

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \mathbb{P}(Y = k), \text{ където } Y \in Pois(\lambda).$$

$\oplus$ При $n \geq 100$ и $np \leq 20$ може да считаме, че $\mathbb{P}(X_n = k) \approx \mathbb{P}(X = k)$, където $X \in Pois(\lambda)$ и $\lambda = np$.

Нека например $n = 1\,000$ души се завръщат в България за някакъв период от време и вероятността за всеки един от тях да внесе някакъв вирус е $p = 0.001$. Искаме да приближим относително добре вероятността да влязат точно $k$ на брой заразени.

$$\tilde{X} \in Bin\left(n = 1\,000, p = \frac{1}{1\,000}\right); \mathbb{P}(\tilde{X} = 3) = \binom{1\,000}{3}\left(\frac{1}{1\,000}\right)^3\left(\frac{999}{1\,000}\right)^{997}$$

$$\lambda = np = 1 \leq 20, n = 1000 \geq 100 \Rightarrow$$

$$\mathbb{P}(\tilde{X} = 3) \approx \frac{1}{3!} \times e^{-1} = \frac{1}{6e} \approx \frac{1}{6} \times \frac{1}{2.7182...}$$

**Доказателство.** $X_n \in Bin(n, p_n)$, тогава $g_{X_n} = (1 - p_n + p_n s)^n$.

Ако $g_{X_n}(s) \xrightarrow[n\to\infty]{} g_X(s)$, където $g_X(s) = e^{-\lambda} \times e^{\lambda s}$, то може да заключим искания резултат, тъй като $e^{-\lambda} \times e^{\lambda s}$ е пораждаща функция на $X \in Pois(\lambda)$.

От $g_{X_n}(s) \xrightarrow[x\to\infty]{} g_X(s) \Rightarrow \lim\limits_{n\to\infty} \mathbb{P}(X_n = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}, \forall k \geq 0$

$$\lim_{n\to\infty} g_{X_n}(s) = \lim_{n\to\infty}\left(1 - \frac{\lambda}{n} - \frac{V_n}{n} + \underbrace{\frac{\lambda}{n}s + \frac{V_n}{n}s}_{<o\left(\frac{1}{n}\right)}\right)^n = \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}(1-s)\right)^n = e^{-\lambda(1-s)} = e^{\lambda s}e^{-\lambda}$$

## E. Хипергеометрично разпределение

Имаме $N$ обекта, от които $M$ са маркирани $(0 \leq M \leq N)$. Избират се $n$ обекта и случайната величина $X$ е броя маркирани измежду тези $N$ $(n \leq N)$. Тогава казваме, че $X$ е разпределено хипергеометрично с параметри $N, M$ и $n$ и бележим $X \in HG(N, M, n)$.

**Твърдение.** Нека $X \in HG(N, M, n)$. Тогава:

a) $\mathbb{P}(X = k) = \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, като $\max\big(0, n-(N-M)\big) \leq k \leq \min(n, M)$

b) $\mathbb{E}X = n \times \dfrac{M}{N}$, $\mathbb{D}X = n \times \dfrac{M}{N} \times \dfrac{N-M}{N} \times \dfrac{N-n}{N-1}$.

**Доказателство.**

a) $\mathbb{P}(X = k) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$.

b)

$$\underbrace{\underbrace{\underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \overset{i}{\underline{\quad}}\ \underline{\quad}\ \underline{\quad}}_{n}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}\ \underline{\quad}}_{N}$$

$X = \displaystyle\sum_{i=1}^{n} X_i$, където $X_i = \begin{cases} 1, & \text{ако на } i\text{-тата позиция \textbf{има} маркиран обект} \\ 0, & \text{ако на } i\text{-тата позиция \textbf{няма} маркиран обект} \end{cases}$

$X_i$ е бернулиево разпределено: $X_i \in Ber(p_i)$. $p_i = \mathbb{P}(X_i = 1) = \dfrac{\binom{N-1}{M-1}}{\binom{N}{M}} = \dfrac{M}{N}$.

$\mathbb{E}X = \displaystyle\sum_{i=1}^{n} \mathbb{E}X_i = n \times \dfrac{M}{N}$.

$\oplus$ $X_i \in Ber(p)$ – поведението на $i$-тия клиент в даден магазин (купува или не)

$X = \displaystyle\sum_{i=1}^{n} X_i$, $X \in Bin(n, p)$ – броя на извършилите покупка клиенти от първите $n$ клиента.

$Y = \min\left\{ \displaystyle\sum_{j=1}^{n} X_j = 1 \right\} - 1 \in Ge(p)$ – броя на не извършили покупка клиенти, до идването на първия извършил покупка клиент.

$Z = \min\left\{ \sum_{j=1}^{n} X_j = r \right\} - r \in NB(r,p)$ – броя на не извършили покупка клиенти, до идването на $r$-тия извършил покупка клиент.

## Съвместно разпределение на дискретни случайни величини

Нека $X, Y$ са дискретни случайни величини. Интересуваме се от $(X, Y)$.
**Дефиниция.** Нека $(X, Y)$ са две дискретни случайни величини. Тогава таблицата по-долу се наричаме съвместно разпределение на $X$ и $Y$:

| $Y \backslash X$ | $x_1$ | $x_2$ | $\dots$ | $x_n$ | $\dots$ | |
|---|---|---|---|---|---|---|
| $y_1$ | $p_{11}$ | $p_{21}$ | $\dots$ | $p_{n1}$ | $\dots$ | $\sum_i p_{i1}$ |
| $y_2$ | $p_{12}$ | $p_{22}$ | $\dots$ | $p_{n2}$ | $\dots$ | $\sum_i p_{i2}$ |
| $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | |
| $y_k$ | $p_{1k}$ | $p_{2k}$ | $\dots$ | $p_{nk}$ | $\dots$ | $\sum_i p_{ik}$ |
| $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | |
| | $\sum_j p_{1j}$ | $\sum_j p_{2j}$ | | $\sum_j p_{nj}$ | | |

Където $0 \leq p_{ij} = \mathbb{P}(X = x_i;\ Y = y_j)$, за $\forall i, j \in$ {Table Indexes}. $\sum_{i,j} p_{ij} = 1$.

Маргинално разпределение на $X$:

| $X$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ | $\dots$ |
|---|---|---|---|---|---|
| | $\sum_j p_{1j}$ | $\sum_j p_{2j}$ | $\dots$ | $\sum_j p_{nj}$ | $\dots$ |

Разпределенията само на $X$ или само на $Y$ се наричат маргинални разпределения.

$\oplus$ Хвърляме два зара $(1, \dots, 6)$. Нека $X$ е броят шестици, а $Y$ е броят единици. Търси се $(X, Y)$. За едно хвърляне може да имаме $0,\ 1,\ 2$ шестици или единици.

| $Y \backslash X$ | 0 | 1 | 2 | |
|---|---|---|---|---|
| 0 | $\left(\frac{4}{6}\right)^2$ | $\binom{2}{1}\cdot\frac{1}{6}\cdot\frac{4}{6}$ | $\left(\frac{1}{6}\right)^2$ | $\frac{25}{36}$ |
| 1 | $\binom{2}{1}\cdot\frac{1}{6}\cdot\frac{4}{6}$ | $\binom{2}{1}\cdot\frac{1}{6}\cdot\frac{1}{6}$ | 0 | $\frac{10}{36}$ |
| 2 | $\left(\frac{1}{6}\right)^2$ | 0 | 0 | $\frac{1}{36}$ |
| | $\frac{25}{36}$ | $\frac{10}{36}$ | $\frac{1}{36}$ | $1\backslash 1$ |

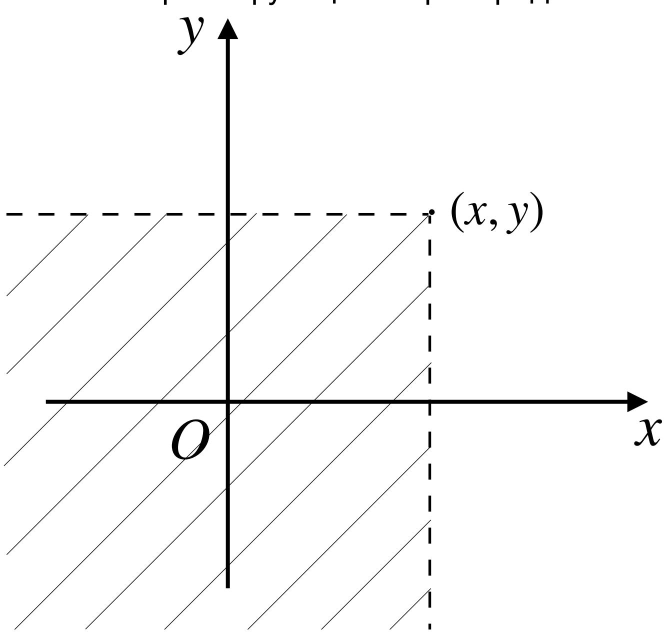
**Дефиниция: (Функция на разпределение на случайни величини).** Нека $(X, Y)$ се състои от *произволни* случайни величини. Тогава ,

$$F_{XY}(x, y) = \mathbb{P}(X < x \cap Y < y) = \mathbb{P}(X < x;\; Y < y),$$

дефинирана за $\forall (x, y) \in \mathbb{R}^2$ се нарича функция на разпределение.

$F_X(x) = F_{XY}(x, \infty)$

$F_Y = F_{XY}(\infty, y)$




**Дефиниция (Независимост $\perp\!\!\!\perp$ ).** Произволни случайни величини $X$ и $Y$ са независими ( $X \perp\!\!\!\perp Y$ ) $\Leftrightarrow$

$$F_{XY}(x, y) = F_X(x)F_Y(y) = F_{XY}(x, \infty)F_{XY}(\infty, y),\ \forall (x, y) \in \mathbb{R}^2.$$

**Ковариация на $X$ и $Y$**

Линейна зависимост $Y = aX + b$.
Показател за това до колко имаме линейна зависимост между произволни случайни величини $X$ и $Y$ е ковариацията.

**Дефиниция (Ковариация).** Нека $X$ и $Y$ са случайни величини с $DX < \infty$ и $\mathbb{D}Y < \infty$. Тогава $cov(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right]$ се нарича ковариация.

**Твърдение.**

a) $cov(X,Y) = \mathbb{E}\tilde{X}\tilde{Y}$, където $\tilde{X} = X - \mathbb{E}X,\ \tilde{Y} = Y - \mathbb{E}Y$
b) $cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$

**Доказателство.**

a) $cov(X,Y) = \mathbb{E}\left[\underbrace{(X - \mathbb{E}X)}_{\tilde{X}}\underbrace{(Y - \mathbb{E}Y)}_{\tilde{Y}}\right] = \mathbb{E}\tilde{X}\tilde{Y}.$

b)

$$cov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right] = \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}XY) \overset{\text{лин. функц.}}{\overset{\mathbb{E}}{=}} \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$$

**Следствие.** Ако $X \perp\!\!\!\perp Y$, то $cov(X,Y) = 0$
**Доказателство.** Ако $X \perp\!\!\!\perp Y$, то $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y \Rightarrow \underbrace{\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y}_{cov(X,Y)} = 0.$

$X \longmapsto 10X,\ Y \longmapsto 10Y\ \Rightarrow \tilde{X} \longmapsto 10\tilde{X},\ \tilde{Y} \longmapsto 10\tilde{Y}$
$\Rightarrow cov(10X,\ 10Y) = 100 \times cov(X,\ Y).$

**Дефиниция (Корелация).** Нека $X, Y$ са случайни величини и $\mathbb{D}X < \infty$ и $\mathbb{D}Y < \infty$. Тогава $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{\mathbb{D}X}\sqrt{\mathbb{D}Y}}$ се нарича коефициент на корелация между $X$ и $Y$.

Целта на корелацията е да мери някаква степен на линейност между $X$ и $Y$.

**Твърдение.** Нека $X$ и $Y$ са случайни величини с крайни дисперсии и нека

$$\overline{X} = \frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}} \text{ и } \overline{Y} = \frac{Y - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}}.$$

Тогава $\rho(X,Y) = \mathbb{E}\overline{X}\,\overline{Y}$ и $\mathbb{E}\overline{X} = \mathbb{E}\overline{Y} = 0,\ D\overline{X} = \mathbb{E}\overline{X}^2 = D\overline{Y} = \mathbb{E}\overline{Y}^2 = 1.$

**Доказателство.**

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{\mathbb{D}X}\sqrt{\mathbb{D}Y}} = \frac{\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right]}{\sqrt{\mathbb{D}X}\sqrt{\mathbb{D}Y}} = \mathbb{E}\left[\underbrace{\left(\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}}\right)}_{=\overline{X}}\underbrace{\left(\frac{Y - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}}\right)}_{=\overline{Y}}\right] = \mathbb{E}\overline{X}\,\overline{Y}$$

$$\mathbb{E}\overline{X} = \mathbb{E}\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}} = \frac{1}{\sqrt{\mathbb{D}X}}\mathbb{E}(X - \mathbb{E}X) = 0$$

$$\mathbb{D}\overline{X} = \mathbb{D}\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}} = \frac{1}{(\sqrt{DX})^2}\mathbb{D}(X - \underbrace{\mathbb{E}X}_{const.}) = \frac{\mathbb{D}X}{\mathbb{D}X} = 1,$$ тъй като $\mathbb{D}(X-c) = \mathbb{D}X$, защото $\mathbb{E}\big(X - c - \mathbb{E}(X-c)\big)^2 = \mathbb{E}(X - \mathbb{E}X)^2$.

$$\mathbb{D}\overline{X}^2 = \mathbb{D}\overline{X} + (\underbrace{\mathbb{E}\overline{X}}_{=0})^2 = \mathbb{D}\overline{X} = 1.$$

**Твърдение.** Нека $X$ и $Y$ са случайни величини с крайни дисперсии. Тогава:

a) $|\rho(X,Y)| \leq 1$

b) $|\rho(X,Y)| = 1 \Leftrightarrow \exists a, b : a, b \in \mathbb{R} \ \& \ Y = aX + b$

(Колкото е по-голям коефициента на корелация, толкова по-добра е линейната зависимост между $X$ и $Y$)

**Доказателство.**

a) $(\overline{X} - \overline{Y})^2$ е неотрицателно случайна величина

$$\Rightarrow 0 \leq \mathbb{E}(\overline{X} - \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 - 2\mathbb{E}\overline{X}\,\overline{Y} = 2\left(1 - \rho(X,Y)\right) \Rightarrow \rho(X,Y) \leq 1$$

Аналогично,
$0 \leq \mathbb{E}(\overline{X} + \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 + 2\mathbb{E}\overline{X}\,\overline{Y} = 2\big(1 + \rho(X,Y)\big) \Rightarrow -1 \leq \rho(X,Y)$

b) $Y = aX + b \Leftrightarrow Y = \mathbb{E}X = a(X - \mathbb{E}X) + a\mathbb{E}X + b - \mathbb{E}Y \Leftrightarrow$

$$\Leftrightarrow \frac{Y - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}} = \underbrace{\frac{a}{\sqrt{\mathbb{D}Y}}\sqrt{\mathbb{D}X}}_{v}\,\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}} + \underbrace{\frac{a\mathbb{E}X + b - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}}}_{w} \Leftrightarrow \overline{Y} = v\overline{X} + w$$

$$0 = \mathbb{E}\overline{Y} = \mathbb{E}(v\overline{X} + w) = v\underbrace{\mathbb{E}\overline{X}}_{=0} + w \Rightarrow w = 0 \Rightarrow \overline{Y} = v\overline{X}$$

$\mathbb{D}\overline{Y} = v^2\mathbb{D}\overline{X} \Rightarrow v^2 = 1$ или $v = \pm 1$.

$Y = aX + b \Leftrightarrow \overline{Y} = v\overline{X}$ и $v^2 = 1$

c) Нека $Y = aX + b$ или $\overline{Y} = v\overline{X}$. Тогава

$$\rho(X,Y) \overset{\text{твърдение}}{=} \mathbb{E}\overline{X}\,\overline{Y} = v\mathbb{E}\overline{X}^2 = v \Rightarrow \rho(X,Y) = \pm 1.$$

Нека $|\rho(X,Y)| = 1$. Тогава $|\mathbb{E}(\overline{X}\,\overline{Y})| = 1$. Да допуснем, че $\mathbb{E}\overline{X}\,\overline{Y} = 1$ (аналогично и за другия случай: $\mathbb{E}\overline{X}\,\overline{Y} = -1$).

Тогава $\mathbb{E}(\overline{X} - \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 - 2\mathbb{E}\overline{X}\,\overline{Y} = 0 \Rightarrow \mathbb{E}(\overline{X} - \overline{Y}^2) = 0 \Rightarrow \overline{X} = \overline{Y}$, защото и $\mathbb{D}(\overline{X} - \overline{Y}) = 0$.