Съвместно разпределение, ковариация и корелация

Дефиниция: (съвместно разпределение): Нека (X, Y) са две дискретни случайни величини. Тогава таблицата се нарича съвместно разпределение на X, Y.

$X\backslash Y$	x_1	$\dots$	x_k	$\dots$	x_n	Y
y_1	p_{11}	$\cdots$	p_{1k}	$\cdots$	p_{1n}	$\sum_k p_{1k}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_j	p_{j1}	$\dots$	p_{jk}	$\dots$	p_{jn}	$\sum_k p_{jk}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_m	p_{m1}	$\cdots$	p_{mk}	$\cdots$	p_{mn}	$\sum_k p_{mk}$
X	$\sum_j p_{j1}$	$\cdots$	$\sum_j p_{jk}$	$\cdots$	$\sum_j p_{jn}$	1 = $\sum_{j,k} p_{jk}$

Където $p_{jk} = \mathbb{P}(X = x_k, Y = y_j)$ и $p_{jk} \geq 0$, $\sum_{j,k} p_{jk} = 1$

- В последния стълб и последния ред са маргиналните разпределения на X и Y съответно.

Дефиниция: (съвместна функция на разпределение): Нека X, Y са произволни случайни величини. Тогава функцията $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ за всяко $(x, y) \in \mathbb{R}^2$ се нарича съвместна функция на разпределение на (X, Y)

Дефиниция: (маргинална функция на разпределение): Нека $F_{X,Y}(x, y)$ е съвместна функция на разпределение на (X, Y) - произволни случайни величини. Тогава функцията $F_X(x) = F_{X,Y}(x, \infty)$ се нарича маргинална функция на разпределение на X. Аналогично се дефинира и $F_Y(y)$.

Дефиниция: (независимост): Случайните величини X, Y, дефинирани в едно вероятностно пространство са независими $(X \perp Y)$, ако за всяко $(x, y) \in \mathbb{R}^2$ е изпълнено, че $F_{X,Y}(x, y) = F_X(x) F_Y(y)$

- Ако X, Y - дискретни, то това е еквивалентно на $\mathbb{P}(X = x_k, Y = y_j) = \mathbb{P}(X = x_k)\mathbb{P}(Y = y_j)$

Дефиниция: (ковариация): Нека X, Y - случайни величини и $\mathbb{D}[X], \mathbb{D}[Y] < \infty$ са дефинирани. Тогава ковариация наричаме

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Твърдение: $cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

- **Доказателство:**
  $cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] =$
  $\mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] =$
  $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] =$
  $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- **Следствие:** Ако X, Y са независими, то $cov(X, Y) = 0$
  - **Доказателство:** $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \Rightarrow cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$

Дефиниция: (корелация): Нека X, Y - случайни величини и $\mathbb{D}[X], \mathbb{D}[Y] < \infty$ са дефинирани. Тогава коефициент на корелация наричаме

$$\rho(X, Y) = cor(X, Y) = \frac{cov(X, Y)}{\sqrt{\mathbb{D}[X]}\sqrt{\mathbb{D}[Y]}}$$

- Корелацията показва степента на линейност между X, Y.

Твърдение: Нека X, Y - случайни величини и $\mathbb{D}[X], \mathbb{D}[Y] < \infty$ са дефинирани. Тогава, ако $\overline{X} = \dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}$ и $\overline{Y} = \dfrac{Y - \mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}}$ е изпълнено, че

- $\mathbb{E}[\overline{X}] = \mathbb{E}[\overline{Y}] = 0$ - от центрирането
- $\mathbb{D}[\overline{X}] = \mathbb{D}[\overline{Y}] = 1$ - от нормирането
- $\rho(X, Y) = \mathbb{E}[\overline{X} \cdot \overline{Y}]$
- **Доказателство:**
  - $\mathbb{E}[\overline{X}] = \mathbb{E}\left[\dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}\right] = \dfrac{1}{\sqrt{\mathbb{D}[X]}}\mathbb{E}[X - \mathbb{E}[X]] = \dfrac{\mathbb{E}[X] - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}} = 0$
  - $\mathbb{D}[\overline{X}] = \mathbb{D}\left[\dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}\right] = \dfrac{1}{\sqrt{\mathbb{D}[X]}^2}\mathbb{D}[X - \mathbb{E}[X]] = \dfrac{\mathbb{D}[X]}{\mathbb{D}[X]} = 1$
  - $\rho(X, Y) = \dfrac{cov(X, Y)}{\sqrt{\mathbb{D}[X]}\sqrt{\mathbb{D}[Y]}} = \dfrac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{D}[X]}\sqrt{\mathbb{D}[Y]}} =$
    $\mathbb{E}\left[\dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}} \cdot \dfrac{Y - \mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}}\right] = \mathbb{E}[\overline{X} \cdot \overline{Y}]$

Теорема: Нека X, Y - случайни величини и $\mathbb{D}[X], \mathbb{D}[Y] < \infty$ са дефинирани. Тогава:

- $|\rho(X,Y)| \leq 1$
- $|\rho(X,Y)| = 1 \iff \exists a, b \in \mathbb{R} : Y = aX + b$
- $|\rho(X,Y)| = 0 \iff X \perp Y$
- **Доказателство:**
  - Ще използваме, че $\rho(X,Y) = \mathbb{E}[\overline{X} \cdot \overline{Y}]$
    - $\mathbb{E}[(\overline{X} + \overline{Y})^2] = \mathbb{E}[\overline{X}^2] + \mathbb{E}[\overline{Y}^2] + 2\mathbb{E}[\overline{X} \cdot \overline{Y}] =$
      $\mathbb{D}[X] - \mathbb{E}[X] + \mathbb{D}[Y] - \mathbb{E}[Y] + 2\rho(X,Y) =$
      $2 + 2\rho(X,Y) \geq 0 \Rightarrow \rho(X,Y) \geq -1$
    - $\mathbb{E}[(\overline{X} - \overline{Y})^2] = \mathbb{E}[\overline{X}^2] + \mathbb{E}[\overline{Y}^2] - 2\mathbb{E}[\overline{X} \cdot \overline{Y}] =$
      $\mathbb{D}[X] - \mathbb{E}[X] + \mathbb{D}[Y] - \mathbb{E}[Y] - 2\rho(X,Y) =$
      $2 - 2\rho(X,Y) \geq 0 \Rightarrow \rho(X,Y) \leq 1$
    - $\implies -1 \leq \rho(X,Y) \leq 1 \Leftrightarrow |\rho(X,Y)| \leq 1$
  - Ще разгледаме двете посоки:
    - Нека $Y = aX + b$
      $Y - \mathbb{E}[Y] = aX + b - \mathbb{E}[Y]$
      $Y - \mathbb{E}[Y] = aX + b - \mathbb{E}[Y] + a\mathbb{E}[X] - a\mathbb{E}[X]$
      $\dfrac{Y - \mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}} = \dfrac{aX - a\mathbb{E}[X]}{\sqrt{\mathbb{D}[Y]}} + \dfrac{b - \mathbb{E}[Y] + a\mathbb{E}[X]}{\sqrt{\mathbb{D}[Y]}}$
      $\overline{Y} = \dfrac{a(X - \mathbb{E}[X])\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}\sqrt{\mathbb{D}[X]}} + w$, където $w = \dfrac{b - \mathbb{E}[Y] + a\mathbb{E}[X]}{\sqrt{\mathbb{D}[Y]}}$
      $\overline{Y} = a\overline{X}\dfrac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}} + w$
      $\overline{Y} = v\overline{X} + w$, където $v = a\dfrac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}}$
      $0 = \mathbb{E}[\overline{Y}] = \mathbb{E}[v\overline{X} + w] = w \Rightarrow \overline{Y} = v\overline{X}$
      $1 = \mathbb{D}[\overline{Y}] = \mathbb{D}[v\overline{X}] = v^2 \Rightarrow |v| = 1$
      $\rho(X,Y) = \mathbb{E}[\overline{X} \cdot \overline{Y}] = \mathbb{E}[\overline{X} \cdot v\overline{X}] = v\mathbb{E}[\overline{X}^2] = v = \pm 1$
      $|\rho(X,Y)| = 1$
    - Нека $\rho(X,Y) = 1 = \mathbb{E}[\overline{X} \cdot \overline{Y}]$
      $\mathbb{E}[(\overline{X} - \overline{Y})^2] = \mathbb{E}[\overline{X}^2] + \mathbb{E}[\overline{Y}^2] - 2\mathbb{E}[\overline{X} \cdot \overline{Y}] = 1 + 1 - 2 \cdot 1 = 0 \Rightarrow$
      $\overline{X} - \overline{Y} = 0 \Rightarrow \overline{X} = \overline{Y}$
      $\overline{X} = \overline{Y} \Leftrightarrow \dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}} = \dfrac{Y - \mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}}$
      $Y = \dfrac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}}X - \mathbb{E}[X]\dfrac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}} + \mathbb{E}[Y]$

$Y = aX + b$, където $a = \frac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}}$ и $b = \mathbb{E}[Y] - \mathbb{E}[X]\frac{\sqrt{\mathbb{D}[X]}}{\sqrt{\mathbb{D}[Y]}}$

За $\rho(X, Y) = -1$ аналогично $\overline{Y} = -\overline{X}$

Твърдение: $\mathbb{D}[X + Y] = \mathbb{D}[X] + \mathbb{D}[Y] + 2cov(X, Y)$