9. cvičení z STP

15. - 19. duben 2019

Příklad 9.1 *Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě* 67 *m a* 75 *m a směrodatné odchylky* 6 *m a* 3 *m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.*

Řešení:
Náhodná veličina

$A =$"*délka hodu Anny*"

má rozdělení $N(67, 6^2)$ a veličina

$B =$"*délka hodu Barbory*"

má rozdělení $N(75, 3^2)$.

Zajímá nás $P(A > B) = P(A-B > 0)$. Protože veličiny A a B jsou nezávislé, tak veličina $Z := A-B$ má také normální rozdělení, a sice

$$Z \sim N(67-75, 6^2+3^2) = N(-8, 45)\ .$$

Takže

$$P(A > B) = P(Z > 0) = P\left(\underbrace{\frac{Z-(-8)}{\sqrt{45}}}_{norm(Z)} > \frac{0-(-8)}{\sqrt{45}}\right) = 1 - P\left(norm(Z) \leq \frac{8}{\sqrt{45}}\right) =$$

$$= 1 - \Phi\left(\tfrac{8}{\sqrt{45}}\right) \doteq 1 - \Phi(1.1926) \doteq 1 - 0.883 = 0.117\,.$$

POZOR! Zatímco střední hodnota je lineární zobrazení, tak rozptyl se chová jinak! Konkrétně je to takto:

Nechť X a Y jsou veličiny se střední hodnotou a konečným rozptylem. Pak

- $E(X \pm Y) = E(X) \pm E(Y)$
- $\mathrm{var}(X \pm Y) = \mathrm{var}(X) + \mathrm{var}(Y) \pm 2 \cdot \mathrm{cov}(X, Y)\ (\geq 0)$

Zde $\mathrm{cov}(X, Y)$ je tzv. kovariance (viz poznámky níže). Speciálně, pokud X a Y jsou nezávislé, je $\mathrm{cov}(X, Y) = 0$. Máme tedy:

- X a Y nezávislé $\Rightarrow$ $\mathrm{var}(X \pm Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

Tedy v tomto případě se rozptyly VŽDY sčítají!

Poznámky ke kovarianci a korelaci: Náhodné veličiny (jako funkce na pravděpodobnostním prostoru Ω) tvoří přirozené (reálný) vektorový prostor (kde dvě veličiny navíc ještě budeme pokládat za totožné, pokud se rovnají s pravděpodobností 1). Na vektorovém podprostoru veličin s konečnou střední hodnotou a konečným rozptylem pak můžeme přirozeným způsobem zavést skalární součin jako

$$\langle X|Y\rangle := E(XY)$$

Díky němu můžeme přirozeně zavést *normu* $\|X\|$ (neboli "délku" vektoru X) jako

$$\|X\| := \sqrt{\langle X|X\rangle} = \sqrt{E(X^2)} \ .$$

Mimo jiné si všimněme, že pro X je $\mathrm{var}(X) = \|X - E(X)\|^2$, takže platí

$$\|norm(X)\| = \left\| \frac{X - EX}{\sqrt{\mathrm{var}(X)}} \right\| = \frac{\|X - EX\|}{\sqrt{\mathrm{var}(X)}} = 1$$

neboli $norm(X)$ má délku skutečně znormovanou na hodnotu 1.

Skalární součin nám dále umožňuje měřit také úhel mezi dvěma vektory. Pro veličiny X a Y je užitečné znát, jestli jejich výchylky vůči středním hodnotám (tj. veličiny $X - EX$ a $Y - EY$) mají podobné chování (tj. jestli korelují). Zavádíme proto korelaci mezi veličinami X a Y jako kosinus úhlu α mezi vektory $X - EX$ a $Y - EY$, tedy

$$\mathrm{corr}(X,Y) := \frac{\langle X - EX \mid Y - EY\rangle}{\|X - EX\| \cdot \|Y - EY\|} = \cdots = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{\mathrm{var}(X)\cdot \mathrm{var}(Y)}} \ .$$

(Veličiny $X - EX$ a $Y - EY$ mají nulovou střední hodnotu).

Kovariance pro X a Y je pak definována jako

$$\mathrm{cov}(X,Y) := \langle X - EX \mid Y - EY\rangle = \cdots = E(XY) - E(X)\cdot E(Y) \ .$$

Speciálně je

$$\mathrm{cov}(X,X) = \mathrm{var}(X) \quad \text{a} \quad \mathrm{corr}(X,X) = 1 \ .$$

Praktické použití korelace:

Pokud máme dvě veličiny X a Y takové, že

$$X - EX \geq 0 \iff Y - EY \geq 0 \quad \Big(\text{což implikuje, že} \quad (X - EX)(Y - EY) \geq 0\Big)$$

pak je $\mathrm{corr}(X,Y) \geq 0$.

Obdobně platí: Jestliže

$$X - E(X) \geq 0 \iff Y - E(Y) \leq 0$$

pak je $\mathrm{corr}(X,Y) \leq 0$.

Ačkoliv zpětné implikace v obou případech neplatí, přesto nám korelace umožňuje nějakým způsobem zachytit jistou míru kauzální závislosti dvou veličin.

Příklad 9.2 *Sdružené pravděpodobnosti náhodných veličin X a Y jsou dány následující tabulkou:*

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	1/4	1/8	0
$Y = 1$	1/4	1/4	1/8

(a) *Jaká jsou jejich marginální rozdělení?*

(b) *Určete varianční a korelační matici.*

(c) *Jsou veličiny X a Y nezávislé? Zdůvodněte.*

Řešení:

(a) Marginální (česky: okrajová) rozdělení náhodného vektoru (X, Y) jsou rozdělení jeho jednotlivých složek, tedy veličin X a Y. Vektor (X, Y) má diskrétní rozdělení a protože je zadané tabulkou, musí platit, že součet všech pravděpodobností v tabulce musí být 1.

Obě veličiny X a Y budou mít také diskrétní rozdělení a pro jejich rozdělení platí:

$$P(X = i) = P(X = i,\ Y \in \mathbb{R}) = \sum_{j \in \mathbb{R}} P(X = i,\ Y = j)$$

$$P(Y = j) = P(X \in \mathbb{R},\ Y = j) = \sum_{i \in \mathbb{R}} P(X = i,\ Y = j)$$

Pravděpodobnosti pro jednotlivé hodnoty tak získáme sečtením pravděpodobností ve sloupcích (pro X) a v řádcích (pro Y) naší tabulky:

	$X = 0$	$X = 1$	$X = 2$	$P(Y = j)$
$Y = 0$	1/4	1/8	0	3/8
$Y = 1$	1/4	1/4	1/8	5/8
$P(X = i)$	1/2	3/8	1/8	

Tedy

$$P(X = i) = \begin{cases} 1/2, & i = 0 \\ 3/8, & i = 1 \\ 1/8, & i = 2 \\ 0, & \text{jinak} \end{cases} \quad \text{a} \quad P(Y = j) = \begin{cases} 3/8, & j = 0 \\ 5/8, & j = 1 \\ 0, & \text{jinak.} \end{cases}$$

(b) Kovarianci vypočteme ze vztahu

$$\operatorname{cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

$$\begin{aligned} E(X) &= 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{1}{8} = \tfrac{5}{8} \\ E(Y) &= 0 \cdot \tfrac{3}{8} + 1 \cdot \tfrac{5}{8} = \tfrac{5}{8} \\ E(XY) &= \sum_{(i,j) \in \mathbb{R}^2} i \cdot j \cdot P(X = i,\ Y = j) = \\ &= 0 \cdot 0 \cdot \tfrac{1}{4} + 0 \cdot 1 \cdot \tfrac{1}{8} + 0 \cdot 2 \cdot 0 + 1 \cdot 0 \cdot \tfrac{1}{4} + 1 \cdot 1 \cdot \tfrac{1}{4} + 1 \cdot 2 \cdot \tfrac{1}{8} = \tfrac{1}{2} \end{aligned}$$

Takže

$$\operatorname{cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = \tfrac{1}{2} - \tfrac{5}{8} \cdot \tfrac{5}{8} = \tfrac{7}{64}.$$

Pro korelaci

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

potřebujeme ještě znát rozptyly, takže si je dopočítáme:

$$\begin{aligned} E(X^2) &= 0^2 \cdot \tfrac{1}{2} + 1^2 \cdot \tfrac{3}{8} + 2^2 \cdot \tfrac{1}{8} = \tfrac{7}{8}, \\ E(Y^2) &= 0^2 \cdot \tfrac{3}{8} + 1^2 \cdot \tfrac{5}{8} = \tfrac{5}{8}, \\ \mathrm{var}(X) &= E(X^2) - \Big(E(X)\Big)^2 = \tfrac{7}{8} - (\tfrac{5}{8})^2 = \tfrac{31}{64}, \\ \mathrm{var}(Y) &= E(Y^2) - \Big(E(Y)\Big)^2 = \tfrac{5}{8} - (\tfrac{5}{8})^2 = \tfrac{15}{64}. \end{aligned}$$

Všimněme si ještě, že $Y \sim \mathrm{Alt}(\frac{5}{8})$, takže rozptyl jsme mohli spočítat jako $\mathrm{var}(Y) = \frac{5}{8} \cdot (1 - \frac{5}{8}) = \frac{15}{64}$.
Korelace tedy je

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)} \cdot \sqrt{\mathrm{var}(Y)}} = \frac{\frac{7}{64}}{\sqrt{\frac{31}{64}} \cdot \sqrt{\frac{15}{64}}} = \frac{7}{\sqrt{465}} \doteq 0.32462,$$

Varianční matice je tudíž

$$\mathrm{Var}(X,Y) = \begin{pmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} \mathrm{var}(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & \mathrm{var}(Y) \end{pmatrix} = \begin{pmatrix} \frac{31}{64} & \frac{7}{64} \\ \frac{7}{64} & \frac{15}{64} \end{pmatrix}$$

a korelační matice je

$$\mathrm{Corr}(X,Y) = \begin{pmatrix} \mathrm{corr}(X,X) & \mathrm{corr}(X,Y) \\ \mathrm{corr}(Y,X) & \mathrm{corr}(Y,Y) \end{pmatrix} = \begin{pmatrix} 1 & \mathrm{corr}(X,Y) \\ \mathrm{corr}(X,Y) & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{7}{\sqrt{465}} \\ \frac{7}{\sqrt{465}} & 1 \end{pmatrix}.$$

Pro zajímavost si ještě můžeme zjistit úhel $\alpha \in \langle 0, \pi \rangle$ mezi našimi náhodnými veličinami $X - EX$ a $Y - EY$:

$$\alpha = \arccos\big(\mathrm{corr}(X,Y)\big) \doteq \arccos\big(0.32462\big) \doteq 71.06^\circ\ .$$

(c) Jestliže náhodný vektor (X,Y) má diskrétní rozdělení, pak:

veličiny X a Y jsou nezávislé $\quad\Leftrightarrow\quad P(X = i, Y = j) = P(X = i) \cdot P(Y = j)$ pro všechna $i, j \in \mathbb{R}$

Naše náhodné veličiny tudíž NEJSOU nezávislé, protože např.

$$P(X = 2, Y = 0) = 0 \neq \tfrac{1}{8} \cdot \tfrac{3}{8} = P(X = 2) \cdot P(Y = 0).$$

Jiným důvodem je také fakt, že $\mathrm{cov}(X,Y) \neq 0$. (Zdůrazněme ale, že pokud je kovariance nulová, nemůžeme o nezávislosti obecně nic říct!)

Příklad 9.3 *Sdružená hustota náhodných veličin X a Y je*

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{2} e^{-x-\frac{y}{2}}\,, & x > 0, y > 0\,, \\ 0\,, & \textit{jinak}. \end{cases}$$

(a) Jaká jsou jejich marginální rozdělení?

(b) Jsou veličiny X a Y nezávislé? Zdůvodněte.

(c) Jak vypadá jejich korelační matice?

Řešení:

(a) Marginální hustoty (tj. hustoty jednotlivých veličin X a Y) jsou

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \begin{cases} \int_0^{\infty} \frac{1}{2} e^{-x-\frac{y}{2}} dy = e^{-x} \cdot \left[-e^{-\frac{y}{2}} \right]_0^{\infty} = e^{-x} & \text{pro } x > 0, \\ \int_{-\infty}^{\infty} 0\ dy = 0 & \text{pro } x \leq 0\ . \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \begin{cases} \int_0^{\infty} \frac{1}{2} e^{-x-\frac{y}{2}} dx = \frac{1}{2} e^{-\frac{y}{2}} \cdot \left[-e^{-x} \right]_0^{\infty} = \frac{1}{2} e^{-\frac{y}{2}} & \text{pro } y > 0, \\ \int_{-\infty}^{\infty} 0\ dx = 0 & \text{pro } y \leq 0\ . \end{cases}$$

Vidíme tedy, že obě rozdělení jsou exponenciální, konkrétně $X \sim \text{Exp}(1)$ a $Y \sim \text{Exp}(\frac{1}{2})$.

(b) Složky X a Y jsou nezávislé právě tehdy, když

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \quad \text{pro skoro všechna} \quad (x,y) \in \mathbb{R}^2\ ,$$

což znamená, že množina bodů, kde uvedená rovnost neplatí má nulový plošný obsah.

(Podmínce "skoro všude" se nelze vyhnout z toho důvodu, že hustoty nejsou jednoznačně definovány svými hodnotami, ale svými integrály.)

Jak je hned vidět, v našem případě je rovnost splněna dokonce všude, takže X a Y JSOU nezávislé.

(c) Z nezávislosti X, Y plyne okamžitě $\text{cov}(X,Y) = 0$, tedy také $\text{corr}(X,Y) = 0$ a korelační matice je tak

$$\text{Corr}(X,Y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$