7. cvičení z PST

15. listopadu 2017

7.1 (korelace)
Náhodný vektor (X, Y) má následující parametry:

$$E(X) = 10, \quad \sigma_X = 5, \quad E(Y) = 150, \quad \sigma_Y = 20, \quad \varrho(X, Y) = 0.5 \text{ (korelace)}.$$

Stanovte střední hodnotu a rozptyl náhodných veličin

$$T = 2X + 3, \quad U = 200 - Y, \quad V = X + Y \ .$$

Řešení:
Příklad 1: `https://math.feld.cvut.cz/prucha/psi/wcv4.pdf`

7.2 (kovariance)
Pro náhodné veličiny X a Y platí, že

$$D(X) = 3, \quad D(Y) = 4, \quad \text{cov}(X, Y) = -2.$$

Pro náhodné veličiny

$$U = 3X + 4Y - 1 \quad \text{a} \quad V = -2X + 2Y + 3$$

určete

(a) koeficient kovariance $\text{cov}(U, V)$.

(b) rozptyl $D(X + Y)$.

Řešení:
Kovariance $\text{cov}(\cdot, \cdot)$ má tyto vlastnosti:

- je lineární v každé složce zvlášť (tj. bilineární),
- $\text{cov}(Z + c, W + d) = \text{cov}(Z, W)$ pro všechna $c, d \in \mathbb{R}$,
- symetrická (tj. $\text{cov}(Z, W) = \text{cov}(W, Z)$) a
- $\text{cov}(Z, Z) = D(Z)$.

(a) Díky tomu můžeme tedy jednotlivé složky "roznásobit":

$$\text{cov}(U, V) = \text{cov}(3X + 4Y - 1, -2X + 2Y + 3) = \text{cov}(3X + 4Y, -2X + 2Y) =$$

$$= 3 \cdot (-2) \cdot \text{cov}(X, X) + 3 \cdot 2 \cdot \text{cov}(X, Y) + 4 \cdot (-2) \cdot \text{cov}(Y, X) + 4 \cdot 2 \cdot \text{cov}(Y, Y) =$$
$$= (-6) \cdot D(X) + (-2) \cdot \text{cov}(X, Y) + 8 \cdot D(Y) = (-6) \cdot 3 + (-2) \cdot (-2) + 8 \cdot 4 = 18 \ .$$

Jak je vidět, znalost $E(X)$ a $E(Y)$ jsme vůbec nepotřebovali!

Poznámka: Vlastnosti kovariance si můžeme zapamatovat pomocí definice

$$\operatorname{cov}(Z, W) = \pi(Z) \bullet \pi(W) \ ,$$

kde $\pi(Z) = Z - E(Z)$.

Je dobré si všimnout, ze díky bilinearitě můžeme také používat přehlednější maticový zápis:

$$\operatorname{cov}(aX + bY, cX + dY) = (a, b) \begin{pmatrix} \operatorname{cov}(X,X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & \operatorname{cov}(Y,Y) \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix}$$

V našem případě tedy

$$\operatorname{cov}(U, V) = (3, 4) \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ 2 \end{pmatrix} = (3, 4) \begin{pmatrix} -10 \\ 12 \end{pmatrix} = 18 \ .$$

(b) Využijeme vlastnosti kovariance:

$$D(X + Y) = \operatorname{cov}(X + Y, X + Y) = \operatorname{cov}(X, X) + \operatorname{cov}(X, Y) + \operatorname{cov}(Y, X) + \operatorname{cov}(Y, Y) =$$

$$= D(X) + 2 \cdot \operatorname{cov}(X, Y) + D(Y) =$$

$$= 3 + 2 \cdot (-2) + 4 = 3 \ .$$

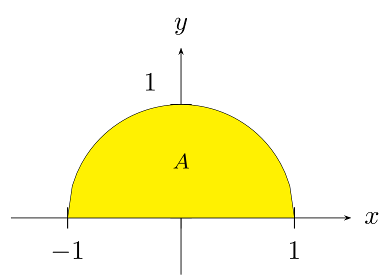
7.3 (kovariance)

Náhodný vektor (X, Y) má spojité rovnoměrné rozdělení v množině

$$A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1, \ y \geq 0\} \ .$$

Určete koeficient kovariance $cov(X, Y)$.

Řešení:
Množina E je půlkruh



Pro kovarianci máme

$$cov(X, Y) = E(XY) - E(X) \cdot E(Y) \ .$$

Přitom je

$$E(XY) = \iint\limits_{\mathbb{R}^2} xy \cdot f_{X,Y}(x, y) \ dx \ dy = \iint\limits_{A} \tfrac{2}{\pi} xy \ dx \ dy = \begin{bmatrix} x = r \cos\varphi \\ y = r \sin\varphi \\ 0 \leq r \leq 1 \\ 0 \leq \varphi \leq \pi \end{bmatrix} =$$

$$= \tfrac{2}{\pi} \int_0^{\pi} \int_0^{1} r^3 \cos\varphi \sin\varphi \ dr \ d\varphi = \tfrac{2}{\pi} \left(\int_0^{1} r^3 \ \ dr \right) \cdot \left(\int_0^{\pi} \cos\varphi \sin\varphi \ d\varphi \right) =$$

$$= \frac{2}{\pi} \cdot \frac{1}{4} \cdot \left[\frac{\sin^2 \varphi}{2} \right]_{\varphi=0}^{\varphi=\pi} = 0$$

To, že tento integrál vyjde nula bylo vidět už na začátku z lichosti integrované funkce vzhledem k proměnné x. Ze stejných důvodů bude nulový i následující integrál:

$$E(X) = \iint_{\mathbb{R}^2} x \cdot f_{X,Y}(x,y)\ dx\ dy = 0$$

Máme tedy

$$cov(X,Y) = 0$$

a připomeňme si, že veličiny X a Y přitom nezávislé nejsou (viz tvar množiny A). Z cvičných důvodů si ještě spočítáme $E(Y)$:

$$E(Y) = \iint_{\mathbb{R}^2} y \cdot f_{X,Y}(x,y)\ dx\ dy = \iint_{A} \frac{2}{\pi} y\ dx\ dy = \begin{bmatrix} x=r\cos\varphi \\ y=r\sin\varphi \\ 0\leq r\leq 1 \\ 0\leq\varphi\leq\pi \end{bmatrix} =$$

$$= \frac{2}{\pi} \int_0^{\pi} \int_0^{1} r^2 \sin\varphi\ dr\ d\varphi = \frac{2}{\pi} \left(\int_0^1 r^2\ \ dr \right) \cdot \left(\int_0^{\pi} \sin\varphi\ d\varphi \right) = \frac{2}{\pi} \cdot \frac{1}{3} \cdot 2 = \frac{4}{3\pi}\ .$$

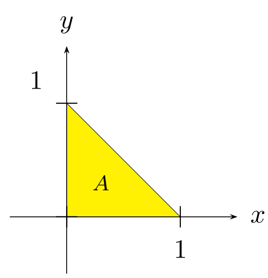
7.4 (korelace)

Náhodný vektor (X,Y) má spojité rovnoměrné rozdělení v množině

$$A = \{(x,y) \mid y \geq 0,\ x+y \leq 1,\ x \geq 0\}\ .$$

Určete koeficient korelace $\varrho(X,Y)$.

Řešení:
Množina A je trojúhelník



Sdružená hustota je tedy dána jako

$$f_{X,Y}(x,y) = \begin{cases} 2 & ,\ (x,y) \in A \\ 0 & ,\ \text{jinak,} \end{cases}$$

tedy tak, aby byla konstantní na A, nulová mimo A a integrál z hustoty byl roven 1.

Pro kovarianci máme

$$\varrho(X,Y) = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{D(X) \cdot D(Y)}}\ .$$

Přitom je

$$E(XY) = \iint\limits_{\mathbb{R}^2} xy \cdot f_{X,Y}(x,y)\ dx\ dy = \iint\limits_{A} 2xy\ dx\ dy =$$

$$= \int_0^1 \int_0^{1-y} 2xy\ dx\ dy = \int_0^1 y(1-y)^2\ dy = \int_0^1 y - 2y^2 + y^3\ dy = \tfrac{1}{2} - \tfrac{2}{3} + \tfrac{1}{4} = \tfrac{1}{12}\ .$$

A podobně máme

$$E(X) = \iint\limits_{A} 2x\ dx\ dy = \int_0^1 \int_0^{1-y} 2x\ dx\ dy = \int_0^1 (1-y)^2\ dy = \tfrac{1}{3}\ .$$

$$E(X^2) = \iint\limits_{A} 2x^2\ dx\ dy = \int_0^1 \int_0^{1-y} 2x^2\ dx\ dy = \tfrac{2}{3} \int_0^1 (1-y)^3\ dy = \tfrac{2}{3} \cdot \tfrac{1}{4} = \tfrac{1}{6}\ .$$

Ze symetrie máme $E(Y) = E(X) = \frac{1}{3}$ a $E(Y^2) = E(X^2) = \frac{1}{6}$. Takže

$$D(Y) = D(X) = E(X^2) - \big(E(X)\big)^2 = \tfrac{1}{6} - (\tfrac{1}{3})^2 = \tfrac{1}{18}$$

a tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{D(X) \cdot D(Y)}} = \frac{\frac{1}{12} - (\frac{1}{3})^2}{\frac{1}{18}} = -\tfrac{1}{2}\ .$$

7.5 (korelace)

Náhodný vektor (X,Y) má spojité rovnoměrné rozdělení v množině

$$A = \{(x,y) \in \mathbb{R}^2 \mid y \geq 0,\ x + y \leq 1,\ y - x \leq 1\}\ .$$

Určete koeficient korelace $\varrho(X,Y)$.

Řešení:

Příklad 3: `https://math.feld.cvut.cz/prucha/psi/wcv4.pdf`

7.6 (kovarianční a korelační matice)

Náhodný vektor (X,Y) má kovarianční matici.

$$\begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

(a) Rozhodněte, zda jsou náhodné veličiny X a Y závislé či nezávislé.

(b) Napište korelační matici náhodného vektoru (X,Y).

(c) Napište korelační a kovarianční matice náhodných vektorů $(X,-Y)$ a $(X, 2Y-1)$.

Řešení:
Příklad 9: `https://math.feld.cvut.cz/prucha/psi/wcv4.pdf`

7.7 (náhodný vektor - diskrétní)
Dvourozměrný náhodný vektor (X, Y) má pravděpodobnosti hodnot dané tabulkou:

X \ Y	0	1	2
0	1/18	1/9	1/6
1	1/9	1/18	1/9
2	1/6	1/6	1/18

(a) Stanovte pravděpodobnost $P(\frac{1}{2} \leq X \leq 3 \,\&\, Y \geq 1)$.

(b) Určete marginální rozdělení veličin X a Y.

(c) Zjistěte, zda X a Y jsou nezávislé. Pokud ne, vypočtěte korelaci a popište rozdělení náhodného vektoru (X', Y') se stejnými marginálními rozděleními, jehož složky jsou nezávislé veličiny.

Řešení:
(a) Zřejmě

$$P\left(\frac{1}{2} \leq X \leq 3 \,\&\, Y \geq 1\right) = P\Big((X, Y) \in \{(1, 1), (1, 2), (2, 1), (2, 2)\}\Big) = \frac{1}{18} + \frac{1}{9} + \frac{1}{6} + \frac{1}{18} = \frac{7}{18}\ .$$

(b) Marginální pravděpodobnostní funkce p_X a p_Y (tj. pravděpodobnostní funkce jednotlivých složek vektoru) získáme pro jednotlivé hodnoty sečtením pravděpodobností v řádcích (p_X) a sloupcích (p_Y) naší tabulky:

X \ Y	0	1	2	p_X
0	1/18	1/9	1/6	1/3
1	1/9	1/18	1/9	5/18
2	1/6	1/6	1/18	7/18
p_Y	1/3	1/3	1/3	

(c) Protože např. $p_{X,Y}(0, 0) = \frac{1}{18} \neq \frac{1}{3} \cdot \frac{1}{3} = p_X(0) \cdot p_Y(0)$, tak X a Y jsou závislé. Spočítáme tedy jejich korelaci (můžeme si pomoci i maticí):

$$E(X) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{5}{18} + 2 \cdot \frac{7}{18} = \frac{19}{18}$$

$$E(X^2) = 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{5}{18} + 2^2 \cdot \frac{7}{18} = \frac{11}{6}$$

$$E(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$$

$$E(Y^2) = 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} = \frac{5}{3}$$

$$E(XY) = (0,1,2) \cdot \begin{pmatrix} 1/18 & 1/9 & 1/6 \\ 1/9 & 1/18 & 1/9 \\ 1/6 & 1/6 & 1/18 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} =$$

$$= (1,2) \cdot \begin{pmatrix} 1/18 & 1/9 \\ 1/6 & 1/18 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{15}{18}$$

Korelace tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{E(X^2)-E(X)^2}\cdot\sqrt{E(Y^2)-E(Y)^2}} = \frac{\frac{15}{18} - \frac{19}{18}\cdot 1}{\sqrt{\frac{11}{6} - (\frac{19}{18})^2}\cdot\sqrt{\frac{5}{3} - 1^2}} =$$

$$= -4\sqrt{\frac{3}{466}} \doteq -0,32094 \doteq \arccos(108,72^\circ)\ .$$

Nechť (X', Y') je nyní náhodný vektor s nezávislými složkami takovými, že $p_{X'} = p_X$ a $p_{Y'} = p_Y$. Pak pro jeho pravděpodobnostní funkci máme $p_{X',Y'}(i,j) = p_{X'}(i)\cdot p_{Y'}(j)$ a můžeme ji tak popsat následující tabulkou:

X' \ Y'	0	1	2	$p_{X'}$
0	1/9	1/9	1/9	1/3
1	5/54	5/54	5/54	5/18
2	7/54	7/54	7/54	7/18
$p_{Y'}$	1/3	1/3	1/3	

7.8 (náhodný vektor - diskrétní)

Náhodný vektor (X,Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}$, která je daná tabulkou

$p_{X,Y}(x,y)$: y / x	-1	0	1
1	1/6	0	1/3
2	1/8	1/4	1/8

(a) Vypočtěte pravděpodobnost $P\Big(|Y| \geq X\Big)$, střední hodnotu $\mathrm{E}(X^2\,Y)$ a rozptyl $D(X)$.

(b) Rozhodněte, zda jsou náhodné veličiny X a Y závislé či nezávislé.

Řešení:

(a) Nenulové pravděpodobnosti pro případ $|Y| \geq X$ nastávají právě když $(X,Y) \in \{(1,-1),\ (1,1)\}$. Takže

$$P\big(|Y| \geq X\big) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}\ .$$

Pro střední hodnotu veličiny XY^2 použijeme obvyklý vzorec s pravděpodobnostní funkcí

$$E(X^2Y) = \sum_{(i,j)\in\mathbb{R}^2} i^2 \cdot j \cdot p_{X,Y}(i,j) =$$

$$= 1^2\cdot(-1)\cdot\frac{1}{6} + 1^2\cdot 0\cdot 0 + 1^2\cdot 1\cdot\frac{1}{3} + 2^2\cdot(-1)\cdot\frac{1}{8} + 2^2\cdot 0\cdot\frac{1}{4} + 2^2\cdot 1\cdot\frac{1}{8} =$$

$$= -\frac{1}{6} + \frac{1}{3} - \frac{1}{2} + \frac{1}{2} = -\frac{1}{6}\ .$$

Pro výpočet rozptylu veličiny X je nejjednodušší si určit její pravděpodobnostní funkci

$$p_X(1) = \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}$$

$$p_X(2) = 1 - p_X(1) = \frac{1}{2}\ .$$

Takže máme

$$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}$$

$$E(X^2) = 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{2} = \frac{5}{2}$$

a tudíž

$$D(X) = E(X^2) - \Big(E(X)\Big)^2 = \frac{5}{2} - \left(\frac{3}{2}\right)^2 = \frac{1}{4}\ .$$

(b) Protože nějaké dva řádky (zde tedy všechny) v tabulce hodnot

1/6	0	1/3
1/8	1/4	1/8

sdružené pravděpodobnostní funkce jsou *lineárně NEzávislé*, tak veličiny X a Y jsou ZÁVISLÉ. Jiné zdůvodnění je například, že máme

$$p_{X,Y}(1,0) = 0 \neq \frac{1}{2} \cdot \left(0 + \frac{1}{4}\right) = p_X(1) \cdot p_Y(0)$$

a tudíž veličiny musí být ZÁVISLÉ.

7.9 (náhodný vektor - diskrétní)

Náhodný vektor (X, Y) má sdruženou pravděpodobnostní funkci $p = p(x, y)$ dánu tabulkou:

$p_{X,Y}(x,y)$: y / x	0	1	2
1	1/8	0	1/8
3	0	1/6	3/8
4	1/8	1/12	0

Určete:

(a) Marginální pravděpodobnostní funkce p_X a p_Y náhodných veličin X a Y ;

(b) zda jsou náhodné veličiny X a Y závislé či nezávislé;

(c) koeficient korelace $\varrho(X, Y)$.

Řešení:

Příklad 5: `https://math.feld.cvut.cz/prucha/psi/wcv4.pdf`

7.10 (normální rozdělení)

Nechť veličina X má normované normální rozdělení $N(0,1)$. Určete $P(X^2 < 3X-2)$ a najděte takové číslo ε, že $P(|X| < \varepsilon) = 0.95$.

Řešení:
Máme

$$P(X^2 < 3X-2) = P(X^2-3X+2<0) = P\Big((X-1)(X-2)<0\Big) = P\Big(1<X<2\Big) =$$

$$= \Phi(2) - \Phi(1) \doteq 0.97725 - 0.84134 = 0.13591 \ .$$

A dále je

$$0.95 = P(|X|<\varepsilon) = P(-\varepsilon < X < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1$$

a tedy

$$\varepsilon = \Phi^{-1}\left(\frac{1+0.95}{2}\right) = \Phi^{-1}(0.975) \doteq 1.96 \ .$$

7.11 (normální rozdělení)

Rozvodné závody dodávaly elektřinu, jejíž napětí ve voltech mělo normální rozdělení $\mathrm{N}(\mu_1, \sigma_1^2)$, kde $\mu_1 = 230$ a $\sigma_1^2 = 25$. Horní mez U_0 dodávaného napětí je nejnižší mez, která je překročena s pravděpodobností nejvýše $\alpha = 10^{-4}$. Nyní se závodům podařilo snížit rozptyl na $\sigma_2^2 = 10$. O kolik mohou zvýšit střední hodnotu μ_2, aby byla zachována horní mez?

Řešení:
Máme tedy veličinu

$$U_1 = \text{"}\textit{hodnota původního dodávaného napětí}\text{"}$$

s rozdělením $N(\mu_1, \sigma_1^2) = N(230, 25)$ a veličinu

$$U_2 = \text{"}\textit{hodnota nově dodávaného napětí}\text{"}$$

s rozdělením $N(\mu_2, \sigma_2^2) = N(\mu_2, 10)$. Horní mez U_0 je v obou případech nejnižší hranice vzhledem k podmínce

$$P(U_i > U_0) \le \alpha = 10^{-4} \quad \text{pro} \ \ i = 1, 2 \ .$$

Takže

$$\alpha \ge P(U_i > U_0) = P\left(\mathrm{norm}(U_i) > \frac{U_0 - \mu_i}{\sigma_i}\right) = 1 - \Phi\left(\frac{U_0-\mu_i}{\sigma_i}\right)$$

a U_0 tudíž splňuje rovnosti

$$U_0 = \mu_1 + \sigma_1 \cdot \Phi^{-1}(1-\alpha)$$
$$U_0 = \mu_2 + \sigma_2 \cdot \Phi^{-1}(1-\alpha) \ .$$

Jejich odečtením dostaneme

$$\mu_2 - \mu_1 = (\sigma_1 - \sigma_2)\cdot\Phi^{-1}(1-\alpha) = (5-\sqrt{10})\cdot\Phi^{-1}(0.9999) \doteq$$

$$\doteq (5-\sqrt{10})\cdot 3.719 \doteq 6.8345 \ .$$

Středního hodnota nově dodávaného napětí tedy může být až o 6.8345 V vyšší oproti původnímu.

Otázkou je, co na to řeknou přístroje, které jsou dimenzovány na 230 V ...

7.12 (normální rozdělení)

Předpokládejme, že chyba, se kterou tachometr ukazuje rychlost, má normální rozdělení se směrodatnou odchylkou $\sigma = 1.5\ \frac{\mathrm{km}}{\mathrm{h}}$. Jaká musí být střední hodnota chyby, aby pravděpodobnost, že tachometr ukazuje menší než skutečnou rychlost, byla nejvýše $\alpha = 0.001$?

Řešení:
Máme veličinu

$$X = \text{"}\textit{chyba, se kterou tachometr ukazuje rychlost}\text{"}$$

s rozdělením $N(\mu, \sigma^2)$. Zajímá nás pro jaké hodnoty parametru μ platí, že

$$P(X < 0) \leq \alpha = 0.001\ .$$

Stačí opět přejít k normalizované veličině $\mathrm{norm}(X) = \frac{X-\mu}{\sigma}$ s rozdělením $N(0,1)$:

$$\alpha \geq P\big(X < 0\big) = P\left(\mathrm{norm}(X) < -\frac{\mu}{\sigma}\right) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

neboli

$$\frac{\mu}{\sigma} \geq \Phi^{-1}(1-\alpha)$$

a tedy

$$\mu \geq \sigma \cdot \Phi^{-1}(1-\alpha) = 1.5 \cdot \Phi^{-1}(0.999) \doteq 1.5 \cdot 3.09 = 4.635\ .$$

Tachometr tak musí mít střední hodnotu chyby alespoň $4.635\ \frac{\mathrm{km}}{\mathrm{h}}$.