7. cvičení z PST

14. listopadu 2018

7.1 Náhodná veličina X má hustotu pravděpodobnosti

$$f_X(t) = \begin{cases} c \cdot 2^{-t} & , \ t \geq 0 \\ 0 & , \ t < 0 \end{cases}$$

kde $c \in \mathbb{R}$ je konstanta.

(a) Určete hodnotu c, typ rozdělení, střední hodnotu $E(X)$ a rozptyl $D(X)$.

(b) Najděte $\varepsilon > 0$ tak, aby $P\big(|X - E(X)| < \varepsilon\big) = 0.9$ (neboli hledáme 90% oboustranný interval spolehlivosti se středem v $E(X)$).

Řešení:

(a) Abychom měli hustotu pravděpodobnosti, musí platit, že $f_X(t) \geq 0$ pro všechna $t \in \mathbb{R}$ a že

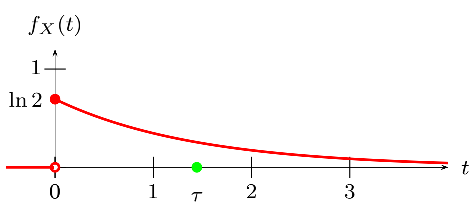
$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = \int_0^{+\infty} c\ 2^{-t}\ dt = c \int_0^{+\infty} e^{-t \ln 2}\ dt = c \left[\frac{e^{-t \ln 2}}{-\ln 2}\right]_0^{+\infty} = \frac{c}{\ln 2}$$

tedy $c = \ln 2 \doteq 0.693$.

Hustota je tedy tvaru

$$f_X(t) = \begin{cases} (\ln 2) \cdot 2^{-t} \ \left(= \frac{1}{\tau} e^{-\frac{t}{\tau}}\right) & , \ t \geq 0 \\ 0 & , \ t < 0 \end{cases}$$

kde $\tau = \frac{1}{\ln 2} \doteq 1.443$ a její graf je:

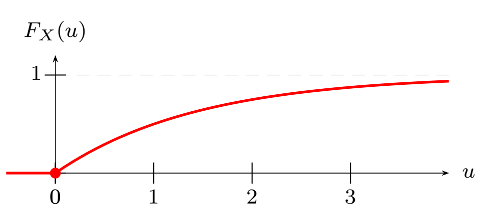


(Zelený bod je vodorovná souřadnice těžiště plochy pod grafem f_X neboli hodnota $E(X)$.)

Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt = \begin{cases} 0 & , \ u \leq 0 \\ \int\limits_0^u \frac{1}{\tau} e^{-\frac{t}{\tau}}\ dt = \left[-e^{-\frac{t}{\tau}}\right]_{t=0}^{t=u} = 1 - e^{-\frac{u}{\tau}} \ \left(= 1 - 2^{-u}\right) & , \ u \geq 0 \end{cases}$$

což znamená, že X má exponenciální rozdělení s parametrem $\tau = \frac{1}{\ln 2}$.



Střední hodnotu i rozptyl exponenciálního rozdělení sice už známe, ale teď si je také explicitně ověříme. Pomocí hustoty spočítáme střední hodnotu

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\ dx = \int_0^{\infty} x \cdot \tfrac{1}{\tau} e^{-\frac{x}{\tau}}\ dx = \underbrace{\Big[- x \cdot e^{-\frac{x}{\tau}} \Big]_{x=0}^{x=\infty}}_{=0} + \int_0^{\infty} e^{-\frac{x}{\tau}}\ dx = \Big[- \tau \cdot e^{-\frac{x}{\tau}} \Big]_{x=0}^{x=\infty} = \tau$$

a rozptyl

$$D(X) = E(X^2) - \big(E(X)\big)^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\ dx = \int_0^{\infty} x^2 \cdot \tfrac{1}{\tau} e^{-\frac{x}{\tau}}\ dx = \underbrace{\Big[- x^2 \cdot e^{-\frac{x}{\tau}} \Big]_{x=0}^{x=\infty}}_{=0} + 2 \cdot \underbrace{\int_0^{\infty} x \cdot e^{-\frac{x}{\tau}}\ dx}_{=\tau \cdot E(X)} = 2\tau^2.$$

Tedy

$$D(X) = E(X^2) - \big(E(X)\big)^2 = 2\tau^2 - \tau^2 = \tau^2 \doteq 0.4806.$$

Poznámka: Obecně pro borelovskou funkci $h : \mathbb{R} \to \mathbb{R}$ (tj. takovou, že vzorem intervalu je borelovská množina - neboli množina, která se dá poskládat z intervalů pomocí spočetného sjednocování a doplňků), např. pro h po částech spojitou funkcí, platí, že

$$E(h(X)) = \int_{-\infty}^{\infty} h(u) \cdot f_X(u)\ du\ .$$

Je ještě jiný způsob, jak spočítat totéž (ovšem pouze za předpokladu, že veličina $Y = h(X)$ má hustotu f_Y !!! - a to velmi často nastat nemusí):

$$E(h(X)) = \int_{-\infty}^{\infty} u \cdot f_{h(X)}(u)\ du\ .$$

Zatímco ale druhý ze způsobů vyžaduje znalost (a existenci) hustoty $f_{h(X)}$, tak první způsob můžeme použít pro (absolutně) spojitou veličinu X vždy.

(b) Hledáme $\varepsilon > 0$ tak, že

$$0.9 = P\big(|X - E(X)| < \varepsilon\big) = P(\tau - \varepsilon < X < \tau + \varepsilon) = F_X(\tau + \varepsilon) - F_X(\tau - \varepsilon)\ .$$

Dále musíme rozlišit případy:

- $\varepsilon \geq \tau$:

$$0.9 = F_X(\tau + \varepsilon) - F_X(\tau - \varepsilon) = 1 - e^{-\frac{\tau+\varepsilon}{\tau}} - 0 \quad \Leftrightarrow \quad e^{-\frac{\varepsilon}{\tau}} = 0.1e \quad \Leftrightarrow \quad \varepsilon = \tau \underbrace{(\ln(10) - 1)}_{\doteq 2.3026} \quad (\geq \tau)$$

  tedy podmínka $\varepsilon \geq \tau$ je splněna a my dostáváme $\varepsilon \doteq 2.3026 \cdot \tau \doteq 1.8792$. Další případ (tj. $\varepsilon < \tau$) už tudíž nemusíme řešit, protože díky tvaru hustoty již nemůže mít řešení (při zmenšování ε by už nutně musela hodnota výrazu $F_X(\tau + \varepsilon) - F_X(\tau - \varepsilon) = \int\limits_{\tau-\varepsilon}^{\tau+\varepsilon} f_X(t)\ dt$ klesat a nemohla by tedy být zachována hodnota 0.9). Přesto si ho z cvičných důvodů vyřešíme.

- $\varepsilon < \tau$:

$$0.9 = F_X(\tau + \varepsilon) - F_X(\tau - \varepsilon) = 1 - e^{-\frac{\tau+\varepsilon}{\tau}} - \left(1 - e^{-\frac{\tau-\varepsilon}{\tau}}\right) = \underbrace{\left(e^{\frac{\varepsilon}{\tau}} - e^{-\frac{\varepsilon}{\tau}}\right)}_{=2\sinh(\frac{\varepsilon}{\tau})} / e$$

  Pak ale dostáváme (díky monotonii hyperbolického sinu a toho, že $\frac{\varepsilon}{\tau} < 1$):

$$0.45e = \sinh\left(\frac{\varepsilon}{\tau}\right) < \sinh(1) = \frac{e^1 - e^{-1}}{2} \quad \Leftrightarrow \quad e^{-1} < 0.1e \quad \Leftrightarrow \quad 10 < e^2 \doteq 7.389$$

  kde poslední nerovnost neplatí. Tudíž tento (druhý) případ nemá řešení.

7.2 Náhodná veličina X má hustotu pravděpodobnosti

$$f_X(t) = \begin{cases} (\ln 2) \cdot 2^{-t} & ,\ t \geq 0 \\ 0 & ,\ t < 0\ . \end{cases}$$

(a) Pro veličinu $Y = \sqrt{|X|}$ určete střední hodnotu $E(Y)$, rozptyl $D(Y)$ a hustotu pravděpodobnosti f_Y.

(b) Spočítejte horní kvartil $q_X(0.75)$.

Řešení:

(a) Pomocí hustoty veličiny X spočítáme střední hodnotu veličiny $Y = \sqrt{|X|}$. Hustotu f_X si pro jednodušší počítání zapíšeme jako

$$f_X(t) = \begin{cases} \frac{1}{\tau} e^{-\frac{t}{\tau}} & , \ t \geq 0 \\ 0 & , \ t < 0 \end{cases}$$

kde $\tau = \frac{1}{\ln 2}$.

$$E(Y) = E(\sqrt{|X|}) = \int_{-\infty}^{\infty} \sqrt{|x|} \cdot f_X(x)\, dx = \int_0^{\infty} \sqrt{x} \cdot \frac{1}{\tau} e^{-\frac{x}{\tau}}\, dx = \left\{ \begin{array}{c} x = y^2 \\ dx = 2ydy \end{array} \right\} =$$

$$= \int_0^{\infty} 2y^2 \cdot \frac{1}{\tau} e^{-\frac{y^2}{\tau}}\, dy = \underbrace{\Big[-y \cdot e^{-\frac{y^2}{\tau}} \Big]_{y=0}^{y=\infty}}_{=0} + \int_0^{\infty} e^{-\frac{y^2}{\tau}}\, dy = \left\{ \begin{array}{c} y = \sqrt{\tau} \cdot u \\ dy = \sqrt{\tau} du \end{array} \right\} =$$

$$= \sqrt{\tau} \underbrace{\int_0^{\infty} e^{-u^2}\, du}_{=\sqrt{\pi}/2} = \frac{\sqrt{\pi}}{2} \cdot \sqrt{\tau} = \frac{1}{2}\sqrt{\frac{\pi}{\ln 2}} \doteq 1.0645$$

Poslední integrál je tabulková hodnota (která se jinak počítá pomocí vhodného dvojného integrálu). Pro rozptyl využijeme toho, že X má exponenciální rozdělení a tudíž $E(|X|) = E(X) = \tau$:

$$D(Y) = E(Y^2) - \big(E(Y)\big)^2 = E\left(\sqrt{|X|}^2\right) - \left(\frac{\sqrt{\pi}}{2} \cdot \sqrt{\tau}\right)^2 = \tau - \frac{\pi}{4} \cdot \tau = \left(1 - \frac{\pi}{4}\right)\tau =$$

$$= \left(1 - \frac{\pi}{4}\right) \frac{1}{\ln 2} \doteq 0.3096$$

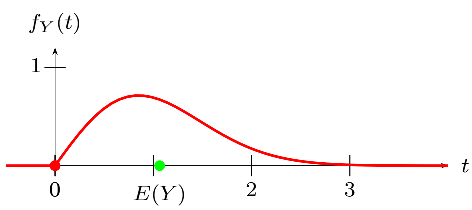
Pro zjištění hustoty f_Y si nejdříve určíme distribuční funkci F_Y:

- $t < 0$: $F_Y(t) = P\left(\sqrt{|X|} \leq t\right) = 0$
- $t \geq 0$: $F_Y(t) = P\left(\sqrt{|X|} \leq t\right) = P(|X| \leq t^2) = P(-t^2 \leq X \leq t^2) = F_X(t^2) - \underbrace{F_X(-t^2)}_{=0} = F_X(t^2)$

Zderivováním (tam, kde lze) pak dostaneme hustotu:

- $t < 0$: $f_Y(t) = F'_Y(t) = 0$
- $t > 0$: $f_Y(t) = F'_Y(t) = (F_X(t^2))' = 2t \cdot F'_X(t^2) = 2t \cdot f_X(t^2) = 2\ln(2) \cdot t\ 2^{-t^2}$

s grafem



(Zelený bod je vodorovná souřadnice těžiště plochy pod grafem f_Y neboli hodnota $E(Y)$.)

Poznamenejme, že se znalostí hustoty f_Y jsme mohli určit $E(Y)$ a $E(Y^2)$ také "standardně", tj.

$$E(Y) = \int_{-\infty}^{\infty} t \cdot f_Y(t)\, dt = \int_0^{\infty} t \cdot 2\ln(2) t\ 2^{-t^2}\, dt = \cdots = \frac{1}{2}\sqrt{\frac{\pi}{\ln 2}}$$

a

$$E(Y^2) = \int_{-\infty}^{\infty} t^2 \cdot f_Y(t)\, dt = \int_0^{\infty} t^2 \cdot 2\ln(2) t\ 2^{-t^2}\, dt = \cdots = \frac{1}{\ln 2}\ .$$

Tento druhý způsob ale pochopitelně vyžaduje, aby hustota f_Y existovala, což se ale nemusí (pro obecnou $Y = h(X)$) vždy stát!!

(b) Hledáme takové $t \in \mathbb{R}$ (evidentně bude $t > 0$), že $q_X(0.75) = t$ neboli $0.75 = F_X\big(q_X(0.75)\big) = F_X(t)$ (protože F_X je na intervalu $(0, +\infty)$ ostře rostoucí a spojitá, takže q_X je tady její inverzí). Tedy

$$0.75 = F_X(t) = 1 - e^{-\frac{t}{\tau}} \quad \Leftrightarrow \quad t = -\tau \ln(0.25) = \frac{\ln 4}{\ln 2} = 2\ .$$

7.3 (náhodný vektor - diskrétní)

Diskrétní náhodný vektor (X, Y) má sdruženou pravděpodobnostní funkci $p_{X,Y}$ dánu tabulkou:

$p_{X,Y}(x,y)$: y / x	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, popište rozdělení náhodného vektoru (X', Y') se stejnými marginálními rozděleními, jehož složky jsou nezávislé veličiny.

(d) Vypočtěte korelaci $\varrho(X, Y)$.

Řešení:
Sdružená pravděpodobnostní funkce $p_{X,Y}$ pro diskrétní náhodný vektor (X, Y) je definována analogicky jako pro náhodnou funkci, tj.

$$p_{X,Y}(i,j) = P(X = i, Y = j)$$

pro $(i,j) \in \mathbb{R}^2$.

(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)

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

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

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) Veličiny X a Y jsou nezávislé právě když

$$F_{X,Y}(i,j) = F_X(i) \cdot F_Y(j) \quad \text{pro všechna} \quad (i,j) \in \mathbb{R}^2$$

což je v případě existence sdružené hustoty ekvivalentní podmínce

$$p_{X,Y}(i,j) = p_X(i) \cdot p_Y(j) \quad \text{pro všechna} \quad (i,j) \in \mathbb{R}^2 \ .$$

Protože v našem případě 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.

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	

(d) Korelace je dána jako

$$\varrho(X,Y) = \frac{cov(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = \frac{E\Big((X-EX)\cdot(Y-EY)\Big)}{\sqrt{E\Big((X-EX)^2\Big)} \cdot \sqrt{E\Big((Y-EY)^2\Big)}} = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(Y^2) - E(Y)^2}} \ .$$

Spočítáme jednotlivé střední hodnoty (můžeme si pomoci i součinem matic):

$$E(X) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{5}{18} + 2 \cdot \frac{7}{18} = \frac{19}{18}$$
$$E(X^2) = 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{5}{18} + 2^2 \cdot \frac{7}{18} = \frac{11}{6}$$
$$E(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$$
$$E(Y^2) = 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} = \frac{5}{3}$$
$$E(XY) = \sum_{(i,j)\in\mathbb{R}^2} ij \cdot p_{X,Y}(i,j) = (0,1,2) \cdot \begin{pmatrix} 1/18 & 1/9 & 1/6 \\ 1/9 & 1/18 & 1/9 \\ 1/6 & 1/6 & 1/18 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} =$$
$$= (1,2) \cdot \begin{pmatrix} 1/18 & 1/9 \\ 1/6 & 1/18 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{15}{18}$$

Korelace tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(Y^2) - E(Y)^2}} = \frac{\frac{15}{18} - \frac{19}{18} \cdot 1}{\sqrt{\frac{11}{6} - (\frac{19}{18})^2} \cdot \sqrt{\frac{5}{3} - 1^2}} =$$
$$= -4\sqrt{\frac{3}{466}} \doteq -0.32094 \ .$$

Pro připomenutí:

Na (reálném) vektorovém prostoru všech náhodných veličin s konečnou střední hodnotou a konečným rozptylem (které ztotožníme, pokud se rovnají s pravděpodobností 1) zavádíme skalární součin

$$X \bullet Y := E(XY)$$

a *normu* $\|X\|$ (neboli "délku"vektoru X) přirozeně zadanou jako $\|X\| := \sqrt{X \bullet X} = \sqrt{E(X^2)}$.

Korelace je pak dána pomocí vztahu

$$\varrho(X,Y) := \frac{(X-EX) \bullet (Y-EY)}{\|X-EX\| \cdot \|Y-EY\|}$$

tedy jako kosinus úhlu α mezi složkami $X-EX$ a $Y-EY$ veličin X a Y (kde složky $X-EX$ a $Y-EY$ jsou v prostoru všech veličin s nulovou střední hodnotou).

Úhel $\alpha \in \langle 0, \pi \rangle$ mezi našimi náhodnými veličinami $X-EX$ a $Y-EY$ je tedy

$$\arccos\big(\varrho(X,Y)\big) \doteq \arccos(-0.32094) \doteq 108,72^\circ \ .$$

Praktické použití korelace:
Pokud máme dvě veličiny X a Y takové, že

- výchylka veličiny X od jejího průměru je nezáporná právě když výchylka Y zase od jejího průměru je také nezáporná,

pak dostaneme nezápornou korelaci.

Neboli platí: Jestliže

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

pak je $\varrho(X, Y) \geq 0$.

Obdobně platí: Jestliže

$$X - E(X) \geq 0 \;\Leftrightarrow\; Y - E(Y) \leq 0$$

pak je $\varrho(X, Y) \leq 0$.

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

7.4 (náhodný vektor - diskrétní)

Diskrétní náhodný vektor (X, Y) má sdruženou pravděpodobnostní funkci $p_{X,Y}$ dánu tabulkou:

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

Určete:

(a) pravděpodobnost $P(X \cdot Y \geq 2.5)$.

(b) marginální pravděpodobnostní funkce p_X a p_Y náhodných veličin X a Y.

(c) zda jsou náhodné veličiny X a Y závislé či nezávislé. Pokud nejsou, popište rozdělení náhodného vektoru (X', Y') se stejnými marginálními rozděleními, jehož složky jsou nezávislé veličiny.

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

Řešení:
Postup je analogický jako v **7.5**.

(a) Máme

$$X \cdot Y \geq 2.5 \quad \Leftrightarrow \quad (X, Y) \in \{(3, 1), (3, 2), (4, 1), (4, 2)\}$$

a tedy

$$P\left(X \cdot Y \geq 2.5\right) = P\Big((X, Y) \in \{(3, 1), (3, 2), (4, 1), (4, 2)\}\Big) = \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + 0 = \frac{5}{8}\,.$$

(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
1	1/8	0	1/8	1/4
3	0	1/4	1/4	1/2
4	1/8	1/8	0	1/4
p_Y	1/4	3/8	3/8	

(c) V tabulce se vyskytla nulová pravděpodobnost, konkrétně

$$p_{X,Y}(4,2) = 0 \neq \frac{1}{4} \cdot \frac{3}{8} = p_X(4) \cdot p_Y(2)$$

takže X a Y jsou závislé.

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	$p_{X'}$
1	1/16	3/32	3/32	1/4
3	1/8	3/16	3/16	1/2
4	1/16	3/32	3/32	1/4
$p_{Y'}$	1/4	3/8	3/8	

(d) Spočítáme korelaci X a Y:

$$E(X) = 1 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{11}{4}$$
$$E(X^2) = 1^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{2} + 4^2 \cdot \frac{1}{4} = \frac{35}{4}$$
$$E(Y) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} = \frac{9}{8}$$
$$E(Y^2) = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} = \frac{15}{8}$$
$$E(XY) = \sum_{(i,j)\in\mathbb{R}^2} ij \cdot p_{X,Y}(i,j) = 1 \cdot 2 \cdot \frac{1}{8} + 3 \cdot 1 \cdot \frac{1}{4} + 3 \cdot 2 \cdot \frac{1}{4} + 4 \cdot 1 \cdot \frac{1}{8} = 3$$

Korelace tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{E(X^2) - E(X)^2}\cdot\sqrt{E(Y^2) - E(Y)^2}} = \frac{3 - \frac{11}{4}\cdot\frac{9}{8}}{\sqrt{\frac{35}{4} - (\frac{11}{4})^2}\cdot\sqrt{\frac{15}{8} - (\frac{9}{8})^2}} =$$
$$= \frac{15}{7\sqrt{19}} \doteq 0.4916 \doteq \arccos(60,55^\circ)\ .$$

Úhel mezi náhodnými veličinami $X - E(X)$ a $Y - E(Y)$ je pak $60,55^\circ$.

7.5 (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:

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

- je lineární v každé složce zvlášť (tj. bilineární),
- $\operatorname{cov}(Z + c, W + d) = \operatorname{cov}(Z, W)$ pro všechna $c, d \in \mathbb{R}$,
- symetrická (tj. $\operatorname{cov}(Z, W) = \operatorname{cov}(W, Z)$) a
- $\operatorname{cov}(Z, Z) = D(Z) = (\sigma_Z)^2$.

Dále je $\varrho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} = \frac{\operatorname{cov}(X,Y)}{\sigma_X\cdot\sigma_Y}$

Střední hodnota:

$E(T) = E(2X+3) = 2E(X)+3 = 2 \cdot 10 + 3 = 23$

$E(U) = E(200-Y) = 200 - E(Y) = 200 - 150 = 50$

$E(V) = E(X+Y) = E(X) + E(Y) = 10 + 150 = 160$

Rozptyl:

$D(T) = D(2X+3) = D(2X) = 2^2 \cdot D(X) = 2^2 \cdot (\sigma_X)^2 = 4 \cdot 25 = 100$

$D(U) = D(200-Y) = (-1)^2 \cdot D(Y) = (\sigma_Y)^2 = 400$

$$D(V) = D(X+Y) = \mathrm{cov}(X+Y, X+Y) = \mathrm{cov}(X,X) + \mathrm{cov}(X,Y) + \mathrm{cov}(Y,X) + \mathrm{cov}(Y,Y) =$$
$$= D(X) + 2 \cdot \mathrm{cov}(X,Y) + D(Y) = (\sigma_X)^2 + 2 \cdot \sigma_X \sigma_Y \varrho(X,Y) + (\sigma_Y)^2 = 25 + 2 \cdot 5 \cdot 20 \cdot 0.5 + 400 = 525\ .$$