8. cvičení z PST

11. listopadu 2020

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

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

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

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

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

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

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

$$\varrho(X,Y) := \frac{(X - EX) \bullet (Y - EY)}{\|X - EX\| \cdot \|Y - EY\|} = \cdots = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{D(X) \cdot D(Y)}}\ .$$

(Veličiny $X - EX$ a $Y - EY$ mají nulovou střední hodnotu).
A kromě toho máme:

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

Připomenutí: Kovariance $\mathrm{cov}(\cdot,\cdot)$ má tyto vlastnosti (X, Y, Z jsou veličiny, $a, b, c, d \in \mathbb{R}$):

- je lineární v každé složce zvlášť (tj. je bilineární), tedy:
$$\mathrm{cov}(aX + bY, Z) = a \cdot \mathrm{cov}(X,Z) + b \cdot \mathrm{cov}(Y,Z)$$
$$\mathrm{cov}(Z, aX + bY) = a \cdot \mathrm{cov}(Z,X) + b \cdot \mathrm{cov}(Z,Y)$$
- symetrická, tj. $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$
- pozitivně semi-definitní, tj. $\mathrm{cov}(X,X) \geq 0$, kde navíc platí, že:
$\mathrm{cov}(X,X) = 0 \ \Rightarrow\ \exists \alpha \in \mathbb{R}$, že $P(X = \alpha) = 1$ (neboli: X odpovídá konstantní veličině)
- $\mathrm{cov}(X + c, Y + d) = \mathrm{cov}(X,Y)$,
- $\mathrm{cov}(X,X) = D(X) = (\sigma_X)^2$.

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

$$\varrho(aX + c, bY + d) = \mathrm{sgn}(ab) \cdot \varrho(X,Y)$$

pro $a \neq 0$ a $b \neq 0$ a c, d libovolné konstanty (kde sgn je znaménková funkce).

Pro rozptyl $D(\cdot)$ díky tomu máme:

- $D(aX + c) = D(aX) = a^2 \cdot D(X)$,
- $D(X + Y) = D(X) + D(Y) + 2 \cdot \mathrm{cov}(X,Y)$ (viz níže).

Poznámka: Uvědomme si, že existuje několik stupňů "nezávislosti" veličin:

$$X \text{ a } Y \text{ jsou nezávislé} \xRightarrow{\text{pokud cov ex.}} \begin{array}{c}\mathrm{cov}(X,Y) = 0 \\ \text{(tj. } X - E(X) \text{ a } Y - E(Y) \text{ jsou kolmé)}\end{array} \xRightarrow{\text{pokud } X, Y \text{ nejsou konst.}} \begin{array}{c} X \text{ a } Y \text{ jsou lineár.} \\ \text{nezáv.}\end{array}$$

Konstantní veličina X spolu s jakoukoliv jinou veličinou Y vždy tvoří vzájemně nezávislé veličiny X a Y (tento případ je ale celkem nezajímavý).

8.1 (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:
(a) Kovarianční matice pro (X, Y) představuje tyto parametry:

$$\begin{pmatrix} \operatorname{cov}(X,X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} D(X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & D(Y) \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}$$

Máme tedy $\operatorname{cov}(X, Y) = 2 \neq 0$ a proto veličiny X a Y musí být závislé. (Jestliže by kovariance byla nulová, nemohli bychom udělat o nezávislosti žádný závěr.)

(b) Korelační matice pro (X, Y):

$$\begin{pmatrix} \varrho(X,X) & \varrho(X,Y) \\ \varrho(Y,X) & \varrho(Y,Y) \end{pmatrix} = \begin{pmatrix} 1 & \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} \\ \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

protože máme $\varrho(X, Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} = \frac{2}{\sqrt{3\cdot 4}} = \frac{1}{\sqrt{3}}$.

(c) Pro náhodný vektor $(X, -Y)$ potřebujeme ještě dopočítat:

$D(-Y) = (-1)^2 \cdot D(Y) = 4$

$\operatorname{cov}(X, -Y) = (-1) \cdot \operatorname{cov}(X, Y) = -2$

$\varrho(X, -Y) = \operatorname{sgn}(-1) \cdot \varrho(X, Y) = -\frac{1}{\sqrt{3}}$

Kovarianční a korelační matice pro $(X, -Y)$ tedy jsou:

$$\begin{pmatrix} D(X) & \operatorname{cov}(X,-Y) \\ \operatorname{cov}(X,-Y) & D(-Y) \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & \varrho(X,-Y) \\ \varrho(X,-Y) & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

Podobně pro náhodný vektor $(X, 2Y - 1)$ potřebujeme ještě dopočítat:

$D(2Y - 1) = D(2Y) = 2^2 \cdot D(Y) = 16$

$\operatorname{cov}(X, 2Y - 1) = \operatorname{cov}(X, 2Y) = 2 \cdot \operatorname{cov}(X, Y) = 4$

$\varrho(X, 2Y - 1) = \operatorname{sgn}(2) \cdot \varrho(X, Y) = \frac{1}{\sqrt{3}}$

Kovarianční a korelační matice pro $(X, 2Y - 1)$ tedy jsou:

$$\begin{pmatrix} D(X) & \operatorname{cov}(X,2Y-1) \\ \operatorname{cov}(X,2Y-1) & D(2Y-1) \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 4 & 16 \end{pmatrix} \qquad \begin{pmatrix} 1 & \varrho(X,2Y-1) \\ \varrho(X,2Y-1) & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

8.2 (kovariance, korelace)

Náhodný vektor (X, Y) má následující parametry:

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

(a) Stanovte střední hodnotu a rozptyl náhodných veličin

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

(b) Určete kovarianci $\mathrm{cov}(T, U)$ a rozhodněte, zda jsou náhodné veličiny T a U závislé či nezávislé.

Řešení:

(a) Střední hodnota je lineární zobrazeni ("veličina"$\mapsto$ "její 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 na druhou stranu NENÍ lineární zobrazení. (Je to tzv. kvadratická forma, tedy vzniká z bilineárního zobrazení - viz výše.):

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

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

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

(b) $\mathrm{cov}(T, U) = \mathrm{cov}(2X + 3, 200 - Y) = \mathrm{cov}(2X, -Y) =$

$$= 2 \cdot (-1) \cdot \mathrm{cov}(X, Y) = (-2) \cdot \sigma_X \cdot \sigma_Y \cdot \varrho(X, Y) = (-2) \cdot 5 \cdot 20 \cdot 0.5 = -100$$

Protože kovariance vyšla nenulová, můžeme hned říct, že U z T musí být *závislé*. V případě, že by kovariance byla nulová, bychom o nezávislosti nic usoudit nemohli!

8.3 (kovariance, kovarianční a korelační matice)

Pro náhodné veličiny X a Y platí, že

$$E(X) = 2, \quad E(Y) = -1, \quad D(X) = 3, \quad D(Y) = 4, \quad \mathrm{cov}(X, Y) = -2.$$

Pro náhodné veličiny

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

určete

(a) koeficient kovariance $\mathrm{cov}(U, V)$ a střední hodnotu $E(U)$.

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

(c) kovarianční a korelační matice náhodných vektoru (X,Y) a $(X,-2Y)$.

Řešení:

(a) Díky bilinearitě kovariance můžeme jednotlivé složky "roznásobit":

$$\mathrm{cov}(U,V) = \mathrm{cov}(3X+4Y-1,\ -2X+2Y+3) = \mathrm{cov}(3X+4Y,\ -2X+2Y) =$$
$$= 3\cdot(-2)\cdot \mathrm{cov}(X,X) + 3\cdot 2\cdot \mathrm{cov}(X,Y) + 4\cdot(-2)\cdot \mathrm{cov}(Y,X) + 4\cdot 2\cdot \mathrm{cov}(Y,Y) =$$
$$= (-6)\cdot D(X) + (-2)\cdot \mathrm{cov}(X,Y) + 8\cdot D(Y) = (-6)\cdot 3 + (-2)\cdot(-2) + 8\cdot 4 = 18\ .$$

Jak je vidět, znalosti středních hodnot jsme zatím vůbec nepotřebovali!

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

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

V našem případě tedy

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

Střední hodnota:

$$E(U) = E(3X+4Y-1) = 3E(X) + 4E(Y) - 1 = 3\cdot 2 + 4\cdot(-1) - 1 = 1$$

(b) Využijeme vlastnosti kovariance:

$$D(X+Y) = \mathrm{cov}(X+Y, X+Y) = \mathrm{cov}(X,X) + \mathrm{cov}(X,Y) + \mathrm{cov}(Y,X) + \mathrm{cov}(Y,Y) =$$
$$= D(X) + 2\cdot \mathrm{cov}(X,Y) + D(Y) =$$
$$= 3 + 2\cdot(-2) + 4 = 3\ .$$

(c) Kovarianční matice pro (X,Y):

$$\begin{pmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} D(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & D(Y) \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix}$$

Korelační matice pro (X,Y):

$$\begin{pmatrix} \varrho(X,X) & \varrho(X,Y) \\ \varrho(Y,X) & \varrho(Y,Y) \end{pmatrix} = \begin{pmatrix} 1 & \frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} \\ \frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

protože máme $\varrho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)\cdot D(Y)}} = \frac{-2}{\sqrt{3\cdot 4}} = -\frac{1}{\sqrt{3}}$.

Podobně pro vektor $(X,-2Y)$ máme korelační matici:

$$\begin{pmatrix} D(X) & \mathrm{cov}(X,-2Y) \\ \mathrm{cov}(X,-2Y) & D(-2Y) \end{pmatrix} = \begin{pmatrix} D(X) & -2\cdot\mathrm{cov}(X,Y) \\ -2\cdot\mathrm{cov}(X,Y) & (-2)^2\cdot D(Y) \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 4 & 16 \end{pmatrix}$$

a korelační matici:

$$\begin{pmatrix} 1 & \frac{4}{\sqrt{3\cdot 16}} \\ \frac{4}{\sqrt{3\cdot 16}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

což není překvapení, protože díky vlastnostem korelace (viz výše) máme, že:

$$\varrho(X,-2Y) = \mathrm{sgn}(-2)\cdot\varrho(X,Y) = -\varrho(X,Y)\ .$$

Připomenutí: Jestliže máme dvě náhodné veličiny $X, Y : \Omega \to \mathbb{R}$, pak zobrazení

$$(X, Y) : \Omega \to \mathbb{R}^2$$

nazýváme náhodný (dvousložkový) vektor.

Tedy náhodnému výsledku ω (tj. elementárnímu jevu) přiřadíme dvojici hodnot $(x, y) \in \mathbb{R}^2$. Např. vybranému člověku z množiny lidí Ω přiřadíme jeho tělesnou výšku a hmotnost.

Náhodný vektor (X, Y) opět umí přenést a vytvořit rozdělení pravděpodobnosti na $\mathbb{R}^2$ - a to tak, že každá "rozumná" množina $A \subseteq \mathbb{R}^2$ (např. otevřená množina nebo interval atd.) bude mít prostě pravděpodobnost

$$P_{(X,Y)}(A) := P\Big((X, Y)^{-1}(A)\Big) \ .$$

Rozdělení této pravděpodobnosti $P_{(X,Y)}$ na $\mathbb{R}^2$ můžeme opět úplně popsat pokud známe pravděpodobnosti jen některých speciálních intervalů a ty nám definují tzv. *sdruženou distribuční funkci* $F_{(X,Y)} : \mathbb{R}^2 \to \langle 0, 1\rangle$ jako

$$F_{(X,Y)}(a, b) := P(X \leq a, Y \leq b)$$

Pojem nezávislosti pro jevy umožňuje přirozeně definovat nezávislost veličin:

Definice: Veličiny X a Y jsou nezávislé $\Leftrightarrow$ $P(X \in I, Y \in J) = P(X \in I) \cdot P(Y \in J)$ pro libovolné intervaly $I, J \subseteq \mathbb{R}$.

Věta: X a Y jsou nezávislé $\Leftrightarrow$ $F_{(X,Y)}(a, b) = F_X(a) \cdot F_Y(b)$ pro všechna $a, b \in \mathbb{R}$.

Definice: Náhodný vektor (X, Y) má *diskrétní rozdělení* $\Leftrightarrow$ existuje $A \subseteq \mathbb{R}^2$, která je konečná nebo spočetná a taková, že $P((X, Y) \in A) = 1$. (tedy vektor má nejvýše spočetně mnoho "zajímavých" hodnot)

V tomto případě pak rozdělení vektoru (X, Y) popisuje *sdružená pravděpodobnostní funkce* $p_{(X,Y)} : \mathbb{R}^2 \to \langle 0, +\infty)$ definovaná jako

$$p_{(X,Y)}(a, b) := P(X = a, Y = b)$$

a platí

$$F_{(X,Y)}(a, b) = P(X \leq a, Y \leq b) = \sum_{\substack{u \leq a \\ t \leq b}} p_{(X,Y)}(u, t) \ .$$

Věta: Nechť náhodný vektor (X, Y) má diskrétní rozdělení. Pak jsou veličiny X a Y nezávislé $\Leftrightarrow$

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

8.4 (diskrétní náhodný vektor)

Mějme urnu s 1 bílou a 2 černými koulemi. Dvakrát (nezávisle) za sebou vždy jednu kouli vytáhneme a zase vrátíme zpět.

Nechť veličina X_i představují počet vytažených bílých koulí při i-tém tahu $i = 1, 2$. Uvažujme veličiny $X = X_1$ a $Y = X_1 + X_2$.

(a) Jaké jsou sdružené pravděpodobnosti náhodných veličin X a Y a jejich marginální rozdělení?

(b) Určete pravděpodobnost $P(X + Y \geq 1)$.

(c) Určete rozdělení veličiny $Z = X + Y$.

(d) 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.

(e) Určete kovarianční matici a koeficient korelace $\varrho(X, Y)$.

Řešení:
Veličiny X_1 a X_2 jsou nezávislé s alternativním rozdělením $\mathrm{Alt}(\frac{1}{3})$.

Veličina $X = X_1 \sim \mathrm{Alt}(\frac{1}{3})$ tak nabývá hodnot 0 a 1. Veličina

$$Y = X_1 + X_2 = \text{"}\textit{počet bílých koulí během 2 tahů}\text{"}$$

má binomické rozdělení $\mathrm{Bi}(2, \frac{1}{3})$ a nabývá hodnot $0, 1, 2$. Známe tak marginální rozdělení, tj. rozdělení veličin X a Y, popsaná v tomto případě např. pravděpodobnostními funkcemi p_X a p_Y.

Snadno dále (díky nezávislosti X_1 a X_2) získáme sdružené pravděpodobnosti v následující tabulce pro $p_{(X,Y)}$, např.

$$p_{(X,Y)}(0,1) = P(\underbrace{X = 0, Y = 1}_{\Leftrightarrow\ X_1=0, X_2=1}) = P(X_1 = 0, X_2 = 1) = P(X_1 = 0) \cdot P(X_2 = 1) = \tfrac{2}{3} \cdot \tfrac{1}{3}$$

$$p_{(X,Y)}(0,2) = P(\underbrace{X = 0, Y = 2}_{\Leftrightarrow\ X_1=0, X_2=2}) = P(\emptyset) = 0$$

$$p_{(X,Y)}(1,0) = P(\underbrace{X = 1, Y = 0}_{\Leftrightarrow\ X_1=0, X_2=-1}) = P(\emptyset) = 0$$

$p_{X,Y}(x,y):$ y / x	0	1	2	p_X
0	4/9	2/9	0	2/3
1	0	2/9	1/9	1/3
p_Y	4/9	4/9	1/9	

(b) & (c) Tyto dvě části můžeme udělat společně. Hodnoty z veličiny $Z = X + Y$ si pro přehlednost zapíšeme také do tabulky

$z(x,y) = x + y$ y / x	0	1	2
0	0	1	2
1	1	2	3

a pravděpodobnostní funkce p_Z vznikne nasčítáním pravděpodobností pro jednotlivé případy

$$p_Z(z) = P({\color{red}X + Y} = z) = \sum_{\substack{{\color{red}x+y=z} \\ x,y\in\mathbb{R}}} \underbrace{P(X = x,\ Y = y)}_{p_{X,Y}(x,y)} = \begin{cases} \frac{4}{9} & , z = 0 \\ 0 + \frac{2}{9} = \frac{2}{9} & , z = 1 \\ \frac{2}{9} + 0 = \frac{2}{9} & , z = 2 \\ \frac{1}{9} & , z = 3 \\ 0 & , \text{jinak.} \end{cases}$$

Takže $P(X + Y \geq 1) = P(Z \geq 1) = \frac{2}{9} + \frac{2}{9} + \frac{1}{9} = \frac{5}{9}$.

(d) Stačí si všimnout např. toho, že v tabulce se vyskytla nulová pravděpodobnost, konkrétně

$$p_{X,Y}(0,2) = 0 \neq \frac{2}{3} \cdot \frac{1}{9} = p_X(0) \cdot p_Y(2)$$

takže X a Y jsou závislé (podle věty výše).

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	$\frac{4}{9} \cdot \frac{2}{3} = \frac{8}{27}$	$\frac{4}{9} \cdot \frac{2}{3} = \frac{8}{27}$	$\frac{1}{9} \cdot \frac{2}{3} = \frac{2}{27}$	2/3
1	$\frac{4}{9} \cdot \frac{1}{3} = \frac{4}{27}$	$\frac{4}{9} \cdot \frac{1}{3} = \frac{4}{27}$	$\frac{1}{9} \cdot \frac{1}{3} = \frac{1}{27}$	1/3
$p_{Y'}$	4/9	4/9	1/9	

(e) Spočítáme kovarianci a korelaci X a Y:

$$X \sim \mathrm{Alt}(\tfrac{1}{3}) \;\Rightarrow\; E(X) = \tfrac{1}{3} \;\;\&\;\; D(X) = \tfrac{1}{3} \cdot \tfrac{2}{3} = \tfrac{2}{9}$$

$$Y \sim \mathrm{Bi}(2, \tfrac{1}{3}) \;\Rightarrow\; E(Y) = 2 \cdot \tfrac{1}{3} = \tfrac{2}{3} \;\;\&\;\; D(Y) = 2 \cdot \tfrac{1}{3} \cdot \tfrac{2}{3} = \tfrac{4}{9}$$

$$E(XY) = \sum_{(i,j) \in \mathbb{R}^2} ij \cdot p_{X,Y}(i,j) = 1 \cdot 1 \cdot \tfrac{2}{9} + 1 \cdot 2 \cdot \tfrac{1}{9} = \tfrac{4}{9}$$

Takže

$$\mathrm{cov}(X,Y) = E(XY) - E(X) \cdot E(Y) = \tfrac{4}{9} - \tfrac{1}{3} \cdot \tfrac{2}{3} = \tfrac{2}{9}$$

a kovarianční matice pro (X,Y) je tedy

$$\begin{pmatrix} D(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & D(Y) \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} \end{pmatrix} .$$

Pro korelaci pak máme

$$\varrho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{D(X) \cdot D(Y)}} = \frac{\frac{2}{9}}{\sqrt{\frac{2}{9} \cdot \frac{4}{9}}} = \frac{\sqrt{2}}{2} .$$

(Úhel mezi náhodnými veličinami $X - E(X)$ a $Y - E(Y)$ je pak $\arccos(\frac{\sqrt{2}}{2}) = \frac{\pi}{4}$.)

8.5 (náhodný vektor - diskrétní)

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

$p_{X,Y}(x,y)$: x \ y	0	1	2
0	1/6	1/9	1/9
1	1/9	2/9	0
2	1/6	0	1/9

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

(b) Určete rozdělení veličiny $Z = X - Y$.

(c) Určete marginální rozdělení veličin X a Y.

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

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

Řešení:
Na začátku bychom si měli pro pořádek ještě ověřit, že součet všech pravděpodobností v tabulce je $= 1$ (pokud by byl např. < 1, pak nemáme úplnou informaci o rozdělení a nemůžeme dál pokračovat).

(a) Zřejmě

$$P\left(\frac{1}{2} \leq X \leq 3 \;\&\; Y \geq 1\right) = P\Big((X,Y) \in \{(1,1),(1,2),(2,1),(2,2)\}\Big) = \frac{2}{9} + 0 + 0 + \frac{1}{9} = \frac{1}{3}\,.$$

(b) U veličin X a Y předpokládáme obory hodnot určené tabulkou (ostatní hodnoty mohou být také případně nabyty, ale s nulovou pravděpodobností). Hodnoty z veličiny $Z = X - Y$ si pro přehlednost zapíšeme také do tabulky

$z(x,y) = x - y$ $\quad y$ x	0	1	2
0	0	-1	-2
1	1	0	-1
2	2	1	0

a pravděpodobnostní funkce p_Z vznikne nasčítáním pravděpodobností pro jednotlivé případy

$$p_Z(z) = P(X - Y = z) = \sum_{\substack{x-y=z \\ x,y\in\mathbb{R}}} \underbrace{P(X = x,\ Y = y)}_{p_{X,Y}(x,y)} = \begin{cases} \frac{1}{9} & , z = -2 \\ \frac{1}{9} + 0 = \frac{1}{9} & , z = -1 \\ \frac{1}{6} + \frac{2}{9} + \frac{1}{9} = \frac{1}{2} & , z = 0 \\ \frac{1}{9} + 0 = \frac{1}{9} & , z = 1 \\ \frac{1}{6} & , z = 2 \\ 0 & , \text{jinak.} \end{cases}$$

(c) Marginální pravděpodobnostní funkce p_X a p_Y (tj. pravděpodobnostní funkce jednotlivých složek vektoru) jsou

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

a jejich hodnoty tudíž získáme sečtením pravděpodobností v řádcích (p_X) a sloupcích (p_Y) naší tabulky:

$X \backslash Y$	0	1	2	p_X
0	1/6	1/9	1/9	7/18
1	1/9	2/9	0	1/3
2	1/6	0	1/9	5/18
p_Y	4/9	1/3	2/9	

(d) Veličiny X a Y jsou nezávislé právě když

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

což je v případě existence sdružené hustoty ekvivalentní podmínce

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

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	$\frac{4}{9} \cdot \frac{7}{18} = \frac{14}{81}$	$\frac{1}{3} \cdot \frac{7}{18} = \frac{7}{54}$	$\frac{2}{9} \cdot \frac{7}{18} = \frac{7}{81}$	7/18
1	$\frac{4}{9} \cdot \frac{1}{3} = \frac{4}{27}$	$\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$	$\frac{2}{9} \cdot \frac{1}{3} = \frac{2}{27}$	1/3
2	$\frac{4}{9} \cdot \frac{5}{18} = \frac{10}{81}$	$\frac{1}{3} \cdot \frac{5}{18} = \frac{5}{54}$	$\frac{2}{9} \cdot \frac{5}{18} = \frac{10}{81}$	5/18
$p_{Y'}$	4/9	1/3	2/9	

(e) Korelace je dána jako

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

$$= \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(Y^2) - E(Y)^2}} \ .$$

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

$$E(X) = 0 \cdot \frac{7}{18} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{5}{18} = \frac{8}{9}$$

$$E(X^2) = 0^2 \cdot \frac{7}{18} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{5}{18} = \frac{13}{9}$$

$$E(Y) = 0 \cdot \frac{4}{9} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{9} = \frac{7}{9}$$

$$E(Y^2) = 0^2 \cdot \frac{4}{9} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{2}{9} = \frac{11}{9}$$

$$E(XY) = \sum_{(i,j) \in \mathbb{R}^2} ij \cdot p_{X,Y}(i,j) = (0,1,2) \cdot \begin{pmatrix} 1/6 & 1/9 & 1/9 \\ 1/9 & 2/9 & 0 \\ 1/6 & 0 & 1/9 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} =$$

$$= (1,2) \cdot \begin{pmatrix} 2/9 & 0 \\ 0 & 1/9 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{2}{3}$$

Korelace tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(Y^2) - E(Y)^2}} = \frac{\frac{2}{3} - \frac{8}{9} \cdot \frac{7}{9}}{\sqrt{\frac{13}{9} - (\frac{8}{9})^2} \cdot \sqrt{\frac{11}{9} - (\frac{7}{9})^2}} =$$

$$= -\frac{1}{5}\sqrt{\frac{2}{53}} \doteq -0.0389\ .$$

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

$$\arccos\big(\varrho(X,Y)\big) \doteq \arccos(-0.0389) \doteq 92,23^\circ\ .$$

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

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

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.

8.6 (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
0	0	1/4	1/4
1	1/8	1/8	0

Určete:

(a) pravděpodobnost $P(X + Y > 1)$.

(b) rozdělení veličiny $Z = X^2 \cdot Y$.

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

(d) 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.

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

Řešení:
Postup je analogický jako v **8.5**.

(a) Máme

$$X \cdot Y > 1 \quad \Leftrightarrow \quad (X,Y) \in \{(0,2),(1,1),(1,2)\}$$

a tedy

$$P\,(X \cdot Y > 1) = P\Big((X,Y) \in \{(0,2),(1,1),(1,2)\}\Big) = \frac{1}{4} + \frac{1}{8} + 0 = \frac{3}{8}\;.$$

(b) U veličin X a Y předpokládáme obory hodnot určené tabulkou (ostatní hodnoty mohou být také případně nabyty, ale s nulovou pravděpodobností). Hodnoty z veličiny $Z = X^2 \cdot Y$ si pro přehlednost zapíšeme také do tabulky

$z(x,y) = x^2 \cdot y$ (x \ y)	0	1	2
-1	0	1	2
0	0	0	0
1	0	1	2

a pravděpodobnostní funkce p_Z vznikne nasčítáním pravděpodobností pro jednotlivé případy

$$p_Z(z) = P(X^2 \cdot Y = z) = \sum_{\substack{x^2 \cdot y = z \\ x,y \in \mathbb{R}}} \underbrace{P(X = x,\ Y = y)}_{p_{X,Y}(x,y)} = \begin{cases} \frac{1}{8} + 0 + \frac{1}{8} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} & ,\ z = 0 \\ 0 + \frac{1}{8} = \frac{1}{8} & ,\ z = 1 \\ \frac{1}{8} + 0 = \frac{1}{8} & ,\ z = 2 \\ 0 & ,\ \text{jinak.} \end{cases}$$

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

(d) 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	$\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	1/4
0	$\frac{1}{4}\cdot\frac{1}{2}=\frac{1}{8}$	$\frac{3}{8}\cdot\frac{1}{2}=\frac{3}{16}$	$\frac{3}{8}\cdot\frac{1}{2}=\frac{3}{16}$	1/2
1	$\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	1/4
$p_{Y'}$	1/4	3/8	3/8	

(e) Spočítáme korelaci X a Y:

$$E(X) = (-1)\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} = 0$$

$$E(X^2) = (-1)^2\cdot\frac{1}{4} + 0^2\cdot\frac{1}{2} + 1^2\cdot\frac{1}{4} = \frac{1}{2}$$

$$E(Y) = 0\cdot\frac{1}{4} + 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} = \frac{9}{8}$$

$$E(Y^2) = 0^2\cdot\frac{1}{4} + 1^2\cdot\frac{3}{8} + 2^2\cdot\frac{3}{8} = \frac{15}{8}$$

$$E(XY) = \sum_{(i,j)\in\mathbb{R}^2} ij\cdot p_{X,Y}(i,j) = (-1)\cdot 2\cdot\frac{1}{8} + 1\cdot 1\cdot\frac{1}{8} = -\frac{1}{8}$$

Korelace tedy je

$$\varrho(X,Y) = \frac{E(XY) - E(X)\cdot E(Y)}{\sqrt{E(X^2)-E(X)^2}\cdot\sqrt{E(Y^2)-E(Y)^2}} = \frac{-\frac{1}{8} - 0\cdot\frac{9}{8}}{\sqrt{\frac{1}{2}-0^2}\cdot\sqrt{\frac{15}{8}-(\frac{9}{8})^2}} =$$

$$= -\sqrt{\frac{2}{39}} \doteq -0.2264 \doteq \arccos(103,09°)\ .$$

Úhel mezi náhodnými veličinami $X - E(X)$ a $Y - E(Y)$ je pak $103,09°$.

Definice: Náhodný vektor (X,Y) má spojité rozdělení se *sdruženou hustotou pravděpodobnosti* $f_{X,Y} : \mathbb{R}^2 \to \langle 0,+\infty) \Leftrightarrow$ $f_{X,Y}$ je integrabilní funkce a pro každou "rozumnou" množinu $A \subseteq \mathbb{R}^2$ (tj. takovou, která se dá získat z intervalu v $\mathbb{R}^2$ pomocí sjednocování, průniku a doplňku) platí, že

$$P((X,Y)\in A) = \iint_A f_{X,Y}(x,y)\ \mathrm{d}x\mathrm{d}y\ .$$

To nastává právě když

$$F_{X,Y}(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)\ \mathrm{d}x\mathrm{d}y$$

pro každé $a,b\in\mathbb{R}$.

Sdružená hustota $f_{X,Y}$ opět (jako u veličin) NENÍ zdaleka určena jednoznačně, co se týče její funkční hodnoty, ale pouze hodnotami integrálů z této funkce (např. její změnou v konečně mnoha bodech nebo na nějaké hladké křivce se nezmění příslušné integrály, takže i změněná funkce bude také hustotou). Přesněji, dvě nezáporné funkce $f_{X,Y}$ a $g_{X,Y}$ (s integrálem rovným jedné) jsou hustotami pro tutéž sdruženou distribuční funkci $F_{X,Y}$ právě když se rovnají *skoro všude* a zapisuje se to jako

$$f_{X,Y} = g_{X,Y} \quad \text{(s.v.)}\ .$$

(tj. mohou se lišit jen na takové množině $A\subseteq\mathbb{R}^2$, že $\iint_A 1\ \mathrm{d}x\mathrm{d}y = 0$, tj. pokud A má nulový plošný obsah).

Je dobré si uvědomit, že náhodný vektor (X,Y) můžeme snadno sestavit z libovolných dvou náhodných veličin X a Y. Zatímco ale k počítání s veličinou X nám stačí znát jen její distribuční funkci F_X, k práci s vektorem nám NESTAČÍ znalost distribučních funkcí jeho složek! Potřebujeme totiž znát, jaký je vztah mezi veličinami X a Y, a ten je schovaný právě ve sdružené distribuční funkci.

8.7 (spojitý náhodný vektor)
Sdružená hustota náhodných veličin X a Y je

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

(a) Jaká jsou jejich marginální rozdělení?

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

(c) Jak vypadá jejich korelační matice?

(d) Určete pravděpodobnost $P(X > Y)$.

Řešení:

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

$$f_X(x) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \begin{cases} \int\limits_0^\infty \frac{1}{2}e^{-x-\frac{y}{2}}dy = e^{-x}\cdot\left[-e^{-\frac{y}{2}}\right]_0^\infty = e^{-x} & \text{pro } x>0, \\ \int\limits_{-\infty}^\infty 0\ dy = 0 & \text{pro } x\le 0\ . \end{cases}$$

$$f_Y(y) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \begin{cases} \int\limits_0^\infty \frac{1}{2}e^{-x-\frac{y}{2}}dx = \frac{1}{2}e^{-\frac{y}{2}}\cdot\left[-e^{-x}\right]_0^\infty = \frac{1}{2}e^{-\frac{y}{2}} & \text{pro } y>0, \\ \int\limits_{-\infty}^\infty 0\ dx = 0 & \text{pro } y\le 0\ . \end{cases}$$

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.

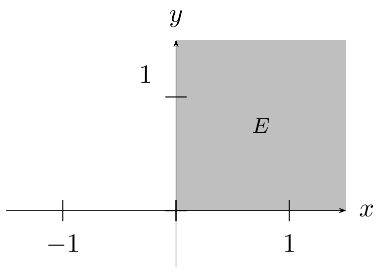
(Podmínce “skoro všude” se nelze vyhnout z toho důvodu, že hustoty nejsou jednoznačně definovány svými hodnotami, ale svými integrály.)

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

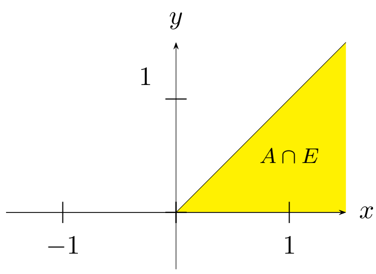
$$\text{Corr}(X,Y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(d) Sdružená hustota $f_{(X,Y)}$ je nenulová na množině $E := (0, +\infty)^2$ (vyznačena šedě):



Jev "$X > Y$" je popsán jako $(X,Y) \in \underbrace{\{(x,y) \in \mathbb{R}^2 \mid x > y\}}_{A}$. Pravděpodobnost tohoto jevu tak dostaneme zintegrováním hustoty přes množinu A, neboli

$$P(X > Y) = P\Big((X,Y) \in A\Big) = \iint\limits_{A} f_{X,Y}(x,y)\ dx\ dy = \iint\limits_{A\cap E} f_{X,Y}(x,y)\ dx\ dy =$$



$$= \Big\{A \cap E:\ 0 < y < x\Big\} = \int_0^\infty \int_0^x \tfrac{1}{2} e^{-x-\frac{y}{2}}\ dy\ dx = \int_0^\infty e^{-x} \left[-e^{-\frac{y}{2}}\right]_{y=0}^{y=x} dx =$$

$$= \int_0^\infty e^{-x}(1 - e^{-\frac{x}{2}})\ dy = \int_0^\infty e^{-x} - e^{-\frac{3}{2}x}\ dx = \left[-e^{-x} + \tfrac{2}{3} e^{-\frac{3}{2}x}\right]_{x=0}^{x=\infty} = 1 - \tfrac{2}{3} = \tfrac{1}{3}\ .$$

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

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

Určete:

(a) sdruženou hustotu $f_{X,Y}$.

(b) marginální hustoty f_X, f_Y.

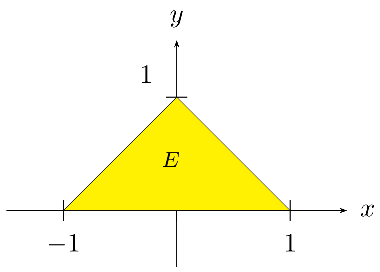
(c) zda jsou náhodné veličiny X a Y závislé či nezávislé.

(d) hodnotu $P(X + Y \geq \frac{1}{2})$.

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

Řešení:

Množina E je trojúhelník



(a) Spojité rovnoměrné rozdělení na E odpovídá modelu geometrické pravděpodobnosti na E (můžeme si pro větší názornost představovat E jako terč, do kterého se trefujeme všude se stejnou “intenzitou”). Sdružená hustota je tedy dána jako

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

kde c je konstanta taková, aby integrál z hustoty byl roven 1, tedy

$$1 = \iint\limits_{\mathbb{R}^2} f_{X,Y}(x,y)\ dx\ dy = \iint\limits_{E} c\ dx\ dy = c \iint\limits_{E} 1\ dx\ dy = c \cdot \text{obsah}(E) = c \cdot 1$$

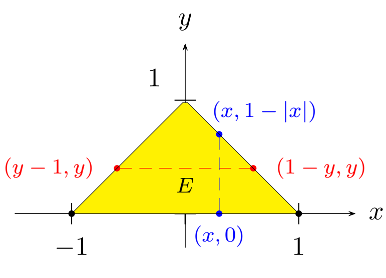
neboli

$$c = 1\ .$$

Integrál $\iint\limits_{E} 1\ dx\ dy$ jsme mohli buď skutečně spočítat nebo prostě využít jeho geometrickou interpretaci, tj. že je to obsah trojúhelníka.

(b) Marginální rozdělení vektoru (X,Y) jsou rozdělení jeho jednotlivých složek, tj. veličin X a Y. Marginální hustoty f_X a f_Y jsou tedy hustoty náhodných veličin X a Y. V tomto případě je nejsnadněji získáme částečným zintegrováním sdružené hustoty, tedy integrací podél vhodného řezu množiny E, kterou si ještě vyjádříme jako

$$E: \quad -1 \leq x \leq 1\ \ \&\ \ 0 \leq y \leq 1 - |x|$$



Pro $x \in \langle -1, 1\rangle$ tak máme

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \int_0^{1-|x|} 1\ dy = 1 - |x|$$

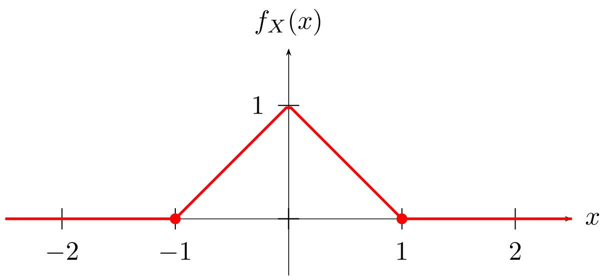
a pro $y \in \langle 0, 1\rangle$ máme

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_{y-1}^{1-y} 1\ dx = 2(1-y)\ .$$

Celkově tedy

$$f_X(x) = \begin{cases} 1 - |x| & ,\ x \in \langle -1, 1\rangle \\ 0 & ,\ \text{jinak,} \end{cases}$$

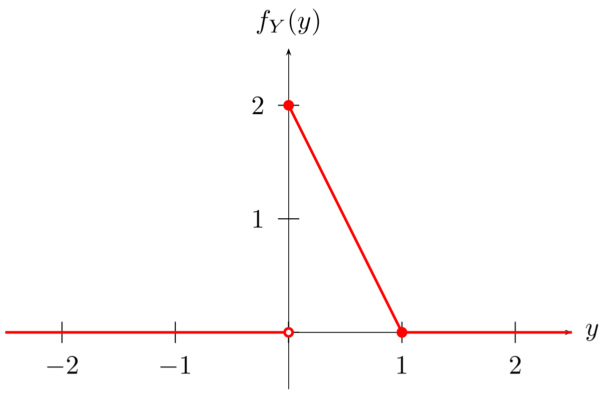
s grafem



a podobně

$$f_Y(y) = \begin{cases} 2(1-y) & ,\ y \in \langle 0, 1\rangle \\ 0 & ,\ \text{jinak,} \end{cases}$$

s grafem



(c) Veličiny X a Y jsou nezávislé právě když

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

V případě existence sdružené hustoty je to ekvivalentní tomu, že

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

(tj. množina bodů, kde rovnost NENASTÁVÁ, má nulový plošný obsah). Obě strany se mohou lišit např. v konečně mnoha bodech, nebo na nějaké křivce atd. Rovnost hustot *skoro všude* se velmi často opomíjí a nevyznačuje se. Nicméně to nic nemění na tom, že je potřeba tuhle věc mít na paměti.

Speciálně, pro nezávislé veličiny musí platit následující (pouze nutná podmínka!):

Nechť

$$S = \{(x,y) \in \mathbb{R}^2 \mid f_{X,Y}(x,y) \neq 0\}$$

a $\pi_1, \pi_2 : \mathbb{R}^2 \to \mathbb{R}$ jsou projekce na jednotlivé souřadné osy, tj. $\pi_1(x,y) = x$ a $\pi_2(x,y) = y$. Pokud jsou veličiny X a Y *nezávislé*, má množina

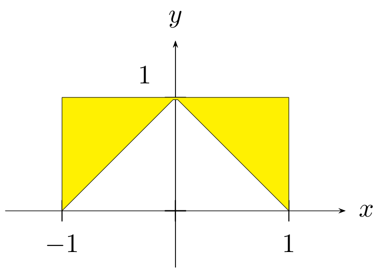
$$\Big(\pi_1(S) \times \pi_2(S)\Big) \setminus S$$

nulový obsah.

V našem konkretním případě $S = E$ a $\pi_1(E) = \langle -1, 1\rangle$ a $\pi_2(E) = \langle 0, 1\rangle$. Ovšem množina

$$\Big(\pi_1(E) \times \pi_2(E)\Big) \setminus E \ = \ \text{“obdélník”} \ \setminus \ \text{“trojúhelník”}$$

tj.



zřejmě nulový obsah NEMÁ. Veličiny X a Y tudíž NEJSOU nezávislé.

Naopak, pokud bychom hledali sdruženou hustotu vektoru (X', Y'), který by měl mít stejná marginální rozdělení jako vektor (X, Y), ale složky X' a Y' by byly nezávislé, stačilo by (podobně jako u diskrétních vektorů a jejich tabulek) položit

$$f_{X',Y'}(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} 2(1-|x|)(1-y), & (x,y) \in \langle -1,1\rangle \times \langle 0,1\rangle \\ 0, & \text{jinak.} \end{cases}$$

(d) Jev “$X + Y \geq \frac{1}{2}$” je množina $\Phi^{-1}(A)$, kde $\Phi = (X,Y) : \Omega \to \mathbb{R}^2$ je náš náhodný vektor a

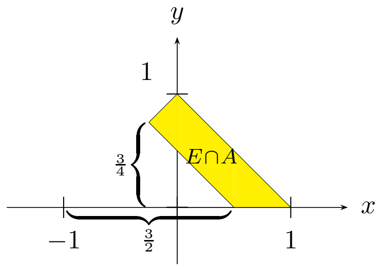
$$A = \{(x,y) \in \mathbb{R}^2 \mid x + y \geq \frac{1}{2}\}$$

tj. $\Phi^{-1}(A)$ je vzor množiny $A \subseteq \mathbb{R}^2$ při zobrazení Φ. Pravděpodobnost tohoto jevu tak dostaneme zintegrováním hustoty přes množinu A, neboli

$$P(X + Y \geq \tfrac{1}{2}) = \iint\limits_{A = \{(a,b) \in \mathbb{R}^2 \ \mid \ a+b \geq \frac{1}{2}\}} f_{X,Y}(x,y)\ dx\ dy = \iint\limits_{E \cap A} 1\ dx\ dy = \text{obsah}(E \cap A)\ .$$

Množina $E \cap A$ je tvaru

$$E \cap A : \quad \tfrac{1}{2} \le x + y \le 1 \quad \wedge \quad y - x \le 1 \quad \wedge \quad y \ge 0$$



Velikost plochy $E \cap A$ spočítáme snadněji pomocí doplňkové plochy (tj. trojúhelníka s výškou $\frac{3}{4}$ a základnou $\frac{3}{2}$) do původního trojúhelníka E (s plochou 1):

$$P(X + Y \ge \tfrac{1}{2}) = 1 - \tfrac{1}{2} \cdot \tfrac{3}{4} \cdot \tfrac{3}{2} = \tfrac{7}{16} \ .$$

(e) Pro kovarianci máme

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

Přitom je

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

$$= \int_0^1 \int_{y-1}^{1-y} 2xy \ dx \ dy = \int_0^1 y \underbrace{\left[x^2\right]_{x=y-1}^{x=1-y}}_{=0} \ dy = 0.$$

To, že tento integrál vyjde nula bylo vidět už na začátku z lichosti integrované funkce (tj. $xy \cdot f_{X,Y}(x, y)$) vzhledem k proměnné x (lichost této funkce je pochopitelně určena i tím, že množina A je symetrická podle osy y). Ze stejných důvodů bude nulový i následující integrál:

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

což bylo vidět už z toho, že hustota f_X pro X byla sudá.

Takže $\varrho(X, Y) = 0$, ačkoliv veličiny X a Y nezávislé nejsou (viz (c)).

Poznamenejme ještě, že $E(Y) = \frac{1}{3}$, což snadno zjistíme integrováním pomocí hustoty f_Y a nebo z toho, že plocha pod hustotou je (pravoúhlý) trojúhelník, jehož x-ová složka těžiště je právě ve vzdálenosti $\frac{1}{3}$ od svislé odvěsny.

8.9 (spojitý náhodný vektor)

Náhodný vektor (X, Y) má sdruženou hustotu $f_{X,Y}$. Jeho sdružená distribuční funkce je

$$F_{X,Y}(x, y) = \frac{1}{250}(20xy - x^2y - xy^2) \quad \text{pro} \quad (x, y) \in \langle 0, 5 \rangle \times \langle 0, 5 \rangle.$$

Určete:

(a) sdruženou hustotu $f_{X,Y}$.

(b) $F_{X,Y}(x,y)$ v ostatních bodech.

(c) marginální distribuční funkce F_X, F_Y.

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

(e) hodnotu $P(X > 2)$.

(f) střední hodnotu vektoru $E(X,Y)$.

Řešení:
Podobně jako u diskrétního vektoru zadaného tabulkou potřebujeme i zde ověřit, že máme všechny informace pro popis pravděpodobnosti na $\mathbb{R}^2$. Znamená to, že musí platit, že $P\big((X,Y) \in \langle 0,5\rangle^2\big) = 1$.

Proč to potřebujeme vědět: V opačném případě by totiž pro $D = \mathbb{R}^2 \setminus \langle 0,5\rangle^2$ bylo $P\big((X,Y) \in D\big) > 0$. Pak bychom ovšem měli $\iint\limits_D f_{X,Y} \mathrm{d}S = P\big((X,Y) \in D\big) > 0$ a přitom bychom nevěděli, jaký tvar hustota $f_{X,Y}$ na D má!

K tomuto ověření využijeme existence hustoty a vlastnosti sdružené distribuční funkce:

$$P\big((X,Y) \in \langle 0,5\rangle^2\big) = \iint\limits_{\langle 0,5\rangle^2} f_{X,Y} \mathrm{d}S = \iint\limits_{(0,5\rangle^2} f_{X,Y} \mathrm{d}S = P\big((X,Y) \in (0,5\rangle^2\big) =$$
$$= F_{X,Y}(5,5) - F_{X,Y}(0,5) - F_{X,Y}(5,0) - F_{X,Y}(0,0) =$$
$$= \frac{1}{250}(20 \cdot 5 \cdot 5 - 5^2 \cdot 5 - 5 \cdot 5^2) - 0 - 0 - 0 = 1 \ .$$

Připomenutí odvození vlastnosti pro $F_{X,Y}$: Označme si pro jednoduchost $U_{a,b} := (-\infty, a\rangle \times (-\infty, b\rangle$ Pro pravděpodobnost $P_{X,Y}$ na $\mathbb{R}^2$ máme:

$$(0,5\rangle^2 = U_{5,5} \setminus (U_{0,5} \cup U_{5,0})$$

a tedy

$$P\Big((X,Y) \in (0,5\rangle^2\Big) = P_{X,Y}\big((0,5\rangle^2\big) = P_{X,Y}\big(U_{5,5}\big) - P_{X,Y}\big(U_{0,5} \cup U_{5,0}\big) =$$
$$= P_{X,Y}\big(U_{5,5}\big) - P_{X,Y}\big(U_{0,5}\big) - P_{X,Y}\big(U_{5,0}\big) - P_{X,Y}\big(\underbrace{U_{0,5} \cap U_{5,0}}_{=U_{0,0}}\big) =$$
$$= F_{X,Y}(5,5) - F_{X,Y}(0,5) - F_{X,Y}(5,0) - F_{X,Y}(0,0) \ .$$

(a) Vzhledem k tomu, že už víme, že $P\Big((X,Y) \in (0,5)^2\Big) = 1$ (okraje intervalu díky hustotě nehrají roli), musí být $P\Big((X,Y) \in \mathbb{R}^2 \setminus (0,5)^2\Big) = 0$. Na množině $\mathbb{R}^2 \setminus (0,5)^2$ se proto hustota nemůže významně lišit od nulové funkce. Konkrétněji nám to popisuje následující věta, která zachytí i situaci pro body $(x,y) \in (0,5)^2$:

Podmínka pro sdruženou hustotu: Nechť $U \subseteq \mathbb{R}^2$ je otevřená množina a (X,Y) je náhodný vektor. Jestliže

- $P\Big((X,Y) \in \mathbb{R}^2 \setminus U\Big) = 0$ a
- sdružená distribuční funkce $F_{X,Y}$ je třídy C^2 na U (tj. má v U spojité druhé parciální derivace)

pak funkce

$$f_{X,Y}(x,y) = \begin{cases} \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) & , \ (x,y) \in U \\ 0 & , \ (x,y) \in \mathbb{R}^2 \setminus U \end{cases}$$

je hustotou pravděpodobnosti pro (X,Y).

Pro $(x, y) \in (0,5)^2$ tedy máme

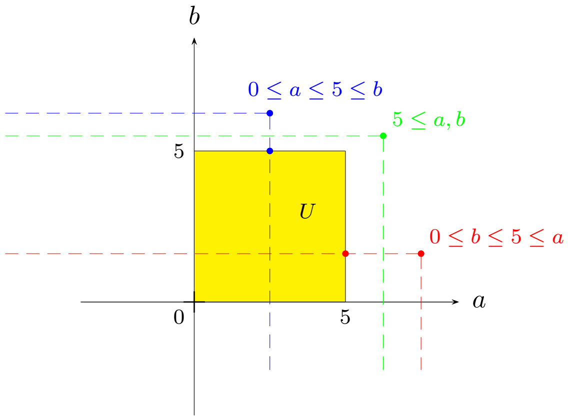
$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{20xy - x^2y - xy^2}{250}\right)\right) =$$

$$= \frac{\partial}{\partial x}\left(\frac{20x - x^2 - 2xy}{250}\right) = \frac{10 - x - y}{125}$$

Celkem jsme tedy zjistili, že hustota pravděpodobnosti je

$$f_{X,Y}(x, y) = \begin{cases} \frac{10-x-y}{125} & , (x, y) \in (0,5)^2 \\ 0 & , \text{jinak.} \end{cases}$$

(b) Ostatní hodnoty $F_{X,Y}(a, b)$ už snadno určíme z obrázku (funkce $F_{X,Y}(a, b)$ je spojitá a hustota $f_{X,Y}$ je nulová mimo žlutou oblast $U = (0,5)^2$):



Např. pro $0 \leq b \leq 5 \leq a$ máme, že

$$F_{X,Y}(a, b) = \iint_{U \cap U_{a,b}} f_{X,Y}\ dS = \iint_{U \cap U_{5,b}} f_{X,Y}\ dS = F_{X,Y}(5, b)$$

Celkem máme

$$F_{X,Y}(a, b) = \begin{cases} 0 & , a \leq 0 \text{ nebo } b \leq 0 \\ \frac{1}{250}(20ab - a^2b - ab^2) & , 0 \leq a, b \leq 5 \\ F_{X,Y}(5, b) = \frac{1}{50}(15b - b^2) & , 0 \leq b \leq 5 \leq a \\ F_{X,Y}(a, 5) = \frac{1}{50}(15a - a^2) & , 0 \leq a \leq 5 \leq b \\ 1 & , 5 \leq a, b. \end{cases}$$

Z cvičných důvodů si ještě zpětně odvodíme hodnoty $F_{X,Y}$ z hustoty $f_{X,Y}$ pro $(a, b) \in (0,5)^2$:

$$F_{X,Y}(a, b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x, y)\ \mathrm{d}y\mathrm{d}x = \int_0^a\int_0^b \frac{10 - x - y}{125}\ \mathrm{d}y\mathrm{d}x =$$

$$= \frac{1}{125}\int_0^a (10 - x)b - \left[\frac{y^2}{2}\right]_0^b\ \mathrm{d}x = \frac{1}{125}\int_0^a 10b - xb - \frac{b^2}{2}\ \mathrm{d}x =$$

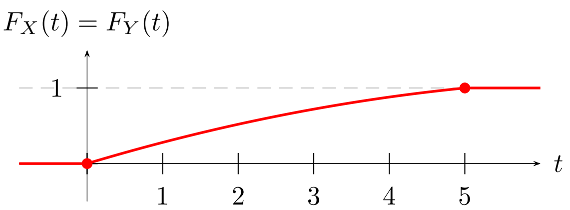
$$= \frac{1}{125}\left(10ba - \frac{a^2}{2}b - \frac{b^2}{2}a\right) = \frac{1}{250}(20ab - a^2b - ab^2)$$

(c) Marginální distribuční funkce F_X a F_Y představují distribuční funkce jednotlivých složek vektoru, tj. (samostatných) náhodných veličin X a Y. Získáme je jednoduše jako limity sdružené distribuční funkce (také s využitím obrázků):

$$F_X(a) = P(X \leq a) = \lim_{b\to\infty} P(X \leq a, Y \leq b) = \lim_{b\to\infty} F_{X,Y}(a,b) = \begin{cases} 0 & , \ a \leq 0 \\ \frac{1}{50}(15a - a^2) & , \ 0 \leq a \leq 5 \\ 1 & , \ 5 \leq a \end{cases}$$

$$F_Y(b) = P(Y \leq b) = \lim_{a\to\infty} P(X \leq a, Y \leq b) = \lim_{a\to\infty} F_{X,Y}(a,b) = \begin{cases} 0 & , \ b \leq 0 \\ \frac{1}{50}(15b - b^2) & , \ 0 \leq b \leq 5 \\ 1 & , \ 5 \leq b \end{cases}$$

s grafy



(d) Veličiny X a Y jsou nezávislé právě když

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \ \text{ pro všechna } \ (x,y) \in \mathbb{R}^2 \ .$$

V našem případě pro $(x,y) \in \langle 0,5\rangle^2$ máme

$$F_X(x) \cdot F_Y(y) = \frac{1}{2500}(15x - x^2)(15y - y^2)$$

což se zjevně liší od $F_{X,Y}(x,y) = \frac{1}{250}(20xy - x^2y - xy^2)$ např. v bodě $(x,y) = (1,1)$.

Veličiny X a Y tedy jsou *závislé*.

(e)

$$P(X > 2) = 1 - P(X \leq 2) = 1 - \lim_{b\to\infty} P(X \leq 2, Y \leq b) =$$

$$= 1 - \lim_{b\to\infty} F_{X,Y}(2,b) = 1 - F_{X,Y}(2,5) = 1 - \frac{130}{250} = \frac{12}{25} = 0.48$$

(f) Střední hodnota náhodného vektoru (X,Y) popisuje jakou průměrnou hodnotu dvojic (x,y) můžeme při opakovaných měřeních očekávat. Střední hodnota je přirozeně definována jako

$$E(X,Y) := (E(X), E(Y)) \ .$$

Protože X a Y mají stejná rozdělení, tak máme

$$E(Y) = E(X) = -\int\limits_{-\infty}^{0} F_X(t)\ \mathrm{d}t + \int\limits_{0}^{\infty} \big(1 - F_X(t)\big)\ \mathrm{d}t = \int\limits_{0}^{5} 1 - \frac{1}{50}(15t - t^2)\ \mathrm{d}t =$$

$$= \left[t - \frac{3t^2}{20} + \frac{t^3}{150}\right]_0^5 = 5 - \frac{15}{4} + \frac{5}{6} = \frac{25}{12} \doteq 2.08\ .$$