8. cvičení z STP

13. - 17. duben 2020

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.

(Obdobně vznikne náhodný vektor s více složkami. My se teď zaměříme hlavně na dvousložkový případ.)

Náhodný vektor (X, Y) 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

$$P_{(X,Y)}(A) := P\Big(\underbrace{(X, Y)^{-1}(A)}_{\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A\}} \Big) .$$

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

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

Opět si připomeňme, že 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 pro *sdruženou distribuční funkci* máme

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

Věta: Nechť náhodný vektor (X, Y) má *diskrétní* rozdělení. Pak:

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

- -

Příklad 8.1 *Diskrétní náhodný vektor (X, Y) má sdružené pravděpodobnosti dány tabulkou:*

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

(a) Určete pravděpodobnost $P(X + Y > 1)$.

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

(c) *Určete marginální rozdělení vektoru (X, Y). Jsou veličiny X a Y nezávislé?*

(d) *Pokud X **a** Y **jsou závislé**, popište (jednoznačně určené) rozdělení náhodného vektoru (X', Y') se stejnými marginálními rozděleními jako (X, Y), jehož složky **jsou nezávislé**.*

*Pokud X **a** Y **jsou nezávislé**, najděte rozdělení nějakého náhodného vektoru (X'', Y'') se stejnými marginálními rozděleními jako (X, Y), jehož složky **jsou závislé**.*

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) U veličin X a Y předpokládáme obory hodnot určené tabulkou. Pak máme

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

a tedy

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

(b) Hodnoty z veličiny $Z = X^2 \cdot Y$ si pro přehlednost zapíšeme také do tabulky

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

a pravděpodobnosti hodnot veličiny Z vzniknou nasčítáním pravděpodobností pro jednotlivé případy

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

(c) Marginální (česky: okrajová) rozdělení náhodného vektoru (X, Y) jsou rozdělení jeho jednotlivých složek, tedy veličin X a Y. Vektor (X, Y) má diskrétní rozdělení a obě veličiny X a Y budou proto mít také diskrétní rozdělení a pro jejich rozdělení platí:

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

Hodnoty pravděpodobností získáme tedy sečtením v řádcích (pro X) a sloupcích (pro Y) naší tabulky:

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

Protože nyní je např.

$$P(X=-1,\ Y=1)=0\neq \tfrac{1}{4}\cdot\tfrac{3}{8}=P(X=-1)\cdot P(Y=1)$$

jsou X a Y **závislé**. (Nejjednodušší k tomu účelu je najít si v tabulce právě nulovou hodnotu.)

(d) Nechť (X',Y') je nyní náhodný vektor s **nezávislými** složkami, které mají stejná marginální rozdělení jako má vektor (X,Y) tedy

$$P(X'=i)=P(X=i)\quad \text{a}\quad P(Y'=j)=P(Y=j)\quad \text{pro všechna}\quad i,j\in\mathbb{R}\ .$$

Pro sdružené pravděpodobnosti vektoru (X',Y') pak tedy platí, že

$$P(X'=i,\ Y'=j)=P(X'=i)\cdot P(Y'=j)=P(X=i)\cdot P(Y=j)$$

a můžeme je tak popsat následující tabulkou:

	$Y'=0$	$Y'=1$	$Y'=2$	$P(X'=i)$
$X'=-1$	$\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	1/4
$X'=0$	$\frac{1}{4}\cdot\frac{1}{2}=\frac{1}{8}$	$\frac{3}{8}\cdot\frac{1}{2}=\frac{3}{16}$	$\frac{3}{8}\cdot\frac{1}{2}=\frac{3}{16}$	1/2
$X'=1$	$\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	$\frac{3}{8}\cdot\frac{1}{4}=\frac{3}{32}$	1/2
$P(Y'=j)$	1/4	3/8	3/8	

Abychom ukázali, jak řešit druhou možnost, předpokládejme nyní naopak, že takovouto tabulku pro nezávislé složky (X',Y') dostaneme. Speciálně vidíme, že všechny sdružené pravděpodobnosti v tabulce budou nenulové. Jak teď vyrobit nějaké jiné rozdělení (X'',Y''), se **závislými** složkami, které budou mít stejné součty v řádcích a sloupcích? Stačí si vybrat dva řádky a dva sloupce (pro přehlednost např. první a druhý)

	$Y'=j_1$	$Y'=j_2$	
$X'=i_1$	a	b	
$X'=i_2$	c	d	

a tam, kde se protínají (celkově tedy jen ve 4 buňkách), udělat úpravu o hodnotu ε a tím vytvořit tabulku pro (X'',Y'') (zbylé hodnoty v tabulce necháme stejné):

	$Y''=j_1$	$Y''=j_2$	
$X''=i_1$	$a-\varepsilon$	$b+\varepsilon$	
$X''=i_2$	$c+\varepsilon$	$d-\varepsilon$	

Přitom pochopitelně musíme dodržet, aby upravené hodnoty byly nezáporné (protože to musí být pravděpodobnosti), takže máme toto omezení:

$$-\min(b,c)\leq\varepsilon\leq\min(a,d)\ .$$

V našem případě tedy

$$-3/32=-\min(3/32,\ 1/8)\leq\varepsilon\leq\min(1/16,\ 3/16)=1/16\ .$$

a my si můžeme zvolit např. $\varepsilon = 1/16$, čímž si v první buňce vyrobíme nulu.

Protože jsme změnou hodnot v buňkách alespoň na jednom místě (dokonce však na čtyřech místech) porušili původní rovnosti pro nezávislé složky X' a Y', tedy konkrétně

$$P(X'' = i_1, Y'' = j_1) = a - \varepsilon \neq a = P(X'' = i_1) \cdot P(Y'' = j_1)$$

budou veličiny X'' a Y'' závislé. Výsledná tabulka tedy bude následující:

	$Y'' = 0$	$Y'' = 1$	$Y'' = 2$	$P(X'' = i)$
$X'' = -1$	$\frac{1}{16} - \frac{1}{16} = 0$	$\frac{3}{32} + \frac{1}{16} = \frac{5}{32}$	$\frac{3}{32}$	$1/4$
$X'' = 0$	$\frac{1}{8} + \frac{1}{16} = \frac{3}{16}$	$\frac{3}{16} - \frac{1}{16} = \frac{2}{16}$	$\frac{3}{16}$	$1/2$
$X'' = 1$	$\frac{1}{16}$	$\frac{3}{32}$	$\frac{3}{32}$	$1/2$
$P(Y'' = j)$	$1/4$	$3/8$	$3/8$	

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

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

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 $\operatorname{var}(X) = \|X - E(X)\|^2$, takže platí

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

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

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

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

A kromě toho máme:

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

Praktický důsledek korelace:

Pokud máme dvě veličiny X a Y takové, že

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

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

Obdobně platí: Jestliže
$$X - E(X) \geq 0 \;\Leftrightarrow\; Y - E(Y) \leq 0$$
pak je $\operatorname{corr}(X, Y) \leq 0$.

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

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

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

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

Příklad 8.2 *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 střední hodnotu $E(X \cdot Y)$. Určete varianční a korelační matici.

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

Řešení:

(a) Marginální (diskrétní) rozdělení (tj. rozdělení jednotlivých složek vektoru) získáme pro jednotlivé hodnoty sečtením pravděpodobností v řádcích (pro Y) a sloupcích (pro X) 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

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

(b) Pro diskrétní vektor (X, Y) a borelovskou (tj. téměř každou, např. spojitou) funkci $h : \mathbb{R}^2 \to \mathbb{R}$ platí podobná věc jako jsme měli už u náhodné veličiny a sice:

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

Odsud tedy snadno spočítáme střední hodnotu $E(XY)$:

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

Kovarianci pak vypočteme ze vztahu

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

K tomu potřebujeme znát také

$$\begin{aligned} E(X) &= 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{1}{8} = \tfrac{5}{8} \\ E(Y) &= 0 \cdot \tfrac{3}{8} + 1 \cdot \tfrac{5}{8} = \tfrac{5}{8} \end{aligned}$$

Takže

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

Pro korelaci

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

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 $\;\Leftrightarrow\;$ $P(X=i, Y=j) = P(X=i)\cdot P(Y=j)$ 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 (pouze na základě její znalosti) o nezávislosti obecně nic říct!

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

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

To nastává právě když

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

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 $\iint\limits_A 1\ \mathrm{d}x\mathrm{d}y = 0$, 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.

Příklad 8.3 *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?

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

Řešení:

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

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

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

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

(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.

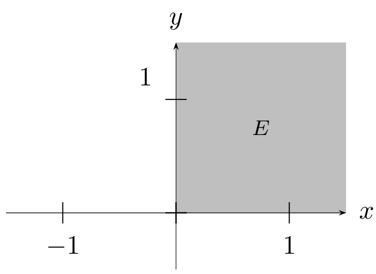
(Podmínce "skoro všude" se nelze vyhnout z toho důvodu, že hustoty nejsou jednoznačně definovány svými hodnotami, ale svými integrály.)

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 $\mathrm{cov}(X,Y) = 0$, tedy také $\mathrm{corr}(X,Y) = 0$ a korelační matice je tak

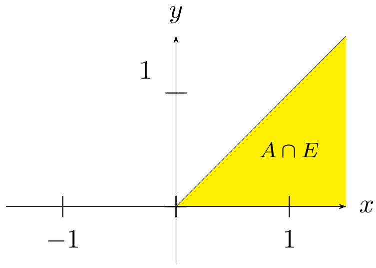
$$\mathrm{Corr}(X,Y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(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

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



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

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

Připomenutí: Pro veličinu X s konečným rozptylem $\operatorname{var}(X)$ se definuje tzv. *směrodatná odchylka*

$$\sigma_X := \sqrt{\operatorname{var}(X)}$$

Směrodatná odchylka je často vhodnější pro popis parametru veličiny, protože má stejný fyzikální rozměr (např. kg) jako původní veličina (zatímco fyzikální rozměr rozptylu je druhou mocninou původní veličiny, zde např. kg^2). Kromě toho, jak bylo napsáno výše, význam směrodatné odchylky je norma jistého vektoru, konkrétně $\sigma_X = \|X - E(X)\|$.

Pro korelaci dvou veličin pak máme

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

a tedy také

$$\operatorname{cov}(X,Y) = \sigma_X \cdot \sigma_Y \cdot \operatorname{corr}(X,Y)\ .$$

Příklad 8.4 *Náhodný vektor (X,Y) má následující parametry:*

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

(a) Stanovte střední hodnotu a rozptyl náhodných veličin

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

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

Řešení:

(a) Střední hodnota je lineární zobrazení (“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, co vzniká ze skalárního součinu - viz výše.)

$$\mathrm{var}(T) = \mathrm{var}(2X+3) = \mathrm{var}(2X) = 2^2 \cdot \mathrm{var}(X) = 2^2 \cdot (\sigma_X)^2 = 4 \cdot 25 = 100$$

$$\mathrm{var}(U) = \mathrm{var}(200 - Y) = \mathrm{var}(-Y) = (-1)^2 \cdot \mathrm{var}(Y) = (\sigma_Y)^2 = 400$$

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

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

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

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!

Příklad 8.5 *Pro náhodné veličiny X a Y platí, že*

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

Pro náhodné veličiny

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

určete

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

(b) rozptyl $\mathrm{var}(X+Y)$.

(c) kovarianční a korelační matice náhodného vektoru (X,Y).

Řešení:

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

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

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

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

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

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

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:

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

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

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

(c) Kovarianční matice pro (X,Y):

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

Korelační matice pro (X,Y):

$$\operatorname{Corr}(X,Y) = \begin{pmatrix} \operatorname{corr}(X,X) & \operatorname{corr}(X,Y) \\ \operatorname{corr}(Y,X) & \operatorname{corr}(Y,Y) \end{pmatrix} = \begin{pmatrix} 1 & \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}} \\ \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 1 \end{pmatrix}$$

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