4 NÁHODNÝ VEKTOR

Čas ke studiu kapitoly: 60 minut

Cíl: Po prostudování této kapitoly budete umět

- popsat náhodný vektor a jeho sdružené rozdělení
- vysvětlit pojmy marginální a podmíněné rozdělení pravděpodobnosti
- popsat stochastickou nezávislost náhodných veličin

Průvodce studiem:

V předcházející kapitole jsme se seznámili s popisem náhodné veličiny. Nyní se zaměříme na závažnější a komplikovanější problém, kterým je hledání, zkoumání a hodnocení závislosti mezi dvěma a více náhodnými veličinami. K poznání a matematickému popisu těchto závislosti slouží metody **regresní a korelační analýzy**.

Výklad:

4.1. Náhodný vektor

Jsou situace, kdy je při náhodném pokusu získáno několik veličin. (*Například můžeme měřit pouze teplotu, ale také celý vektor meteorologických veličin (výška, tlak, teplota, rosný bod). Nebo při volbě Miss ČR jsou jednotlivé krásky popsány trojicí čísel (např. 92, 60, 95). Každé z těchto čísel je samo o sobě náhodnou veličinou, abychom si však udělali o dané dívce představu, potřebujeme znát celou trojici těchto náhodných veličin.*) Soubor těchto veličin nazýváme **náhodným vektorem**.

Jednotlivé náhodné veličiny v rámci náhodného vektoru mohou být naprosto nezávislé, mohou však také mít silnou vazbu.

Náhodným vektorem budeme dále rozumět sloupcový vektor složený z náhodných veličin $\boldsymbol{X} = (X_1, X_2, \dots X_n)$.

Pro názornou ilustraci se omezíme na studium vztahu mezi dvěmi náhodnými veličinami (tj. budeme se zabývat dvousložkovým náhodným vektorem) s tím, že učiněné závěry lze jednoduše zobecnit na n-složkový náhodný vektor.

Sdružená (simultánní) distribuční funkce náhodných veličin X a Y je definována předpisem:

$$F(x, y) = P(X < x; Y < y)$$

Sdružená distribuční funkce má obdobné vlastnosti jako distribuční funkce jedné proměnné.

Vlastnosti sdružené distribuční funkce:

1. $\lim\limits_{x, y \to -\infty} F(x, y) = 0$
2. $\lim\limits_{x, y \to \infty} F(x, y) = 1$
3. funkce F(x, y) je neklesající v každé proměnné
4. funkce F(x, y) je zleva spojitá v každé proměnné

Pravděpodobnost, že náhodný vektor je z obdélníkové oblasti lze vyjádřit pomocí distribuční funkce:

$$P(a_1 \le X < b_1; a_2 < Y < b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$$

4.1.1 Rozdělení náhodného vektoru (dvousložkového)

Jestliže existuje nejvýše spočetně mnoho hodnot náhodného vektoru, jde o **náhodný vektor s diskrétním rozdělením** a sdružená distribuční funkce je pak definována jako:

$$F(x, y) = \sum_{(i,j)\ x_i < x, y_j < y} P(X = x_i, Y = y_j)$$

Veličina $P(x_i, y_j)$ se nazývá **sdružená pravděpodobnostní funkce**, popř. (vícerozměrná) pravděpodobnostní funkce náhodného vektoru X.

O **náhodném vektoru se spojitým rozdělením** mluvíme v případě, že má absolutně spojitou distribuční funkci F(x,y), tj. pokud existuje nezáporná funkce f(x,y) taková, že:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x^1, y^1) dx^1 dy^1$$

Pokud existuje druhá smíšená derivace distribuční funkce, pak:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Pravděpodobnost, že spojitý náhodný vektor je z obdélníkové oblasti lze vyjádřit jako:

$$P(a_1 \le X < b_1; a_2 < Y < b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dx dy$$

4.2 Marginální rozdělení pravděpodobnosti

Chceme-li určit distribuční funkci složky X (resp. složky Y) náhodného vektoru, mluvíme o **marginální distribuční funkci**.

Marginální distribuční funkce dvousložkového náhodného vektoru definujeme takto:

$$F_X(x) = P(X < x) = \lim_{y \to \infty} F(x, y)$$
$$F_Y(y) = P(Y < y) = \lim_{x \to \infty} F(x, y)$$

Z tohoto vyjádření dále plyne, že v případě diskrétního náhodného vektoru s pravděpodobnostní funkci $P(x_i, y_j)$ můžeme získat následující vztahy pro **marginální pravděpodobnosti**:

$$P_X(x) = \sum_{y_j} P(X = x, Y = x_j)$$
$$P_Y(y) = \sum_{x_i} P(X = x_i, Y = y)$$

Obdobně pro spojitý náhodný vektor s hustotou f(x,y) získáme vztahy pro **marginální hustoty pravděpodobnosti**:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$$

4.3 Nezávislost složek náhodného vektoru

Složky X,Y náhodného vektoru jsou navzájem nezávislé právě tehdy, jsou-li nezávislé náhodné veličiny X, Y.

Platí tedy:

$$F(x, y) = F_X(x).F_Y(y)$$

Konkrétně pro **náhodný vektor s diskrétním rozdělením** platí, že složky X, Y náhodného vektoru jsou nezávislé právě když platí:

$$P(X = x_i, Y = y_j) = P_X(X = x_i).P_Y(Y = y_j)$$

A obdobně pro **náhodný vektor se spojitým rozdělením** platí, že složky X, Y náhodného vektoru jsou nezávislé právě když platí:

$$f(x, y) = f_X(x_i).f_Y(y)$$

4.4 Korelační tabulka

V případě diskrétního dvousložkového náhodného vektoru s konečným počtem hodnot se sdružená pravděpodobnostní funkce často prezentuje prostřednictvím **korelační tabulky**. V této tabulce se mimo sdružené pravděpodobnostní funkce uvádí rovněž v posledním řádku (sloupci) marginální pravděpodobnostní funkce.

X\Y	**y_1**	**y_2**	**...**	**y_m**	$\sum_j P(X=x_i, Y=y_j)$
x_1	$P(X=x_1,Y=y_1)$	$P(X=x_1,Y=y_2)$	...	$P(X=x_1,Y=y_m)$	$P_X(x_1)$
x_2	$P(X=x_2,Y=y_1)$	$P(X=x_2,Y=y_2)$	...	$P(X=x_2,Y=y_m)$	$P_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
x_n	$P(X=x_n,Y=y_1)$	$P(X=x_n,Y=y_2)$	...	$P(X=x_n,Y=y_m)$	$P_X(x_n)$
$\sum_i P(X=x_i, Y=y_j)$	$P_Y(y_1)$	$P_Y(y_2)$	...	$P_Y(y_m)$	$\sum_{j=1}^{m} P_Y(y_j) = \sum_{i=1}^{n} P_X(x_i) = 1$

4.5 Podmíněné rozdělení pravděpodobnosti

Pro náhodný vektor s diskrétním rozdělením pravděpodobnosti je definována **podmíněná pravděpodobnostní funkce**:

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P_Y(Y = y)}, \qquad pro\ P_Y(Y = y) \neq 0$$

Obdobně pro náhodný vektor se spojitým rozdělením pravděpodobnosti je definována **podmíněná hustota pravděpodobnosti**:

$$f(x| y) = \frac{f(x, y)}{f_Y(y)}, \qquad pro\ f_Y(y) \neq 0$$

4.6 Charakteristiky náhodného vektoru

- **Obecné a centrální momenty**

Kromě momentů náhodných veličin X,Y (složek náhodného vektoru) jsou ještě definovány tzv. smíšené momenty.

Nejpoužívanějším obecným momentem je **střední hodnota**: $\mu_1^{'} = EX$

Nejpoužívanějším centrálním momentem je **rozptyl**: $\mu_2 = DX$

Smíšený obecný moment řádu k,n je definován jako: $\mu_{kn}^{'} = E(X^k . Y^n)$

Smíšený centrální moment řádu k,n je definován jako: $\mu_{kn} = E\left[(X - EX)^k \left(Y - EY^n\right)\right]$

- **Kovariance**

Kovariance je nejjednodušším ukazatelem souvislosti dvou náhodných veličin. Je definována jako smíšený centrální moment 2. řádu (1,1).

$$Cov(X,Y) = \mu_{11} = E\left[(X - EX)(Y - EY)\right]$$

Kladná hodnota kovariance znamená, že se zvětšením hodnoty X se pravděpodobně zvýší i hodnota Y.

Oproti tomu **záporná hodnota kovariance** znamená, že se zvětšením hodnoty X se pravděpodobně sníží hodnota Y.

- **Kovarianční matice**

V praxi se čato setkáváme s reprezentací centrálních momentů 2.řádu ve formě tzv. kovarianční matice:

$$\begin{pmatrix} DX & Cov(X,Y) \\ Cov(X,Y) & DY \end{pmatrix}$$

V této souvislosti se někdy označuje $DX \approx Cov(X,X)$, resp. $DY \approx Cov(Y,Y)$.

- **Jednoduchý korelační koeficient**

Jednoduchý **korelační koeficient je mírou lineární závislosti** dvou náhodných veličin. (**POZOR!!!!** Je mírou pouze **lineární** závislosti, žádné jiné, tj. je-li jednoduchý korelační koeficient nulový, neznamená to, že mezi náhodnými veličinami neexistuje závislost!!!!)

Korelační koeficient definován jako:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} \qquad DX, DY \neq 0$$

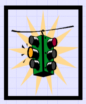
Pro jednoduchý korelační koeficient platí:

1. $-1 \leq \rho_{XY} \leq 1$
2. $\rho_{XY} = \rho_{YX}$

Předpokládejme pro X a Y vztah lineární závislosti $Y = a + bX$. Pak lze dále odvodit:

1. $\rho_{XY} = 0 \Rightarrow$ X a Y jsou **nekorelované** (= **lineárně nezávislé,** b=0) (POZOR!!! $\rho_{XY} = 0$ **<u>neznamená</u> nezávislé** náhodné veličiny)
2. $\rho_{XY} > 0 \Rightarrow$ X a Y jsou **pozitivně korelované** (b>0, s rostoucím X roste i Y)
3. $\rho_{XY} < 0 \Rightarrow$ X a Y jsou **negativně korelované** (b<0, s rostoucím X klesá Y)

Je tedy zřejmé, že pokud se hodnota korelačního koeficientu blíží „1“ (resp. „(-1)“), značí to přímou (resp. nepřímou) lineární závislost, pokud se hodnota korelačního koeficientu blíží „0“, jsou náhodné veličiny X,Y nekorelované (= lineárně nezávislé). V mnoha případech však nelze na první pohled určit zda hodnotu korelačního koeficientu už můžeme považovat za blízkou „1“ (resp. „-1“, popř. „0“) a potom je nutné významnost („blízkost „0“) korelačního koeficientu testovat (viz kapitola Testování hypotéz).



Řešený příklad:

Představme si, že budeme třikrát opakovat pokus u nějž známe pravděpodobnost úspěchu (např. hody mincí, p = 0,5). Zvolme tyto náhodné veličiny:

Y ... počet pokusů do prvního úspěchu
Z ... počet po sobě jdoucích úspěchů

Náhodný vektor X = (Y, Z).

a) Určete marginální pravděpodobnostní funkce $P_Y(y)$, $P_Z(z)$
b) Sestavte sdruženou pravděpodobnostní funkci P(Y=y, Z=z)
c) Určete, zda jsou náhodné veličiny Y, Z nezávislé.
d) Určete střední hodnoty a rozptyly složek Y,Z
e) Určete kovarianční matici Y, Z
f) Určete jednoduchý korelační koeficient
g) Určete podmíněné pravděpodobnostní funkce a P(Y=y | Z=z), P(Z=z | Y=y)

Řešení:

Vypišme si všechny možné kombinace, k nimž by mohlo dojít (S - úspěch, F - neúspěch):

{ FFF; SFS; SSF; FSS; FSF; FFS; SFF; SSS }

A uvažujme, že pravděpodobnost úspěchu P(S) = p, pravděpodobnost neúspěchu P(F) = 1-p.

Jedná se o diskrétní dvourozměrný náhodný vektor, přičemž:

složka Y může nabývat hodnot: 0, 1, 2, 3
složka Z může nabývat hodnot: 0, 1, 2, 3

Pojmenujme si všechny elementární jevy základního prostoru a určeme pravděpodobnost jejich výskytu (pro výpočet jednotlivých pravděpodobností využijeme poznatku, že jevy F a S jsou nezávislé).

A1 ... FFF — $P(A1) = (1 - p)^3 = 0{,}125$
A2 ... SFS — $P(A2) = p^2.(1 - p) = 0{,}125$
A3 ... SSF — $P(A3) = p^2.(1 - p) = 0{,}125$
A4 ... FSS — $P(A4) = p^2.(1 - p) = 0{,}125$
A5 ... FSF — $P(A5) = p.(1 - p)^2 = 0{,}125$
A6 ... FFS — $P(A6) = p.(1 - p)^2 = 0{,}125$

A7 ... SFF $\quad P(A7) = p.(1-p)^2 = 0{,}125$

A8 ... SSS $\quad P(A8) = p^3 = 0{,}125$

ada) Zapišme si nyní do pomocných tabulek, které jevy vyhovují daným hodnotám náhodných veličin Y a Z.

Y ... počet pokusů do prvního úspěchu			
0	1	2	3
A2, A3, A7, A8	A4, A5	A6	A1

Z ... počet po sobě jdoucích úspěchů			
0	1	2	3
A1	A2, A5, A6, A7	A3, A4	A8

protože jevy A1, ... , A8 jsou neslučitelné, můžeme marginální pravděpodobnostní funkce jednoduše určit.

Např. $P_Y(0) = P(A2) + P(A3) + P(A7) + P(A8) = 0{,}125 + 0{,}125 + 0{,}125 + 0{,}125 = 0{,}500$

Y ... počet pokusů do prvního úspěchu			
$P_Y(0)$	$P_Y(1)$	$P_Y(2)$	$P_Y(3)$
0.500	0.250	0.125	0.125

Z ... počet po sobě jdoucích úspěchů			
$P_Z(0)$	$P_Z(1)$	$P_Z(2)$	$P_Z(3)$
0.125	0.500	0.250	0.125

V našem případě (máme určovat zároveň sdruženou pravděpodobnostní funkci) by bylo rychlejší pro určeni marginálních pravděpodobnostních funkcí využít korelační tabulku, kterou budeme vytvářet pro zápis sdružené pravděpodobnosti.

adb) Zkonstruujeme korelační tabulku. (nejdříve si do ní vypíšeme jevy, které vyhovují příslušným podmínkám a poté na základě jejich neslučitelnosti určíme pravděpodobnosti výskytu příslušných skupin jevů)

		Z			
		0	1	2	3
Y	0	-	A2, A7	A3	A8
	1	-	A5	A4	-
	2	-	A6	-	-
	3	A8	-	-	-

Tabulka sdružené pravděpodobnostní funkce

		Z			
		0	1	2	3
Y	0	0	0,250	0,125	0,125
	1	0	0,125	0,125	0
	2	0	0,125	0	0
	3	0,125	0	0	0

Např. P(Y = 0, Z = 2) = 0,125; P(Y = 0, Z = 1) = 0,250

Chceme-li získat korelační tabulku v klasickém tvaru, tj. včetně marginálních pravděpodobnosti, stačí sečíst příslušné řádky (sloupce).

		Z				
		0	1	2	3	$P_Y(y)$
Y	0	0	0,250	0,125	0,125	0,500
	1	0	0,125	0,125	0	0,250
	2	0	0,125	0	0	0,125
	3	0,125	0	0	0	0,125
	$P_Z(z)$	0,125	0,500	0,250	0,125	1

Pro srovnání si porovnejte takto získané marginální pravděpodobnosti s marginálními pravděpodobnostmi získanými v ada)

adc) Pro náhodný vektor s diskrétním rozdělením platí, že složky Y, Z náhodného vektoru jsou nezávislé právě když platí:

$$P(Y = y_i, Z = z_j) = P_Y(Y = y_i).P_Z(Z = z_j)$$

Tento předpoklad v našem případě splněn není. (např. $P(\acute{Y} = 2, Z = 1) \neq P_Y(2).P_Z(1)$; $0{,}125 \neq 0{,}125.0{,}500$). Z toho plyne, že **náhodné veličiny Y, Z nejsou nezávislé**.

add) Střední hodnoty a rozptyly získáme z definičních vztahů pomocí marginálních pravděpodobnostních funkcí:

$$\underline{\underline{EY}} = \sum_{i=1}^{4} y_i.P_Y(y_i) = 0.0{,}500 + 1.0{,}250 + 2.0{,}125 + 3.0{,}125 = \frac{7}{8} = \underline{\underline{0{,}875}}$$

$$EY^2 = \sum_{i=1}^{4} y_i^2.P_Y(y_i) = 0^2.0{,}500 + 1^2.0{,}250 + 2^2.0{,}125 + 3^2.0{,}125 = 1{,}875$$

$$\underline{\underline{DY}} = EY^2 - (EY)^2 = 1{,}875 - (0{,}875)^2 = \frac{71}{64} \cong \underline{\underline{1{,}109}}$$

$$\underline{\underline{EZ}} = \sum_{i=1}^{4} z_i.P_Z(z_i) = 0.0{,}125 + 1.0{,}500 + 2.0{,}250 + 3.0{,}125 = \frac{11}{8} = \underline{\underline{1{,}375}}$$

$$EZ^2 = \sum_{i=1}^{4} z_i^2.P_Z(z_i) = 0^2.0{,}125 + 1^2.0{,}500 + 2^2.0{,}250 + 3^2.0{,}125 = 2{,}625$$

$$\underline{\underline{DZ}} = EZ^2 - (EZ)^2 = 2{,}625 - (1{,}375)^2 = \frac{47}{64} \cong \underline{\underline{0{,}734}}$$

ade) Kovarianční matice má obecný tvar:

$$\begin{pmatrix} DY & Cov(Y,Z) \\ Cov(Y,Z) & DZ \end{pmatrix}$$

Pro její zápis musíme určit kovarianci.

$$\underline{\underline{Cov(Y,Z)}} = E[(Y-EY)(Z-EZ)] = \sum_{i,j}\left(y_i - \frac{7}{8}\right)\left(z_j - \frac{11}{8}\right)P(Y = y_i, Z = z_j) =$$

$$= \left(0-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}250 + \left(0-\frac{7}{8}\right)\left(2-\frac{11}{8}\right).0{,}125 + \left(0-\frac{7}{8}\right)\left(3-\frac{11}{8}\right).0{,}125 +$$

$$+ \left(1-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}125 + \left(1-\frac{7}{8}\right)\left(2-\frac{11}{8}\right).0{,}125 + \left(2-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}125 +$$

$$+ \left(3-\frac{7}{8}\right)\left(0-\frac{11}{8}\right).0{,}125 = -\frac{296}{512} \cong \underline{\underline{-0{,}422}}$$

Kovarianční matice má tvar:

$$\begin{pmatrix} 1{,}109 & -0{,}422 \\ -0{,}422 & 0{,}734 \end{pmatrix}$$

adf) Jednoduchý korelační koeficient určíme z definičního vztahu:

$$\underline{\underline{\rho_{X,Y}}} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} = \frac{-0{,}422}{\sqrt{1{,}109.0{,}734}} \cong \underline{\underline{-0{,}468}}$$

Na základě této hodnoty korelačního koeficientu můžeme říci, že mezi náhodnými veličinami Y a Z existuje **středně silná negativní korelace**, tj. že pravděpodobně s růstem Y bude Z klesat (lineárně).

adg) Podmíněné pravděpodobnosti budeme opět zapisovat do tabulky a jejich hodnoty určíme z definice:

$$P(Y = y|Z = z) = \frac{P(Y = y, Z = z)}{P_Z(z)}$$

Tabulka podmíněné pravděpodobnostní funkce P(Y = y| Z = z)

		Z			
		0	1	2	3
Y	0	0/0,125	0,250/0,500	0,125/0,250	0,125/0,125
	1	0/0,125	0,125/0,500	0,125/0,250	0/0,125
	2	0/0,125	0,125/0,500	0/0,250	0/0,125
	3	0,125/0,125	0/0,500	0/0,250	0/0,125

		Z			
		0	1	2	3
Y	0	0	0,500	0,500	1,000
	1	0	0,250	0,500	0
	2	0	0,250	0	0
	3	1,000	0	0	0

Např. P(Y = 2| Z = 1) = 0,250

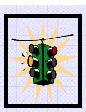
$$P\left(Z = y \middle| Y = y\right) = \frac{P(Y = y, Z = z)}{P_Y(y)}$$

Tabulka podmíněné pravděpodobnostní funkce P(Z = z| Y = y)

		Z			
		0	1	2	3
Y	0	0/0,500	0,250/0,500	0,125/0,500	0,125/0,500
	1	0/0,250	0,125/0,250	0,125/0,250	0/0,250
	2	0/0,125	0,125/0,125	0/0,125	0/0,125
	3	0,125/0,125	0/0,125	0/0,125	0/0,125

		Z			
		0	1	2	3
Y	0	0	0,500	0,250	0,250
	1	0	0,500	0,500	0
	2	0	1,000	0	0
	3	1,000	0	0	0

Např. P(Z = 2| Y= 1) = 0,500



Řešený příklad:

Sdružená hustota pravděpodobnosti dvousložkového náhodného vektoru je definována jako:

$$f(x, y) = \begin{matrix} x + y & \quad pro\ (x, y) \in \langle 0;1 \rangle \times \langle 0;1 \rangle \\ 0 & \quad jinde \end{matrix}$$

Určete:

a) Marginální hustoty pravděpodobnosti $f_X(x)$, $f_Y(y)$
b) Marginální distribuční funkce $F_X(x)$, $F_Y(y)$
c) Střední hodnoty a rozptyly složek X, Y
d) Hodnotu jednoduchého korelačního koeficientu, výsledek dejte do souvislosti s mírou lineární závislosti

Řešení:

ada) Jde o spojitý náhodný vektor, proto:

$$f_X(x) = \begin{cases} \int_0^1 (x+y)dy = \left[xy + \frac{y^2}{2} \right]_0^1 = x + \frac{1}{2} & pro\ x \in \langle 0;1 \rangle \\ \int_0^1 0dy = 0 & jinde \end{cases}$$

Ze symetrie sdružené pravděpodobnostní funkce vyplývá i obdobný tvar $f_Y(y)$.

$$f_Y(y) = \begin{cases} y + \frac{1}{2} & pro\ y \in \langle 0;1 \rangle \\ 0 & jinde \end{cases}$$

adb) Marginální distribuční funkce jednotlivých složek určíme z marginálních hustot pravděpodobnosti:

$$F_X(x) = \begin{cases} \int_{-\infty}^{x} 0dt = 0 & pro\ x \in (-\infty;0) \\ \int_{-\infty}^{0} 0dt + \int_0^x \left(t + \frac{1}{2} \right) dt = 0 + \left[\frac{t^2}{2} + \frac{t}{2} \right]_0^x = \frac{1}{2} x(x+1) & pro\ x \in \langle 0;1 \rangle \\ \int_{-\infty}^{0} 0dt + \int_0^1 \left(t + \frac{1}{2} \right) dt + \int_0^x 0dt = 0 + 1 + 0 = 1 & pro\ x \in (1;\infty) \end{cases}$$

ze symetrie f(x,y) můžeme opět odvodit:

$$F_Y(y) = \begin{cases} 0 & pro\ y \in (-\infty;0) \\ \frac{1}{2} y(y+1) & pro\ y \in \langle 0;1 \rangle \\ 1 & pro\ y \in (1;\infty) \end{cases}$$

adc) Střední hodnoty a rozptyly jednotlivých složek určíme pomocí marginálních hustot pravděpodobnosti, na základě znalosti definičních vztahů pro oba momenty:

$$\underline{\underline{EX}} = \int_{-\infty}^{\infty} x.f_X(x)dx = \int_0^1 x(x + \frac{1}{2})dx = \left[\frac{x^3}{3} + \frac{x^2}{4} \right]_0^1 = \underline{\underline{\frac{7}{12}}}$$

$$EX^2 = \int_{-\infty}^{\infty} x^2.f_X(x)dx = \int_0^1 x^2(x + \frac{1}{2})dx = \left[\frac{x^4}{4} + \frac{x^3}{6} \right]_0^1 = \frac{5}{12}$$

$$\underline{\underline{DX}} = EX^2 - (EX)^2 = \frac{5}{12} - \left(\frac{7}{12} \right)^2 = \frac{60-49}{144} = \underline{\underline{\frac{11}{144}}}$$

Opět využijeme symetrie sdružené hustoty pravděpodobnosti f(x,y) a můžeme tvrdit, že:

$$\underline{\underline{EY = \frac{7}{12}}}; \quad \underline{\underline{\mathrm{DY} = \frac{11}{144}}}$$

add) Pro výpočet jednoduchého korelačního koeficientu potřebujeme znát hodnotu kovariance a proto začneme jejím výpočtem:

$$\underline{\underline{Cov(X,Y)}} = E[(X-EX)(Y-EY)] = \int_0^1\int_0^1 \left(x-\frac{7}{12}\right).\left(y-\frac{7}{12}\right).(x+y)dxdy =$$

$$= \int_0^1\int_0^1 \left[\left(x^2-\frac{7}{12}x\right).\left(y-\frac{7}{12}\right)+\left(x-\frac{7}{12}\right)\left(y^2-\frac{7}{12}y\right)\right]dxdy =$$

$$= \int_0^1 \left[\left(\frac{x^3}{3}-\frac{7x^2}{24}\right).\left(y-\frac{7}{12}\right)+\left(\frac{x^2}{2}-\frac{7x}{12}\right).\left(y^2-\frac{7}{12}y\right)\right]_0^1 dy =$$

$$= \int_0^1 \left[\frac{1}{24}\left(y-\frac{7}{12}\right)-\frac{1}{12}.\left(y^2-\frac{7}{12}y\right)\right]dy = \left[\frac{1}{24}\left(\frac{y^2}{2}-\frac{7y}{12}-\frac{2y^3}{3}+\frac{7y^2}{12}\right)\right]_0^1 = \underline{\underline{\frac{-1}{144}}}$$

Dále již stačí jen dosadit do definičního vztahu pro jednoduchý korelační koeficient:

$$\underline{\underline{\rho_{X,Y}}} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} = \frac{\frac{-1}{144}}{\sqrt{\frac{11}{144}.\frac{11}{144}}} = \underline{\underline{\frac{-1}{11}}} \cong -0{,}091$$

Z velikosti korelačního koeficientu můžeme usuzovat na to, že mezi X a Y pravděpodobně neexistuje lineární závislost, tj, X a Y jsou **nekorelované** náhodné veličiny.

Σ Shrnutí:

Náhodným vektorem rozumíme sloupcový vektor složený z náhodných veličin $X = (X_1, X_2, \ldots X_n)$, který je charakterizován **sdruženou (simultánní) distribuční funkcí.**

Ze sdruženého rozdělení náhodného vektoru můžeme snadno najít **marginální rozdělení pravděpodobnosti** jednotlivých náhodných veličin, z nichž je vektor sestaven.

Podmíněné rozdělení pak chápeme jako podíl sdruženého a marginálního rozdělení pravděpodobnosti (má-li tento podíl smysl), v souladu s definicí podmíněné pravděpodobnosti.

Nezávislost náhodných veličin se projevuje tím, že jejich sdružená distribuční funkce (sdružená pravděpodobnostní funkce, resp. sdružená hustota pravděpodobnosti) se dá matematicky vyjádřit jako součin marginálních distribučních funkcí (marginálních pravděpodobností, resp. marginálních hustot pravděpodobnosti) jednotlivých náhodných veličin.

Mezi nejvýznamnější smíšené momenty náhodného vektoru patří **kovariance**.

Mírou lineární závislosti je **korelační koeficient**.

? Otázky

1. Vysvětlete pojmy simultánní, marginální a podmíněné rozdělení pravděpodobnosti
2. Vysvětlete pojem stochastické nezávislosti náhodných veličin
3. Definujte podmíněné rozdělení pravděpodobnosti.
4. Co nám napovídá hodnota koeficientu korelace ?

Úlohy k řešení

1. Nezávisle hodíme dvěma symetrickými mincemi. Pro každou minci zaznamenáme výsledek „1“, když padne panna, „0“, když padne orel. Označme S součet výsledků na obou mincích, R rozdíl výsledků na obou mincích. Definujme dvousložkový náhodný vektor X = (S, R). Určete:

   a) Typ náhodného vektoru (diskrétní, spojitý)
   b) Sdruženou pravděpodobnostní funkci
   c) Marginální pravděpodobnostní funkce
   d) Střední hodnoty a rozptyly jednotlivých složek
   e) Kovarianční matici
   f) Jednoduchý korelační koeficient
   g) Jsou náhodné veličiny S, R nezávislé?

2. Při průzkumu příčin dopravních nehod byl měřen systolický tlak řidičů autobusů v závislosti na teplotě ovzduší. Vypočtěte jednoduchý korelační koeficient a pouze z jeho hodnoty odhadněte, zda teplota ovzduší spíše zvyšuje či spíše snižuje systolický tlak řidičů.

Teplota ovzduší [^{o}C]	-10,5	-5,4	0,2	6,4	10,2	15,6	18,5	25,5	31,5	35,8
Systolický tlak [mm Hg]	76	78	81	81	74	72	76	81	82	83

3. Náhodný vektor X = (R, S) má sdruženou hustotu pravděpodobnosti:

$$f(r,s) = \begin{cases} \frac{2}{3}(r+2s) & pro\ (r,s) \in \left(\langle 0;1\rangle \times \langle 0;1\rangle\right) \\ 0 & jinde \end{cases}$$

   Určete:
   a) Marginální hustoty pravděpodobnosti $f_R(r)$, $f_S(s)$
   b) Marginální distribuční funkce $F_R(r)$, $F_S(s)$
   c) Střední hodnoty a rozptyly složek R, S
   d) Hodnotu jednoduchého korelačního koeficientu, výsledek dejte do souvislosti s mírou lineární závislosti

Řešení:

1. S … součet výsledků na obou mincích
   R … rozdíl výsledků na obou mincích

   a) Diskrétní náhodný vektor

   b) Korelační tabulka (sdružené pravděpodobnosti, marginální pravděpodobnosti)

		S			
		0	**1**	**2**	**$P_R(r)$**
R	**-1**	0	0,25	0	0,25
	0	0,25	0	0,25	0,50
	1	0	0,25	0,25	0,25
	$P_S(s)$	0,25	0,50	0,50	1,00

   c) Marginální pravděpodobnosti najdete ve výše korelační tabulce.

   d) ES = 1,00; DS = 0,50; ER = 0; DR = 0,50

   e) Cov (S, R) = 0;

   Kovarianční matice: $\begin{pmatrix} 0{,}5 & 0 \\ 0 & 0{,}5 \end{pmatrix}$

   f) $\rho_{S,R} = 0 \Rightarrow S, R \text{ } \textit{jsou nekorelované}$

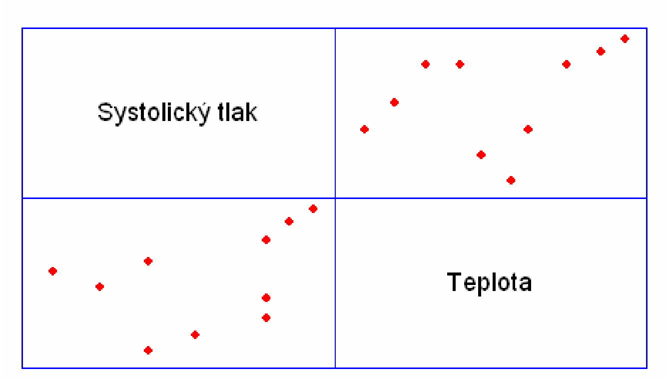
   g) S, R nejsou nezávislé náhodné veličiny

2. Výsledky tohoto příkladu byly získány pomocí programu Statgraphics:

<u>Correlations:</u>

	Systolický tlak	Teplota
Systolický tlak		0,3823
Teplota	0,3823	

$\Rightarrow \rho_{teplota, systolický\ tlak} = 0{,}382 \quad \Rightarrow$ středně silná pozitivní korelace, tj. pravděpodobně s rostoucí teplotou roste i systolický tlak řidičů autobusů. To potvrzuje i grafický záznam naměřených hodnot uvedený níže (graf vpravo nahoře).



3. X je spojitý náhodný vektor

a) $f_R(r) = \begin{cases} \frac{2}{3}(r+1) & r \in \langle 0;1 \rangle \\ 0 & jinde \end{cases}$, $\quad f_S(s) = \begin{cases} \frac{1}{3}(4s+1) & s \in \langle 0;1 \rangle \\ 0 & jinde \end{cases}$

b) $F_R(r) = \begin{cases} 0 & r \in (-\infty;0) \\ \frac{1}{3}r(r+2) & r \in \langle 0;1 \rangle \\ 1 & r \in (1;\infty) \end{cases}$, $\quad F_S(s) = \begin{cases} 0 & s \in (-\infty;0) \\ \frac{1}{3}s(2s+1) & s \in \langle 0;1 \rangle \\ 1 & s \in (1;\infty) \end{cases}$

c) $ER = \frac{5}{9}, \quad DR = \frac{13}{162}, \quad ES = \frac{11}{18}, \quad DS = \frac{23}{324}$

d) $Cov(R,S) = \left(-\frac{8}{81}\right), \quad \rho_{R,s} = \frac{-24.\sqrt{598}}{598} = (-0{,}98) \Rightarrow$ silná negativní korelace, tj. s vysokou pravděpodobností s rostoucím R lineárně klesá S