MV011 Statistika I

5. Číselné charakteristiky náhodné veličiny

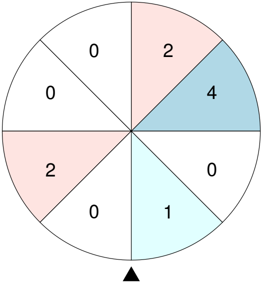
Jan Koláček (kolacek@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

Motivační příklad

Příklad 1

Uvažujme hru, kde účastník hry roztočí „kolo štěstí“. Každé pole tohoto kola definuje výhru (v Kč), která bude vyplacena hráči v případě, že na toto pole ukazuje šipka po zastavení kola. Za každou hru zaplatí hráč provozovateli 1 Kč. Budeme hrát? Tj. jaká je „očekávaná“ výhra?



Y ... **zisk** z jedné hry, X ... **částka**, kterou si vytočíme
Zřejmě $Y = X - 1$.

X	0	1	2	4
$p(x)$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

„Očekávaná“ výhra

$$EY = EX - 1 = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} - 1$$
$$= \frac{1}{8} = \boxed{\mathbf{0{,}125}} > 0.$$

Střední hodnota

Definice 1

Nechť X je náhodná veličina definovaná na $(\Omega, \mathcal{A}, P)$ a nechť existuje integrál $\int\limits_{\Omega} X(\omega)\, dP(\omega) < \infty$. Potom číslo

$$EX = \int\limits_{\Omega} X(\omega)\, dP(\omega)$$

nazýváme **střední hodnotou náhodné veličiny** X (**Expected Value, Mean**).

Značení: $\mathcal{L}_1(\Omega, \mathcal{A}, P)$ . . . množina všech náhodných veličin definovaných na $(\Omega, \mathcal{A}, P)$, které mají **konečné střední hodnoty**.

Střední hodnota

Věta 2 (Výpočet)

Nechť X *je náhodná veličina na* $(\Omega, \mathcal{A}, P)$. *Pak platí*

$X \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Leftrightarrow \int\limits_{-\infty}^{\infty} x dF(x) < \infty$. *V tomto případě je* $\boxed{EX = \int\limits_{-\infty}^{\infty} x dF(x)}$.

- *Nechť* $X \sim (M, p)$ *je diskrétního typu, pak platí* $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Leftrightarrow \sum\limits_{x \in M} xp(x)$ *absolutně konverguje. V tomto případě* $\boxed{EX = \sum\limits_{x \in M} xp(x)}$.
- *Nechť* $X \sim f(x)$ *je absolutně spojitého typu, pak platí* $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Leftrightarrow xf(x)$ *je integrovatelná vzhledem k Lebesgueově míře.*

  V tomto případě je $\boxed{EX = \int\limits_{-\infty}^{\infty} xf(x)dx}$.

Střední hodnota

Věta 3 (Vlastnosti)

Nechť X, X_1, X_2 *jsou náhodné veličiny definované na pravděpodobnostním prostoru* $(\Omega, \mathcal{A}, P)$, $a, a_1, a_2 \in \mathbb{R}$. *Potom*

- EX *existuje* $\quad \Leftrightarrow \quad E|X|$ *existuje.*
- *Jestliže* $P(X = a) = 1 \quad \Rightarrow \quad EX = a$.
- *Existují-li* EX_1, $EX_2 \quad \Rightarrow \quad E(a_1 X_1 + a_2 X_2) = a_1 EX_1 + a_2 EX_2$.
- *Nechť existují* EX_1, EX_2 *a platí* $X_1 \leq X_2 \quad \Rightarrow \quad EX_1 \leq EX_2$.
- *Nechť* $|X_1| \leq X_2$ *a* EX_2 *existuje* $\quad \Rightarrow \quad EX_1$ *existuje.*
- *Nechť* $P(X \geq 0) = 1 \quad \Rightarrow \quad EX \geq 0$.

Věta 4 (Střední hodnota součinu nezávislých náhodných veličin)

Nechť $X_1, \ldots, X_n$ *jsou* **nezávislé** *náhodné veličiny definované na* $(\Omega, \mathcal{A}, P)$ *a nechť existují střední hodnoty* $EX_1, \ldots, EX_n$. *Pak platí*

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} EX_i.$$

Příklad

Příklad 2 (Střední hodnota Alternativního rozdělení)

Mějme náhodnou veličinu $X \sim A(\theta)$, $\theta \in (0,1)$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} 1-\theta & x = 0 \\ \theta & x = 1 \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte střední hodnotu.

$$\boxed{EX} = \sum_{x=0}^{1} x p(x) = 0 \cdot (1-\theta) + 1 \cdot \theta = \boxed{\theta}.$$

Příklad

Příklad 3 (Střední hodnota Binomického rozdělení)

Mějme náhodnou veličinu $X \sim Bi(n,\theta)$, $n \in \mathbb{N}$, $\theta \in (0,1)$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x} & x \in M = \{0,1,\dots,n\} \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte střední hodnotu.

$$X = \sum_{i=1}^{n} Y_i, \qquad Y_i \sim A(\theta),\ i = 1,\dots,n$$

$$\boxed{EX} = E\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} EY_i = \boxed{n\theta}.$$

Nebo

$$\boxed{EX} = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x\binom{n}{x}\theta^x(1-\theta)^{n-x} = \cdots = \boxed{n\theta}.$$

Příklad

Příklad 4

Biatlonista střílí nezávisle na sobě do terče, přičemž pravděpodobnost zásahu při každém výstřelu je $2/3$. Jaká je očekávaná hodnota počtu zasažených terčů ze 300 pokusů?

X ... počet zásahů, $X \sim Bi(300, 2/3)$

$$\boxed{EX} = n\theta = 300 \cdot 2/3 = \boxed{\mathbf{200}}.$$

Příklad

Příklad 5 (Střední hodnota Poissonova rozdělení)

Mějme náhodnou veličinu $X \sim Po(\lambda)$, $\lambda > 0$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & x \in M = \{0, 1, \dots\} \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte střední hodnotu.

$$\boxed{EX} = \sum_{x=0}^{\infty} x p(x) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \left| subst.\, y = x - 1 \right| = \lambda \underbrace{\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!}}_{1 = \sum_{y \in M} p(y)} = \boxed{\lambda}.$$

Příklad 6 (Střední hodnota Rovnoměrného rozdělení)

Mějme náhodnou veličinu $X \sim Rs(a,b)$, $a < b$, $a,b \in \mathbb{R}$ *s hustotou*

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte střední hodnotu.

$$\boxed{EX} = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2}$$

$$= \frac{1}{b-a} \frac{(b-a)(b+a)}{2} = \boxed{\frac{a+b}{2}}.$$

Příklad

Příklad 7 (Střední hodnota Normálního rozdělení)

Mějme náhodnou veličinu s normálním rozdělením $X \sim N(\mu, \sigma^2)$ *s hustotou*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

$$EX = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Položíme-li $y = \frac{x-\mu}{\sigma}$, tj. $x = \sigma y + \mu$ a $dx = \sigma dy$, pak

$$\begin{aligned} \boxed{EX} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-\frac{1}{2}y^2} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy}_{=0 \text{ (lichá funkce)}} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy}_{=1 \text{ (hustota } Y \sim N(0,1))} = \boxed{\mu}. \end{aligned}$$

Motivační příklad

Příklad 8

Chceme koupit do naší továrny novou linku, která bude balit mouku do 1 kg sáčků. Navštívili jsme dva výrobce těchto linek. U každého jsme si nechali vyrobit 5 balíčků a ty pak zvážili, abychom zjistili přesnost balení.

linka A	*975*	*960*	*1030*	*990*	*1045*
linka B	*965*	*965*	*1020*	*995*	*1055*

Pro kterého výrobce se rozhodneme?

$$E(A) = (975 + 960 + 1030 + 990 + 1045)/5 = 1000$$

$$E(B) = (965 + 965 + 1020 + 995 + 1055)/5 = 1000$$

$$E(A) = E(B) \Rightarrow ?$$

Motivační příklad

$$
\begin{aligned}
S_A^2 = &\Big\{(975-1000)^2+(960-1000)^2+(1030-1000)^2+(990-1000)^2 \\
&+(1045-1000)^2\Big\}/5 = 1050 \\
S_B^2 = &\Big\{(965-1000)^2+(965-1000)^2+(1020-1000)^2+(995-1000)^2 \\
&+(1055-1000)^2\Big\}/5 = 1180
\end{aligned}
$$

$$S_A^2 < S_B^2 \Rightarrow \text{volíme } \boxed{A}$$

Rozptyl

Definice 5

Nechť X je náhodná veličina definovaná na $(\Omega, \mathcal{A}, P)$. Potom čísla

$\mu'_k = EX^k$			**obecným**		
$\mu_k = E(X - EX)^k$	nazýváme	k**-tým**	**centrálním**	**momentem**	n. v. X
$\bar{\mu}_k = E\|X\|^k$			**absolutním**		

za předpokladu, že uvedené střední hodnoty pro $k = 1, 2, \ldots$ existují.

Poznámka Je-li k-tý moment konečný, tj. $EX^k < \infty$, píšeme $X \in \mathcal{L}_k(\Omega, \mathcal{A}, P)$.

Definice 6

Druhý centrální moment nazýváme **rozptyl** (**Variance, Dispersion**) a značíme $DX = E(X - EX)^2$. Číslo $\sigma_X = \sqrt{DX}$ nazýváme **směrodatnou odchylkou** náhodné veličiny X (**Standard Deviation**).

Rozptyl

Věta 7 (Vlastnosti rozptylu)

Nechť X, X_1, X_2 *jsou náhodné veličiny definované na* $(\Omega, \mathcal{A}, P)$ *s konečnými druhými momenty,* $a, a_1, a_2 \in \mathbb{R}$. *Pak*

- $DX \geq 0$
- $\boxed{DX = EX^2 - (EX)^2}$
- *Jestliže* $P(X = a) = 1$, *pak* $DX = 0$
- $D(a_1 + a_2 X) = a_2^2 DX$
- *Nechť* X_1, X_2 *jsou* **nezávislé** *náhodné veličiny, pak* $D(X_1 + X_2) = DX_1 + DX_2$.

Příklad

Příklad 9

Vypočtěte rozptyl pro náhodný pokus točení kola štěstí z Příkladu 1.

X	0	1	2	4
$p(x)$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

$EX = \frac{9}{8}$, $DY = D(X-1) = DX$

1) $\boxed{DX = E(X - EX)^2}$

$$DX = \left(0 - \frac{9}{8}\right)^2 \frac{1}{2} + \left(1 - \frac{9}{8}\right)^2 \frac{1}{8} + \cdots = \frac{119}{64} = \mathbf{1{,}8594}$$

2) $\boxed{DX = EX^2 - (EX)^2}$

$$EX^2 = 0^2 \frac{1}{2} + 1^2 \frac{1}{8} + \cdots = \frac{25}{8} \Rightarrow DX = \frac{25}{8} - \left(\frac{9}{8}\right)^2 = \mathbf{1{,}8594}$$

Příklad

Příklad 10 (Rozptyl Alternativního rozdělení)

Mějme náhodnou veličinu $X \sim A(\theta)$, $\theta \in (0,1)$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} 1-\theta & x = 0 \\ \theta & x = 1 \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte rozptyl.

$$EX^2 = \sum_{x=0}^{1} x^2 p(x) = 0^2 \cdot (1-\theta) + 1^2 \cdot \theta = \theta. \qquad EX = \theta \; \ldots \text{viz Příklad 2.}$$

$$\boxed{DX} = EX^2 - E^2X = \theta - \theta^2 = \boxed{\theta(1-\theta)}.$$

Nebo

$$\boxed{DX} = E(X-EX)^2 = \sum_{x=0}^{1} (x-\theta)^2 p(x) = (0-\theta)^2(1-\theta) + (1-\theta)^2\theta = \boxed{\theta(1-\theta)}.$$

Příklad

Příklad 11 (Rozptyl Binomického rozdělení)

Mějme náhodnou veličinu $X \sim Bi(n, \theta)$, $n \in \mathbb{N}$, $\theta \in (0, 1)$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x} & x \in M = \{0, 1, \ldots, n\} \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte rozptyl.

$$X = \sum_{i=1}^{n} Y_i, \qquad Y_i \sim A(\theta),\ i = 1, \ldots, n$$

$$\boxed{DX} = D\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} DY_i = \boxed{n\theta(1-\theta)}.$$

Příklad 12 (Rozptyl Poissonova rozdělení)

Mějme náhodnou veličinu $X \sim Po(\lambda)$ *s pravděpodobnostní funkcí*

$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & x \in M = \{0, 1, \dots, \} \\ 0 & \textit{jinak.} \end{cases}$$

Vypočtěte rozptyl.

$DX = EX^2 - (EX)^2$, $EX = \lambda$ (viz příklad 5)

$$EX^2 = \sum_{x=0}^{\infty} x^2 p(x) = \sum_{x=0}^{\infty} x^2 e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\left[x(x-1)+x\right]\frac{\lambda^x}{x!}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x(x-1)(x-2)!} + \underbrace{e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!}}_{=EX=\lambda}$$

Příklad

$$EX^2 = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} + \lambda = e^{-\lambda} \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$= \lambda^2 \sum_{y=0}^{\infty} \underbrace{e^{-\lambda} \frac{\lambda^y}{(y)!}}_{=1=\sum\limits_{y \in M} p(y)} + \lambda = \lambda^2 + \lambda,$$

takže

$$\boxed{DX} = \lambda^2 + \lambda - \lambda^2 = \boxed{\lambda}.$$

Příklad

Příklad 13 (Rozptyl Rovnoměrného rozdělení)

Mějme náhodnou veličinu $X \sim Rs(a,b)$, $a < b$, $a,b \in \mathbb{R}$ *s hustotou*

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{jinak.} \end{cases}$$

Vypočtěte rozptyl.

$$EX^2 = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3}$$

$$= \frac{1}{b-a} \frac{(b-a)(b^2+ab+a^2)}{3} = \frac{b^2+ab+a^2}{3}.$$

$$\boxed{DX} = EX^2 - E^2X = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2} \right)^2 = \boxed{\frac{(b-a)^2}{12}}.$$

Příklad 14 (Rozptyl normálního (Gaussova) rozdělení)

Mějme náhodnou veličinu s normálním rozdělením $X \sim N(\mu, \sigma^2)$ *s hustotou*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

$$DX = E(X-EX)^2 = \int\limits_{-\infty}^{\infty} (x-EX)^2 f(x) dx = \int\limits_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}\sigma} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Položíme-li $y = \frac{x-\mu}{\sigma}$, tj. $x - \mu = \sigma y$ a $dx = \sigma dy$, potom

$$\begin{aligned} DX &= \int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}\sigma} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \tfrac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{y^2 e^{-\frac{1}{2}y^2}}_{\text{sudá funkce}} dy \\ &= 2 \int_0^{\infty} \tfrac{\sigma^2}{\sqrt{2\pi}} y^2 e^{-\frac{1}{2}y^2} dy. \end{aligned}$$

Příklad

Položme $\frac{1}{2}y^2 = t$, tj. $y = \sqrt{2t}$ a $ydy = dt$. Pak

$$\boxed{DX} = \sigma^2 \int_0^\infty \frac{2}{\sqrt{2\pi}} \sqrt{2t} e^{-t} dt = \sigma^2 \frac{2}{\sqrt{\pi}} \int_0^\infty t^{\frac{3}{2}-1} e^{-t} dt = \boxed{\sigma^2},$$

protože

$$\int_0^\infty t^{\frac{3}{2}-1} e^{-t} = \Gamma\left(\tfrac{3}{2}\right) = \Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}.$$

Čebyševova nerovnost

Věta 8 (Čebyševova nerovnost, (Chebyshev's inequality))

Nechť X *je náhodná veličina s konečným druhým momentem. Potom pro libovolné* $\varepsilon > 0$ *platí*

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

Příklad 15

Biatlonista střílí nezávisle na sobě do terče, přičemž pravděpodobnost zásahu při každém výstřelu je $2/3$. Odhadněte pravděpodobnost, že ze 300 pokusů bude mít 185 až 215 zásahů.

X ... počet zásahů, $X \sim Bi(300, 2/3)$
$EX = 300 \cdot 2/3 = 200$, $DX = 300 \cdot 2/3 \cdot 1/3 = 66,66$

Odhad

$$\begin{aligned}
P(185 \leq X \leq 215) &= P(185 - 200 \leq X - 200 \leq 215 - 200) \\
&= P(-15 \leq X - EX \leq 15) = P(|X - EX| \leq 15) \\
&= 1 - P(|X - EX| \geq 16) \geq 1 - \frac{66,66}{16^2} = \boxed{\mathbf{0,7396}}
\end{aligned}$$

Přesně

$$\begin{aligned}
P(185 \leq X \leq 215) = \binom{300}{185}\left(\frac{2}{3}\right)^{185}\left(\frac{1}{3}\right)^{115} + \binom{300}{186}\left(\frac{2}{3}\right)^{186}\left(\frac{1}{3}\right)^{114} + \dots \\
\dots + \binom{300}{215}\left(\frac{2}{3}\right)^{225}\left(\frac{1}{3}\right)^{75} = \boxed{\mathbf{0,9694}}
\end{aligned}$$

Kovariance a korelační koeficient

Definice 9

Kovariancí (**Covariance**) dvou náhodných veličin X a Y nazýváme číslo

$$C(X,Y) = E(X - EX)(Y - EY)$$

číslo

$$R(X,Y) = \frac{C(X,Y)}{\sqrt{DXDY}}$$

nazýváme **korelační koeficient** (**Correlation coefficient**).

Kovariance a korelační koeficient

Věta 10

Nechť náhodné veličiny X a Y mají sdruženou distribuční funkci $F(x,y)$. Pak

$$C(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-EX)(y-EY)dF(x,y)$$

- *Nechť náhodné veličiny jsou diskrétního typu, tj. $(X,Y)' \sim (M, p(x,y))$, pak platí*
$$C(X,Y) = \sum_{(x,y)\in M}(x-EX)(y-EY)p(x,y)$$
- *Nechť náhodné veličiny jsou absolutně spojitého typu, tj. $(X,Y)' \sim f(x,y)$, pak platí*
$$C(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-EX)(y-EY)f(x,y)dxdy$$

Kovariance a korelační koeficient

Věta 11 (Vlastnosti kovariance a korelace)

Nechť X *a* Y *jsou náhodné veličiny,* $a_1, a_2, b_1, b_2 \in \mathbb{R}$*. Potom*

- $C(X,X) = DX$ *a* $R(X,X) = 1$.
- $C(X,Y) = C(Y,X)$ *a* $R(X,Y) = R(Y,X)$.
- $\boxed{C(X,Y) = E(XY) - (EX)(EY)}$.
- *Jsou–li náhodné veličiny* X *a* Y *nezávislé, pak* $C(X,Y) = R(X,Y) = 0$.
- $|C(X,Y)| \leq \sqrt{DXDY}$ *a* $|R(X,Y)| \leq 1$.
- $C(a_1 + a_2X, b_1 + b_2Y) = a_2b_2C(X,Y)$

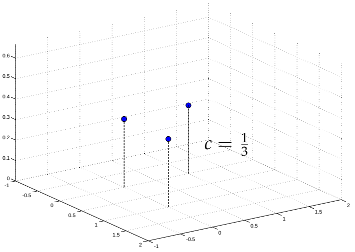
  $R(a_1 + a_2X, b_1 + b_2Y) = R(X,Y)sign(a_2b_2)$, *je–li* $a_2 \neq 0$ *a* $b_2 \neq 0$.
- $D(X+Y) = DX + DY + 2C(X,Y)$.
- $R(X,Y) = 1 \Leftrightarrow$ *existují konstanty* a *a* $b > 0$ *takové, že* $P(Y = a + bX) = 1$

  $R(X,Y) = -1 \Leftrightarrow$ *existují konstanty* a *a* $b < 0$ *takové, že* $P(Y = a + bX) = 1$

Příklad

Příklad 16

Náhodný vektor (X, Y) *má rovnoměrné diskrétní rozdělení na množině* $G = \{[0,0];[1,0];[0,1]\}$. *Vypočtěte* $C(X,Y)$ *a* $R(X,Y)$.



$$p(x,y) = \begin{cases} \frac{1}{3} & \text{pro } (x,y) \in G \\ 0 & \text{jinak} \end{cases}$$

$X \backslash Y$	0	1	$p_X(x)$
0	1/3	1/3	2/3
1	1/3	0	1/3
$p_Y(y)$	2/3	1/3	1

Příklad

$EX = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3} = EY$

$E(XY) = \sum_{x=0}^{1} \sum_{y=0}^{1} x \cdot y \cdot p(x,y) = 0 \cdot 0 \cdot \frac{1}{3} + 0 \cdot 1 \cdot \frac{1}{3} + 1 \cdot 0 \cdot \frac{1}{3} + 1 \cdot 1 \cdot 0 = 0$

$\boxed{C(X,Y)} = E(XY) - (EX)(EY) = 0 - \frac{1}{3}\frac{1}{3} = \boxed{-\frac{1}{9}}$

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

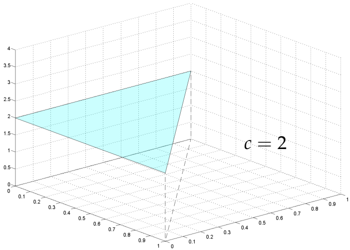
$DX = E(X^2) - (EX)^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9} = DY$

$\boxed{R(X,Y)} = \frac{C(X,Y)}{\sqrt{DX\,DY}} = \frac{\frac{-1}{9}}{\frac{2}{9}} = \boxed{-\frac{1}{2}}$

Příklad

Příklad 17

Náhodný vektor (X,Y) *má rovnoměrné spojité rozdělení na množině* $G = \{(x,y) \in \langle 0;1\rangle \times \langle 0;1\rangle; x+y \leq 1\}$. *Vypočtěte* $C(X,Y)$ *a* $R(X,Y)$.



$$f(x,y) = \begin{cases} 2 & \text{pro } (x,y) \in G \\ 0 & \text{jinak} \end{cases}$$

$$f_X(x) = \int_0^{1-x} 2\,dy = 2(1-x)$$

$$f_Y(y) = \int_0^{1-y} 2\,dx = 2\,(1-y)$$

$$EX = \int_0^1 x \cdot 2\,(1-x)\,dx = 2\int_0^1 (x - x^2)\,dx = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{3} = EY$$

Příklad

$$
\begin{aligned}
E(XY) &= \int_0^1 \int_0^{1-x} xy \cdot 2\,dy\,dx = \int_0^1 x[y^2]_0^{1-x}\,dx \\
&= \int_0^1 x(1-x)^2\,dx = \int_0^1 (x^3 - 2\,x^2 + x)\,dx \\
&= \left[\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}\right]_0^1 = \frac{1}{12}
\end{aligned}
$$

$\boxed{C(X,Y)} = E(XY) - (EX)(EY) = \frac{1}{12} - \frac{1}{3}\frac{1}{3} = \boxed{-\frac{1}{36}}$

$$E\left(X^2\right) = \int\limits_0^1 x^2 \cdot 2\,(1-x)\,dx = 2\int\limits_0^1 (x^2 - x^3)\,dx = 2\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{1}{6} = E\left(Y^2\right)$$

$DX = E(X^2) - (EX)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18} = DY$

$\boxed{R(X,Y)} = \frac{C(X,Y)}{\sqrt{DX\,DY}} = \frac{\frac{-1}{36}}{\frac{1}{18}} = \boxed{-\frac{1}{2}}$

Příklad

Příklad 18

Mějme dvourozměrný diskrétní náhodný vektor $(X,Y)' \sim (M,p)$, *kde* $M = M_X \times M_Y = \{0,1\} \times \{-1,0,1\}$

$$p(x,y) = \begin{cases} \frac{1}{3} & (x,y) \in \{(0,0),(1,-1),(1,1)\}, \\ 0 & jinak. \end{cases}$$

Vypočtěte korelační koeficient a marginální pravděpodobnostní funkce.

$X \backslash Y$	-1	0	1	$p_X(x)$
0	0	$\frac{1}{3}$	0	$\frac{1}{3}$
1	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$$EX = 0 \cdot \tfrac{1}{3} + 1 \cdot \tfrac{2}{3} = \tfrac{2}{3}$$
$$EY = (-1) \cdot \tfrac{1}{3} + 0 \cdot \tfrac{1}{3} + 1 \cdot \tfrac{1}{3} = 0$$

$$E(XY) = 0 \cdot 0 \cdot \frac{1}{3} + 1 \cdot (-1) \cdot \frac{1}{3} + 1 \cdot 1 \cdot \frac{1}{3} = -\frac{1}{3} + \frac{1}{3} = 0$$

Tj. $C(X,Y) = E(XY) - (EX)(EY) = 0 - 0 = 0 \Rightarrow X, Y$ jsou **nekorelované**.

Příklad

Avšak **nejsou nezávislé**, neboť např.

$$p(0,0) = \frac{1}{3} \;\neq\; p_X(0) \cdot p_Y(0) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

Pokud bychom si ihned všimli, že platí vztah

$$\boxed{X = Y^2},$$

lze ihned počítat

$$C(X,Y) = E(XY) - (EX)\underbrace{(EY)}_{=0} = EY^3 = \sum_{y \in M_Y} y^3\, p_Y(y)$$
$$= (-1)^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0$$

Je třeba si uvědomit, že

- **korelace** je mírou **lineárního vztahu**;
- **nulová korelace** neimplikuje nezávislost, ale značí pouze, že mezi náhodnými veličinami **neexistuje lineární vztah**, což nevylučuje možnost jiného funkčního vztahu.

Kvantily a další číselné charakteristiky

Definice 12

Nechť F je distribuční funkcí a $\alpha \in (0, 1)$. Potom funkce

$$F^{-1}(\alpha) = Q(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

se nazývá **kvantilová funkce** (**Quantile function**) a číslo

$$x_\alpha = Q(\alpha)$$

se nazývá **α-kvantilem** (**α-quantile**) rozdělení s distribuční funkcí $F(x)$.

Poznámka 13

Pokud je distribuční funkce F spojitá a rostoucí, pak kvantilová funkce F^{-1} je inverzní funkcí k distribuční funkci F. Za těchto předpokladů také platí vztah

$$P(x_{\alpha/2} < X \leq x_{1-\alpha/2}) = 1 - \alpha.$$

Kvantily a další číselné charakteristiky

Mezi často používané kvantily patří

$$
\begin{array}{rcll}
x_{0,25} & = & Q(0,25) \quad \text{se nazývá} & \textbf{dolní kvartil} \text{ (1st Quartile)} \\
x_{0,5} & = & Q(0,5) & \textbf{medián} \text{ (Median)} \\
x_{0,75} & = & Q(0,75) & \textbf{horní kvartil} \text{ (3rd Quartile)}
\end{array}
$$

V souvislosti s kvantily se také často uvádí **interkvartilové rozpětí** (Interquartile Range) $IQR = x_{0,75} - x_{0,25}$ jako charakteristika variability náhodné veličiny X. Nejznámějším kvantilem je medián $\widetilde{x} = x_{0,5}$, který udává polohu poloviny rozdělení. Další charakteristikou míry polohy je **modus** $\hat{x}$ (Mode).

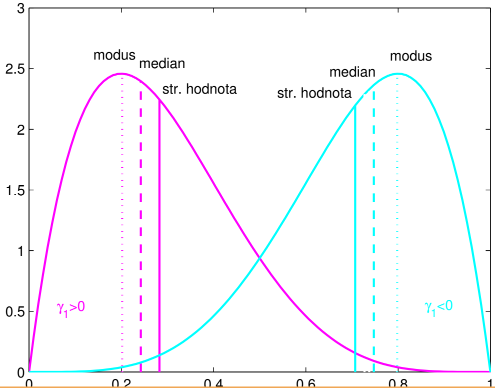
Definice 14

- Nechť $X \sim (M, p)$ je diskrétního typu, pak $\hat{x}$ značí libovolné $x_j \in M$, pro které platí $P(X = \hat{x}) \geq P(X = x_i)$, $i = 1, 2, \ldots$
- Nechť $X \sim f(x)$ je absolutně spojitého typu, pak $\hat{x}$ značí libovolné $x \in \mathbb{R}$, pro které platí $f(\hat{x}) \geq f(x)$, $x \in \mathbb{R}$.

Definice 15

Koeficient šikmosti (Skewness) je definován jako

$$\gamma_1 = \frac{\mu_3}{(DX)^{3/2}} = \frac{E(X-EX)^3}{(DX)^{3/2}}.$$



Kvantily a další číselné charakteristiky

Definice 16

Koeficient špičatosti (**Kurtosis**) je definován jako

$$\gamma_2 = \frac{\mu_4}{(DX)^2} - 3 = \frac{E(X - EX)^4}{(DX)^2} - 3.$$