Rozložení transformované diskrétní náhodné veličiny

11.1. Motivace: Máme náhodnou veličinu X s distribuční funkcí $\Phi(x)$ (resp. pravděpodobnostní funkcí $\pi(x)$ v diskrétním případě resp. hustotou $\varphi(x)$ ve spojitém případě) a borelovskou funkci $g: R \to R$. Zavedeme transformovanou náhodnou veličinu $Y = g(X)$ a hledáme její distribuční funkci $\Phi_*(y)$ (resp. pravděpodobnostní funkcí $\pi_*(y)$ v diskrétním případě resp. hustotu $\varphi_*(y)$ ve spojitém případě).

11.2. Věta: Nechť X je diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$ a g je borelovská ryze monotónní funkce, tedy v oblasti $C \subseteq R$ existuje inverzní funkce $g^{-1} = \tau$. Pak pravděpodobnostní funkce $\pi_*(y)$ transformované náhodné veličiny $Y = g(X)$ má tvar:

$$\boxed{\pi_*(y) = \begin{cases} \pi(\tau(y)) & \text{pro } y \in C \\ 0 & \text{jinak} \end{cases}}.$$

Důkaz: $\pi_*(y) = P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y)) = P(X = \tau(y)) = \pi(\tau(y))$ pro $y \in C$, $\pi_*(y) = 0$ jinak.

11.3. Příklad: $X \sim \pi(x)$, $Y = a + bX$, $\pi_*(y) = ?$
Řešení:

a) $b \neq 0$: $\pi_*(y) = P(Y = y) = P(a + bX = y) = P\left(X = \frac{y-a}{b}\right) = \pi\left(\frac{y-a}{b}\right)$

b) $b = 0$: $Y = a \Rightarrow Y \sim Dg(a)$

Rozložení transformované spojité náhodné veličiny

11.4. Věta: Nechť X je spojitá náhodná veličina s hustotou $\varphi(x)$ a g je borelovská ryze monotónní funkce se spojitou a nenulovou derivací v R, tedy v oblasti $C \subseteq R$ existuje inverzní funkce $g^{-1} = \tau$ se spojitou a nenulovou derivací. Pak hustota $\varphi_*(y)$ transformované náhodné veličiny $Y = g(X)$ má tvar:

$$\boxed{\varphi_*(y) = \begin{cases} \varphi(\tau(y))\left|\tau'(y)\right| \text{pro} y \in C \\ 0 \, \text{jinak} \end{cases}.}$$

Důkaz:

$$\Phi_*(y) = P(Y \le y) = P(g(X) \le y) = \begin{cases} P(X \le \tau(y)) = \Phi(\tau(y)) \text{ pro g rostoucí} \\ P(X \ge \tau(y)) = 1 - \Phi(\tau(y)) \text{ pro g klesající} \end{cases}$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = \left.\begin{cases} \varphi(\tau(y))\tau'(y) \text{ pro g rostoucí} \\ -\varphi(\tau(y))\tau'(y) \text{ pro g klesající} \end{cases}\right\} = \varphi(\tau(y))\left|\tau'(y)\right| \text{ pro } y \in C,\ \varphi_*(y) = 0 \text{ jinak}$$

11.5. Příklad: $X \sim \mathrm{Rs}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $Y = \mathrm{tg}\, X$, $\varphi_*(y) = ?$

Řešení:

$$\varphi(x) = \begin{cases} \frac{1}{\pi} \text{ pro } x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \\ 0 \text{ jinak} \end{cases}$$

$$\Phi_*(y) = P(Y \le y) = P(\mathrm{tg}(X) \le y) = P(X \le \mathrm{arctg}(y)) = \Phi(\mathrm{arctg}(y))$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = \varphi(\mathrm{arctg}(y))\frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}$$

Říkáme, že Y má Cauchyovo rozložení, píšeme $Y \sim t(1)$.

Nemonotónní transformace

11.6. Věta: Není-li transformační funkce g ryze monotónní, pak mezi X a Y neexistuje vzájemně jednoznačný vztah. Distribuční funkce transformované náhodné veličiny Y se vypočte podle vzorce: $\Phi_*(y) = P(X \in \Delta_1) + P(X \in \Delta_2) + \ldots$, kde $\Delta_1, \Delta_2, \ldots$ jsou ty intervaly, pro které $Y \leq y$.

11.7. Příklad: $X \sim N(0,1)$, $Y = X^2$, $\varphi_*(y) = ?$

Řešení:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\Phi_*(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \Phi(\sqrt{y}) - [1 - \Phi(\sqrt{y})] =$$

$$= 2\Phi(\sqrt{y}) - 1$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = 2\varphi(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \text{ pro } y > 0, \quad \varphi_*(y) = 0 \text{ jinak}$$

Y má χ^2 rozložení s jedním stupněm volnosti, píšeme $Y \sim \chi^2(1)$.

$X \sim \chi^2(k)$:

$$\varphi(x,k) = \begin{cases} \frac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{k/2-1} \cdot e^{-x/2} & x > 0 \\ 0 & \textit{jinak} \end{cases} \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Rozložení transformovaného náhodného vektoru

11.8. Věta (transformace náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)$ na skalární náhodnou veličinu $Y = g(X_1, \dots, X_n)$)

a) Diskrétní případ: $\mathbf{X} = (X_1, \dots, X_n) \sim \pi(x_1, \dots, x_n)$, $g: R^n \to R$ je borelovská funkce $\Rightarrow$

$$\boxed{\begin{array}{l} Y = g(x_1, \dots, x_n) \sim \pi_*(y_1, \dots, y_n) = \displaystyle\sum_{(x_1,\dots x_n)\in S(y)} \dots \sum \pi(x_1,\dots,x_n), \text{ kde} \\ S(y) = \{(x_1,\dots,x_n) \in R^n; g(x_1,\dots,x_n) = y\} \end{array}}$$

b) Spojitý případ: $\mathbf{X} = (X_1, \dots, X_n) \sim \varphi(x_1, \dots, x_n)$, $g: R^n \to R$ je borelovská funkce $\Rightarrow$

$$\boxed{\begin{array}{l} Y = g(x_1, \dots, x_n) \sim \varphi_*(y_1, \dots, y_n) = \dfrac{d}{dy} \displaystyle\int_{S(y)} \dots \int \varphi(x_1,\dots,x_n) dx_1 \dots dx_n, \text{ kde} \\ S(y) = \{(x_1,\dots,x_n) \in R^n; g(x_1,\dots,x_n) \le y\} \end{array}}$$

Věta o konvoluci

11.9. Věta (věta o konvoluci)

a) Diskrétní případ: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \pi_i(x_i)$, $i = 1, 2 \Rightarrow$

$$Y = X_1 + X_2 \sim \pi_*(y) = \sum_{x_1=-\infty}^{\infty} \pi_1(x_1)\pi_2(y - x_1) = \sum_{x_2=-\infty}^{\infty} \pi_1(y - x_2)\pi_2(x_2)$$

$\pi_*(y)$ se nazývá konvoluce funkcí $\pi_1(x_1)$, $\pi_2(x_2)$.

b) Spojitý případ: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \varphi_i(x_i)$, $i = 1, 2 \Rightarrow$

$$Y = X_1 + X_2 \sim \varphi_*(y) = \int_{-\infty}^{\infty} \varphi_1(x_1)\varphi_2(y - x_1)\,dx_1 = \int_{-\infty}^{\infty} \varphi_1(y - x_2)\varphi_2(x_2)\,dx_2$$

$\varphi_*(y)$ se nazývá konvoluce funkcí $\varphi_1(x_1)$, $\varphi_2(x_2)$.

Příklad

11.10. Příklad: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim Po(\lambda_i)$, $i = 1, 2$, $Y = X_1 + X_2$, $\pi_*(y) = ?$
Řešení:

$$\pi_i(x_i) = \begin{cases} \dfrac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i} & \text{pro } x_i = 0, 1, \ldots \\ 0 \text{ jinak} \end{cases}$$

$$\pi_*(y) = \sum_{x_1=-\infty}^{\infty} \pi_1(x_1)\pi_2(y-x_1) = \left| x_1 \geq 0, y - x_1 \geq 0 \Rightarrow 0 \leq x_1 \leq y \right| = \sum_{x_1=0}^{y} \pi_1(x_1)\pi_2(y-x_1) =$$

$$= \sum_{x_1=0}^{y} \frac{\lambda_1^{x_1}}{x_1!} e^{-\lambda_1} \frac{\lambda_2^{y-x_1}}{(y-x_1)!} e^{-\lambda_2} = e^{-(\lambda_1+\lambda_2)} \frac{1}{y!} \sum_{x_1=0}^{y} \binom{y}{x_1} \lambda_1^{x_1} \lambda_2^{y-x_1} = \frac{(\lambda_1+\lambda_2)^y}{y!} e^{-(\lambda_1+\lambda_2)} \text{ pro } y = 0, 1, \ldots$$

$\pi_*(y) = 0$ jinak.
Vidíme, že $Y \sim Po(\lambda_1 + \lambda_2)$.
Zobecnění: $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim Po(\lambda_i)$, $i = 1, 2, \ldots, n \Rightarrow$
$Y = X_1 + \ldots + X_n \sim Po(\lambda_1 + \ldots + \lambda_n)$.

Lineární transformace náhodného vektoru

11.11. Věta (lineární transformace n-rozměrného náhodného vektoru)
Nechť $\mathbf{X} = (X_1, \ldots, X_n)'$ je náhodný vektor, $\mathbf{a} = (a_1, \ldots, a_n)'$ je reálný vektor a $\mathbf{B} = (b_{ij})_{i,j=1, \ldots, n}$ je reálná čtvercová pozitivně definitní matice (tj. $\forall \mathbf{x} \in \mathbf{R}^n$ je kvadratická funkce $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$). Pak pro rozložení pravděpodobností transformovaného náhodného vektoru $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$ platí:

a) Diskrétní případ: $\boxed{\pi_*(\mathbf{y}) = \pi(\mathbf{B}^{-1}(\mathbf{y} - \mathbf{a}))}$

b) Spojitý případ: $\boxed{\varphi_*(y) = \det(B)^{-1}\varphi\left(B^{-1}\left(y - a\right)\right)}$

11.12. Věta: Nechť náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ má n-rozměrné normální rozložení $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Položme $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$. Pak $\mathbf{Y} \sim N_n(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$.

Číselné charakteristiky náhodných veličin

12.1. Motivace: Doposud jsme pracovali s funkcionálními charakteristikami náhodných veličin (např. distribuční funkce, pravděpodobnostní funkce, hustota pravděpodobnosti), které plně popisují pravděpodobnostní chování náhodné veličiny. Číselné charakteristiky vystihují pouze některé rysy tohoto chování, např. popisují polohu realizací náhodné veličiny na číselné ose či jejich proměnlivost (variabilitu). Jsou jednodušší než číselné charakteristiky, ale nesou jen částečnou informaci. Podobně jako v popisné statistice volíme vhodnou číselnou charakteristiku podle toho, jakého typu je daná náhodná veličina - zda je ordinální nebo intervalová či poměrová. Číselné charakteristiky znaků mají své teoretické protějšky v číselných charakteristikách náhodných veličin.

12.2. Definice:

Nechť X je náhodná veličina aspoň ordinálního charakteru a $\alpha \in (0,1)$. Číslo $K_\alpha(X)$ se nazývá α-kvantil náhodné veličiny X, jestliže splňuje nerovnosti:

$$\boxed{P(X \le K_\alpha(X)) \ge \alpha \wedge P(X \ge K_\alpha(X)) \ge 1-\alpha}$$

Jiné možné označení kvantilu: x_α

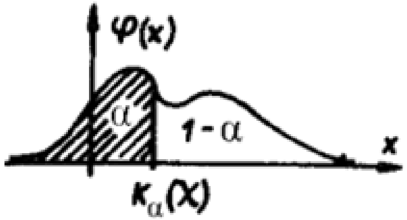
Kvantil $K_{0,50}(X)$ se nazývá medián,
$K_{0,25}(X)$ dolní kvartil,
$K_{0,75}(X)$ horní kvartil,
kvantily $K_{0,10}(X)$, ..., $K_{0,90}(X)$ jsou decily,
$K_{0,01}(X)$, ..., $K_{0,99}(X)$ jsou percentily.
Kterýkoliv α-kvantil je charakteristikou polohy číselných realizací náhodné veličiny na číselné ose.
Jako charakteristika variability slouží kvartilová odchylka $q = K_{0,75}(X) - K_{0,25}(X)$.

Kvantil spojité NV

12.3. Důsledek: (pro spojitou náhodnou veličinu)

Je-li X spojitá náhodná veličina, pak $K_\alpha(X)$ je takové číslo, pro které platí: $\alpha = \Phi(K_\alpha(X)) = \int_{-\infty}^{K_\alpha(X)} \varphi(x)dx$.

Ilustrace:



Příklad

12.4. Příklad: Nechť X ~ Ex(1). Určete medián a kvartilovou odchylku.

Řešení: $\varphi(x) = \begin{cases} e^{-x} \text{pro} x > 0 \\ 0 \text{jinak} \end{cases}$, $\Phi(x) = \begin{cases} 1 - e^{-x} \text{pro} x > 0 \\ 0 \text{jinak} \end{cases}$

$$\alpha = \Phi(K_\alpha(X)) = 1 - e^{-K_\alpha(X)} \Rightarrow K_\alpha(X) = -\ln(1-\alpha)$$

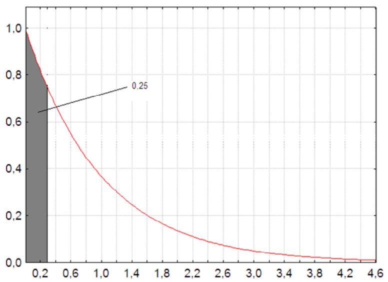
$$K_{0,50}(X) = -\ln(1-0{,}5) = -\ln\frac{1}{2} = \ln 2 = 0{,}693$$

$$K_{0,25}(X) = -\ln(1-0{,}25) = -\ln\frac{3}{4} = \ln 4 - \ln 3 = 0{,}288$$

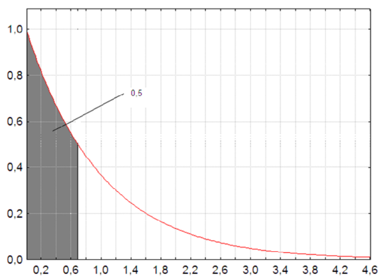
$$K_{0,75}(X) = -\ln(1-0{,}75) = -\ln\frac{1}{4} = \ln 4 = 1{,}386$$

$$q = K_{0,75}(X) - K_{0,25}(X) = 1{,}386 - 0{,}288 = 1{,}098$$

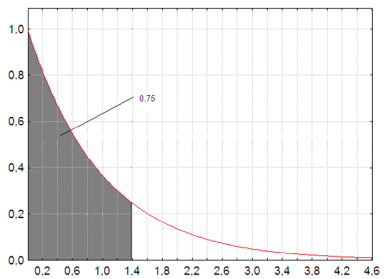
Dolní kvartil



Medián



Horní kvartil

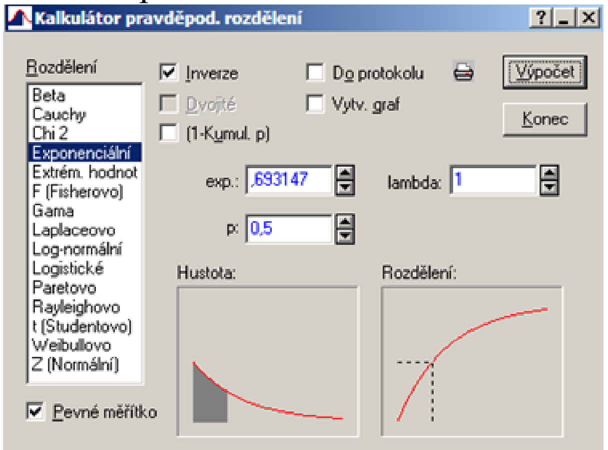


Příklad

Řešení pomocí systému STATISTICA:

První možnost: Použijeme Pravděpodobnostní kalkulátor. Vybereme Rozdělení Exponenciální. Do okénka lambda napíšeme 1, do okénka p napíšeme pro medián 0,5, pro dolní kvartil 0,25 a pro horní kvartil 0,75. V okénku exp se objeví 0,693147 pro medián, 0,287682 pro dolní kvartil a 1,386294 pro horní kvartil.

Ilustrace pro horní kvartil:



Šedá plocha pod grafem hustoty má velikost 0,5 a hodnota distribuční funkce v bodě 0,693147 je 0,5 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o třech proměnné a jednom případu.
Do dlouhého jména první proměnné napíšeme =VExpon(0,5;1). Dostaneme 0,693147.
Do dlouhého jména druhé proměnné napíšeme = VExpon(0,25;1). Dostaneme 0,287682.
Do dlouhého jména třetí proměnné napíšeme = VExpon(0,75;1). Dostaneme 1,386294.

Kvantily vybraných rozložení NV

12.5. Označení:

$X \sim N(0, 1) \Rightarrow K_\alpha(X) = u_\alpha$,
$X \sim \chi^2(n) \Rightarrow K_\alpha(X) = \chi^2_\alpha(n)$,
$X \sim t(n) \Rightarrow K_\alpha(X) = t_\alpha(n)$,
$X \sim F(n_1, n_2) \Rightarrow K_\alpha(X) = F_\alpha(n_1, n_2)$.

Tyto kvantily najdeme ve statistických tabulkách. Při jejich hledání používáme vztahy:

$u_\alpha = - u_{1-\alpha}$,
$t_\alpha(n) = - t_{1-\alpha}(n)$,
$F_\alpha(n_1, n_2) = \dfrac{1}{F_{1-\alpha}(n_2, n_1)}$.

Kvantily lze také vypočítat pomocí statistického software.

12.6. Příklad:

a) Nechť $U \sim N(0, 1)$. Najděte medián a horní a dolní kvartil.
b) Určete $\chi^2_{0,025}(25)$.
c) Určete $t_{0,99}(30)$ a $t_{0,05}(14)$.
d) Určete $F_{0,975}(5, 20)$ a $F_{0,05}(2, 10)$.

Řešení:

ad a) $u_{0,50} = 0$, $u_{0,25} = -0,67449$, $u_{0,75} = 0,67449$
ad b) $\chi^2_{0,025}(25) = 13,12$
ad c) $t_{0,99}(30) = 2,4573$, $t_{0,05}(24) = -1,7613$
ad d) $F_{0,975}(5, 20) = 3,2891$, $F_{0,05}(2, 10) = 0,05156$

Příklad

Řešení pomocí systému STATISTICA:

ad a)

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení Normální. Do okénka průměr napíšeme 0, do okénka Sm. Odch. napíšeme 1, do okénka p napíšeme 0,5 pro medián, 0,75 pro horní kvartil a 0,25 pro dolní kvartil. V okénku X se objeví 0 resp. 0,67449 resp. -0,67449.

Druhá možnost: Otevřeme nový datový soubor o třech proměnných a jednom případu.

Do dlouhého jména první resp. druhé resp. třetí proměnné napíšeme =VNormal(0,5;0;1) resp. =VNormal(0,75;0;1) resp. =VNormal(0,25;0;1). Dostaneme 0 resp. 0,67449 resp. -0,67449.

ad b)

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení Chi 2. Do okénka sv. napíšeme 25 a do okénka p napíšeme 0,025. V okénku Chi 2 se objeví 13,11972.

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.

Do dlouhého jména této proměnné napíšeme =VChi2(0,025;25). Dostaneme 13,1197.

ad c)

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení t (Studentovo). Do okénka sv. napíšeme 25 (resp. 14) a do okénka p napíšeme 0,99 (resp. 0,05). V okénku t se objeví 2,457262 (resp. -1,761310).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.

Do dlouhého jména této proměnné napíšeme =VStudent(0,99;30) (resp. VStudent(0,05;14)). Dostaneme 13,1197 (resp. -1,76131).

ad d)

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení F (Fisherovo). Do okénka sv1 napíšeme 5 (resp. 2), do okénka sv2 napíšeme 20 (resp. 10) a do okénka p napíšeme 0,975 (resp. 0,05). V okénku F se objeví 3,289056 (resp. 0,05156).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a dvou případech

Do dlouhého jména první proměnné napíšeme =VF(0,975;5;20), do dlouhého jména druhé proměnné napíšeme =VF(0,05;2;10).Dostaneme 3,2891 (resp. 0,05156).

Kvantily transformované NV

12.7. Věta:

Nechť X je spojitá náhodná veličina s distribuční funkcí $\Phi(x)$, $\alpha \in (0,1)$ a $g: R \to R$ ryze monotónní borelovská funkce. Pak pro α-kvantil transformované náhodné veličiny Y = g(X) platí:

a) Je-li g všude rostoucí funkce, pak $\boxed{K_\alpha(Y) = g(K_\alpha(X)).}$

b) Je-li g všude klesající funkce, pak $\boxed{K_\alpha(Y) = g(K_{1-\alpha}(X)).}$

Důkaz:

ad a) $\alpha = \Phi(K_\alpha(X)) = P(X \le K_\alpha(X)) = P(g(X) \le g(K_\alpha(X))) = P(Y \le g(K_\alpha(X))) = \Phi_*(g(K_\alpha(X))) \Rightarrow g(K_\alpha(X)) = K_\alpha(Y)$

ad b) $1-\alpha = \Phi(K_{1-\alpha}(X)) = P(X \le K_{1-\alpha}(X)) = P(g(X) \ge g(K_{1-\alpha}(X))) = 1 - P(Y \le g(K_{1-\alpha}(X))) = 1 - \Phi_*(g(K_{1-\alpha}(X))) \Rightarrow g(K_{1-\alpha}(X)) = K_\alpha(Y)$

12.8. Příklad:

Nechť U ~ N(0, 1). Najděte 9. decil transformované náhodné veličiny Y = 3 + 2U.

Řešení:

Funkce y = 3 + 2u je všude rostoucí funkce, tedy $K_{0,90}(Y) = 3 + 2\,u_{0,90} = 3 + 2 \times 1,28155 = 5,5631$.

Střední hodnota NV

12.9. Definice:

Nechť (Ω, $\mathcal{A}$, P) je pravděpodobnostní prostor, X náhodná veličina aspoň intervalového typu definovaná na měřitelném prostoru (Ω, $\mathcal{A}$).

a) Je-li X diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$, pak její střední hodnota (vzhledem k P) je číslo $\boxed{E(X) = \sum_{x=-\infty}^{\infty} x\pi(x)}$, pokud suma vpravo je konečná nebo absolutně konverguje. Jinak řekneme, že střední hodnota neexistuje.

b) Je-li X spojitá náhodná veličina s hustotou pravděpodobnosti $\varphi(x)$, pak její střední hodnota (vzhledem k P) je číslo $\boxed{E(X) = \int_{-\infty}^{\infty} x\varphi(x)dx}$, pokud integrál vpravo je konečný nebo absolutně konverguje. Jinak řekneme, že střední hodnota neexistuje.

(Střední hodnota je číslo, které charakterizuje polohu realizací náhodné veličiny na číselné ose spřihlédnutím k jejich pravděpodobnostem. V diskrétním případě představuje střední hodnota těžiště soustavy hmotných bodů, jejichž hmotnost je popsána pravděpodobnostní funkcí $\pi(x)$ a ve spojitém případě je střední hodnota těžištěm hmotné přímky, na níž je rozprostření hmoty popsáno hustotou pravděpodobnosti $\varphi(x)$. Střední hodnota je teoretickým protějškem váženého aritmetického průměru.)

Příklad

12.10. Příklad a):

Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte její střední hodnotu.

Řešení:

$$\pi(x) = \begin{cases} \frac{1}{6} \text{ pro } x = 1, \ldots, 6 \\ 0 \text{ jinak} \end{cases}, \quad E(X) = \sum_{x=1}^{6} x\pi(x) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3{,}5.$$

12.10. Příklad b):

Rozložení náhodné veličiny X je dáno hustotou φ(x) = 2x+2 na (−1, 0) a nulovou jinde. Vypočtěte její střední hodnotu.

Řešení

$$E(X) = \int_R x\varphi(x)\,dx = \int_{-1}^{0} x(2x+2)\,dx = \left[2\frac{x^3}{3} + x^2\right]_{-1}^{0} = \frac{2}{3} - 1 = -\frac{1}{3}$$

Střední hodnota transformované NV

12.11. Věta:

a) Diskrétní případ: Nechť X je diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$ (resp. $(X_1, \ldots, X_n)$ je diskrétní náhodný vektor s pravděpodobnostní funkcí $\pi(x_1, \ldots, x_n)$). Nechť $g : R \mapsto R$ je borelovská funkce, $Y = g(X)$ je transformovaná náhodná veličina (resp. $g : R^n \mapsto R$ je borelovská funkce, $Y = g(X_1, \ldots, X_n)$ je transformovaná náhodná veličina). Pak $\boxed{E(Y) = \sum_{x=-\infty}^{\infty} g(x)\pi(x)}$, pokud součet vpravo je konečný nebo absolutně konvergentní (resp. $E(Y) = \sum_{x_1=-\infty}^{\infty} \ldots \sum_{x_n=-\infty}^{\infty} g(x_1, \ldots, x_n)\pi(x_1, \ldots, x_n)$, pokud součet vpravo je konečný nebo absolutně konvergentní).

a) Spojitý případ: Nechť X je spojitá náhodná veličina s hustotou $\varphi(x)$ (resp. $(X_1, \ldots, X_n)$ je spojitý náhodný vektor s hustotou $\varphi(x_1, \ldots, x_n)$). Nechť $g : R \mapsto R$ je borelovská funkce, $Y = g(X)$ je transformovaná náhodná veličina (resp. $g : R^n \mapsto R$ je borelovská funkce, $Y = g(X_1, \ldots, X_n)$ je transformovaná náhodná veličina). Pak $\boxed{E(Y) = \int_{-\infty}^{\infty} g(x)\varphi(x)dx,}$ pokud integrál vpravo je konečný nebo absolutně konvergentní (resp.
$E(Y) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n)\varphi(x_1, \ldots, x_n)dx_1 \ldots dx_n$, pokud integrál vpravo je konečný nebo absolutně konvergentní).

Příklad

12.12. Příklad:

Nechť X ~ Ex(λ), Y = $e^{-\gamma X}$, kde γ > 0 je konstanta. Vypočtěte E(Y).

Řešení:

$$\varphi(x)=\begin{cases}\lambda e^{-\lambda x} & \text{pro } x>0\\ 0 & \text{jinak}\end{cases},\quad E(Y)=\int_0^{\infty} e^{-\gamma x}\lambda e^{-\lambda x}\,dx=\frac{\lambda}{\lambda+\gamma}.$$

Rozptyl NV

12.13. Definice:

Nechť (Ω, $\mathcal{A}$, P) je pravděpodobnostní prostor, X náhodná veličina aspoň intervalového typu definovaná na měřitelném prostoru (Ω, $\mathcal{A}$), která má střední hodnotu E(X). Rozptylem náhodné veličiny X rozumíme číslo $\boxed{D(X) = E([X-E(X)]^2),}$ pokud střední hodnota vpravo existuje. Číslo $\sqrt{D(X)}$ se nazývá směrodatná odchylka.

(Rozptyl je číslo, které charakterizuje proměnlivost realizací náhodné veličiny kolem její střední hodnoty s přihlédnutím k jejich pravděpodobnostem. Je teoretickým protějškem váženého rozptylu. Je vhodnější počítat rozptyl podle vzorce $D(X) = E(X^2) - [E(X)]^2$, jak bude ukázáno později.)

12.14. Důsledek:

V diskrétním případě je rozptyl dán vzorcem $\boxed{D(X) = \sum_{x=-\infty}^{\infty} [x - E(X)]^2 \pi(x) = \sum_{x=-\infty}^{\infty} x^2 \pi(x) - [E(X)]^2}$

a ve spojitém případě vzorcem $\boxed{D(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 \varphi(x)\,dx = \int_{-\infty}^{\infty} x^2 \varphi(x)\,dx - [E(X)]^2}$

(pokud suma či integrál vpravo absolutně konvergují).

Centrovaná a standardizovaná NV

12.15. Definice:
Transformovaná náhodná veličina X– E(X) se nazývá centrovaná náhodná veličina.
Transformovaná náhodná veličina $\frac{X - E(X)}{\sqrt{D(X)}}$ se nazývá standardizovaná náhodná veličina.

12.16. Příklad: Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte její rozptyl.

Řešení: $\pi(x) = \begin{cases} \frac{1}{6} \text{ pro } x = 1, \ldots, 6 \\ 0 \text{ jinak} \end{cases}$, E(X) = 3,5 (viz př. 12.10.), $D(X) = \sum_{x=1}^{6} x^2 \frac{1}{6} - 3{,}5^2 = \ldots = \frac{35}{12} = 2{,}92$.

Kovariance a korelace NV

12.17. Definice: Kovariancí náhodných veličin X_1, X_2, které mají střední hodnoty $E(X_1)$, $E(X_2)$, rozumíme číslo $\boxed{C(X_1, X_2) = E([X_1 - E(X_1)]\,[X_2 - E(X_2)])}$ (pokud střední hodnoty vpravo existují).

Kovariance je číslo, které charakterizuje proměnlivost realizací náhodných veličin X_1, X_2 kolem jejich středních hodnot s přihlédnutím k jejich pravděpodobnostem. Je-li kovariance kladná (záporná), pak to svědčí o existenci jistého stupně přímé (nepřímé) lineární závislosti mezi realizacemi náhodných veličin X_1, X_2. Je-li kovariance nulová, pak říkáme, že náhodné veličiny X_1, X_2 jsou nekorelované a znamená to, že mezi jejich realizacemi není žádný lineární vztah. Pozor – z nekorelovanosti nevyplývá stochastická nezávislost, zatímco ze stochastické nezávislosti plyne nekorelovanost. Kovariance je teoretickým protějškem vážené kovariance. Je vhodnější počítat kovarianci podle vzorce

$$\boxed{C(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2).}$$

Koeficientem korelace náhodných veličin X_1, X_2 rozumíme číslo

$$R(X_1, X_2) = \begin{cases} E\left(\dfrac{X_1 - E(X_1)}{\sqrt{D(X_1)}} \cdot \dfrac{X_2 - E(X_2)}{\sqrt{D(X_2)}}\right) & \text{pro } \sqrt{D(X_1)}\sqrt{D(X_2)} > 0 \\ 0 & \text{jinak} \end{cases}$$

, pokud střední hodnoty vpravo existují.

Koeficient korelace je číslo, které charakterizuje těsnost lineární závislosti realizací náhodných veličin X_1, X_2. Čím bližší je 1, tím těsnější je přímá lineární závislost, čím bližší je -1, tím těsnější je nepřímá lineární závislost. Je vhodnější počítat koeficient korelace podle vzorce $\boxed{R(X_1, X_2) = \dfrac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}}$

Kovariance NV

12.18. Důsledek:

V diskrétním případě je kovariance dána vzorcem

$$C(X_1, X_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_2)]\, \pi(x_1, x_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 \pi(x_1, x_2) - E(X_1)\, E(X_2)$$

a ve spojitém případě vzorcem

$$C(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_1)] \varphi(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \varphi(x_1, x_2) dx_1 dx_2 - E(X_1)\, E(X_2)$$

Příklad

12.19. Příklad:
Náhodná veličina X udává příjem manžela (v tisících dolarů) a náhodná veličina Y příjem manželky (v tisících dolarů. Je známa simultánní pravděpodobnostní funkce $\pi(x,y)$ diskrétního náhodného vektoru (X,Y): $\pi(10,10) = 0{,}2$, $\pi(10,20) = 0{,}04$, $\pi(10,30) = 0{,}01$, $\pi(10,40) = 0$, $\pi(20,10) = 0{,}1$, $\pi(20,20) = 0{,}36$, $\pi(20,30) = 0{,}09$, $\pi(20,40) = 0$, $\pi(30,10) = 0$, $\pi(30,20) = 0{,}05$, $\pi(30,30) = 0{,}1$, $\pi(30,40) = 0$, $\pi(40,10) = 0$, $\pi(40,20) = 0$, $\pi(40,30) = 0$, $\pi(40,40) = 0{,}05$, $\pi(x,y) = 0$ jinak. Vypočtěte koeficient korelace příjmů manžela a manželky.
Řešení: Náhodná veličina X i náhodná veličina Y nabývají hodnot 10, 20, 30, 40. Sestavíme kontingenční tabulku:

X	Y				
	10	20	30	40	$\pi_1(x)$
10	0,20	0,04	0,01	0,00	0,25
20	0,10	0,36	0,09	0,00	0,55
30	0,00	0,05	0,10	0,00	0,15
40	0,00	0,00	0,00	0,05	0,05
$\pi_2(y)$	0,30	0,45	0,20	0,05	1,00

Spočteme
$E(X) = 10.0{,}25+20.0{,}55+30.0{,}15+40.0{,}05 = 20$, $E(Y) = 10.0{,}30+20.0{,}45+30.0{,}20+40.0{,}05 = 20$,
$D(X) = 10^2.0{,}25+20^2.0{,}55+30^2.0{,}15+40^2.0{,}05 - 20^2 = 60$, $D(Y) = 10^2.0{,}30+20^2.0{,}45+30^2.0{,}20+40^2.0{,}05 - 20^2 = 70$,
$C(X,Y) = 10.10.0{,}20 + 10.20.0{,}04 + \ldots\ 40.40.0{,}05 - 20.20 = 49$,
$R(X,Y) = 49/\sqrt{60}\sqrt{70} = 0{,}76$.

Příklad

Postup ve STATISTICE:

Vytvoříme nový datový soubor o třech proměnných X, Y, cetnost a 16 případech. Do proměnné X napíšeme 10, 10, 10, 10, 20, 20, 20, 20, 30, 30, 30, 30, 40 ,40 40, 40 , do proměnné Y 4x pod sebe 10, 20, 30, 40 a do proměnné cetnost 20, 4, 1, 0, 10, 36, 9, 0, 0, 5, 10, 0, 0, 0, 0, 5.

Statistiky - Základní statistiky/tabulky – zavedeme proměnnou vah cetnost – OK - Korelační matice – OK – 1 seznam proměnných – X, Y – OK.

	Korelace (Tabulka6) Označ. korelace jsou významné na hlad. p < ,05000 N=100 (Celé případy vynechány u ChD)			
Proměnná	Průměry	Sm.odch.	X	Y
X	20,00000	7,784989	1,000000	0,756086
Y	20,00000	8,408750	0,756086	1,000000

Střední hodnota a rozptyl vybraných typů rozložení NV

12.20. Poznámka: Uvedeme střední hodnoty a rozptyly vybraných typů diskrétních a spojitých rozložení:

a) $X \sim Dg(\mu) \Rightarrow E(X) = \mu,\ D(X) = 0$

b) $X \sim A(\vartheta) \Rightarrow E(X) = \vartheta,\ D(X) = \vartheta\,(1-\vartheta)$

c) $X \sim Bi(n, \vartheta) \Rightarrow E(X) = n\vartheta,\ D(X) = n\vartheta\,(1-\vartheta)$

d) $X \sim Ge(\vartheta) \Rightarrow E(X) = \dfrac{1-\vartheta}{\vartheta},\ D(X) = \dfrac{1-\vartheta}{\vartheta^2}$

e) $X \sim Hg(N,M,n) \Rightarrow E(X) = \dfrac{M}{N}n,\ D(X) = \dfrac{Mn}{N}\left(1-\dfrac{M}{N}\right)\dfrac{N-n}{N-1}$

f) $X \sim Rd(G) \Rightarrow E(X) = \dfrac{n-1}{2},\ D(X) = \dfrac{n^2-1}{12}$

g) $X \sim Po(\lambda) \Rightarrow E(X) = \lambda,\ D(X) = \lambda$

h) $X \sim Rs(a, b) \Rightarrow E(X) = \dfrac{a+b}{2},\ D(X) = \dfrac{(b-a)^2}{12}$

i) $X \sim Ex(\lambda) \Rightarrow E(X) = \dfrac{1}{\lambda},\ D(X) = \dfrac{1}{\lambda^2}$

j) $X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu,\ D(X) = \sigma^2$

k) $X \sim \chi^2(n) \Rightarrow E(X) = n,\ D(X) = 2n$

l) $X \sim t(n) \Rightarrow E(X) = 0$ pro $n \geq 2$, pro $n = 1$ $E(X)$ neexistuje, $D(X) = \dfrac{n}{n-2}$ pro $n \geq 3$, pro $n = 1, 2$ $D(X)$ neexistuje

m) $X \sim F(n_1, n_2) \Rightarrow E(X) = \dfrac{n_2}{n_2-2}$ pro $n_2 \geq 3$, pro $n_2 = 1, 2$ $E(X)$ neexistuje, $D(X) = \dfrac{2n_2^{\,2}(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ pro $n_2 \geq 5$, pro $n_2 = 1, 2, 3, 4$ $D(X)$ neexistuje.

Příklad

12.21. Příklad:

V sadě 15 výrobků je 5 zmetků. Náhodně vybereme 4 výrobky. Určete střední hodnotu a rozptyl náhodné vekičiny X, která udává počet zmetků, jestliže výběr provádíme

a) bez vracení,

b) s vracením.

Řešení:

ad a) $X \sim Hg(N, M, n)$, $N = 15$, $M = 5$, $n = 4$

$$E(X) = \frac{M}{N} n = \frac{5}{15} 4 = \frac{4}{3} = 1{,}\bar{3}, \quad D(X) = \frac{Mn}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1} = \frac{4}{3}\left(1 - \frac{5}{15}\right)\frac{11}{14} = \frac{44}{63} = 0{,}6984$$

ad b) $X \sim Bi(n, \vartheta)$, $n = 4$, $\vartheta = \frac{5}{15} = \frac{1}{3}$

$$E(X) = n\vartheta = 4\frac{1}{3} = 1{,}\bar{3}, \quad D(X) = n\vartheta(1-\vartheta) = \frac{4}{3}\left(1 - \frac{1}{3}\right) = \frac{8}{9} = 0{,}\bar{8}$$

Příklad

Najděte medián rozložení určeného hustotou $\varphi(x) = 1 - x/2$, $0 < x < 2$.

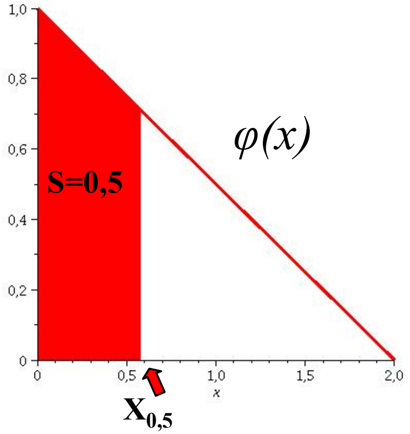
Řešení: Distribuční funkce: $F(x) = 0$ pro $x \le 0$, $F(x) = 1$ pro $x \ge 2$ a

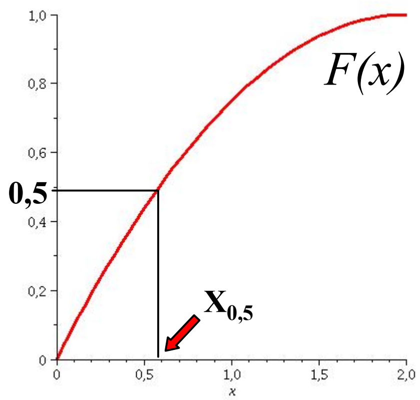
$$F(x) = \int_0^x 1 - \frac{t}{2}\,dt = x - \frac{x^2}{4} \qquad \text{pro } x \in (0, 2)$$

Medián $x_{0,5}$ je řešením rovnice $F(x) = 0,5$, tedy

$$x^2 - 4x + 2 = 0 \quad \Rightarrow x_{1,2} = \frac{4 \pm \sqrt{16-8}}{2} = 2 \pm \sqrt{2} = \begin{cases} 3,4142 \\ 0,5857 \end{cases}$$

Protože 3,4142 > 2, je hledaným řešením $\underline{\underline{x_{0,5} = 0,5857}}$.



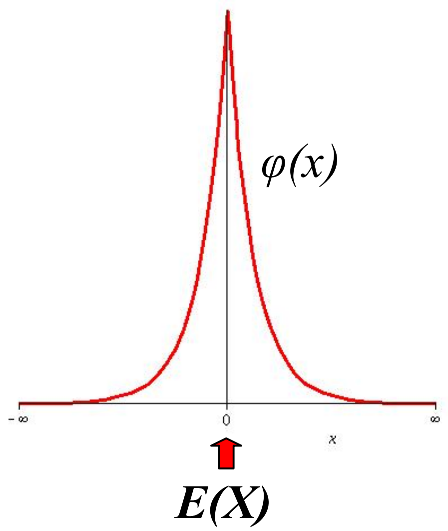


Příklad

Náhodná veličina má hustotu $\varphi(x) = a \cdot e^{-|x|} \quad x \in (-\infty, \infty)$. Určete a, střední hodnotu a rozptyl.

Řešení: $a = 1 \Big/ \left(2\int_0^{\infty} e^{-x}\, dx \right) = \dfrac{1}{2}$

$$E(X) = \int_{-\infty}^{\infty} x\varphi(x)\, dx = \int_{-\infty}^{\infty} x \frac{1}{2} e^{-|x|}\, dx = 0 \qquad D(X) = E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|}\, dx = 2$$

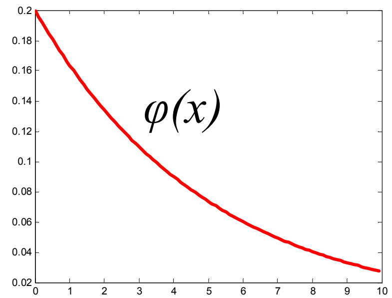


Příklad

Nechť životnost (v letech) výrobků se řídí exponenciálním rozložením s distribuční funkcí $F(x) = 1\text{-}e^{-x/5}$, $x > 0$. Tj. střední doba životnosti je 5 let. Tvar distribuční funkce znamená, že k poruše výrobku dojde s velkou pravděpodobností velmi brzy po jeho prodeji. Jakou záruční dobu stanoví výrobce, nemá-li počet reklamovaných výrobků překročit 10%?

Řešení: Náhodná veličina X udává životnost výrobku. Hledáme takové x, aby platilo

$P(X \le x) = 0{,}1$ Tedy hledáme 10% kvantil.



$$\Rightarrow \quad 0{,}1 = 1 - e^{-x/5}$$

$$\Rightarrow \quad x = -5\ln(0{,}9) = -5 \cdot (-0{,}10536) = 0{,}5268$$

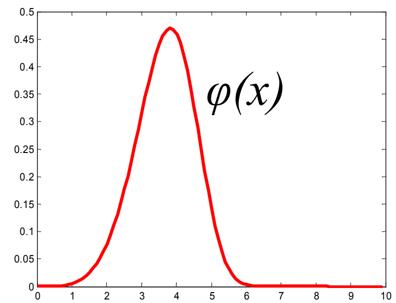
Pro splnění požadované podmínky je třeba stanovit záruční dobu na cca ½ roku.

Příklad

Nechť životnost (v letech) výrobků se řídí Weibullovým rozložením s distribuční funkcí *F(x) = 1-e*$^{-(x/4)^\wedge 5}$ *, x > 0*. Tj. střední doba životnosti je cca 3.67 let. Tvar distribuční funkce znamená, že k poruše výrobku pravděpodobně nedojde hned po jeho prodeji, ale až po nějaké době. Jakou záruční dobu stanoví výrobce, nemá-li počet reklamovaných výrobků překročit 10%?

Řešení: Náhodná veličina *X* udává životnost výrobku. Hledáme takové *x*, aby platilo

$P(X \le x) = 0{,}1$ Tedy hledáme 10% kvantil.



$$0{,}1 = 1 - e^{-(x/4)^5}$$

$$x = 4 \cdot \sqrt[5]{-\ln(0{,}9)} = 2{,}55$$

Pro splnění požadované podmínky je třeba stanovit záruční dobu na cca 2,5 roku.

Momenty, šikmost a špičatost NV

12.22. Definice:

Nechť X, X_1, X_2 jsou náhodné veličiny, $k, k_1, k_2 \in R, r, s \in N$.

a) Číslo $E([X-k]^r)$ se nazývá r**-tý moment** náhodné veličiny X kolem konstanty k. Je-li $k = 0$, jde o r**-tý počáteční moment**, je-li $k = E(X)$, jedná se o r**-tý centrální moment**.

b) Číslo $E([X_1-k_1]^r[X_2-k_2]^s)$ se nazývá $r \times s$**-tý moment** náhodných veličin X_1, X_2 kolem konstant k_1, k_2. Je-li $k_1 = k_2 = 0$, jde o $r \times s$**-tý počáteční moment**, je-li $k_1 = E(X_1), k_2 = E(X_2)$, jedná se o $r \times s$**-tý centrální moment**.

Číslo $\boxed{A_3(X) = \frac{E\left([X-E(X)]^3\right)}{\left[\sqrt{D(X)}\right]^3}}$ se nazývá šikmost náhodné veličiny X.

Číslo $\boxed{A_4(X) = \frac{E\left([X-E(X)]^4\right)}{\left[\sqrt{D(X)}\right]^4} - 3}$ se nazývá špičatost náhodné veličiny X.

Je-li $A_3(X) = 0$, jde o symetrické rozložení. Je-li $A_3(X) > 0$, jde o kladně sešikmené rozložení a je-li $A_3(X) < 0$, jde o záporně sešikmené rozložení.

Je-li $A_4(X) = 0$, jde o rozložení s normální špičatostí. Je-li $A_4(X) > 0$, jde o špičaté rozložení a je-li $A_4(X) < 0$, jde o ploché rozložení.

Vektor středních hodnot, variační a korelační matice náhodného vektoru

12.23. Definice:

Nechť $\mathbf{X} = (X_1, \ldots, X_n)'$ je náhodný vektor. Reálný vektor $E(\mathbf{X}) = (E(X_1), \ldots, E(X_n))'$ se nazývá **vektor středních hodnot**. Reálná čtvercová symetrická matice

$$var(\mathbf{X}) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \ldots & C(X_1, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ C(X_n, X_1) & C(X_n, X_2) & \ldots & D(X_n) \end{pmatrix}$$

se nazývá **varianční matice** náhodného vektoru $\mathbf{X}$ a reálná čtvercová symetrická matice

$$cor(\mathbf{X}) = \begin{pmatrix} 1 & R(X_1, X_2) & \ldots & R(X_1, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ R(X_n, X_1) & R(X_n, X_2) & \ldots & 1 \end{pmatrix}$$

se nazývá **korelační matice** náhodného vektoru $\mathbf{X}$.

Příklad

12.24. Příklad:

Pro náhodný vektor (X, Y) z příkladu 12.19. najděte vektor středních hodnot, varianční a korelační matici.

Řešení:

Bylo spočteno, že

$E(X) = 20,\ \ E(Y) = 20,$

$D(X) = 60,\ \ D(Y) = 70,$

$C(X,Y) = 49,$

$R(X,Y) = 0{,}76.$

$$E(\mathbf{X}) = \begin{pmatrix} 20 \\ 20 \end{pmatrix}, \mathrm{var}(\mathbf{X}) = \begin{pmatrix} 60 & 49 \\ 49 & 70 \end{pmatrix}, \mathrm{cor}(\mathbf{X}) = \begin{pmatrix} 1 & 0{,}76 \\ 0{,}76 & 1 \end{pmatrix}$$