4. Vícenásobné náhodné proměnné (náhodné vektory)

Často se stává, že různé objekty charakterizujeme různými veličinami. Tak například u deště lze sledovat jeho vydatnost, dobu, místo výskytu atd. Dostáváme tak celý vektor náhodných veličin (*vector of random variables*), kterému říkáme náhodný vektor.

Jednotlivé náhodné veličiny v rámci náhodného vektoru jsou buď zcela závislé, částečně závisle (*partially dependent*), nebo zcela nezávislé. Tak například měříme-li u rovnostranného trojúhelníka stranu a obsah, jsou na sobě tyto veličiny zcela závislé. Zjišťujeme-li výskyt zajíců a lišek (*hares and foxes*) v určité lokalitě, existuje jistě jakási závislost mezi jejich množstvím, nejsme však schopni určit ze znalosti jedné náhodné veličiny druhou. Konečně zjišťujeme-li rychlost větru (*wind speed*) v dané oblasti a současně roční přírůstek (*annual growth*) obyvatel tamtéž, budou složky příslušného náhodného vektoru prakticky nezávislé.

Def. 04.01: *Zobrazení* $\boldsymbol{X}: \Omega \to R^{\mathrm{n}}$ *takové, že*

$$\boldsymbol{X}^{-1}\left((-\infty, X_1\rangle \times (-\infty, X_2\rangle \times \ldots \times (-\infty, X_{\mathrm{n}}\rangle\right) \in \delta$$

nazveme náhodným vektorem nebo n-rozměrnou náhodnou veličinou. Obvykle se zapisuje $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_{\mathrm{n}}\right)$, *kde* $X_i, i = 1, \ldots, n$ *jsou náhodné veličiny.*

Def. 04.02: *Funkci F definovanou pro každý bod z* R^{n} *vztahem*

$$F\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right) = P\left[X_1 \le x_1, X_2 \le x_2, \ldots, X_{\mathrm{n}} \le x_{\mathrm{n}}\right]$$

nazýváme sdruženou distribuční funkcí (*joint distribution function*) *náhodného vektoru* ***X****.*

Z předchozích definic lze vyvodit několik dalších tvrzení:

Věta. 04.01: *Nechť F je sdružená distribuční funkce náhodného vektoru* ***X****. Pak*

- $\lim_{x_i \to -\infty} F\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right) = 0$ pro $\forall i$,
- $\lim_{x_1 \to \infty, \ldots, x_{\mathrm{n}} \to \infty} F\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right) = 1$,
- *funkce F je v každé proměnné neklesající* (*nondecreasing*),
- *funkce F je v každé proměnné spojitá zprava*.

Dále je nutno definovat, co je diskrétní a co spojitý náhodný vektor.

Def. 04.03: *Řekneme, že náhodný vektor* $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_{\mathrm{n}}\right)$ *je diskrétní, existuje-li konečná či spočetná množina n-tic*

$$\left\{\left(x_{11}, x_{21}, \ldots, x_{\mathrm{n}1}\right), \left(x_{12}, x_{22}, \ldots, x_{\mathrm{n}2}\right), \left(x_{13}, x_{23}, \ldots, x_{\mathrm{n}3}\right), \ldots\right\}$$

taková, že $P\left(x_{1i}, x_{2i}, \ldots, x_{\mathrm{n}i}\right) = P\left[X_1 = x_{1i}, X_2 = x_{2i}, \ldots, X_{\mathrm{n}} = x_{\mathrm{n}i}\right] > 0$ *pro všechna i a dále*

$$\sum_i P\left(x_{1i}, x_{2i}, \ldots, x_{\mathrm{n}i}\right) = 1 .$$

Funkci $P\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right)$ *nazveme pravděpodobnostní funkcí náhodného vektoru* ***X****.*

Def. 04.04: *Náhodný vektor* $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_{\mathrm{n}}\right)$ *má spojité rozdělení, jestliže existuje nezáporná reálná funkce* $f\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right)$ *taková, že pro každý vektor* $\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right) \in R^n$ *platí*

$$F\left(x_1, x_2, \ldots, x_{\mathrm{n}}\right) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_{\mathrm{n}}} f\left(t_1, t_2, \ldots, t_{\mathrm{n}}\right) \mathrm{d}t_1 \mathrm{d}t_2 \ldots \mathrm{d}t_{\mathrm{n}} .$$

Funkce f se nazývá hustota pravděpodobnosti náhodného vektoru ***X****, nebo sdružená hodnota pravděpodobnosti náhodných veličin* $X_1, X_2, \ldots, X_{\mathrm{n}}$.

Nakonec budeme definovat marginální náhodný vektor a marginální rozdělení.

Def. 04.05: ***Nechť*** $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_{\mathrm{n}}\right)$ *je náhodný vektor. Náhodný vektor* $\boldsymbol{Y} = \left(X_{i_1}, X_{i_2}, \ldots, X_{i_k}\right)$ *kde* $k < n$, $i_j \in \{1, 2, \ldots, n\}$, $i_u \ne i_v$ pro $u \ne v$ *nazveme marginální náhodný vektor* (*marginal random vector*). *Speciálně* X_i *je pro každé* $i \in \langle 1, \ldots, n \rangle$ *marginální náhodná veličina. Rozdělení náhodného vektoru* ***Y*** *nazýváme marginálním rozdělením.*

O distribuční funkci marginálního náhodného vektoru lze mluvit jako o marginální distribuční funkci. Analogicky se hovoří o marginální hustotě pravděpodobnosti. Pro marginální distribuční funkci platí

$$F_{\mathrm{Y}}\left(x_1, x_2, ..., x_k\right) = P\left[X_1 \le x_1, X_2 \le x_2, ..., X_k \le x_k, X_{k+1} \in R, ..., X_{\mathrm{n}} \in R\right] =$$

$$\lim_{x_{k+1} \to \infty, ..., x_{\mathrm{n}} \to \infty} F\left(x_1, x_2, ..., x_{\mathrm{n}}\right).$$

Tuto limitu formálně zapisujeme jako $F\left(x_1, x_2, ..., x_k, \infty, ..., \infty\right)$. Pro marginální hustotu spojitého rozdělení tedy platí

$$f_{\mathrm{Y}}\left(x_1, x_2, ..., x_k\right) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f_{\mathrm{X}}\left(x_1, x_2, ..., x_{\mathrm{n}}\right) \mathrm{d}x_{k+1} ... \mathrm{d}x_{\mathrm{n}},$$

zatímco v případě rozdělení diskrétního pro pravděpodobnostní funkci platí

$$P_{\mathrm{Y}}\left(x_1, x_2, ..., x_k\right) = \sum_{x_{k+1}, ..., x_{\mathrm{n}}} P_{\mathrm{X}}\left(x_1, x_2, ..., x_{\mathrm{n}}\right).$$

Konečně je třeba uvážit, že ve vícerozměrném případě může být některá složka náhodného vektoru diskrétní, zatímco jiná spojitá; takových případů může být mnoho a některé z nich jsou velmi složité.

Příklad 4.1

Diskrétní náhodný vektor $\boldsymbol{X} = \left(X_1, X_2\right)$ je charakterizován těmito pravděpodobnostmi:

X_2 \ X_1	1	3	5
2	0.15	0.30	0.35
4	0.05	0.12	0.03

Stanovte jeho distribuční funkci $F(x_1,x_2)$ a určete rozdělení marginálních náhodných veličin.

Distribuční funkci $F(x_1,x_2)$ lze opět sestavit ve formě tabulky

	$-\infty < x_1 < 1$	$1 \le x_1 < 3$	$3 \le x_1 < 5$	$5 \le x_1 < \infty$
$-\infty < x_2 < 2$	0	0	0	0
$2 < x_2 < 4$	0	0.15	0.45	0.80
$4 < x_2 < \infty$	0	0.20	0.62	1.00

Je zřejmé, že tato funkce je zprava spojitá.

Náhodná veličina X_1 bude mít rozdělení

X_1	1	3	5
$P(X_1)$	0.2	0.42	0.38

a náhodná veličina X_2

X_2	2	4
$P(X_2)$	0.8	0.2

Výpočet marginálních distribučních funkcí je již triviální (*trivial*).

Příklad 4.2

Náhodný vektor $\boldsymbol{X} = (X_1, X_2)$ se spojitým rozdělením má hustotu

$$f(x_1, x_2) = 4 \cdot x_1 x_2 \quad \text{pro} \quad 0 < x_1 < 1,\ 0 < x_2 < 1,$$
$$f(x_1, x_2) = 0 \quad \text{jinde.}$$

Je třeba určit distribuční funkci $F(x_1,x_2)$ a dále pravděpodobnost $P[X_1 < 2X_2]$.

Distribuční funkci určíme následujícím způsobem:

$$F(x_1, x_2) = 0 \quad \text{pro} \quad x_1 \le 0 \quad \text{nebo} \quad x_2 \le 0,$$
$$F(x_1, x_2) = \int_0^{x_1}\int_0^{x_2} 4uv\,\mathrm{d}u\mathrm{d}v = x_1^2 x_2^2 \quad \text{pro} \quad 0 < x_1, x_2 < 1,$$
$$F(x_1, x_2) = \int_0^{x_1}\int_0^{1} 4uv\,\mathrm{d}u\mathrm{d}v = x_1^2 \quad \text{pro} \quad 0 < x_1 < 1,\ x_2 \ge 1,$$
$$F(x_1, x_2) = \int_0^{1}\int_0^{x_2} 4uv\,\mathrm{d}u\mathrm{d}v = x_2^2 \quad \text{pro} \quad x_1 \ge 1,\ 0 < x_2 < 1,$$
$$F(x_1, x_2) = \int_0^{1}\int_0^{1} 4uv\,\mathrm{d}u\mathrm{d}v = 1 \quad \text{pro} \quad x_1 \ge 1,\ x_2 \ge 1,$$

a požadovanou pravděpodobnost obdržíme jako

$$P[X_1 < 2X_2] = \int_0^1 \mathrm{d}u \int_{\frac{u}{2}}^1 4uv\,\mathrm{d}v = \int_0^1 \left(2u - \frac{u^3}{2}\right)\mathrm{d}u = \frac{7}{8}.$$

Podmíněná rozdělení

Nechť X a Y jsou náhodné veličiny, přičemž $P[X = t] > 0$ pro nějaké $t \in R$. Pak má smysl pravděpodobnost $P[Y \le y \mid X = t]$, což je podmíněná pravděpodobnost, že dojde k jevu $[Y \le y]$ za podmínky, že nastane jev $[X = t]$. Označme tuto pravděpodobnost $F(y \mid t)$ a pokládejme ji za funkci proměnné y. Pak můžeme o funkci $F(y \mid t)$ hovořit jako o podmíněné distribuční funkci náhodné veličiny Y za předpokladu, že $X = t$.

Co když ale $P[X = t] = 0$, jak je tomu vždy v případě spojitých rozdělení? Označme A_α jev, kdy náhodná veličina $X \in (t-\alpha, t+\alpha)$. Zde má zřejmě význam mluvit o pravděpodobnosti

$$P[Y \le y \mid X \in (t-\alpha, t+\alpha)] = \frac{P[Y \le y \mid A_\alpha]}{P(A_\alpha)}.$$

Existuje-li limita tohoto výrazu pro $\alpha \to 0_+$ (jedná se o výraz typu 0/0), budeme ji pokládat za $P[Y \le y \mid X = t] = F(y \mid t)$. Důležité je zjistit, kdy tato limita existuje. Ukazuje se, že je tomu tak vždy, kdy $f(t) > 0$ (tedy hustota pravděpodobnosti náhodné proměnné X je větší než 0. Vše lze shrnout do následující definice:

Def. 04.06: *Nechť X, Y jsou náhodné veličiny a nechť* $P[X = x] > 0$, *resp.* $f(x) > 0$. *Pak výraz*

$$F(y \mid x) = P[Y \le y \mid X = x] \quad \textit{resp.} \quad F(y \mid x) = \int_{-\infty}^{y} \frac{f(x,\xi)}{f(x)}\mathrm{d}\xi$$

nazýváme podmíněnou (*conditioned*) *distribuční funkcí náhodné veličiny Y za podmínky, že X = x. Ve druhém případě nazýváme funkci* $F(y \mid x)$ *podmíněnou hustotou.*

S podmíněným rozdělením úzce souvisí useknuté rozdělení.

Def. 04.07: *Nechť* $\langle a,b\rangle \subset R$ *je uzavřený interval. Podmíněné rozdělení náhodné veličiny X za podmínky* $X \in \langle a,b\rangle$, *tedy*

$$F_*(x) = P\left[X \le x \mid X \in \langle a,b\rangle\right]$$

nazveme useknuté rozdělení (*trimmed distribution*) *náhodné veličiny X.*

Konečně do této problematiky patří i otázky statistické závislosti a nezávislosti. Za tím účelem definujeme

Def. 04.08: *Dvě náhodné veličiny X a Y se nazývají statisticky nezávislé* (*statistically independent*)*, jestliže pro jejich distribuční funkce platí*

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

Věta 04.02: *Dvě diskrétní náhodné veličiny X a Y jsou nezávislé právě tehdy, jestliže*

$$P\left[X = x_i \;\; a \;\; Y = y_j\right] = P\left[X = x_i\right]\cdot P\left[Y = y_j\right].$$

Dvě spojité náhodné veličiny X a Y jsou nezávislé právě tehdy, když pro jejich hustoty platí

$$f(x,y) = f_X(x)\cdot f_Y(y),$$

kde $f_X(x)$ *a* $f_Y(y)$ *jsou příslušné marginální hustoty pravděpodobnosti.*

Charakteristiky náhodných vektorů

Def. 04.09: *Nechť* $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ *je náhodný vektor. Vektor*

$$E\boldsymbol{X} = (EX_1, EX_2, ..., EX_n)$$

pak nazveme střední hodnotou náhodného vektoru $\boldsymbol{X}$.

Podobně jako střední hodnotu lze zavést do vícerozměrného případu i další charakteristiky, kterými je popsán případ jednorozměrný. Takové charakteristiky však cosi vypovídají jen o chování dílčích složek a nikoli o vztazích mezi nimi. To je postačující tehdy, jsou-li náhodné veličiny nezávislé. Pokud jsou však závislé, je výhodné znát i jejich interakci. Z těchto charakteristik je nejužívanější kovariance (*covariance*) či korelace (*correlation*) (normovaná kovariance).

Def. 04.10: *Nechť Y a Z jsou dvě náhodné veličiny. Číslo*

$$\mathrm{cov}(Y,Z) = E\left[(Y - EY)(Z - EZ)\right]$$

nazveme jejich kovariancí. Pokud je tato kovariance nulová, říkáme, že Y a Z jsou nekorelované (*uncorrelated*).

Z této definice okamžitě vyplývá tvrzení (*statement*):

Věta 04.03: *Nechť Y a Z jsou dvě náhodné veličiny. Potom*

$$\mathrm{cov}(Y,Z) = \mathrm{cov}(Z,Y),$$

$$\mathrm{cov}(Y,Z) = E(YZ) - E(Y)\cdot E(Z),$$

$$\mathrm{cov}(Y,Y) = \mathrm{var}(Y).$$

Pokud jsou náhodné veličiny Y a Z nezávislé, je $\mathrm{cov}(Y,Z) = 0$.

Def. 04.11: *Nechť* $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ *je náhodný vektor. Označme* $\mathrm{cov}(X_i, X_j) = \sigma_{ij}$.

Matici $\boldsymbol{\sigma}$ *nazveme kovarianční maticí náhodného vektoru* $\boldsymbol{X}$.

Kovarianční matice je zřejmě symetrická a hlavní diagonála obsahuje rozptyly jednotlivých složek. Jsou-li složky náhodného vektoru nezávislé, jsou všechny mimodiagonální prvky kovarianční matice nulové.

Def. 04.12: *Nechť Y a Z jsou dvě náhodné veličiny. Číslo*

$$\rho_{YZ} = \frac{\mathrm{cov}(Y,Z)}{\sqrt{\mathrm{var}(Y)\cdot \mathrm{var}(Z)}} = \frac{\sigma_{YZ}}{\sigma_Y \cdot \sigma_Z}$$

nazveme koeficientem korelace (*correlation coefficient*) *náhodných veličin Y a Z.*

Obecně lze ukázat, že $-1 \le \rho_{YZ} \le 1$, jsou-li náhodné proměnné *Y* a *Z* nezávislé, je koeficient korelace nulový a je-li roven 1 či -1, jsou lineárně závislé s kladným či záporným koeficientem.

Příklad 4.3

Pomocné napáječky v jaderné elektrárně představují důležitou část nouzových ochranných systémů. Napáječky jsou obvykle v pohotovostním stavu a aktivují se při poruchách reaktoru. Musí se proto často kontrolovat, přičemž drobné opravy se provádí na místě. Uvažujme jednu napáječku a zaveďme náhodnou veličinu *X* označující počet poruch během roku. Dále zaveďme náhodnou veličinu *Y* označující počet prohlídek za rok. Následující tabulka udává příslušné pravděpodobnostní rozdělení.

Počet poruch *X*	Počet prohlídek *Y*			Marginální pravděpodobnost *X*
	0	1	2	
0	0.06	0.05	0.03	0.14
1	0.10	0.08	0.08	0.26
2	0.12	0.12	0.03	0.27
3	0.14	0.06	0.01	0.21
4	0.08	0.04	0.00	0.12
Marginální pravděpodobnost *Y*	0.50	0.35	0.15	1.00

Tak například podmíněná pravděpodobnost

$$P[X=1 \mid Y=1] = \frac{p_{XY}(X=1, Y=1)}{p_Y(Y=1)} = \frac{0.08}{0.35} = 0.229\,.$$

Dále spočítejme podmíněné střední hodnoty $E(X \mid Y)$ (ve jmenovateli příslušné marginální hodnoty):

$$E(X \mid Y=0) = \frac{0.06}{0.5}\cdot 0 + \frac{0.10}{0.5}\cdot 1 + \frac{0.12}{0.5}\cdot 2 + \frac{0.14}{0.5}\cdot 3 + \frac{0.08}{0.5}\cdot 4 = 2.16,$$

$$E(X \mid Y=1) = \frac{0.05}{0.35}\cdot 0 + \frac{0.08}{0.35}\cdot 1 + \frac{0.12}{0.35}\cdot 2 + \frac{0.06}{0.35}\cdot 3 + \frac{0.04}{0.35}\cdot 4 = 1.886,$$

$$E(X \mid Y=2) = \frac{0.03}{0.15}\cdot 0 + \frac{0.08}{0.15}\cdot 1 + \frac{0.03}{0.15}\cdot 2 + \frac{0.01}{0.15}\cdot 3 + \frac{0.00}{0.15}\cdot 4 = 1.133.$$

Celková střední hodnota *E*(*X*) se určí jako (není to ovšem jediný způsob)

$$E(X) = \sum_{Y=0,1,2} E(X \mid Y)\, p_Y = 2.16\cdot 0.5 + 1.886\cdot 0.35 + 1.133\cdot 0.15 = 1.91.$$

Příklad 4.4

Normální rozdělení dvou spojitých náhodných veličin *X* a *Y* je dáno vztahem

$$f_{\mathrm{XY}}(x,y)=\frac{1}{2\pi\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2-2\rho\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\frac{y-\mu_{\mathrm{Y}}}{\sigma_{\mathrm{Y}}}+\left(\frac{y-\mu_{\mathrm{Y}}}{\sigma_{\mathrm{Y}}}\right)^2\right]},$$

kde ρ je činitel korelace mezi X a Y. Je třeba určit

- marginální rozložení hustoty pravděpodobnosti X a Y,
- najít podmíněnou hustotu pravděpodobnosti $f_{\mathrm{XY}}(y|x)$,
- při jakém ρ budou X a Y nezávislé.

Poznámka: V případě vektoru náhodných proměnných $\boldsymbol{X}$ je multinormální rozdělení (*multinormal distribution*) dáno vztahem

$$f_{\mathbf{X}}(\boldsymbol{x})=\frac{1}{\sqrt{(2\pi)^{\mathrm{n}}\det(\boldsymbol{C})}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})},$$

kde $\boldsymbol{\mu}$ je vektor středních hodnot vektoru $\boldsymbol{X}$ a $\boldsymbol{C}$ je jeho kovarianční matice (*covariance matrix*). Jakákoli marginální hustota pravděpodobnosti je v tomto případě rovněž normální.

- $f_{\mathrm{X}}(x)=\int_{-\infty}^{\infty}f_{\mathrm{XY}}(x,y)\,\mathrm{d}y$; po pracném výpočtu, jehož první krok spočívá v doplnění jisté části exponentu na čtverec se získá výsledek

$$f_{\mathrm{X}}(x)=\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{X}}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2}$$

a samozřejmě

$$f_{\mathrm{Y}}(y)=\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{Y}}}\cdot e^{-\frac{1}{2}\left(\frac{y-\mu_{\mathrm{Y}}}{\sigma_{\mathrm{Y}}}\right)^2},$$

- podmíněná pravděpodobnost

$$f_{\mathrm{Y|X}}(y\,|\,x)=\frac{f(x,y)}{f_{\mathrm{X}}(x)}=\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{Y}}\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2}\left(\frac{y-\mu_{\mathrm{Y}}-\rho\frac{\sigma_{\mathrm{Y}}}{\sigma_{\mathrm{X}}}(x-\mu_{\mathrm{X}})}{\sigma_{\mathrm{Y}}\sqrt{1-\rho^2}}\right)^2},$$

- náhodné veličiny X a Y jsou nezávislé, jestliže $\rho=0$; pak platí

$$f_{\mathrm{XY}}(x,y)=f_{\mathrm{X}}(x)\cdot f_{\mathrm{Y}}(y).$$

Příklad 4.5

Během výpadku výroby lze část energie dodávat z dražších lokálních jednotek (*local units*) a část nakoupit od sousedních výrobců. Pokud se má energie nakoupit od sousedních výrobců, musí se počítat s vícenáklady ve výši 1,000.000,- Kč a více. Na druhé straně, maximální hodinové náklady na výrobu v lokálních zdrojích představují rovněž 1,000.000,- Kč. Hustota pravděpodobnosti zde lze modelovat funkcí

$$f_{\mathrm{XY}}(x,y)=\frac{4}{5}\cdot\frac{x+y}{x^3},$$

kde náhodná veličina X označuje cenu nakoupené energie v 1,000.000,- Kč $(x\geq 1)$ a náhodná veličina Y vícenáklady (taktéž v milionech Kč) nutné pro provoz lokálních zdrojů $(0\leq y\leq 1)$. Za předpokladu, že výpadek trvá 1 hodinu, má se určit:

- jaká je pravděpodobnost, že náklady na lokální zdroje během následujícího výpadku budou menší jednak než 500.000,- Kč a náklady na nákup menší než 2,000.000,- Kč,

- jaké jsou marginální hustoty pravděpodobnosti,
- zda jsou obě náhodné veličiny nezávislé či nekorelované,
- je-li známo, že náklady na zakoupenou energii činí 2,000.000,- Kč, jaká je pravděpodobnost, že cena za lokální zdroje nepřesáhne 500.000,- Kč.

Nejprve určíme sdruženou kumulativní distribuční funkci.

$$F_{\mathrm{XY}}(x,y)=\frac{4}{5}\int_{u=1}^{x}\int_{v=0}^{y}\frac{u+v}{u^3}\,\mathrm{d}v\mathrm{d}u=\frac{4}{5}\cdot\left(y+\frac{y^2}{4}-\frac{y}{x}-\frac{y^2}{4x^2}\right).$$

Pak $P(X<2,Y<0.5)=F_{\mathrm{XY}}(2,0.5)=0.2375.$ Marginální hustoty pravděpodobnosti se vypočtou z výrazů:

$$f_{\mathrm{X}}(x)=\int_{y=0}^{1}f_{\mathrm{XY}}(x,y)\,\mathrm{d}y=\frac{4}{5}\left(\frac{1}{x^2}+\frac{1}{2x^3}\right),\ x\geq 1,$$

$$f_{\mathrm{Y}}(y)=\int_{x=1}^{\infty}f_{\mathrm{XY}}(x,y)\,\mathrm{d}x=\frac{4}{5}\left(1+\frac{y}{2}\right),\ 0\leq y\leq 1.$$

Dále je zřejmé, že obě náhodné veličiny X a Y jsou závislé, neboť

$$f_{\mathrm{XY}}(x,y)\neq f_{\mathrm{X}}(x)f_{\mathrm{Y}}(y).$$

Zda jsou veličiny korelované či nikoli, musíme určit ze vztahu (přitom $E(X)$ a $E(Y)$ se počítají z marginálních hustot)

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

$$\frac{4}{5}\int_{x=1}^{\infty}\int_{y=0}^{1}xy\frac{x+y}{x^3}\,\mathrm{d}y\mathrm{d}x-\frac{4}{5}\int_{x=1}^{\infty}x\left(\frac{1}{x^2}+\frac{1}{2x^3}\right)\mathrm{d}x\cdot\frac{4}{5}\int_{y=0}^{1}y\left(1+\frac{y}{2}\right)\mathrm{d}y=$$

$$\int_{x=1}^{\infty}\left(\frac{2}{5x}+\frac{4}{15x^2}\right)\mathrm{d}x-\int_{x=1}^{\infty}\left(\frac{32}{75x}+\frac{16}{75x^2}\right)\mathrm{d}x=\int_{x=1}^{\infty}\left(\frac{-2}{15x}+\frac{4}{75x^2}\right)\mathrm{d}x\to-\infty.$$

Náhodné veličiny X a Y jsou tedy korelované.

Konečně se musí spočítat podmíněná pravděpodobnost ze vztahu

$$f_{\mathrm{Y|X}}(y\,|\,x)=\frac{f(x,y)}{f_{\mathrm{X}}(x)}=\frac{\frac{4}{5}\cdot\frac{x+y}{x^3}}{\frac{4}{5}\cdot\left(\frac{1}{x^2}+\frac{1}{2x^3}\right)}=\frac{2(x+y)}{2x+1},\ 0\leq y\leq 1.$$

Odtud

$$P[Y<0.5\,|\,X=2]=\int_{y=0}^{0.5}f_{\mathrm{Y|X}}(y\,|\,x=2)\,\mathrm{d}y=$$

$$\int_{y=0}^{0.5}\frac{2(2+y)}{5}\,\mathrm{d}y=\frac{2}{5}\left[2y+\frac{y^2}{2}\right]_0^{0.5}=0.45.$$

Výběry a jejich zpracování

Každý, kdo se bude zabývat problémy matematické statistiky, se bude setkávat s řadou situací přinášejících experimentální nejistotu. Každý experiment zpravidla mívá jednu nebo více následujících vlastností:

- zákony, jimiž se experiment řídí, nejsou ještě dokonale popsány či prozkoumány,
- experiment se uskutečňuje poprvé, takže není doposud známa korektní metodika jeho provádění (nevyhovující aparatura, nedodržení určitých podmínek apod.),
- existuje tlak na provedení co nejjednodušších a nejlevnějších testů v co nejkratší době,
- zhodnocení experimentu a jeho výsledků nemusí být zcela objektivní,
- získané výsledky jsou neočekávané (*unexpected*),
- přestože experiment přináší další nejistotu, existuje tlak, aby na jeho základě bylo přijato další rozhodnutí, aniž by se tato nejistota hlouběji teoreticky analyzovala.

V takových případech může pomoci matematická statistika, jejíž metody jsou schopny na základě experimentu nalézt nejpravděpodobnější řešení. Dvě základní oblasti matematické statistiky jsou

- teorie odhadu (*estimation theory*),
- testování hypotéz (*testing of hypotheses*).

V teorii odhadu hledáme na základě experimentu co nejpřesněji hodnotu nějakého parametru (ať už bodově - nejpravděpodobnější hodnota, či intervalově - daný parametr leží s určitou pravděpodobností v určitém intervalu). Při testování hypotéz jde o odpověď typu ano - ne na položenou otázku (otázka na určitý parametr nebo rozdělení souboru - v tomto případě se jedná o testy dobré shody).

Populace a výběrový soubor

Celkové množství objektů, o nichž chceme vypovídat, se nazývá populace (*population*) nebo též základní soubor. Informace o celé populaci se však získávají velmi obtížně. Populace může být neúnosně rozsáhlá (*large*), její výzkum by byl časově náročný a nákladný, testy mají destrukční charakter, a konečně zdaleka celá populace nemusí být k dispozici. Proto obvykle ze zkoumané populace vybíráme nějakou její část (nazývanou výběrový soubor) a na základě poznatků získaných vyšetřováním tohoto souboru děláme závěry platné pro celou populaci. Přesnost těchto závěrů závisí zejména na velikosti výběrového souboru, překvapivě ne však na velikosti celé populace. Dokonce existují metody, umožňující určit minimální rozsah výběrového souboru tak, aby výsledky měly požadovanou přesnost. Na druhou stranu je nutno si uvědomit, že můžeme vycházet pouze s existujících dat, nebo že si rozsah výběrového souboru nemůžeme diktovat.

Velikost výběrového souboru však není jediným kritériem k tomu, abychom získali hodnověrné výsledky. Dalším kritériem je jeho reprezentativnost, což lze zajistit náhodným výběrem. Náhodný výběr (*random selection*) spočívá ve stejné šanci (*equal chance*) každého jedince být zahrnut (*be included*) do výběrového souboru. K tomu by ovšem bylo zapotřebí, aby existoval jakýsi úplný seznam populace (*complete list of population*) a možnost, jak nahodile z tohoto seznamu vylosovat výběrový soubor. To je však často neschůdné. Provádějí se proto dvoj- i vícestupňové (*multistep*) výběry, rozvrstvení apod.

Náhodný výběr

Předpokládejme, že základní soubor má v daném znaku nějaké rozdělení (hledáme je). Nyní zjistíme rozložení v daném znaku na výběrovém souboru (jeho vyšetření může ovšem být zatíženo chybou). Na základě získaných poznatků nyní chceme vyvozovat závěry platné pro celou populaci. Na každý prvek výběrového souboru nyní můžeme pohlížet jako na náhodnou veličinu, jejíž rozdělení je dáno pravděpodobnostními vlastnostmi celé populace. Zá-

věry pro celou populaci lze nyní vyvodit za předpokladu, že prvky výběrového souboru byly vybrány náhodně, jinými slovy, získané náhodné veličiny jsou na sobě nezávislé.

Def. 04.13: *Náhodný výběr z rozdělení $F(x)$ je vektor náhodných veličin, které jsou navzájem nezávislé a mají stejnou distribuční funkci $F(x)$.*

Řada pokusů má za výsledek celý vektor čísel či jiných údajů. Výsledkem jednoho experimentu je pak *p*-rozměrný náhodný výběr. Ten naměříme *n*-krát.. Pak hovoříme o náhodném výběru z *p*-rozměrného rozdělení. Existuje i řada složitějších experimentů, kdy se získá více náhodných výběrů se stejným či odlišným rozdělením.

Jakmile je experiment dokončen, stává se náhodný výběr jen vektorem naměřených či jinak zjištěných dat. To znamená, že už nemáme co do činění s náhodným vektorem $\boldsymbol{X}$, ale jen s jednou jeho realizací. Ta se zpravidla označuje malým písmenem.

Statistiky

Z experimentálně zjištěných dat se určují hodnoty různých ukazatelů (*descriptors*) (průměrná hodnota, maximální a minimální pozorovaná hodnota atd. Častější opakování experimentů (*repetition of experiments*) má leckdy za cíl zmenšení chyb či vyloučení hrubých omylů. Zmíněné ukazatele nazýváme statistikami (*statistics*), jde vlastně o funkce náhodných veličin $X_1,X_2,\ldots,X_n$.

Def. 04.14: Nechť $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ *je náhodný výběr. Statistikou je jakákoli měřitelná funkce náhodných veličin $X_1,X_2,\ldots,X_n$, k jejímuž určení není třeba znát konkrétní hodnoty parametrů příslušného rozdělení.*

Příkladem je součet všech hodnot $X_1,X_2,\ldots,X_n$, aritmetický průměr, maximální a minimální hodnota, rozdíl mezi maximální a minimální hodnotou a další funkce (tyto statistiky se často nazývají výběrové). Vzhledem k tomu, že jakákoli statistika je funkcí náhodných veličin, je sama rovněž náhodnou veličinou a s jako takovou s ní lze zacházet.

Odhad parametrů a stanovení distribučního modelu

Klíčovým krokem ve většině statistických metod je nalezení typu matematického modelu, tedy rozdělení, jímž se řídí základní soubor (populace). Dalším krokem je pak nalézt parametry tohoto rozdělení. Výběr rozdělení je obvykle založen na řadě kritérií, jež zahrnují teoretické a experimentální výsledky, a zejména zkušenost. Výchozí data mohou mít například formu pozorovaných hodnot určité proměnné (doba trvání poruchy, typ poruchy), veličin polí (rychlost větru, jeho směr), výsledků laboratorních testů (dielektrická pevnost izolace, nastavení relé). Příslušná informace pak musí být zahrnuta do funkce hustoty pravděpodobnosti. Tento proces může probíhat čtyřmi způsoby:

- návrh parametricky nebo numericky definovaného rozdělení,
- přizpůsobení standardních teoretických rozdělení,
- stanovení rozložení maximální entropie,
- subjektivní posouzení.

Ve většině případů je výhodné nalézt nejprve střední hodnotu a rozptyl či směrodatnou odchylku. Návrh hustoty pravděpodobnosti je pak snazší.

Tak například mnohé energetické podniky shromažďují statistická data o chování klíčových zařízení v energetických systémech a také o počasí a klimatických podmínkách v dané oblasti. Jako příklad lze uvést: doba trvání výpadků, životnosti generátorů, transformátorů a přenosových tras, rychlost a směr větru, intenzitu srážek, časový průběh zatížení, napětí na vybraných přípojnicích, toky výkonů po určitých linkách atd. Tato data lze doplnit o další údaje získané měřením v laboratořích.

Pokud se jedná o odhad parametrů rozdělení, existují tři základní metodiky:

- metoda momentů,

- metoda maximální věrohodnosti,
- Bayesovská metoda.

O všech uvedených postupech bude pojednáno později. Nyní se budeme zabývat druhým krokem, a to odhadem parametrů.

Odhad parametrů

Běžně se používají dva typy odhadů: bodový a intervalový. Jestliže odhad nějakého parametru (nebo charakteristiky) rozdělení základního souboru vyjádříme číslem, mluvíme o bodovém odhadu. Příkladem je statistika $\overline{X}$ jakožto bodový odhad střední hodnoty μ. Zde však nevíme nic o přesnosti takového odhadu. Nejsme např. schopni říci, jak velký rozdíl je mezi $\overline{X}$ a μ. Proto se leckdy preferuje tzv. intervalový odhad, který udává, v jakém intervalu se s určitou pravděpodobností může daná hodnota parametru očekávat.

Hlavní údaje popisující náhodnou veličinu jsou její průměrná hodnota a rozptyl. Obě tyto hodnoty mají úzkou souvislost s parametry rozložení, jak již bylo naznačeno dříve. I když pak nevyžadujeme skutečné rozdělení pravděpodobnosti náhodné veličiny, můžeme o něm takto získat dobrou představu.

Podívejme se tedy nejprve na statistiku $\overline{X}$. Jedná se o intuitivní odhad střední hodnoty základního souboru μ. Jde o to, jak je tento odhad přesný. Vytvoříme-li postupně několik stejně rozsáhlých výběrových souborů, jistě bude příslušná střední hodnota $\bar{x}$ kolísat a vytvářet jakési rozdělení. V praxi se ovšem zpravidla vytvoří jen jeden výběr a vlastnosti výběrového rozdělení se neurčují experimentálně, ale na základě teoretických úvah. Vychází se přitom z několika základních vět.

Věta 04.04*: Je-li náhodný výběr $X_1,X_2,\ldots,X_n$ rozsahu n vybrán ze základního souboru o střední hodnotě μ a rozptylu σ^2, pak výběrový průměr $\overline{X}$ bude mít rozdělení se stejnou střední hodnotou μ a rozptylem σ^2/n.*

Nyní jde o to, jaké rozdělení bude mít vlastně $\bar{X}$.

Věta 04.05: *Je-li náhodný výběr $X_1,X_2,\ldots,X_n$ rozsahu n vybrán z normálně rozdělené populace, má $\overline{X}$ rovněž normální rozdělení.*

Pokud tedy náhodné veličiny $X_1,X_2,\ldots,X_n$ mají rozdělení $N(\mu,\sigma^2)$, má $\bar{X}$ rozdělení $N(\mu,\sigma^2/n)$. Nejdůležitější je však nyní tzv. centrální limitní věta, která říká

Věta 04.06: *Nechť $X_1,X_2,\ldots,X_n$ je posloupnost vzájemně nezávislých náhodných veličin, které mají totéž rozdělení se střední hodnotou μ a s konečným rozptylem σ^2. Pak*

$$\lim_{n\to\infty} P\left[\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le t\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}\,\mathrm{d}u\,.$$

Věta vlastně vypovídá, že pro základní soubor, jehož rozdělení nemusí být normální, bude mít hodnota $\overline{X}$ výběru o rozumném rozsahu ($n > 30$) normální rozdělení.

Obecně nelze získat o určitém souboru exaktní informace. Ty se získávají na základě testování vzorků. Statistika spočtená na základě testování vzorků se nazývá estimátor a představuje rovněž náhodnou proměnnou. Estimátor se nazývá neovlivněný, je-li jeho očekávaná hodnota totožná se skutečnou hodnotou parametru.

Metoda momentů

V tomto případě jsou momenty vzorků využívány přímo jako celkové momenty. Střední hodnota vzorku se označuje $\bar{x}$ (což nahrazuje μ) a rozptyl vzorku s^2 (aproximuje σ^2) Tyto veličiny se určují jako

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,,$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i-1}^{n} x_i^2 - n\bar{x}^2\right).$$

V posledním vztahu se ve jmenovateli místo n užívá n - 1 z důvodu potlačení ovlivněnosti estimátorem.

Příklad 4.6

Odhad namáhání přenosových struktur. V jisté oblasti se obdržely za posledních 31 let tyto roční extrémní rychlosti větru (km/hod): 76, 92, 80, 90, 64, 77, 97, 61, 69, 60, 77, 80, 61, 61, 58, 64, 55, 64, 61, 69, 58, 74, 74, 71, 59, 70, 81, 70, 65, 70, 59. Kumulativní distribuční funkce má tvar

$$F_{\mathrm{x}}(x) = e^{-e^{-\alpha(x-u)}} \,.$$

Je třeba určit parametry α a u tohoto rozdělení.

Již dříve byly pro uvedené rozdělení odvozeny tyto vztahy:

$$\mu = u + \frac{0.557}{\alpha}, \quad \sigma = \frac{\pi}{\sqrt{6}\alpha},$$

kde μ označuje střední hodnotu a σ směrodatnou odchylku. Nahradíme-li tyto momenty hodnotami $\bar{x}$ a s odpovídajícími dané populaci, dostaneme

$$\bar{u} = \bar{x} - 0.577\frac{\sqrt{6}}{\pi}s, \quad \bar{\alpha} = \frac{\pi}{\sqrt{6}s}.$$

Podle vztahů platí:
$\bar{x}$=69.9, s = 10.7 a odtud $\bar{u}$=65.1 a $\bar{\alpha}$=0.12.

5. Funkce náhodné proměnné

5.1. Úvod

V mnoha energetických aplikacích se zajímáme o funkční závislost mezi závislými a nezávislými náhodnými proměnnými. Například chceme spočítat toky výkonů ve vedeních které jsou dány skutečnými dodávkami výkonů do sběrnic. Tyto dodávky jsou rovněž funkce dvou náhodných proměnných, tedy výroby a spotřeby. Výsledné toky jsou proto také náhodné veličiny.

Velikost zkratových proudů závisí na typu a místě zkratu. Poněvadž obě tyto veličiny jsou náhodné proměnné, je i velikost zkratového proudu náhodnou proměnnou. V této kapitole ukážeme, že rozdělení pravděpodobnosti závislé náhodné proměnné a jejich momentů lze odvodit z rozdělení základní náhodné proměnné.

5.2. Pravděpodobnostní rozdělení funkce náhodné proměnné

Bude kladen důraz zejména na problematiku spojité náhodné proměnné, neboť problém transformace diskrétní proměnné se lépe zpracovává s použitím základních principů.

Uvažujme nejprve funkci jednoduché náhodné proměnné

$$Y = g(X).$$

Problémem je nalézt rozdělení pravděpodobnosti $f_{\mathrm{y}}(Y)$, je-li známo rozdělení $f_{\mathrm{x}}(X)$. Obecně technika nalezení takové funkce sestává ze tří kroků.

- vyjádří se jev $(Y \le y)$ pomocí jevu zahrnujícího náhodnou proměnnou X,

- nalezne se $F_y(Y)$,
- tato funkce se zderivuje za účelem nalezení $f_y(Y)$ a určí se obor její platnosti.

Existuje určitá třída funkcí, kde lze nalézt řešení explicitně. Uvedená technika tří kroků však obecně platí pro jakoukoli funkci g.

Jedna třída funkcí, kde lze nalézt řešení explicitně, jsou spojité monotónní funkce. Předpokládejme nyní, že g je monotónně rostoucí funkce x, pro niž existuje jednoznačná inverzní funkce $g^{-1}(y)$. Je-li $Y = y$ a $X = x = g^{-1}(y)$ a dále

$$P(Y \le y) = P\left[X \le g^{-1}(y)\right],$$

obdržíme

$$F_y(y) = F_x\left[g^{-1}(y)\right] = \int_{-\infty}^{g^{-1}(y)} f_x(x) \cdot \mathrm{d}x$$

a odtud

$$f_y(y) = f_x\left[g^{-1}(y)\right] \cdot \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

Pokud funkce g monotónně klesá, pak jev $(Y \le y)$ odpovídá jevu $\left(X \ge g^{-1}(y)\right)$ a odtud

$$F_y(y) = 1 - F_x\left(g^{-1}(y)\right) \Rightarrow f_y(y) = -f_x\left[g^{-1}(y)\right] \cdot \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

Lze tedy shrnout, že v případě monotónních spojitých funkcí obdržíme

$$f_y(y) = f_x\left[g^{-1}(y)\right] \cdot \left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right|.$$

Příklad 5.1

Střední odběr elektrického výkonu (v kW) v domácnostech jistého sídliště o populaci P se mění podle vztahu

$$Y = 6 \cdot \ln\left(\frac{P}{50}\right) - 15 \quad \text{pro} \quad P > 1000.$$

Předpokládejme, že populace v této části města v roce 1998 může být popsána logaritmicko-normálním rozdělením s průměrem μ = 10.000 a kovariancí δ = 5%. Očekává se, že medián populace bude narůstat z uvedené populace o 10% ročně, zatímco kovariance zůstane prakticky konstantní.

Předpokládejme i nadále, že rozdělení populace zůstane logaritmicko-normální i v dalších letech. Je třeba stanovit rozložení Y (středního odběru výkonu) v roce 2008.

Řešení zahájíme stanovením parametrů rozdělení popisujícího velikost populace v letech 1998 a 2008. Pro logaritmicko-normální rozdělení platí

$$\xi \approx \delta = 0.05, \qquad \lambda = \ln\mu - 0.5\xi^2$$

a odtud pro rok 1998

$$\xi = 0.05, \quad \lambda = 9.21.$$

V roce 2008 bude opět $\delta' = 0.05$, ale $\mu' = 1.1^9 \cdot \mu = 23579$. Odtud $\lambda' = 10.07$.

Střední požadavek na výkon v roce 2008 plyne pak z rovnice odvozené v předchozím odstavci:

$$p = 50 \cdot e^{\frac{y+15}{6}}, \qquad \frac{\mathrm{d}p}{\mathrm{d}y} = \frac{50}{6} \cdot e^{\frac{y+15}{6}},$$

$$f_{\mathrm{y}}(y) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{0.05 \cdot 50 \cdot e^{\frac{y+15}{6}}} \cdot e^{-\frac{1}{2}\left(\frac{\ln 50 + \frac{y+15}{6} - 10.07}{0.05}\right)^2} \cdot \frac{50}{6} \cdot e^{\frac{y+15}{6}} = \frac{1}{0.3 \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{y-21.95}{0.3}\right)^2},$$

Rozdělení Y je tedy normální, s $\mu_{\mathrm{y}} = 21.95$ a standardní odchylkou $\sigma = 0.3$.

Problémy mohou nastat tehdy, jestliže funkce $y = g(x)$ není monotónní, ale i ty lze velmi dobře obejít. Předpokládejme tedy, že $y = g(x)$ je monotónní po částech, takže $g^{-1}(y) = x_1, x_2, \ldots, x_{\mathrm{n}}$. Pak

$$(Y = y) = \bigcup_{i=1}^{n} (X = x_i)$$

a odtud

$$f_{\mathrm{y}}(y) = \sum_{i=1}^{n} f_{\mathrm{x}}\left[g_i^{-1}(y)\right] \cdot \left|\frac{\mathrm{d}g_i^{-1}(y)}{\mathrm{d}y}\right|.$$

Příklad 5.2

Při konstrukci stožárů vvn a jejich základů hrají velkou roli přechodná živá namáhání vodičů vyvolaná klimatickými ději. Nejkritičtější zátěž může být vyvolána silným větrem a pro vodič daného průměru a rozpětí je tato zátěž dána vztahem

$$Q = 0.0467 \cdot S \cdot D \cdot L \cdot V_{\mathrm{g}}^2,$$

kde Q je zatížení v kN působící na vodič o rozpětí L (m), S je součinitel rozpětí, D je průměr vodiče v mm a V_{g} je nárazová rychlost daná jako $V_{\mathrm{g}} = 2.08 \cdot V_{\mathrm{m}} + 9.3$, přičemž V_{m} je střední rychlost větru (km/hod). Určete rozdělení Q, pokud je známo rozdělení V_{m}.

Při vyšetřování účinků větru se musí vzít v úvahu především hodnoty jeho nejvyšší rychlosti. Meteorologická měření ukazují, že roční maxima mohou být velmi dobře reprezentována rozdělením

$$f_{\mathrm{v_m}}(v) = \alpha \cdot e^{-\alpha(v-u)} \cdot e^{-e^{-\alpha(v-u)}}, \qquad v, \alpha, u > 0,$$

kde hodnoty u a $1/\alpha$ představují parametry místa a rozptylu. Dále

$$v_{\mathrm{m}} = \frac{v_{\mathrm{g}}}{2.08} - 9.3 \Rightarrow \frac{\mathrm{d}v_{\mathrm{m}}}{\mathrm{d}v_{\mathrm{g}}} = \frac{1}{2.08}$$

a odtud

$$f_{\mathrm{v_g}}(v) = \alpha' \cdot e^{-\alpha'(v-u')} \cdot e^{-e^{-\alpha'(v-u')}}, \qquad \alpha' = \frac{\alpha}{2.08}, \qquad u' = 2.08 \cdot (u + 9.3).$$

Nyní již můžeme psát:

$$Q = cV_{\mathrm{g}}^2 \ (c = 0.0476 \cdot S \cdot D \cdot L) \Rightarrow v_{\mathrm{g}} = \pm\sqrt{\frac{q}{c}} \Rightarrow \left|\frac{\mathrm{d}v_{\mathrm{g}}}{\mathrm{d}q}\right| = \frac{1}{2\sqrt{cq}}$$

a odtud

$$f_{\mathrm{Q}}(q) = \left[f_{\mathrm{v_g}}\left(\sqrt{\frac{q}{c}}\right) + f_{\mathrm{v_g}}\left(-\sqrt{\frac{q}{c}}\right)\right] \cdot \frac{1}{2\sqrt{qc}} = \frac{1}{2\sqrt{qc}} \cdot f_{\mathrm{v_g}}\left(\sqrt{\frac{q}{c}}\right) + 0 = \frac{\alpha' \cdot e^{-\alpha'\left(\sqrt{\frac{q}{c}} - u'\right)} \cdot e^{-e^{-\alpha'\left(\sqrt{\frac{q}{c}} - u'\right)}}}{2\sqrt{qc}}$$

pro $q > 0$.