Vícerozměrná náhodná veličina

Ekonometrie

Jiří Neubauer

Katedra kvantitativních metod FVL UO Brno
kancelář 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

Sdružená distribuční funkce

Zaměříme se především na popis dvourozměrných náhodných veličin (vektorů).

Definice

Nechť X a Y jsou náhodné veličiny, $\mathbf{X} = (X, Y)'$ se nazývá (dvourozměrný) **náhodný vektor**.

Definice

Sdružená distribuční funkce náhodných veličin X a Y je definována jako

$$F(\mathbf{x}) = F(x, y) = P(X \leq x, Y \leq y) \quad \text{pro } x, y \in \mathbb{R}.$$

Funkce $F(\mathbf{x})$ je někdy označována jako distribuční funkce náhodného vektoru $\mathbf{X} = (X, Y)'$.

Sdružená distribuční funkce

Věta

Mají-li X a Y sdruženou distribuční funkci $F(x, y)$, potom X a Y mají distribuční funkce

$$F_X(x) = F(x, \infty) \quad \text{a} \quad F(y) = F(\infty, y).$$

Distribuční funkce náhodných veličin X a Y se nazývají **marginální distribuční funkce**.

Diskrétní náhodný vektor

Definice

Jsou-li X a Y diskrétní náhodné veličiny, potom $\mathbf{X} = (X, Y)'$ se nazývá **diskrétní náhodný vektor**.

Definice

Je-li $(X, Y)'$ diskrétní náhodný vektor, pak funkce

$$p(x, y) = P(X = x, Y = y), \; (x, y) \in \mathbb{R}^2$$

se nazývá **sdružená pravděpodobnostní funkce** náhodných veličin X a Y.

Funkce $p(x, y)$ bývá též nazývána **pravděpodobností funkce náhodného vektoru** $(X, Y)'$.

Diskrétní náhodný vektor

Věta

Jestliže náhodný vektor $(X, Y)'$ má pravděpodobnostní funkci $p(x, y)$ a obor hodnot $M = \{(x, y);\ x \in M_x, y \in M_y\}$, pak **marginální pravděpodobnostní funkce náhodných veličin** X a Y jsou

$$p_X(x) = \sum_{y \in M_y} p(x, y), \quad x \in M_x,$$

$$p_Y(y) = \sum_{x \in M_x} p(x, y), \quad y \in M_y.$$

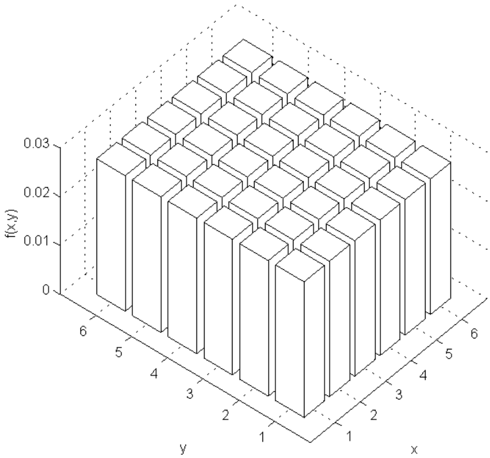
Příklad

Budeme házet dvěma kostkami. Nechť X značí číslo, které padlo na první kostce a Y číslo na kostce druhé. Určete sdruženou pravděpodobnostní funkci.

Obor hodnot je množina $M = \{(j,k);\ j = 1,2,\ldots,6,\ \ k = 1,2,\ldots,6\}$. Pro každou dvojici (j,k) dostáváme

$$p(j,k) = P(X = j, Y = k) = P(X = j)P(Y = k) = \frac{1}{36}.$$

Příklad



Obrázek: Hod 2 kostkami

Příklad

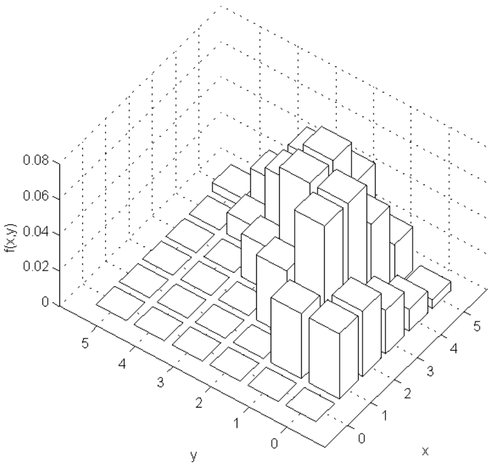
Počet příchozích telefonátů v jedné společnosti za jednu minutu lze popsat jako náhodnou veličinu s Poissonovým rozdělení se střední hodnotou 4. Předpokládejme, že volající je žena s pravděpodobností 0,5 nezávisle na dalších volajících. Označme X náhodnou veličinu udávající počet volajících žen a Y celkový počet hovorů. Určete sdruženou pravděpodobnostní funkci vektoru $(X, Y)'$.

Veličiny X a Y nejsou nezávislé, protože X je vždy menší nebo rovno Y. Nicméně můžeme využít podmíněné pravděpodobnosti. Uvědomíme-li si, že pokud $Y = k$, pak počet volající žen má binomické rozdělení s parametry k a 0,5, dostáváme

$$p(j,k) = P(X = j|Y = k) \cdot P(Y = k) = \binom{k}{j} 0{,}5^j 0{,}5^{k-j} \cdot e^{-4}\frac{4^k}{k!} =$$

$$= e^{-4}\frac{2^k}{j!(k-j)!},\ 0 \le j \le k,\ k = 0, 1, 2, \dots .$$

Příklad



Obrázek: Telefonní hovory

Spojitý náhodný vektor

Definice

Existuje-li taková nezáporná funkce $f(x, y)$, pro kterou platí

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) ds\, dt,$$

potom se funkce f se nazývá **sdružená hustota pravděpodobnosti** náhodných veličin X a Y.

Funkce $f(x, y)$ bývá také nazývána **hustota náhodného vektoru** $(X, Y)'$.

Spojitý náhodný vektor

Věta

Jsou-li X a Y spojité náhodné veličiny se sdruženou distribuční funkcí F a sdruženou hustotou f, potom

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y), \quad x, y \in \mathbb{R}.$$

Sdružená funkce hustoty pravděpodobnosti musí mít následující vlastnosti:

1. $f(x,y) \geq 0$ pro všechna $x, y \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx\, dy = 1$.

Příklad

Náhodný vektor $(X, Y)'$ má hustotu

$$f(x,y) = c(x+2y),\ 0 \leq x \leq 1,\ 0 \leq y \leq 1.$$

Určete konstantu c a sdruženou distribuční funkci $F(x,y)$.

Pro určení konstanty c využijeme druhou vlastnost $f(x,y)$. Dostáváme

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\,dy = c\int_0^1\int_0^1 (x+2y)dx\,dy = c\int_0^1 \left(\frac{1}{2}+2y\right) dy = \frac{3c}{2} = 1,$$

odkud $c = \frac{2}{3}$. Pro sdruženou distribuční funkci potom platí

$$F(x,y) = \int_0^y\int_0^x \frac{2}{3}(s+2t)ds\,dt = \int_0^y \left(\frac{x^2}{3}+\frac{4xt}{3}\right) dt =$$

$$= \left[\frac{tx^2}{3}+\frac{4xt^2}{6}\right]_{t=0}^{y} = \frac{1}{3}(x^2y+2xy^2),\ 0 \leq x \leq 1,\ 0 \leq y \leq 1.$$

Spojitý náhodný vektor

Věta

Nechť X a Y spojité náhodné veličiny se sdruženou hustotou $f(x,y)$, potom jejich marginální hustoty jsou

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad x \in \mathbb{R},$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \quad y \in \mathbb{R}.$$

Náhodný vektor

Podobně, jako jsme definovali dvourozměrný náhodný vektor $(X, Y)'$, lze zavést k-rozměrný náhodný vektor. Nechť $X_1, X_2, \dots, X_k$ jsou náhodné veličiny, vektor $(X_1, X_2, \dots, X_k)'$ nazýváme **k-rozměrným náhodným vektorem**. Sdruženou distribuční funkcí náhodných veličin $X_1, X_2, \dots, X_k$, resp. distribuční funkcí náhodného vektoru $(X_1, X_2, \dots, X_k)'$ je funkce

$$F(x_1, x_2, \dots, x_k) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_k \le x_k).$$

Jsou-li náhodné veličiny $X_1, X_2, \dots, X_k$ diskrétní, je sdružená pravděpodobností funkce definována jako

$$p(x_1, x_2, \dots, x_k) = P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k).$$

Předpokládejme nyní, že náhodné veličiny $X_1, X_2, \dots, X_k$ jsou spojité. Existuje-li taková nezáporná funkce $f(x_1, x_2, \dots x_k)$, pro kterou platí

$$F(x_1, x_2, \dots x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(x_1, x_2, \dots x_k) dx_1\, dx_2 \dots dx_k,$$

potom se funkce f se nazývá sdružená hustota náhodných veličin $X_1, X_2, \dots, X_k$.

Podmíněné rozdělení

Definice

Nechť X a Y jsou diskrétní náhodné veličiny s obory hodnot M_x a M_y, nechť $p(x,y)$ je jejich sdružená pravděpodobnostní funkce. **Podmíněná pravděpodobnostní funkce** Y za podmínky $X = x$ je definována vztahem

$$p_Y(y|x) = \frac{p(x,y)}{p_X(x)} \quad \text{pro } y \in M_y,\ p_X(x) \neq 0.$$

$$p_Y(y_k) = \sum_{j=1}^{\infty} p_Y(y_k|x_j)p_X(x_j)$$

Definice

Nechť spojité náhodné veličiny X a Y mají sdruženou hustotu $f(x,y)$. **Podmíněná funkce hustoty pravděpodobnosti** Y za podmínky $X = x$ je definována vztahem

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)} \quad y \in \mathbb{R},\ f_X(x) \neq 0.$$

Nezávislost

Definice

Řekneme, že náhodné veličiny X a Y jsou **nezávislé**, jestliže mezi sdruženou distribuční funkcí $F(x, y)$ náhodných veličin X a Y a jejich marginálními distribučními funkcemi $F_X(x)$ a $F_Y(y)$ platí vztah

$$F(x, y) = F_X(x) \cdot F_Y(y) \quad \text{pro každé } x, y \in \mathbb{R}.$$

Věta

Diskrétní náhodné veličiny X a Y jsou nezávislé, právě když pro každé $x, y \in \mathbb{R}$ platí

$$p(x, y) = p_X(x) \cdot p_Y(y),$$

spojité náhodné veličiny X a Y jsou nezávislé právě tehdy, když

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

Nezávislost

Definici nezávislosti lze zobecnit i na větší počet náhodných veličin. Řekneme, že náhodné veličiny $X_1, X_2, \dots, X_k$ jsou nezávislé, jestliže mezi sdruženou distribuční funkcí $F(x_1, x_2, \dots, x_k)$ náhodných veličin $X_1, X_2, \dots, X_k$ a jejich marginálními distribučními funkcemi $F_{X_1}(x_1)$, $F_{X_2}(x_2), \dots F_{X_k}(x_k)$ platí vztah

$$F(x_1, x_2, \dots, x_k) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdots F_{X_k}(x_k), \quad \text{pro každé } x_1, x_2, \dots x_k \in \mathbb{R}.$$

Jsou-li uvažované náhodné veličiny diskrétní, pak jsou nezávislé, právě když

$$p(x_1, x_2, \dots, x_k) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdots p_{X_k}(x_k),$$

kde $p_{X_i}(x_i)$, $i = 1, 2, \dots k$ jsou marginální pravděpodobnostní funkce náhodných veličin $X_1, X_2, \dots, X_k$. Jedná-li se o veličiny spojité, pak jsou nezávislé, právě když

$$f(x_1, x_2, \dots, x_k) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_k}(x_k),$$

kde $f_{X_i}(x_i)$, $i = 1, 2, \dots k$ jsou marginální hustoty náhodných veličin $X_1, X_2, \dots, X_k$.

Charakteristiky náhodného vektoru

Definice

Existují-li střední hodnoty $E(X)$, $E(Y)$ náhodných veličin X a Y, pak střední hodnota náhodného vektoru $\mathbf{X} = (X, Y)'$ je vektor

$$E(\mathbf{X}) = (E(X), E(Y))'.$$

Charakteristiky náhodného vektoru

Definice

Jestliže $E(X^2) < \infty$ a $E(Y^2) < \infty$, definujeme **kovarianci** náhodných veličin vztahem

$$C(X, Y) = E\{[X - E(X)][Y - E(Y)]\}.$$

Pro kovarianci platí:

- $C(Y, X) = C(X, Y)$,
- $C(X, X) = D(X)$,
- $C(X, Y) = E(XY) - E(X)E(Y)$,
- $C(a_1 + a_2X, b_1 + b_2Y) = a_2b_2C(X, Y)$, kde $a_1, a_2, b_1, b_2 \in \mathbb{R}$.

Charakteristiky náhodného vektoru

Definice

Korelační koeficient náhodných veličin X a Y je definován vztahem

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{D(X)D(Y)}}.$$

Pro korelační koeficient platí:

- $-1 \leq \rho(X, Y) \leq 1$,
- jestliže jsou X a Y nezávislé, pak $\rho(X, Y) = 0$,
- $\rho(X, Y) = 1$ právě když $Y = aX + b$, kde $a > 0$,
- $\rho(X, Y) = -1$ právě když $Y = aX + b$, kde $a < 0$.

Charakteristiky náhodného vektoru

Definice

Varianční maticí vektoru $\mathbf{X} = (X, Y)'$ rozumíme matici

$$\operatorname{var} \mathbf{X} = E\left\{[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]\right\}' = E(\mathbf{X}\mathbf{X}') - [E(\mathbf{X})][E(\mathbf{X})]'.$$

$$\operatorname{var} \mathbf{X} = \begin{pmatrix} C(X,X) & C(X,Y) \\ C(Y,X) & C(Y,Y) \end{pmatrix} = \begin{pmatrix} D(X) & C(X,Y) \\ C(X,Y) & D(Y) \end{pmatrix}.$$

Charakteristiky náhodného vektoru

Definice

Korelační maticí vektoru $\mathbf{X} = (X, Y)'$ rozumíme matici

$$\operatorname{cor} \mathbf{X} = \begin{pmatrix} 1 & \rho(X, Y) \\ \rho(X, Y) & 1 \end{pmatrix}.$$

Střední hodnota a rozptyl součtu

Věta

Nechť X a Y jsou náhodné veličiny, pak

$$E(X+Y) = E(X) + E(Y).$$

Pro spojité veličiny platí

$$\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dx\,dy \\
&= \int_{-\infty}^{\infty}xf_X(x)dx + \int_{-\infty}^{\infty}yf_Y(y)dy = E(X) + E(Y)
\end{aligned}$$

Střední hodnota a rozptyl součtu

Věta

Nechť X a Y jsou náhodné veličiny, $a, b \in \mathbb{R}$, pak

$$E(aX + bY) = aE(X) + bE(Y).$$

Střední hodnota a rozptyl součtu

Věta

Nechť X a Y jsou náhodné veličiny, pak

$$D(X+Y) = D(X) + D(Y) + 2C(X,Y).$$

$$\begin{aligned}
D(X+Y) &= E[(X+Y)^2] - [E(X+Y)]^2 \\
&= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\
&= E(X^2) - E(X^2) + E(Y) - E(Y^2) + 2(E(XY) - E(X)E(Y)) \\
&= D(X) + D(Y) + 2C(X,Y)
\end{aligned}$$

Střední hodnota a rozptyl součtu

Věta

Nechť X a Y jsou nezávislé náhodné veličiny, pak platí

- $E(XY) = E(X)E(Y)$,
- $D(X+Y) = D(X) + D(Y)$.

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dx\,dy$$
$$= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E(X)E(Y)$$

$$C(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

Střední hodnota a rozptyl součtu

Věta

Nechť X a Y jsou nezávislé náhodné veličiny, $a, b \in \mathbb{R}$, pak

$$D(aX + bY) = a^2 D(X) + b^2 D(Y).$$

Dvourozměrné normální rozdělení

Definice

Má-li náhodný vektor $\mathbf{X} = (X, Y)'$ sdruženou hustotu pravděpodobnosti

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right\}$$

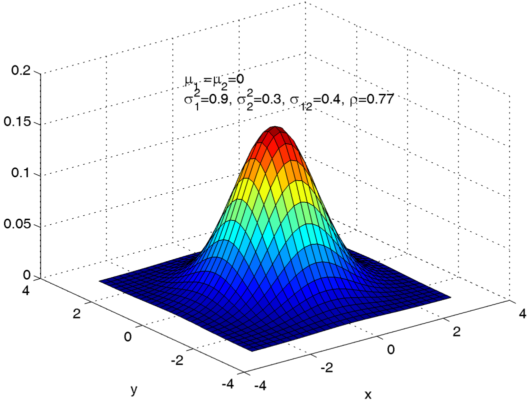
pro $x, y \in \mathbb{R}$, pak říkáme, že má **dvourozměrné normální rozdělení** s parametry $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$.

Věta

Nechť $\mathbf{X} = (X, Y)'$ má dvourozměrné normální rozdělení s parametry $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$, potom

- $X \sim N(\mu_1, \sigma_1^2)$ a Y $\sim N(\mu_2, \sigma_2^2)$,
- ρ je korelační koeficient X a Y.

Dvourozměrné normální rozdělení



Obrázek: Graf dvourozměrného normálního rozdělení

Dvourozměrné normální rozdělení

Věta

Nechť $\mathbf{X} = (X, Y)'$ má dvourozměrné normální rozdělení. Pro pevné $x \in \mathbb{R}$ platí

$$Y|X = x \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2)\right)$$

Z uvedené věty plyne, že

$$E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1).$$

Tato podmíněná střední hodnota se nazývá **regresní přímka.**