PRAVDĚPODOBNOST A STATISTIKA, PŘEDNÁŠKA 30. 3. 2020

Doporučená literatura: Likeš J., Machek J.: Počet pravděpodobosti. SNTL, Praha 1982.

- Náhodné vektory – základní pojmy (spojitý – distr. fce a hustota; diskrétní – pravd. fce). Vlastnosti distr. fce, hustoty. Příklady – dvourozměrné normální, multinomická rozdělení. Marginální distr. fce, hustota a pst. funkce.
- Nezávislost ↔ distr. fce, hustota, pst. fce jako součin marginálních;
- Číselné charakteristiky náh. vektorů (střední hodnota; kovariance, korelace jako míra lineární závislosti, vlastnosti; varianční matice)

Náhodné vektory

Nechť $X_1, \dots, X_n$ jsou náhodné veličiny definované na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. Potom uspořádanou n-tici $\boldsymbol{X} = (X_1, \dots, X_n)$ nazýváme (n-rozměrným) náhodným vektorem (opět zkratka n.v.), resp. (n-rozměrnou) náhodnou veličinou. (náhodnou veličinu X_i $(i = 1, \dots, n)$ nazýváme i-tou složkou n.v. $\boldsymbol{X}$)

Analogický jednorozměrnému případu, definujeme rozdělení n. v. $\boldsymbol{X} = (X_1, \dots, X_n)$ obvykle pomocí vícerozměrné (sdružené) distribuční funkce. Ta je definována vztahem

$$F(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n), \forall (x_1, \dots, x_n) \in R^n$$

Poznámky

- Jestliže jsou všechny složky náhodného vektoru spojité a $F(x_1, \dots, x_n)$ absolutně spojitá, potom lze rozdělení zadat také pomocí vícerozměrné (sdružené) hustot $f(x_1, \dots, x_n)$y definované vztahem

$$f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

  Zřejmě tak platí: $\forall (x_1, \dots, x_n) \in R^n \quad f(x_1, \dots, x_n) \ge 0,$

  $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_1 \dots dx_n = 1,$

  $F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n)\, dt_1 \dots dt_n.$

- Jestliže jsou všechny složky náhodného vektoru diskrétní, potom lze rozdělení zadat také pomocí vícerozměrné (sdružené) pravděpodobnostní funkce $P(X_1 = x_1, \dots, X_n = x_n)$
  (udává pravděpodobnosti všech n-tic $(x_1, \dots, x_n)$, které může náhodný vektor nabývat)
  Zřejmě tak platí

$$\sum_{(x_1, \dots, x_n)} P(X_1 = x_1, \dots, X_n = x_n) = 1$$

<u>V následující části se pro jednoduchost omezíme pouze na dvourozměrné náhodné vektory (X, Y) a jejich distribuční funkce (hustoty, pravděpodobnostní funkce).</u>

Základní vlastnosti dvourozměrné distribuční funkce jsou analogické jednorozměrnému případu, nicméně s několika podstatnými (a očekávanými) doplněními:

- $F(x, y)$ je v každé své proměnné neklesající a spojitá zprava,

- $\lim\limits_{x\to-\infty} F(x,y) = 0$, $\lim\limits_{y\to-\infty} F(x,y) = 0$ a $\lim\limits_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1$

  Jistě není překvapením, že k tomu, aby $\lim F(x,y) = 0$ stačí, aby šla libovolná složka do $-\infty$, kdežto aby $\lim F(x,y) = 1$ musí jít všechny složky současně do $+\infty$.
- Ve vícerozměrném případě je nutné nahradit vlastnost $\forall(x,y) \in R^2$ platí $0 \le F(x,y) \le 1$ následující vlastností

$$\forall(x_1,y_1),(x_2,y_2) \text{ platí } x_1 < x_2, y_1 < y_2 \ \rightarrow F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1) \ge 0$$

Využití distribuční funkce (resp. hustoty, pravděpodobnostní funkce) k výpočtům pravděpodobností je dáno následujícími vztahy:

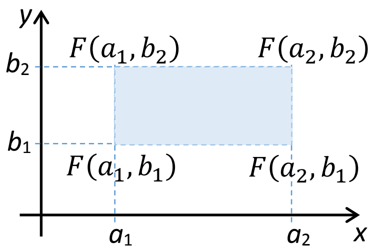
$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F(a_2,b_2) - F(a_2,b_1) - F(a_1,b_2) + F(a_1,b_1) \qquad (*)$$

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

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \sum_{\substack{(x_i,y_j) \\ a_1 < x_i \le a_2 \\ b_1 < y_j \le b_2}} P(X = x_i, Y = y_j)$$

Poznámky

- Interpretace vztahu $(*)$ je patrná z uvedeného obrázku. Pravděpodobnost $P(a_1 < X \le a_2, b_1 < Y \le b_2)$ je míra (zadaná distribuční funkcí) množiny $\{(x,y) | a_1 < x \le a_2, b_1 < y \le b_2\}$.
- **Pozor - neplatí** tedy $P(a_1 < X \le a_2, b_1 < Y \le b_2) = F(a_2,b_2) - F(a_1,b_1)$



Příklad

Dvourozměrné normální rozdělení $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ má hustotu:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}}, \text{ kde } \mu_1,\mu_2 \in R;\ \sigma_1,\sigma_2 > 0;\ |\rho| < 1.$$

Následující grafy jsou ukázkou průběhu hustoty dvourozměrného normálního rozdělení.

a) levý graf zobrazuje hustoty

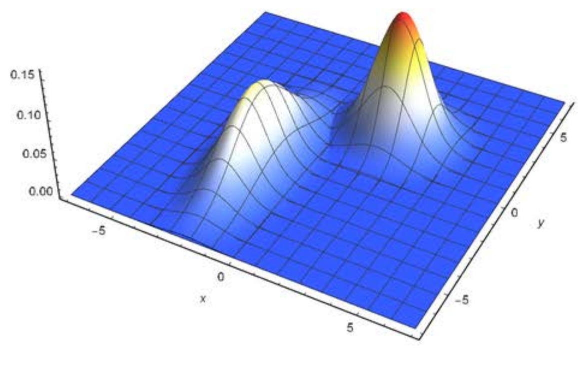
$N(\mu_1 = 2, \mu_2 = 2, \sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0)$

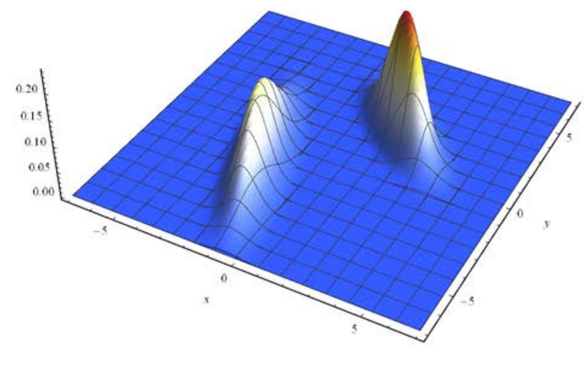
$N(\mu_1 = -2, \mu_2 = -2, \sigma_1^2 = 0{,}8, \sigma_2^2 = 2, \rho = 0)$

b) pravý graf zobrazuje hustoty

$N(\mu_1 = 2, \mu_2 = 2, \sigma_1^2 = 1, \sigma_2^2 = 1, \rho = -0{,}7)$

$N(\mu_1 = -2, \mu_2 = -2, \sigma_1^2 = 0{,}8, \sigma_2^2 = 2, \rho = -0{,}7)$





Příklad

Náhodný vektor $\boldsymbol{X} = (X, Y)$ má dvourozměrné rovnoměrné rozdělení uvnitř trojúhelníka s vrcholy o souřadnicích (0,0); (0,1); (1,1). Určete hustotu $f(x, y)$, distribuční funkci $F(x, y)$ a pravděpodobnost $P(0{,}25 < X \leq 0{,}6; 0{,}3 < Y \leq 0{,}75)$.

Řešení

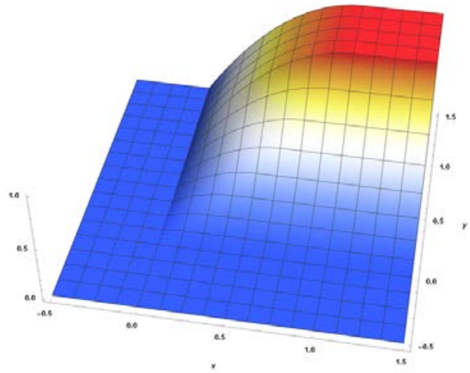
Trojúhelník se zadanými vrcholy má plochu rovnou 0,5 a vzhledem k tomu, že hustota je na něm konstantní, dostáváme:

$$f(x,y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{jinde} \end{cases}$$

Pro sdruženou distribuční funkcí $F(x, y)$ snadno dostáváme:

$$F(x,y) = \begin{cases} 0 & ,(x \leq 0) \vee (y \leq 0) \\ 2xy - x^2 & ,(0 < x < 1) \wedge (x < y < 1) \\ y^2 & ,(0 < y \leq 1) \wedge (y \leq x) \\ 2x - x^2 & ,(0 < x \leq 1) \wedge (1 \leq y) \\ 1 & ,(1 \leq x) \wedge (1 \leq y) \end{cases}$$

Průběh distribuční funkce $F(x, y)$ znázorňuje následující graf.



S ohledem na výše uvedené dostáváme pro hledanou pravděpodobnost:

$$P(0{,}25 < X \leq 0{,}6; 0{,}3 < Y \leq 0{,}75) = F(0{,}6; 0{,}75) - F(0{,}25; 0{,}75) - F(0{,}6; 0{,}3) + F(0{,}25; 0{,}3) =$$
$$= 0{,}54 - 0{,}3125 - 0{,}09 + 0{,}0875 = 0{,}225$$

Uvedenou pravděpodobnost lze v tomto jednoduchém případě spočítat také „přímo“ (bez znalosti distribuční funkce) jako objem „pod“ hustotou.

Marginální rozdělení

Je zřejmé, že každá složka náhodného vektoru je jednorozměrnou náhodnou veličinou a tudíž má své rozdělení pravděpodobnosti (tj. distribuční funkci, resp. hustotu/pravděpodobnostní funkci). Rozdělení složek nazýváme marginální rozdělení. Přirozené vzniká otázka, jak ze sdruženého rozdělení určíme marginální rozdělení. Odpověď (která je intuitivně zřejmá) dává následující:

- Je-li $F(x, y)$ sdružená distribuční funkce n. v. $\boldsymbol{X} = (X, Y)$ a označíme-li $F_X(x)$, resp. $F_Y(y)$ marginální distribuční funkci n.v. X, resp. Y, potom platí

$$F_X(x) = \lim_{y \to +\infty} F(x,y), \text{ resp. } F_Y(y) = \lim_{x \to +\infty} F(x,y).$$

- Je-li $f(x, y)$ sdružená hustota n. v. $\boldsymbol{X} = (X, Y)$ a označíme-li $f_X(x)$, resp. $f_Y(y)$ marginální hustoty n.v. X, resp. Y, potom platí

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

- Je-li $P(X = x_i, Y = y_j), i, j \in N$ sdružená pravděpodobnostní funkce n. v. $\boldsymbol{X} = (X, Y)$ a označíme-li $P_X(X = x_i)$, resp. $P_Y(Y = y_j)$ marginální pravděpodobnostní funkci n.v. X, resp. Y, potom platí

$$P_X(X = x_i) = \sum_{j=0}^{\infty} P(X = x_i, Y = y_j), \text{ resp. } P_Y(Y = y_j) = \sum_{i=0}^{\infty} P(X = x_i, Y = y_j).$$

Poznámky

- Sdružená distribuční funkce (sdružená hustota, sdružená pravděpodobnostní funkce) jednoznačně určuje marginální distribuční funkce (marginální hustoty, marginální pravděpodobnostní funkce). Obrácené tvrzení však obecně neplatí.
- Má-li náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_k)$ již dříve definované multinomické rozdělení $Multi(n, p_1, p_2, \dots, p_k)$, snadno odvodíme, že jednotlivé složky mají binomické rozdělení, tj. platí

$$X_i \sim Bi(n, p_i), \text{ tj. } E(X_i) = np_i, var(X_i) = np_i(1-p_i)$$

- Má-li náhodný vektor $\boldsymbol{X} = (X, Y)$ dvourozměrnéné normální rozdělení $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, lze ukázat, že jednotlivé složky mají jednorozměrné normální rozdělení, tj. platí

$$X \sim N(\mu_X, \sigma_X^2), \text{ tj. } E(X) = \mu_X, var(X) = \sigma_X^2, \quad \text{resp.} \quad Y \sim N(\mu_Y, \sigma_Y^2), \text{ tj. } E(Y) = \mu_Y, var(Y) = \sigma_Y^2$$

Příklad

Náhodný vektor $\boldsymbol{X} = (X, Y)$ má sdruženou hustotu $f(x, y) = 2$ pro $0 < x < y < 1$ a $f(x, y) = 0$ jinde. Určete marginální hustoty $f_X(x), f_Y(y)$ a marginální distribuční funkce $F_X(x), F_Y(y)$.

Řešení

$$f_X(x) = \int_x^1 f(x, y)\, dy = \begin{cases} 2 - 2x & , 0 < x < 1 \\ 0 & , \text{jinde} \end{cases}, \text{ resp.}$$

$$f_Y(y) = \int_0^y f(x, y)\, dx = \begin{cases} 2y & , 0 < y < 1 \\ 0 & , \text{jinde} \end{cases}.$$

Pro marginální distribuční funkce dostáváme

$$F_X(x) = \int_{-\infty}^x f_X(t)\, dt = \left(\lim_{y \to +\infty} F(x, y)\right) = \begin{cases} 0 & , x \le 0 \\ 2x - x^2 & , 0 < x < 1, \text{ resp.} \\ 1 & , 1 \le x \end{cases}$$

$$F_Y(y) = \int_{-\infty}^y f_Y(t)\, dt = \left(\lim_{x \to +\infty} F(x, y)\right) = \begin{cases} 0 & , y \le 0 \\ y^2 & , 0 < y < 1. \\ 1 & , 1 \le y \end{cases}$$

<u>Nezávislost náhodných veličin</u>

Řekneme, že náhodné veličiny $X_1, \dots, X_n$ jsou (sdruženě) nezávislé, jestliže pro sdruženou distribuční funkcí $F(x_1, \dots, x_n)$ a marginálními distribuční funkce $F_{X_i}(x_i), i = 1, \dots, n$ platí

$$\forall (x_1, \dots, x_n) \in R^n \quad F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n).$$

(tj. sdružená distribuční funkce je součinem marginálních distribučních funkcí)

Poznámky

- V případě, kdy jsou náhodné veličiny X, Y nezávislé, platí vztah

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \big(F_X(a_2) - F_X(a_1)\big)\big(F_Y(b_2) - F_Y(b_1)\big),$$

  kde $F_X(x), F_Y(y)$ jsou marginální distribuční funkce.
- Dále platí, že spojité náhodné veličiny X, Y (mající hustoty) jsou nezávislé právě tehdy, jestliže sdružená hustota je součinem marginálních hustot, tj.

$$\forall (x, y) \in R^2 \quad f(x, y) = f_X(x) \cdot f_Y(y).$$

- Analogický pro diskrétní náhodné veličiny X, Y platí, že jsou nezávislé právě tehdy, jestliže sdružená pravděpodobnostní funkce je součinem marginálních, tj.

$$\forall x_i, y_j \quad P\big(X = x_i, Y = y_j\big) = P_X(X = x_i) \cdot P_Y\big(Y = y_j\big).$$

Příklad

Náhodný vektor $\boldsymbol{X} = (X, Y)$ má sdruženou distribuční funkci $F(x, y) = (1 - e^{-x})(1 - e^{-y})$ pro $0 < x < \infty$, $0 < y < \infty$ a $F(x, y) = 0$ jinde. Rozhodněte, zda jsou veličiny X, Y nezávislé. A spočtěte $P(0{,}25 < X \le 1; 0{,}3 < Y \le 0{,}5)$.

Řešení

Pro marginální distribuční funkce snadno dostáváme

$$F_X(x) = \lim_{y\to+\infty} F(x,y) = \begin{cases} 1-e^{-x} & ,0<x<\infty \\ 0 & ,\text{jinde} \end{cases}, \text{ resp.}$$

$$F_Y(y) = \lim_{x\to+\infty} F(x,y) = \begin{cases} 1-e^{-y} & ,0<y<\infty \\ 0 & ,\text{jinde} \end{cases}.$$

Jelikož $F(x,y) = F_X(x)\cdot F_Y(y)$, jsou náhodné veličiny X, Y nezávislé a tedy

$$P(0{,}25 < X \le 1; 0{,}3 < Y \le 0{,}5) = P(0{,}25 < X \le 1)\cdot P(0{,}3 < Y \le 0{,}5) =$$
$$= \big(F_X(1) - F_X(0{,}25)\big)\big(F_Y(0{,}5) - F_Y(0{,}3)\big) = 0{,}055181$$

Alternativně lze využít (a dospět samozřejmě ke stejnému výsledku) i pomocí vztahu

$$P(0{,}25 < X \le 1; 0{,}3 < Y \le 0{,}5) = F(1; 0{,}5) - F(1; 0{,}3) - F(0{,}25; 0{,}5) + F(0{,}25; 0{,}3)$$

Příklad

N. v. $\boldsymbol{X} = (X,Y)$ má sdruženou hustotu $f(x,y) = \frac{1+xy}{4}$ pro $-1<x<1, -1<y<1$ a $f(x,y)=0$ jinde.
Rozhodněte, zda náhodné veličiny X, Y jsou nezávislé. A spočtěte $P(-0{,}3 < X \le 0{,}7; -0{,}9 < Y \le 0{,}6)$.

Řešení

Pro marginální hustoty dostáváme

$$f_X(x) = \int_{-1}^{1} f(x,y)\,dy = \begin{cases} 1/2 & ,-1<x<1 \\ 0 & ,\text{jinde} \end{cases}, \text{ resp.}$$

$$f_Y(y) = \int_{-1}^{1} f(x,y)\,dx = \begin{cases} 1/2 & ,-1<y<1 \\ 0 & ,\text{jinde} \end{cases}.$$

Jelikož $f(x,y) \neq f_X(x)\cdot f_Y(y)$, nejsou náhodné veličiny X, Y nezávislé.

Pro hledanou pravděpodobnost dostáváme

$$P(-0{,}3 < X \le 0{,}7; -0{,}9 < Y \le 0{,}6) = \int_{-0{,}3}^{0{,}7}\int_{-0{,}9}^{0{,}6} f(x,y)dy\,dx = 0{,}36375.$$

Alternativně můžeme vyřešit úlohu s použitím distribuční funkce.

<u>Základní číselné charakteristiky náhodného vektoru</u>

Vektor středních hodnot: $E(\boldsymbol{X}) = \big(E(X_1), \dots, E(X_n)\big)$

Kovariance: $cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big]$

- Po snadné úpravě dostáváme pro kovarianci vztah $cov(X,Y) = E(XY) - E(X)E(Y)$.
- Platí: - $cov(X,X) = var(X)$
  - Jestliže X, Y jsou nezávislé (tj. $E(XY) = E(X)E(Y)$) $\longrightarrow$ $cov(X,Y) = 0$
  - Jestliže $cov(X,Y) = 0$, potom náhodné veličiny X, Yjsou tzv. nekorelované.
    Pozor - v obecném případě je nekorelovanost nutná, nikoliv postačující podmínka pro nezávislosti!
  - Jestliže (X,Y) má dvourozměrné normální rozdělení a $cov(X,Y) = 0$ (tj. parametr $\rho = 0$), potom jsou náhodné veličiny X, Y nezávislé.

Korelační koeficient: $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)\cdot var(Y)}}$

- $\rho(X,Y)$ je mírou lineární závislosti mezi X a Y.
  $\rho(X,Y) = \pm 1$ právě když s pravděpodobností 1 platí $Y = aX + b$
  ($\rho(X,Y) = 1$ pro $a > 0$; $\rho(X,Y) = -1$ pro $a < 0$).

- Platí: - $\rho(X,Y) = \rho(Y,X)$
  - $|\rho(X,Y)| \leq 1$
  - $a \cdot c \neq 0 \longrightarrow \rho(aX+b, cY+d) = \begin{cases} \rho(X,Y) & ,ac > 0 \\ -\rho(X,Y) & ,ac < 0 \end{cases}$

Varianční matice: $var(\boldsymbol{X}) = \left(cov\left(X_i, X_j\right)\right)_{i,j=1}^{n}$