M3121 Pravděpodobnost a statistika I

7. Náhodné vektory

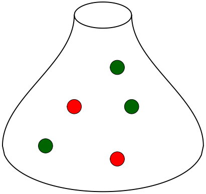
Jan Koláček (kolacek@math.muni.cz)

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

Motivační příklad

Příklad 1

V pytlíku jsou 3 zelené a 2 červené kuličky. Náhodně vybereme jednu kuličku, nevracíme ji a vybereme druhou kuličku. Popište rozdělení pravděpodobnosti tohoto pokusu. Jak se toto rozdělení změní v případě, že před druhým výběrem první kuličku vrátíme do pytlíku?



X . . . počet ●, $X \in \{0,1,2\}$
Y . . . počet ●, $Y \in \{0,1,2\}$

a) 1. kuličku **nevracíme**

$X \backslash Y$	0	1	2
0	0	0	$\frac{3}{5} \cdot \frac{2}{4}$
1	0	$2 \cdot \frac{3}{5} \cdot \frac{2}{4}$	0
2	$\frac{2}{5} \cdot \frac{1}{4}$	0	0

$\leftarrow p(x,y)$

Motivační příklad

b) 1. kuličku **vracíme**

$X \backslash Y$	0	1	2
0	0	0	$\frac{3}{5}\cdot\frac{3}{5}$
1	0	$2\cdot\frac{3}{5}\cdot\frac{2}{5}$	0
2	$\frac{2}{5}\cdot\frac{2}{5}$	0	0

$\leftarrow p(x,y)$

$$p(x,y) = \begin{cases} \frac{2!}{x!y!}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^y & \text{pro } (x,y)\in\{0,1,2\}^2,\ x+y=2 \\ 0 & \text{jinak} \end{cases}$$

$$(X,Y) \sim Mn\left(2,\frac{2}{5},\frac{3}{5}\right) \ldots \text{viz Příklad 4}$$

Náhodné vektory

Definice 1

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $\mathbf{X} = (X_1, \dots, X_n)' : \Omega \to \mathbb{R}^n$ je takové zobrazení, že pro $\forall\, \mathbf{x} \in \mathbb{R}^n$ platí

$$\{\omega \in \Omega : \mathbf{X}(\omega) \leq \mathbf{x}\} \in \mathcal{A}.$$

Pak $\mathbf{X}$ nazýváme **n-rozměrným náhodným vektorem** (random vector).

Definice 2

Nechť $\mathbf{X} = (X_1, \dots, X_n)'$ je n-rozměrný náhodný vektor definovaný na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. Potom reálnou funkci

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(\mathbf{X} \leq \mathbf{x})$$

definovanou pro každý vektor $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$ nazveme **distribuční funkcí náhodného vektoru $\mathbf{X}$**.

Značení: $[\mathbf{X} \in B] = [X_1 \in B_1, \dots, X_n \in B_n] = \bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \in B_i\}$,

kde $B = B_1 \times \cdots \times B_n \in \mathcal{B}^n$.

Distribuční funkce

Věta 3 (Vlastnosti vícerozměrné distribuční funkce)

Pro distribuční funkci náhodného vektoru $\mathbf{X}$ platí

- $F(x_1, \ldots, x_n)$ *je* **neklesající** *v každé z proměnných* $x_1, \ldots, x_n$, *při pevně daných hodnotách ostatních proměnných.*
- $F(x_1, \ldots, x_n)$ *je* **zprava spojitá** *v každé z proměnných* $x_1, \ldots, x_n$, *při pevně daných hodnotách ostatních proměnných.*
- *Pro* $\forall\ i = 1, \ldots, n$ *je* $\lim\limits_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0$, *tj. vícerozměrná distribuční funkce je nulová, jestliže alespoň jedna z proměnných jde k* $-\infty$.
- $\lim\limits_{\substack{x_1 \to \infty \\ \vdots \\ x_n \to \infty}} F(x_1, \ldots, x_n) = 1$, *tj. vícerozměrná distribuční funkce je rovna jedné, jestliže všechny proměnné jdou k* ∞.

Diskrétní náhodné vektory

Definice 4

Řekneme, že náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)'$ je **diskrétního typu**, jestliže existuje nejvýše spočetná množina $M \subset \mathbb{R}^n$ taková, že $P_{\mathbf{X}}(M) = 1$. Funkci $p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$ nazýváme **pravděpodobnostní funkcí** náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)'$ a M nazýváme **oborem hodnot** náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)'$.

Značení: Fakt, že jde o diskrétní náhodný vektor budeme značit $\mathbf{X} \sim (M, p)$.

Věta 5 (Vlastnosti pravděpodobnostní funkce)

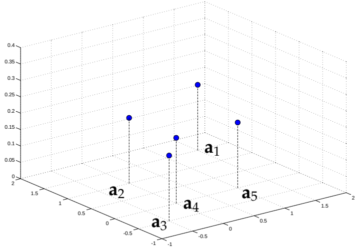
Nechť $\mathbf{X} \sim (M, p)$. *Pak*

- $p(\mathbf{x}) = p(x_1, \dots, x_n) \geq 0$ *pro* $\forall\, \mathbf{x} \in \mathbb{R}^n$ *a* $\sum_{\mathbf{x} \in M} p(\mathbf{x}) = 1$.
- $P(\mathbf{X} \in B) = \sum_{\mathbf{x} \in M \cap B} p(\mathbf{x})$ *pro libovolné* $B \in \mathcal{B}^n$.
- $F(\mathbf{x}) = \sum_{\mathbf{t} \in M, \mathbf{t} \leq \mathbf{x}} p(\mathbf{t})$ *pro* $\forall\, \mathbf{x} \in \mathbb{R}^n$.

Příklady

Příklad 2 (Rovnoměrné diskrétní rozdělení)

Nechť $G = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ *je konečná množina,* $\mathbf{a}_i \in \mathbb{R}^2$, $i = 1, \ldots, n$.
Pravděpodobnost je pro všechny body stejná, (X, Y) *značí souřadnice bodů v* $\mathbb{R}^2$*.*



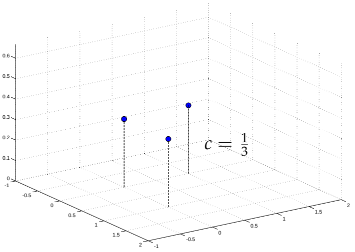
$$p(x,y) = \begin{cases} \frac{1}{n} & \text{pro } (x,y) \in G \\ 0 & \text{jinak} \end{cases}$$

Náhodný vektor (X, Y) má **rovnoměrné diskrétní** rozdělení na množině G.
Značíme $\boxed{(X, Y) \sim Rd_2(G)}$.

Příklady

Příklad 3

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



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

X \ Y	0	1
0	1/3	1/3
1	1/3	0

Příklady

Příklad 4 (Multinomické rozdělení)

Uvažujme pokus, který může mít n disjunktních výsledků $A_1, \ldots, A_n$. Nechť $\theta_i = P(A_i)$ pro $i = 1, \ldots, n$, přičemž $\sum\limits_{i=1}^{n} \theta_i = 1$. Tento pokus budeme k-krát <u>nezávisle</u> opakovat.
X_i ... **počet nastoupení jevu A_i v provedených k pokusech**.
Nalezněte rozdělení pravděpodobností náhodného vektoru $\mathbf{X} = (X_1, \ldots, X_n)'$.

$X_i \in \{0, 1, \ldots, k\}$ pro $i = 1, \ldots, n$ a pravděpodobnostní funkce je rovna

$$p(\mathbf{x}) = \begin{cases} \binom{k}{x_1}\binom{k-x1}{x_2}\cdots\binom{k-x_1-\cdots-x_{n-1}}{x_n}\,\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_n^{x_n} & \\ \quad = \frac{k!}{x_1!\cdots x_n!}\,\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_n^{x_n} & \text{pro } x_i \in \{0,1,\ldots,k\},\ \sum\limits_{i=1}^{n} x_i = k \\ 0 & \text{jinak} \end{cases}$$

Značíme $\boxed{\mathbf{X} \sim Mn(k, \theta_1, \ldots, \theta_n)}$.

Příklad

Příklad 5

V České republice má 45 % populace krevní skupinu A, 20 % populace skupinu B, 30 % populace skupinu 0 a zbytek má krevní skupinu AB. Náhodně vybereme 7 krevních vzorků. Jaká je pravděpodobnost, že 4 vzorky budou skupina A, 1 vzorek bude skupina B, 2 vzorky budou skupina 0 a skupina AB se nebude ve výběru vyskytovat?

X_i . . . počet vzorků i-té krevní skupiny, $X_i \in \{0, 1, \ldots, 7\}$ pro $i = 1, \ldots, 4$.
$\mathbf{X} = (X_1, X_2, X_3, X_4) \sim Mn(7; 0,45; 0,2; 0,3, 0,05)$

$$p(4, 1, 2, 0) = \binom{7}{4}\binom{3}{1}\binom{2}{2}\binom{0}{0} 0,45^4 \cdot 0,2^1 \cdot 0,3^2 \cdot 0,05^0 = 0.0775$$

Příklad

Příklad 6

Hodím kostkou. Poté házím mincí tolikrát, kolik bylo ok na kostce. Náhodná veličina X popisuje počet ok na kostce, náhodná veličina Y popisuje, kolikrát padl „orel". Určete rozdělení pravděpodobnosti náhodného vektoru (X, Y).

X ... **počet ok**, $X \in \{1, 2, \dots, 6\}$, Y ... **počet „orlů"**, $Y \in \{0, 1, \dots, 6\}$

$X \backslash Y$	0	1	2	3	...
1	$\frac{1}{6}\,\frac{1}{2}$	$\frac{1}{6}\,\frac{1}{2}$	0	0	...
2	$\frac{1}{6}\left(\frac{1}{2}\right)^2$	$\frac{1}{6}\binom{2}{1}\frac{1}{2}\,\frac{1}{2}$	$\frac{1}{6}\left(\frac{1}{2}\right)^2$	0	...
3	$\frac{1}{6}\left(\frac{1}{2}\right)^3$	$\frac{1}{6}\binom{3}{1}\frac{1}{2}\left(\frac{1}{2}\right)^2$	$\frac{1}{6}\binom{3}{2}\left(\frac{1}{2}\right)^2\frac{1}{2}$	$\frac{1}{6}\left(\frac{1}{2}\right)^3$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

$$p(x, y) = \begin{cases} \frac{1}{6}\binom{x}{y}\left(\frac{1}{2}\right)^{x-y}\left(\frac{1}{2}\right)^{y} & \text{pro } (x, y) \in \{1, 2, \dots, 6\} \times \{0, 1, \dots, 6\},\ y \leq x \\ 0 & \text{jinak} \end{cases}$$

Spojité náhodné vektory

Definice 6

Řekneme, že náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ definovaný na $(\Omega, \mathcal{A}, P)$ je **absolutně spojitého typu**, jestliže existuje **nezáporná integrovatelná funkce** f taková, že rozdělení pravděpodobností

$$P(\mathbf{X} \in B) = \int\limits_B f(\mathbf{x})\, d\mathbf{x} = \int\limits_{B_1} \cdots \int\limits_{B_n} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

pro každé

$$B = B_1 \times \cdots \times B_n \in \mathcal{B}^n.$$

Funkci f nazýváme **hustotou** rozdělení pravděpodobností náhodného vektoru $\mathbf{X} = (X_1, \ldots, X_n)'$ absolutně spojitého typu, stručněji f je hustotou $\mathbf{X}$.

Značení: Fakt, že jde o spojitý náhodný vektor budeme značit $\mathbf{X} \sim f$.

Spojité náhodné vektory

Věta 7 (Vlastnosti hustoty)

Nechť $\mathbf{X} = (X_1, \dots, X_n)'$ *je náhodný vektor absolutně spojitého typu,* f *jeho hustota a* F *jeho distribuční funkce. Pak*

- $f(\mathbf{x}) \geq 0$ *pro každé* $\mathbf{x} \in \mathbb{R}^n$
- $\int\limits_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = 1$
- *Protože* $P(\mathbf{X} \in B) = \int\limits_B f(\mathbf{x})\, d\mathbf{x}$*, pak pro* $B = (-\infty, x_1\rangle \times \cdots \times (-\infty, x_n\rangle$,

$$F(x_1, \dots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f(t_1, \dots, t_n)\, dt_1 \dots t_n$$

- *Hustotu lze pomocí distribuční funkce vyjádřit takto*

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

přičemž uvedená derivace existuje skoro všude vzhledem k Lebesgueově míře.

Příklad

Příklad 7 (Vícerozměrné rovnoměrné rozdělení)

Řekneme, že náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ *má* **vícerozměrné rovnoměrné rozdělení** *s parametry* $a_1, b_1, \ldots, a_n, b_n \in \mathbb{R}$ *(*$a_i < b_i,\ i = 1, \ldots, n$*), pokud její hustota má tvar*

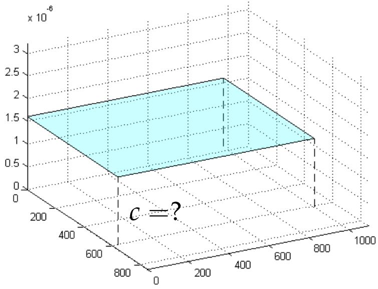
$$f(x_1, \ldots, x_n) = \begin{cases} \prod\limits_{i=1}^{n} \frac{1}{b_i - a_i} & \text{pro } x_i \in (a_i, b_i),\ a_i < b_i,\ i = 1, \ldots, n, \\ 0 & \text{jinak.} \end{cases}$$

Náhodný vektor budeme značit $\boxed{\mathbf{X} \sim Rs_n(a_1, b_1, \ldots, a_n, b_n)}$.

Příklad

Příklad 8 (Squash)

Předpokládáme, že squash hrají dva začátečníci, kterým míček padá zcela náhodně do hřiště. Popište rozdělení pravděpodobnosti náhodného vektoru (X, Y), který označuje souřadnice dopadu míčku. Rozměr squashového kurtu je $640 \times 975\,cm$.



Míček padá „náhodně“ $\Rightarrow$ rovnoměrné rozdělení

$$f(x,y) = \begin{cases} c \text{ pro } (x,y) \in \langle 0;640\rangle \times \langle 0;975\rangle \\ 0 \text{ jinak} \end{cases}$$

$c = ?$

Musí platit $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x,y)dxdy = 1$,

tj. **objem** kvádru $= 1$

$640 \cdot 975 \cdot c = 1 \Rightarrow c = \frac{1}{640 \cdot 975}$

$$(X, Y) \sim Rs_2(0, 640, 0, 975)$$

Příklad

Příklad 9 (Vícerozměrné normální rozdělení)

Řekneme, že náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ *má* n-**rozměrné normální (Gaussovo) rozdělení** *s parametry* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ *a* $\mathbf{\Sigma} > 0$*, pokud její hustota má tvar*

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\mathbf{\Sigma}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

Píšeme

$$\boxed{\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})}$$

$\mathbf{\Sigma} > 0$ *... matice je pozitivně definitní a tedy i regulární.*
Symbol $|\mathbf{\Sigma}|$ *... determinant matice.*

Příklad

Pro $n = 2$:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \rho \in \langle 0, 1 \rangle$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

Píšeme

$$\mathbf{X} = (X_1, X_2)' \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

Příklad

Ukázky hustot $f(x_1, x_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$