Zadání příkladů – 4. 10 2016

Příklad 13 (normální rozdělení). Model pro náhodný výběr $X_1, X_2, \ldots, X_n$ je z $N(\mu, \sigma^2)$ a říkáme, že $X_1, X_2, \ldots, X_n$ pochází z normálního rozdělení, t.j. $X \sim N(\mu, \sigma^2)$. Parametr modelu $N(\mu, \sigma^2)$ je vektor $\boldsymbol{\theta} = (\mu, \sigma^2)$. Hustota tohoto rozdělení má tvar

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Příklad 14 (standardizované normální rozdělení). Model pro náhodný výběr $X_1, X_2, |dots, X_n$ pochází ze standardizovaného normálního rozdělení, t.j. $X \sim N(\mu, \sigma^2)$, kde $\mu = 0$, $\sigma^2 = 1$. Parametr modelu $N(\mu, \sigma^2)$ je vektor $\boldsymbol{\theta} = (0, 1)$. Hustota tohoto rozdělení má tvar

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}, x \in \mathbb{R}.$$

Příklad 15 (dvojrozměrné normální rozdělení). Náhodný vektor $(X, Y)^T$ má dvojrozměrné normální rozdělení

$$N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ kde } \boldsymbol{\mu} = (\mu_1, \mu_2)^T \text{ a } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

s hustotou

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

kde $(x, y)^T \in \mathbb{R}^2$, $\mu_j \in \mathbb{R}$, $\sigma_j^2 > 0$, $j = 1, 2$, $\rho \in \langle -1, 1 \rangle$ jsou parametry. Potom $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Výraz v exponentu můžeme zapsat jako

$$-\frac{1}{2}\begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}.$$

Marginální rozdělení [1] jsou $X \sim N(\mu_1, \sigma_1^2)$ a $Y \sim N(\mu_2, \sigma_2^2)$, ρ je koeficient korelace[2](Viz obrázek 1)

Příklad 16 (dvojrozměrné normální rozdělení)**.** (1) Nakreslete hustotu dvojrozměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pomocí funkce `image()` a superponujte ho s konturovým grafem hustoty toho stejného rozdělení pomocí funkce `contour()`. (2) Nakreslete hustotu dvojrozměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pomocí funkce `persp()`. Hustotu rozsekejte na 12 intervalů, kde hodnoty v těchto intervalech budou odpovídat barvám `terrain.colors(12)`. Použijte následující parametry:

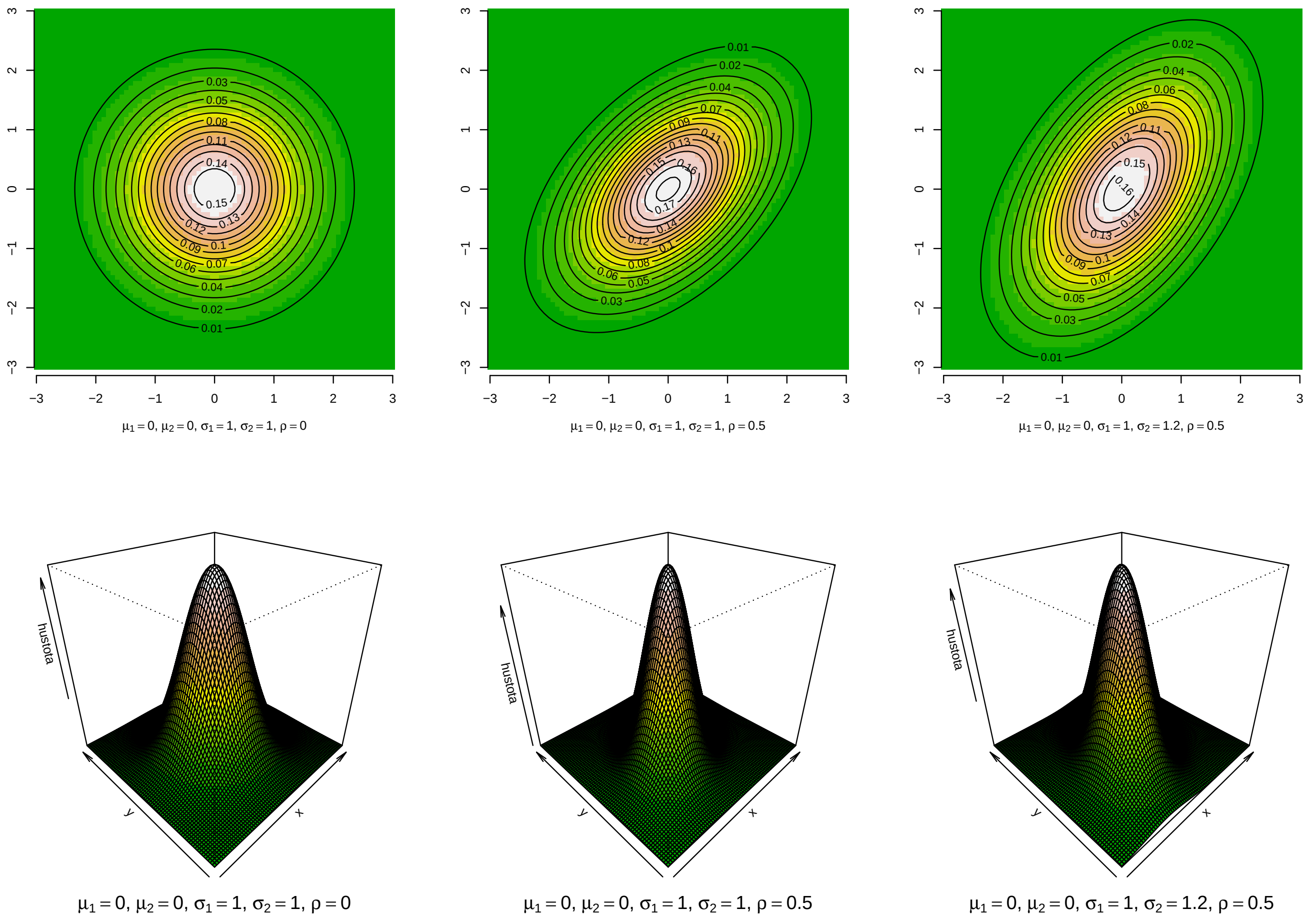
- $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0$;
- $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.5$;

[1] Marginální rozdělení je rozdělení náhodné proměnné, zde X nezávisle na Y a naopak Y nezávisle na X.

[2] Z tohoto příkladu je zřejmé, že na dostatečný popis dvojrozměrného normálního rozdělení potřebujeme pět parametrů, t.j. střední hodnotu a rozptyl pro marginální rozdělení náhodných proměnných X a Ya korelační koeficient $\rho = \rho(X, Y)$ popisující sílu lineárního vztahu X a Y.

- $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1.2, \sigma_2 = 1, \rho = 0.5.$

Vzorové řešení je uvedeno na obrázku 1.



Obrázek 1: Hustoty dvojrozměrného normálního rozdělení při různých parametrech (první řádek – konturový graf; druhý řádek - perspektivní trojrozměrný graf v podobě plochy); čím je ρ odlišnější od nuly, tím více se kontury liší od kruhů (mění se na elipsy); se zvyšujícím se rozdílem mez σ_1 a σ_2 se zvětšuje rozdíl rozptýlení koncentrických kruhů ve směru jednotlivých os (říkáme, že rozdíl variability proměnných X a Y se zvětšuje.)

Příklad 17 (standardizované normální rozdělení)**.** Náhodný vektor $(X, Y)^T$ má dvojrozměrné normální rozdělení

$$N_2(\mathbf{0}, \mathbf{\Sigma}), \text{ kde } \mathbf{0} = (0, 0)^T \text{ a } \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

s hustotou

$$\phi(x, y) = f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\},$$

kde $(x, y)^T \in \mathbb{R}^2$, $\rho \in \langle -1, 1 \rangle$ jsou parametry, potom $\boldsymbol{\theta} = (0, 0, 1, 1, \rho)$. Výraz v exponentu můžeme psát jako

$$-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix},$$

marginální rozdělení jsou obě $N(0,1)$ a ρ je koeficient korelace.

Příklad 18 (standardizované normální rozdělení). Nechť náhodnou proměnnou $X \sim N(\mu_1, \sigma_1^2)$ je největší výška mozkovny (skull.pH; v mm) a náhodnou proměnnou $Y \sim N(\mu_2, \sigma_2^2)$ je morfologická výška tváře (face.H; v mm). Nechť X a Y mají dvojrozměrné normální rozdělení s parametry $(\mu_1, \mu_2)^T$ a σ_1^2, σ_2^2 a ρ jsou parametry kovarianční matice $\mathbf{\Sigma}$. Když od náhodné proměnné X odpočítáme její střední hodnotu μ_1 a tento rozdíl podělíme odmocninou z rozptylu (σ_1), dostaneme náhodnou proměnnou Z_X, která má asymptoticky normální rozdělení se střední hodnotou $\mu_1 = 0$ a rozptylem $\sigma_1^2 = 1$, což zapisujeme jako $Z_X \sim N(0,1)$. Pokud od náhodné proměnné Y odečteme její střední hodnotu μ_2 a tento rozdíl podělíme odmocninou z rozptylu (σ_2), dostaneme náhodnou proměnnou Z_Y, která má asymptoticky normální rozdělení se střední hodnotou $\mu_2 = 0$ a rozptylem $\sigma_2^2 = 1$, což zapisujeme jako $Z_Y \sim N(0,1)$. Potom $(Z_X, Z_Y)^T$ má standardizované dvourozměrné normální rozdělení $N_2(\boldsymbol{\mu}, \mathbf{\Sigma})$ s parametry $\boldsymbol{\mu} = (0,0)^T$ a $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ a ρ jsou parametry kovarianční matice $\mathbf{\Sigma}$.

Příklad 19 (dvourozměrné normální rozdělení). Simulaci pseudonáhodných čísel z $N_2(\boldsymbol{\mu}, \mathbf{\Sigma})$ můžeme v R vytvořit následujícími způsoby:

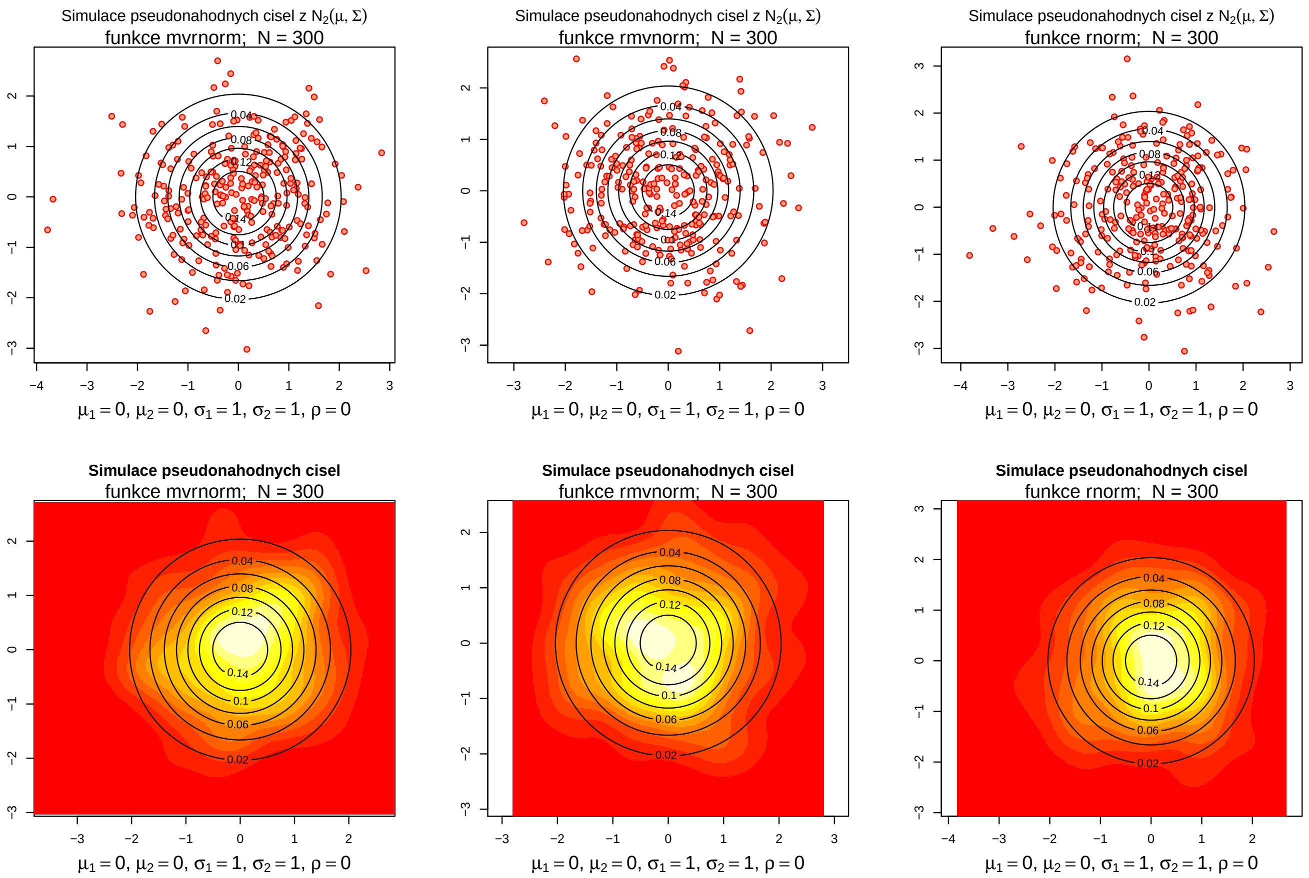
1. použitím funkce mvrnorm() z knihovny MASS;
2. použitím funkce rmvnorm() z knihovny mvtnorm
3. použitím funkce rnorm() a následujícího algoritmu:
   Nechť $X_1 \sim N(0,1)$ a $X_2 \sim N(0,1)$; potom $(Y_1, Y_2)^T \sim N_2(\boldsymbol{\mu}, \mathbf{\Sigma})$, kde $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ je vektor středních hodnot a σ_1^2 a σ_2^2 a ρ jsou parametry kovarianční matice $\mathbf{\Sigma}$, přičemž síla lineárního vztahu Y_1 a Y_2 je daná velikostí a znaménkem ρ; $Y_1 = \sigma_1 X_1 + \mu_1$ a $Y_2 = \sigma_2(\rho X_1 + \sqrt{1-\rho^2} X_2) + \mu_2$. Nasimulujte pseudonáhodná čísla Y_1 a Y_2 z $N_2(\boldsymbol{\mu}, \mathbf{\Sigma})$. Vypočítejte dvourozměrný jádrový odhad hustoty $(Y_1, Y_2)^T$ pomocí funkce kde2d(). Nakreslete jej také pomocí funkce image() a superponujte jej kontúrovým grafem hustoty dvourozměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \mathbf{\Sigma})$ pomocí funkce contour(). Hustotu rozsekejte na 12 intervalů, kde hodnoty v těchto intervalech budou odpovídat barvám terrain.colors(12). Při simulaci použijte následující parametry:

   (a) $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho = 0$; (1) $n = 50$, (2) $n = 500$

   (b) $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho = 0.5$; (1) $n = 50$, (2) $n = 500$

   (c) $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 1.2$, $\rho = 0.5$; (1) $n = 50$, (2) $n = 500$

   Vzorové řešení viz obrázek 2.



Obrázek 2: Hustoty dvourozměrného normálního rozdělení