

5 Pravděpodobnostní modely pro spojité náhodné veličiny

5.1 Spojitá náhodná veličina

Spojité náhodné veličiny nabývají všech hodnot z nějakého intervalu. Její pravděpodobnostní chování je popsáno hustotou pravděpodobnosti, což je po částech spojitá nezáporná funkce $f(x)$, která je normovaná (tj. $\int_{-\infty}^{\infty} f(x)dx = 1$) a s distribuční funkcí je spjata integrálním vztahem: $\forall x \in \mathbb{R} : F(x) = \int_{-\infty}^x f(t)dt$. Je to zidealizovaný protějšek hustoty četnosti. Pravděpodobnost, že spojitá náhodná veličina nabude nějaké konkrétní hodnoty, je 0. Při výpočtu pravděpodobnosti jevu, že spojitá náhodná veličina nabude hodnot z nějakého intervalu, tedy nehraje roli, zda je tento interval uzavřený, polouzavřený nebo otevřený, tj. $\forall a, b \in \mathbb{R}, a < b: \Pr(a \leq X \leq b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \Pr(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$.

5.1.1 Normální rozložení

Normální rozložení se vyskytuje při opakovaných nezávislých měřeních téže charakteristiky za stejných podmínek. Jednotlivé naměřené hodnoty, tj. realizace náhodné veličiny X , se v důsledku působení náhodných vlivů, jejichž proměnlivost je vyjádřena konstantou $\sigma > 0$, poněkud odchylní od skutečné hodnoty μ zkoumané charakteristiky. Píšeme $X \sim N(\mu, \sigma^2)$. Parametr μ má význam střední hodnoty náhodné veličiny X , parametr σ představuje její směrodatnou odchylku (viz kap. 6).

Hustota: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Grafem hustoty normálního rozložení je tzv. Gaussova křivka. Má zvonovitý tvar a je symetrická kolem osy procházející jejím vrcholem.

Věta 5.1: Důležité vlastnosti normálního rozložení:

- Jestliže $X \sim N(\mu, \sigma^2)$ a $Z = a + bX$, pak $Z \sim N(a + b\mu, b^2\sigma^2)$.
- Jestliže X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^2), i = 1, \dots, n, Z = \sum_{i=1}^n X_i$, pak $Z \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.
- Jestliže X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu, \sigma^2), i = 1, \dots, n, M = \frac{1}{n} \sum_{i=1}^n X_i$, pak $M \sim N(\mu, \frac{\sigma^2}{n})$.
- Pravděpodobnost, že náhodná veličina $X \sim N(\mu, \sigma^2)$ se realizuje v nějakém intervalu $\langle a, b \rangle$, nemůžeme vypočítat jako $\Pr(a \leq X \leq b) = \int_a^b f(x)dx$, protože jde o integrál z vyšší transcendentní funkce. Lze využít toho, že $\Pr(a \leq X \leq b) = F(b) - F(a)$, přičemž hodnoty distribuční funkce $F(x)$ získáme pomocí statistického softwaru.

V praxi jsou zvláště užitečné následující pravděpodobnosti:

- $\Pr(\mu - \sigma \leq X \leq \mu + \sigma) \doteq 0,6827$,
- $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \doteq 0,9545$,
- $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \doteq 0,9973$ (pravidlo 3σ).

Poznámka: Uvedené tři rovnosti lze použít k jednoduchému ověření, zda se hodnoty v datovém souboru řídí normálním rozložením. Z dat vypočítáme aritmetický průměr m a směrodatnou odchylku s . Pokud přibližně 68 % hodnot leží v intervalu $(m - s, m + s)$, 95 % hodnot leží v intervalu $(m - 2s, m + 2s)$ a 99,7 % hodnot leží v intervalu $(m - 3s, m + 3s)$, lze soudit, že data pocházejí z normálního rozložení.

Vzorce pro výpočty pravděpodobností pomocí normálního rozložení v softwaru 

- $\Pr(X \leq x) = F(x) = \text{pnorm}(x, \mu, \sigma)$,
- $\Pr(X \geq x) = 1 - F(x) = 1 - \text{pnorm}(x, \mu, \sigma)$,
- $\Pr(a \leq X \leq b) = F(b) - F(a) = \text{pnorm}(b, \mu, \sigma) - \text{pnorm}(a, \mu, \sigma)$.

5.1.2 Standardizované normální rozložení

Pro $\mu = 0$, $\sigma^2 = 1$ se jedná o standardizované normální rozložení, píšeme $U \sim N(0, 1)$.

Hustota: $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.

Hodnoty distribuční funkce $F(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ najdeme pro $u \geq 0$ ve statistických tabulkách, pro $u < 0$ se používá přepočtový vzorec $F(u) = 1 - F(-u)$. Jestliže $X \sim N(\mu, \sigma^2)$, pak $U = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Příklad 5.1. Řešený příklad

Načtete datový soubor `13-two-samples-correlations-trunk.txt` obsahující údaje o délce dolní končetiny (`lowex.L`) a délce trupu (`tru.L`) mladých dospělých mužů a žen. Předpokládejme, že náhodná veličina X popisuje délku dolní končetiny žen. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X , a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X ; pro naměřená data vykreslete histogram a superponujte jej (i) křivkou jádrového odhadu hustoty; (ii) křivkou hustoty nalezeného rozložení. Obě křivky vzájemně porovnejte.

Řešení příkladu 5.1

Délka dolní končetiny žen je spojitý znak, k jeho popisu tedy použijeme spojitou náhodnou veličinu. Protože délka dolní končetiny je tělesný rozměr, mohlo by k popisu tohoto znaku být vhodné normální rozložení. O náhodné veličině X tedy předpokládáme, že pochází z normálního rozložení s parametry μ a σ^2 , kde μ je střední hodnota délky dolní končetiny žen a σ^2 je rozptyl délky dolní končetiny žen okolo střední hodnoty μ . Parametr μ odhadneme pomocí aritmetického průměru naměřených hodnot, tj.

$$\hat{\mu} = m = \frac{1}{n} \sum_{i=1}^n x_i.$$

Parametr σ^2 odhadneme pomocí rozptylu naměřených hodnot, tj.

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2.$$

Aritmetický průměr m vypočítáme pomocí příkazu `mean()`. Rozptyl s^2 budeme častěji reprezentovat ve tvaru umocněné směrodatné odchylky $s = \sqrt{s^2}$, jejíž hodnotu vypočítáme pomocí příkazu `sd()`.

```
1 data <- read.delim('13-two-samples-correlations-trunk.txt')
2 lowex.LF <- na.omit(data[data$sex == 'f', 'lowex.L'])
3 n <- length(lowex.LF) # 100
4 m <- mean(lowex.LF) # 940,5
5 s <- sd(lowex.LF) # 45,47116
```

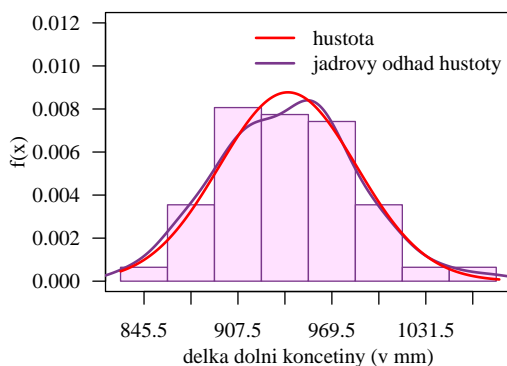
Odhad parametru $\hat{\mu} = 940,50$, odhad parametru $\hat{\sigma}^2 = 45,47^2$. Délka dolní končetiny žen se pohybuje okolo střední hodnoty 940,50 mm se směrodatnou odchylkou 45,47 mm. O náhodné veličině X předpokládáme, že pochází z normálního rozložení s parametry $\mu = 940,50$ a $\sigma^2 = 45,47^2$, tj. $X \sim N(940,50; 45,47^2)$.

Nyní vykreslíme histogram s křivkou jádrového odhadu hustoty, který vypočítáme na základě naměřených dat, a s křivkou hustoty rozložení $N(940,50; 45,47^2)$. Pro potřeby histogramu rozdělíme naměřené hodnoty do osmi třídicích intervalů o ekvidistantní šířce 31 mm stanovením hranic 830, 861, ..., 1078 mm. Histogram vykreslíme v relativní škále příkazem `hist()` s argumentem `prob = T` (viz kapitola 2). Jádrový odhad hustoty vypočítáme příkazem `density()`, jehož jediným vstupním argumentem bude vektor `lowex.LF`. Křivku jádrového odhadu hustoty vykreslíme příkazem `lines()` aplikovaným na výstup funkce `density()`. Barvu křivky jádrového odhadu hustoty volíme v podobném odstínu jako barvu histogramu, protože oba objekty reprezentují naměřená data. Pro vykreslení křivky hustoty rozložení $N(940,50; 45,47^2)$ musíme nejprve příkazem `seq()` vytvořit posloupnost hodnot pokrývající rozpětí naměřených délek dolní končetiny žen. Pro každý prvek této posloupnosti potom příkazem `dnorm()` vypočítáme hodnotu hustoty rozložení $N(940,50; 45,47^2)$. Křivku hustoty následně vykreslíme příkazem `lines()`. Histogram spolu s oběma křivkami je zobrazen na obrázku 5.1.

```

6 r <- round(3.3 * log10(n) + 1) # 8
7 # 1076 - 835 = 241; 248 / 8 = 31 -> seq(830, 1078, by = 31)
8 b.lowex.LF <- seq(830, 1078, by = 31)
9 h <- hist(lowex.LF, breaks = b.lowex.LF, plot = F)
10 hist(lowex.LF, breaks = b.lowex.LF, prob = T, include.lowest = F, ylim = c(0, 0.012),
11       col = 'thistle1', border = 'mediumorchid4', xlab = 'delka dolni koncetiny (v mm)',
12       ylab = 'f(x)', main = '', axes = F)
13 box(bty = 'o')
14 axis(1, at = h$mids)
15 axis(2, las = 1)
16 lines(density(lowex.LF), col = 'mediumorchid4', lwd = 2)
17
18 x <- seq(830, 1080, length = 512)
19 fx <- dnorm(x, mean = m, sd = s)
20 lines(x, fx, col = 'red', lwd = 2)
21 legend('topright', lwd = 2, col = c('red', 'mediumorchid4'), bty = 'n',
22       legend = c('hustota', 'jadrovy odhad hustoty'))

```



Obrázek 5.1: Histogram superponovaný křivkou jádrového odhadu hustoty a křivkou hustoty normálního rozložení

Z obrázku 5.1 vidíme, že křivka jádrového odhadu hustoty a křivka hustoty rozložení $N(940, 50; 45, 47^2)$ jsou si tvarem velmi blízké. Grafická vizualizace tedy podporuje předpoklad, že náhodná veličina X popisující délku dolní končetiny žen pochází z normálního rozložení s parametry $\mu = 940, 50$ a $\sigma^2 = 45, 47^2$.

Poznámka: Grafická vizualizace je nástrojem sloužícím k subjektivnímu posouzení normálního rozložení dat. V kapitole 8 si ukážeme objektivní zhodnocení normality dat prostřednictvím testů normality. ★

Příklad 5.2. Řešený příklad

Za předpokladu, že náhodná veličina X popisující délku dolní končetiny žen pochází z rozložení $N(940, 50; 45, 47^2)$, vypočítejte pravděpodobnost, že délka dolní končetiny žen bude (a) menší než 962,50 mm; (b) větší než 974,13 mm; (c) v rozmezí 930,00 mm až 950,00 mm; (d) rovná 943,25 mm.

Řešení příkladu 5.2

Pravděpodobnost, že délka dolní končetiny žen bude menší než 962,50 mm, tj. $\Pr(X < 962, 50) = \Pr(X \leq 962, 50)$ odpovídá hodnotě distribuční funkce $F(x)$ rozložení $N(940, 50; 45, 47^2)$ v bodě $x = 962, 50$. Tuto hodnotu vypočítáme pomocí příkazu `pnorm()`. Poznamenejme, že k výpočtu pravděpodobností použijeme nezaokrouhlené odhady parametrů μ a σ^2 vypočítané v příkladu 5.1.

```

23 pnorm(962.50, m, s) # 0,6857443

```

Pravděpodobnost, že délka dolní končetiny žen bude větší než 974,13 mm, tj. $\Pr(X > 974, 13) = \Pr(X \geq 974, 13)$, vypočítáme tak, že od jedné odečteme pravděpodobnost, že délka dolní končetiny žen bude nejvýše 974,13 mm.

Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F(x)$ v bodě $x = 974,13$, vypočítáme příkazem `pnorm()`.

```
24 1 - pnorm(974.13, m, s) # 0,2297745
```

Pravděpodobnost, že délka dolní končetiny žen bude v rozmezí 930,00 mm až 950,00 mm, tj. $\Pr(930,00 < X < 950,00) = \Pr(930,00 \leq X < 950,00) = \Pr(930,00 < X \leq 950,00) = \Pr(930,00 \leq X \leq 950,00)$, vypočítáme tak, že od pravděpodobnosti, že délka dolní končetiny žen nejvýše 950,00 mm, odečteme pravděpodobnost, že délka dolní končetiny žen bude nejvýše 930,00 mm. Obě pravděpodobnosti jsou hodnotami distribuční funkce $F(x)$ v bodě $x = 950,00$, resp. v bodě $x = 930,00$ a vypočítáme je pomocí příkazu `pnorm()`.

```
25 pnorm(950.00, m, s) - pnorm(930.00, m, s) # 0,1740559
```

Pravděpodobnost, že délka dolní končetiny žen bude rovná 943,25 mm, tj. $\Pr(X = 943,25)$, je z definice rovná nule, protože náhodná veličina X popisující délku dolní končetiny žen má spojité rozložení.

Pravděpodobnost, že délka dolní končetiny žen bude menší než 962,50 mm, je 68,57 %. Pravděpodobnost, že délka dolní končetiny žen bude větší než 974,13 mm, je 22,98 %. Pravděpodobnost, že délka dolní končetiny žen bude v rozmezí 930,00 mm až 950,00 mm, je 17,41 %. Pravděpodobnost, že délka dolní končetiny žen bude rovná 943,25 mm, je 0,00 %.



Příklad 5.3. Řešený příklad

Za předpokladu, že náhodná veličina X popisující délku dolní končetiny žen pochází z rozložení $N(940, 50; 45, 47^2)$, (a) určete rozložení náhodné veličiny $M = \frac{1}{8} \sum_{i=1}^8 X_i$ popisující průměrnou délku dolní končetiny osmi žen a odhadněte parametry tohoto rozložení; (b) vypočítejte pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude (i) menší než 962,50 mm; (ii) větší než 974,13 mm; (iii) v rozmezí 930,00 mm až 950,00 mm; (iv) rovná 943,25 mm.

Řešení příkladu 5.3

Za předpokladu, že náhodná veličina X pochází z normálního rozložení s parametry $\mu = 940,50$ a $\sigma^2 = 45,47^2$, můžeme z třetího bodu věty 5.1 odvodit, že náhodná veličina M pochází z normálního rozložení s parametry $\mu = 940,50$ a $\sigma^2 = \frac{45,47^2}{8} = 16,08^2$, tj. $M \sim N(940,50; 16,08^2)$. Průměrná délka dolní končetiny osmi žen se pohybuje okolo střední hodnoty 940,50 mm se směrodatnou odchylkou 16,08 mm.

```
26 s8 <- sqrt(s ^ 2 / 8) # 16,07648
```

Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude menší než 962,50 mm, tj. $\Pr(M < 962,50) = \Pr(M \leq 962,50)$ odpovídá hodnotě distribuční funkce $F_M(x)$ rozložení $N(940,50; 16,08^2)$ v bodě $x = 962,50$. Tuto hodnotu vypočítáme pomocí příkazu `pnorm()`. Analogicky jako v příkladu 5.2 budeme i zde k výpočtu pravděpodobností používat nezaokrouhlený odhad parametru μ vypočítaný v příkladu 5.1 a nezaokrouhlený odhad parametru σ^2 náhodné veličiny M vypočítaný výše.

```
27 pnorm(962.50, m, s8) # 0,9144157
```

Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude větší než 974,13 mm, tj. $\Pr(M > 974,13) = \Pr(M \geq 974,13)$, vypočítáme tak, že od jedné odečteme pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude nejvýše 974,13 mm. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F_M(x)$ v bodě $x = 974,13$, vypočítáme příkazem `pnorm()`.

```
28 1 - pnorm(974.13, m, s8) # 0,01822483
```

Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude v rozmezí 930,00 mm až 950,00 mm, tj. $\Pr(930,00 < M < 950,00) = \Pr(930,00 \leq M < 950,00) = \Pr(930,00 < M \leq 950,00) = \Pr(930,00 \leq M \leq 950,00)$, vypočítáme tak, že od pravděpodobnosti, že průměrná délka dolní končetiny osmi žen bude nejvýše

950,00 mm, odečteme pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude nejvýše 930,00 mm. Obě pravděpodobnosti jsou hodnotami distribuční funkce $F_M(x)$ v bodě $x = 950,00$, resp. v bodě $x = 930,00$ a vypočítáme je pomocí příkazu `pnorm()`.

```
29 pnorm(950.00, m, s8) - pnorm(930.00, m, s8) # 0,4658779
```

Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude rovná 943,25 mm, tj. $\Pr(M = 943,25)$, je z definice rovná nule, protože náhodná veličina M popisující průměrnou délku dolní končetiny osmi žen pochází ze spojitého rozložení.

Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude menší než 962,50 mm, je 91,44 %. Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude větší než 974,13 mm, je 1,82 %. Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude v rozmezí 930,00 mm až 950,00 mm, je 46,59 %. Pravděpodobnost, že průměrná délka dolní končetiny osmi žen bude rovná 943,25 mm, je 0,00 %.

★

Příklad 5.4. Řešený příklad

Předpokládejme, že náhodná veličina X popisující délku dolní končetiny žen pochází z rozložení $N(940, 50; 45, 47^2)$ a náhodná veličina M popisující průměrnou délku dolní končetiny osmi žen pochází z rozložení $N(940, 50; 16, 08^2)$. Do jednoho grafu nakreslete (a) křivku funkce hustoty $f(x)$, resp. $f_M(x)$; (b) křivku distribuční funkce $F(x)$, resp. $F_M(x)$. Křivky vzájemně porovnejte.

Řešení příkladu 5.4

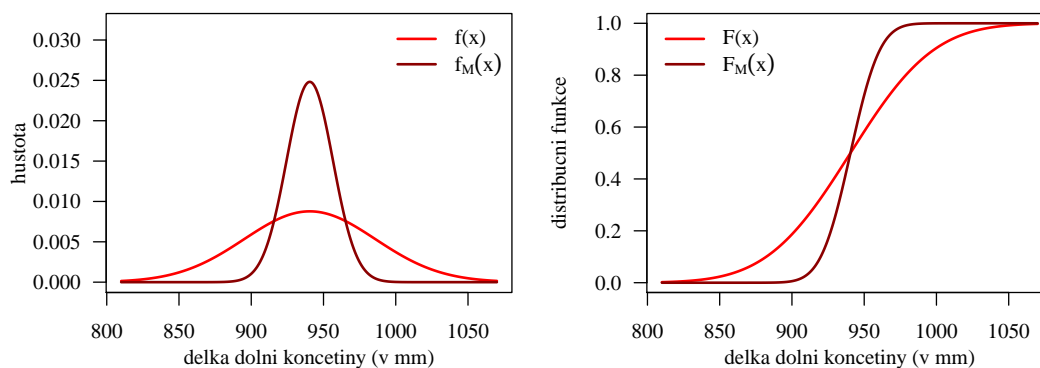
K vykreslení grafu s křivkami hustot si nejprve vytvoříme posloupnost hodnot z intervalu se středem v bodě 940,50 mm, tj. například posloupnost v rozmezí 810,5 mm až 1070,5 mm. Pro každý prvek této posloupnosti potom příkazem `dnorm()` vypočítáme hodnotu hustoty $f(x)$ rozložení $N(940, 50; 45, 47^2)$, resp. hustoty $f_M(x)$ rozložení $N(940, 50; 16, 08^2)$. Následně příkazem `plot()` s argumentem `type = 'l'` vykreslíme graf s křivkou hustoty $f(x)$ a příkazem `lines()` do něj doplníme křivku hustoty $f_M(x)$. Výsledný graf je zobrazen na obrázku 5.2 vlevo.

```
30 x <- seq(810.5, 1070.5, length = 512)
31 fx <- dnorm(x, mean = m, sd = s)
32 fm <- dnorm(x, mean = m, sd = s8)
33 plot(x, fx, type = 'l', ylim = c(0, 0.032), col = 'red', lwd = 2,
34      xlab = 'delka dolni koncetiny (v mm)', ylab = 'hustota', las = 1)
35 lines(x, fm, col = 'darkred', lwd = 2)
36 legend('topright', lwd = 2, col = c('red', 'darkred'), bty = 'n',
37      legend = c('f(x)', expression(f[M](x))))
```

K vykreslení grafu s křivkami distribučních funkcí nejprve pro každý prvek posloupnosti z rozmezí 810,5 mm až 1070,5 mm vypočítáme příkazem `pnorm()` hodnotu distribuční funkce $F(x)$ rozložení $N(940, 50; 45, 47^2)$, resp. distribuční funkce $F_M(x)$ rozložení $N(940, 50; 16, 08^2)$. Následně příkazem `plot()` s argumentem `type = 'l'` vykreslíme graf s křivkou distribuční funkce $F(x)$ a příkazem `lines()` do něj doplníme křivku distribuční funkce $F_M(x)$. Graf s křivkami distribučních funkcí je zobrazen na obrázku 5.2 vpravo.

```
38 Fx <- pnorm(x, mean = m, sd = s)
39 Fm <- pnorm(x, mean = m, sd = s8)
40 plot(x, Fx, type = 'l', col = 'red', lwd = 2,
41      xlab = 'delka dolni koncetiny (v mm)', ylab = 'distribucni funkce', las = 1)
42 lines(x, Fm, col = 'darkred', lwd = 2)
43 legend('topleft', lwd = 2, col = c('red', 'darkred'), bty = 'n',
44      legend = c('F(x)', expression(F[M](x))))
```

Z obrázku 5.2 vlevo vidíme, že vrchol křivek obou hustot se nachází nad střední hodnotou $\mu = 940,50$ mm. Křivka hustoty $f_M(x)$ nabývá zpočátku nízkých hodnot, od určitého okamžiku pak začíná strmě růst až do vrcholu v bodě 940,50 mm, a následně strmě klesá. Tento průběh je důsledkem nižšího rozptylu rozložení náhodné veličiny M . Ve srovnání s křivkou hustoty $f_M(x)$ křivka hustoty $f(x)$ od začátku roste pozvolna až do svého vrcholu v



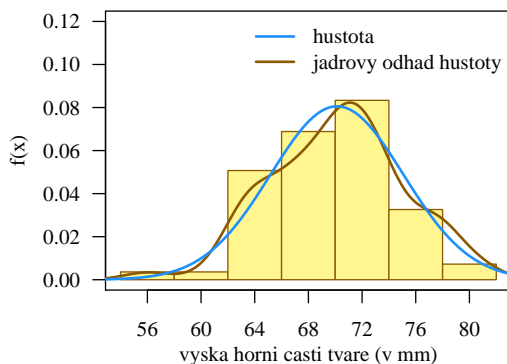
Obrázek 5.2: Graf (a) křivek funkcí hustoty (vlevo); (b) křivek distribučních funkcí (vpravo) rozložení $N(940, 50; 45, 47^2)$, resp. rozložení $N(940, 50; 16, 08^2)$

bodě 940, 50 mm a od tohoto bodu pozvolna klesá. Tento průběh je důsledkem vyššího rozptylu rozložení náhodné veličiny X . Z obrázku 5.2 vpravo vidíme, že křivka distribuční funkce $F_M(x)$ nabývá na poměrně dlouhém úseku nízkých hodnot, následně začíná strmě růst a od určitého okamžiku je její růst k hodnotě 1 velmi pomalý. Ve srovnání s křivkou distribuční funkce $F_M(x)$ křivka distribuční funkce $F(x)$ roste pozvolna po celé délce zobrazeného úseku. Obě křivky se protínají v jediném bodě $[940, 50; 0, 5]$, což značí, že 50 % hodnot (průměrné) délky dolní končetiny (osmi) žen je menší než 940, 50 mm. ★

Příklad 5.5. Neřešený příklad

Načtěte datový soubor 15-anova-means-skull.txt obsahující údaje o výšce horní části tváře (upface.H) mužů bantuské, čínské, malajské, německé a peruánské populace. Předpokládejme, že náhodná veličina X popisuje výšku horní části tváře mužů malajské populace. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X , a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X ; pro naměřená data vykreslete histogram a superponujte jej (i) křivkou jádrového odhadu hustoty; (ii) křivkou hustoty nalezeného rozložení. Obě křivky vzájemně porovnejte.

Výsledky: (a) $X \sim N(70, 13; 4, 95^2)$; (b) histogram s křivkou jádrového odhadu hustoty a s křivkou hustoty rozložení $N(70, 13; 4, 95^2)$ viz obrázek 5.3.



Obrázek 5.3: Histogram superponovaný křivkou jádrového odhadu hustoty a křivkou hustoty normálního rozložení

★

Příklad 5.6. Neřešený příklad

Za předpokladu, že náhodná veličina X popisující výšku horní části tváře mužů malajské populace pochází z rozložení $N(70, 13; 4, 95^2)$, vypočítejte pravděpodobnost, že výška horní části tváře mužů malajské populace bude (a) větší než 70,75 mm; (b) menší než 66,00 mm; (c) rovná 82,30 mm; (d) v rozmezí 64,50 mm až 72,50 mm.

Výsledky: (a) $\Pr(X > 70, 75) = 0, 4502$; (b) $\Pr(X < 66, 00) = 0, 2021$; (c) $\Pr(X = 82, 30) = 0$; (d) $\Pr(64, 50 <$

$$P(X < 72, 50) = 0, 5560.$$

★

Příklad 5.7. Neřešený příklad

Za předpokladu, že náhodná veličina X popisující výšku horní části tváře mužů malajské populace pochází z rozložení $N(70, 13; 4, 95^2)$, (a) určete rozložení náhodné veličiny $M = \frac{1}{5} \sum_{i=1}^5 X_i$ popisující průměrnou výšku horní části tváře pěti mužů malajské populace a odhadněte parametry tohoto rozložení; (b) vypočítejte pravděpodobnost, že průměrná výška horní části tváře pěti mužů malajské populace bude (i) větší než 70,75 mm; (ii) menší než 66,00 mm; (iii) rovná 82,30 mm; (iv) v rozmezí 64,50 mm až 72,50 mm.

Výsledky: (a) $M \sim N(70, 13; 2, 21^2)$; (b-i) $\Pr(M > 70, 75) = 0, 3898$; (b-ii) $\Pr(M < 66, 00) = 0, 0311$; (b-iii) $\Pr(M = 82, 30) = 0$; (b-iv) $\Pr(64, 50 < M < 72, 50) = 0, 8522$.

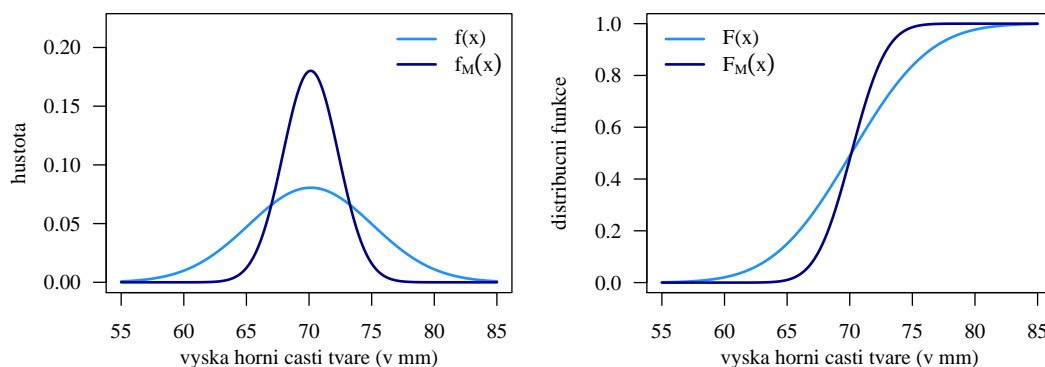
★

Příklad 5.8. Neřešený příklad

Předpokládejme, že náhodná veličina X popisující výšku horní části tváře mužů malajské populace pochází z rozložení $N(70, 13; 4, 95^2)$ a náhodná veličina M popisující průměrnou výšku horní části tváře pěti mužů malajské populace pochází z rozložení $N(70, 13; 2, 21^2)$. Do jednoho grafu nakreslete (a) křivku funkce hustoty $f(x)$, resp. $f_M(x)$; (b) křivku distribuční funkce $F(x)$, resp. $F_M(x)$. Křivky vzájemně porovnejte.

Výsledky: (a) graf s křivkami hustot $f(x)$ a $f_M(x)$ viz obrázek 5.4 vlevo; (b) graf s křivkami distribučních funkcí $F(x)$ a $F_M(x)$ viz obrázek 5.4 vpravo.

★



Obrázek 5.4: Graf (a) křivek funkcí hustoty (vlevo); (b) křivek distribučních funkcí (vpravo) rozložení $N(70, 13; 4, 95^2)$, resp. rozložení $N(70, 13; 2, 21^2)$

5.2 Spojitý dvourozměrný náhodný vektor

Jestliže obě složky dvourozměrného náhodného vektoru $(X, Y)^T$ jsou spojitě náhodné veličiny, hovoříme o spojitém dvourozměrném náhodném vektoru. Jeho pravděpodobnostní chování je popsáno simultánní hustotou pravděpodobnosti $f(x, y)$. Hustota náhodné veličiny X , resp. Y se značí $f_X(x)$, resp. $f_Y(y)$ a nazývá se marginální hustota. Získá se tak, že se simultánní hustota integruje přes přebývající proměnnou.

Náhodné veličiny X, Y jsou stochasticky nezávislé, když $\forall (x, y)^T \in \mathbb{R}^2$ platí $f(x, y) = f_X(x)f_Y(y)$.

Pravděpodobnost, že náhodný vektor $(X, Y)^T$ se bude realizovat v nějaké dvourozměrné oblasti B , se vypočítá jako dvojnásobný integrál ze simultánní hustoty přes tuto oblast:

$$\Pr((X, Y)^T \in B) = \int_{(x, y)^T \in B} \int f(u, v) du dv.$$

5.2.1 Dvourozměrné normální rozložení

Dvourozměrné normální rozložení vzniká ve dvourozměrných situacích za podobných okolností jako normální rozložení. Závisí na reálném vektoru $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ a reálné čtvercové symetrické matici $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Vektor $\boldsymbol{\mu}$ se nazývá vektor středních hodnot, matice $\boldsymbol{\Sigma}$ se nazývá varianční matice. Píšeme $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Přitom $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$ jsou střední hodnoty veličin X, Y , $\sigma_1 > 0$, $\sigma_2 > 0$ jsou směrodatné odchylky a $\rho \in \langle -1, 1 \rangle$ je koeficient korelace veličin X, Y (viz kapitola 6).

Hustota má tvar: $f(x, y) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{q(x, y)}{2}}$, kde $q(x, y) = \frac{1}{1-\rho^2} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \frac{x-\mu_1}{\sigma_1} \frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right]$.

Pro $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0$ se jedná o standardizované dvourozměrné normální rozložení. Píšeme $\mathbf{U} = (U_1, U_2)^T \sim N_2(\mathbf{0}, \mathbf{I})$, kde $\mathbf{0} = (0, 0)^T$ a $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ je jednotková matice druhého řádu.

Věta 5.2: Důležité vlastnosti dvourozměrného normálního rozložení:

- Jestliže $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{a} = (a_1, a_2)^T$ je reálný vektor, $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ je reálná matice, pak pro transformovaný náhodný vektor $\mathbf{Z} = \mathbf{a} + \mathbf{B}(X, Y)^T$ platí, že $\mathbf{Z} \sim N_2(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$.
- Marginální rozložení náhodné veličiny X , resp. Y jsou $X \sim N(\mu_1, \sigma_1^2)$, resp. $Y \sim N(\mu_2, \sigma_2^2)$.
- Je-li koeficient korelace $\rho = 0$, pak řekneme, že veličiny X a Y jsou nekorelované a znamená to, že mezi nimi není žádná lineární závislost. U dvourozměrného normálního rozložení je však nekorelovanost ekvivalentní se stochastickou nezávislostí veličin X a Y .

Příklad 5.9. Řešený příklad

Načtete datový soubor 31-goldman-alaska.csv obsahující údaje o délce pažní kosti z levé strany humer.L a průměru hlavice pažní kosti z levé strany humer.HDL u mužů a žen z kmene Tigara, resp. z kmene Ipituaq. Předpokládejme, že náhodný vektor $(X, Y)^T$ popisuje prostřednictvím náhodné veličiny X délku pažní kosti z levé strany a prostřednictvím náhodné veličiny Y průměr hlavice pažní kosti z levé strany u mužů z kmene Tigara. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodného vektoru $(X, Y)^T$, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodného vektoru $(X, Y)^T$; (i) vykreslete tečkový diagram a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ nalezeného rozložení; (ii) vykreslete vrstevnicový diagram jádrového odhadu hustoty $k(x, y)$ a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty nalezeného rozložení $f(x, y)$; (iii) vykreslete 3D-diagram jádrového odhadu hustoty $k(x, y)$, resp. 3D-diagram hustoty nalezeného rozložení $f(x, y)$.

Řešení příkladu 5.9

Délka pažní kosti i průměr hlavice pažní kosti z levé strany tvoří dvojici spojitých znaků, k jejich společnému popisu tedy použijeme spojitý náhodný vektor $(X, Y)^T$. Protože oba znaky popisují tělesné rozměry, mohlo by k popisu této dvojice znaků být vhodné dvourozměrné normální rozložení. O náhodném vektoru $(X, Y)^T$ tedy předpokládáme, že pochází z dvourozměrného normálního rozložení s vektorem středních hodnot $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ a varianční maticí $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, kde μ_1 a σ_1^2 je střední hodnota a rozptyl délky pažní kosti, μ_2 a σ_2^2 je střední hodnota a rozptyl průměru hlavice pažní kosti a ρ je korelační koeficient. Parametr μ_1 , resp. σ_1^2 odhadneme pomocí aritmetického průměru m_1 , resp. rozptylu s_1^2 naměřených délek pažních kostí. Parametr μ_2 , resp. σ_2^2 odhadneme pomocí aritmetického průměru m_2 , resp. rozptylu s_2^2 naměřených průměrů hlavic pažních kostí. Parametr ρ odhadneme pomocí Pearsonova koeficientu korelace r_{12} (viz kapitola 3). Aritmetické průměry m_1 a m_2 vypočítáme příkazem `mean()`. Rozptyly s_1^2 a s_2^2 budeme častěji reprezentovat ve tvaru umocněných směrodatných odchylek, jejichž hodnoty vypočítáme příkazem `sd()`. Pearsonův korelační koeficient r_{12} vypočítáme příkazem `cor()` s argumentem `method = 'pearson'`.


```
45 data <- read.delim('31-goldman-alaska.csv', sep = ';', dec = '.')
46 data.M <- na.omit(data[data$sex == 'm' & data$pop == 'Tigara',
47                       c('humer.L', 'humer.HDL')])
48 humer.LM <- data.M$humer.L
49 humer.HDLM <- data.M$humer.HDL
50 mL <- mean(humer.LM) # 296,34
51 mHDL <- mean(humer.HDLM) # 43,8984
52 sL <- sd(humer.LM) # 16,30294
53 sHDL <- sd(humer.HDLM) # 3,21493
54 r12 <- cor(humer.LM, humer.HDLM, method = 'pearson') # 0,8523391
```

Odhad parametru $\hat{\mu}_1 = 296,34$, $\hat{\mu}_2 = 43,90$, odhad parametru $\hat{\sigma}_1^2 = 16,30^2$, $\hat{\sigma}_2^2 = 3,21^2$, odhad parametru $\hat{\rho} = 0,85$. O náhodném vektoru $(X, Y)^T$ předpokládáme, že pochází z dvourozměrného normálního rozložení s vektorem středních hodnot $\boldsymbol{\mu} = (296,34, 43,90)^T$ a varianční maticí $\boldsymbol{\Sigma} = \begin{pmatrix} 16,30^2 & 0,85 \times 16,30 \times 3,21 \\ 0,85 \times 16,30 \times 3,21 & 3,21^2 \end{pmatrix}$, tj.

$$(X, Y)^T \sim N_2 \left((296,34, 43,90)^T, \begin{pmatrix} 16,30^2 & 44,67 \\ 44,67 & 3,21^2 \end{pmatrix} \right).$$

K vykreslení grafů je třeba vypočítat jádrový odhad hustoty $k(x, y)$ a dvourozměrnou hustotu $f(x, y)$ rozložení $N_2 \left((296,34, 43,90)^T, \begin{pmatrix} 16,30^2 & 44,67 \\ 44,67 & 3,21^2 \end{pmatrix} \right)$. Jádrový odhad hustoty $k(x, y)$ vypočítáme příkazem `kde2d()` z knihovny MASS. Vstupními argumenty příkazu budou vektory `humer.LM` a `humer.HDLM`, dále argument `n` specifikující počet bodů, pro které bude jádrový odhad počítán (celkem $n \times n$), a argument `lim = c(x_min, x_max, y_min, y_max)` určující rozpětí hodnot z hlediska osy x a z hlediska osy y . Výstupem funkce bude posloupnost bodů osy x , resp. osy y o délce $n = 75$ s hodnotami v rozpětí x_{\min} až x_{\max} , resp. y_{\min} až y_{\max} a matice `z`, která má dimenzi 75×75 a obsahuje pro každý bod $[x_i, y_j]$, $i, j = 1, \dots, 75$, hodnotu jádrového odhadu hustoty.

```
55 n <- 75
56 kxy <- MASS::kde2d(humer.LM, humer.HDLM, n = n, lim = c(255, 348, 35, 55))
```

Dvourozměrnou hustotu $f(x, y)$ vypočítáme pomocí funkce `d2norm()`, implementované v -skriptu `AS-sbirka-funkce.R`, který je součástí této publikace. Syntax a výstupy funkce `d2norm()` jsou analogické jako u funkce `kde2d()`.

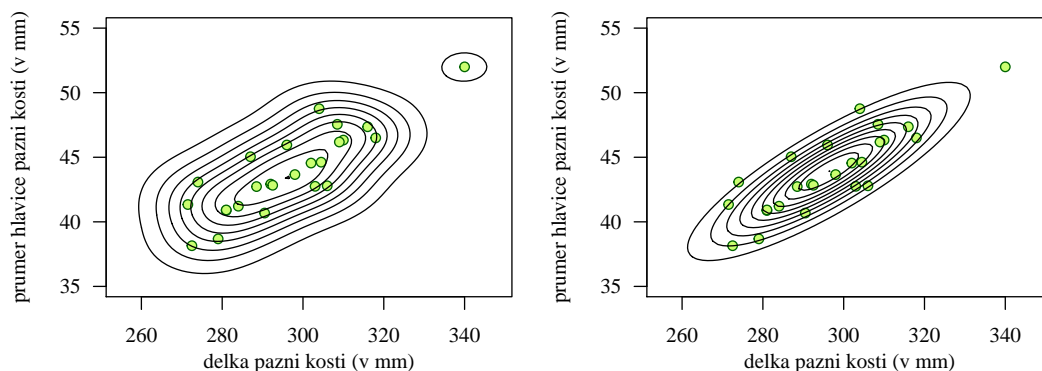
```
57 source('AS-sbirka-funkce.R')
58 fxy <- d2norm(x = humer.LM, y = humer.HDLM, n = n, lim = c(255, 348, 35, 55))
```

Tečkový diagram vykreslíme příkazem `plot()`. Kontury jádrového odhadu hustoty $k(x, y)$ vykreslíme pomocí funkce `contour()`. Prvními třemi vstupními argumenty funkce budou výstupy funkce `kde2d()`, tj. posloupnost `x`, posloupnost `y` a matice `z`. Dále následuje argument `add = T` určující, že se kontury mají vykreslit do stávajícího tečkového diagramu, argument `levels` určující výšky, ve kterých jsou kontury počítány (zde konkrétně dělíme jádrový odhad hustoty do $l = 10$ vrstev oddělených $l - 1$ konturami prostřednictvím posloupnosti o délce l s hodnotami v rozpětí od nuly po maximální hodnotu jádrového odhadu hustoty), a konečně argument `drawlabels = F` potlačující u každé kontury vypsání popisku s výškou kontury. Výsledný graf je zobrazen na obrázku 5.5 vlevo.

```
59 l <- 10
60 plot(humer.LM, humer.HDLM, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1',
61      xlim = c(255, 348), ylim = c(35, 55), xlab = 'delka pazni kosti (v mm)',
62      ylab = 'prumer hlavice pazni kosti (v mm)', las = 1)
63 contour(kxy$x, kxy$y, kxy$z, add = T, levels = seq(0, max(kxy$z), length = l),
        drawlabels = F)
```

Analogickým způsobem vykreslíme tečkový diagram s konturami hustoty $f(x, y)$ dvourozměrného normálního rozložení (viz obrázek 5.5 vpravo).

```
64 plot(humer.LM, humer.HDLM, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1',
65      xlim = c(255, 348), ylim = c(35, 55), xlab = 'delka pazni kosti (v mm)',
66      ylab = 'prumer hlavice pazni kosti (v mm)', las = 1)
67 contour(fxy$x, fxy$y, fxy$z, add = T, levels = seq(0, max(fxy$z), length = l + 1),
68      drawlabels = F)
```



Obrázek 5.5: Dvourozměrný tečkový diagram superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozložení (vpravo)

Vrstevnicový diagram jádrového odhadu hustoty $k(x, y)$ vykreslíme pomocí funkce `image()`. Prvními třemi vstupními argumenty funkce budou výstupy funkce `kde2d()`, tj. posloupnost `x`, posloupnost `y` a matice `z`. Následuje argument `col`, kterým specifikujeme paletu odstínů barev vrstevnicového diagramu. Zde konkrétně volíme $l - 1$ odstínů barev z palety `terrain.colors()`; l -tým odstínem je vždy automaticky přednastavená bílá barva. Volit můžeme i další palety barev, například paletu `heat.colors()` (viz příklad 5.11), `rainbow()`, `topo.colors()`, `cm.colors()`, apod. Konečně argumentem `breaks` specifikujeme výšky oddělující od sebe jednotlivé vrstvy diagramu. Syntax argumentu `breaks` je analogická syntaxi argumentu `levels` příkazu `contour()`. Výsledný graf je zobrazen na obrázku 5.6 vlevo.

```

69 image(kxy$x, kxy$y, kxy$z, xlab = 'delka pazni kosti (v mm)',
70       ylab = 'prumer hlavice pazni kosti (v mm)', las = 1,
71       col = terrain.colors(1 - 1), breaks = seq(0, max(kxy$z), length = 1))
72 contour(kxy$x, kxy$y, kxy$z, add = T, levels = seq(0, max(kxy$z), length = 1),
73         drawlabels = F)

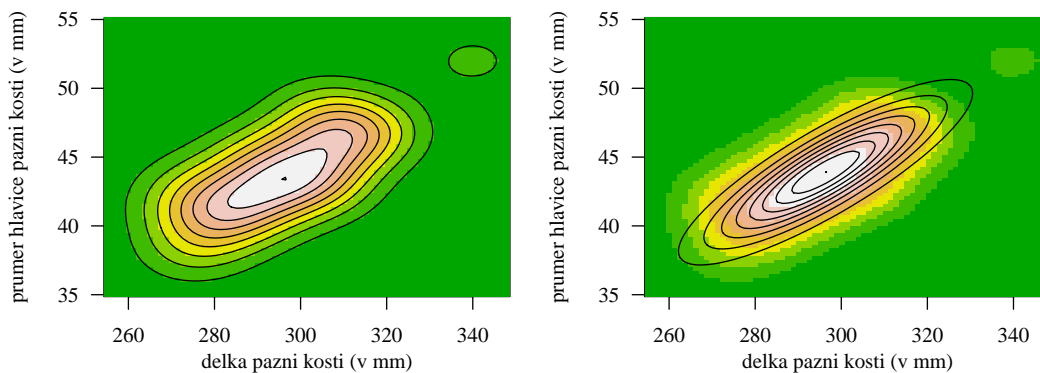
```

Analogickým způsobem vykreslíme vrstevnicový diagram jádrového odhadu hustoty $k(x, y)$ superponovaný konturami hustoty $f(x, y)$ dvourozměrného normálního rozložení (viz obrázek 5.6 vpravo). Takovýto graf slouží k porovnání tvaru dvourozměrných reálných dat s tvarem dvourozměrného normálního rozložení s parametry odhadnutými na základě dat.

```

74 image(kxy$x, kxy$y, kxy$z, xlab = 'delka pazni kosti (v mm)',
75       ylab = 'prumer hlavice pazni kosti (v mm)', las = 1,
76       col = terrain.colors(1 - 1), breaks = seq(0, max(kxy$z), length = 1))
77 contour(fxy$x, fxy$y, fxy$z, add = T, levels = seq(0, max(fxy$z), length = 1),
78         drawlabels = F)

```



Obrázek 5.6: Vrstevnicový diagram jádrového odhadu hustoty superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozložení (vpravo)

3D-diagram jádrového odhadu hustoty $k(x, y)$ vykreslíme pomocí funkce `persp3D()` z knihovny GA. První tři vstupní argumenty funkce jsou stejné jako u příkazů `contour()` a `image()`. Následují argumenty `phi` a `theta` specifikující natočení 3D-diagramu z hlediska osy x a osy y . Dále argumentem `ticktype = 'simple'` nahradíme měřítko os x , y a z ilustrativními šipkami určujícími směr růstu hodnot na osách (toto nastavení používáme pro přehlednější a elegantnější vzhled diagramu), argumentem `col.palette` specifikujeme paletu barev a argumentem `border` stanovíme barvu sítě 3D-diagramu. Výsledný graf je zobrazen na obrázku 5.7 vlevo.

```

79 GA::persp3D(kxy$x, kxy$y, kxy$z, phi = 50, theta = 5, ticktype = 'simple',
80            xlab = 'delka pazni kosti (v mm)',
81            ylab = 'prumer hlavice pazni kosti (v mm)',
82            zlab = 'k(x, y)', col.palette = terrain.colors, border = 'black')

```

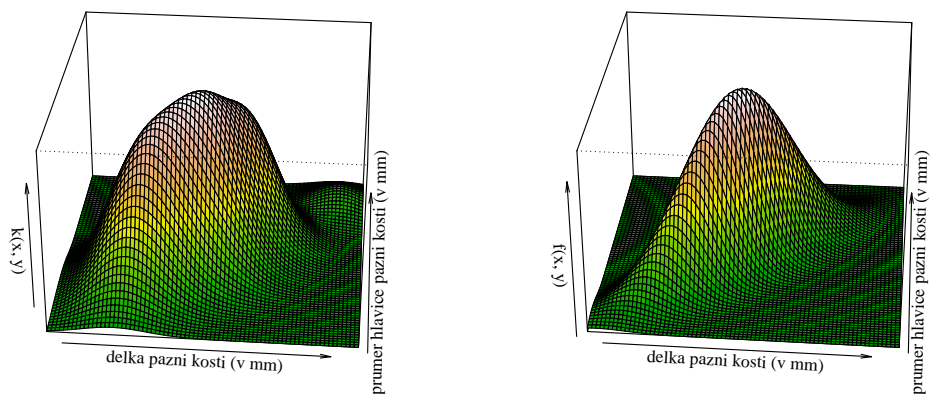
Analogicky vykreslíme 3D-diagram hustoty $f(x, y)$, viz obrázek 5.7 vpravo.

```

83 GA::persp3D(fxy$x, fxy$y, fxy$z, phi = 30, theta = 5, ticktype = 'simple',
84            xlab = 'delka pazni kosti (v mm)',
85            ylab = 'prumer hlavice pazni kosti (v mm)',
86            zlab = 'f(x, y)', col.palette = terrain.colors, border = 'black')

```

Porovnáním grafů zobrazujících jádrový odhad hustoty $k(x, y)$ a hustotu $f(x, y)$ dvourozměrného normálního rozložení na obrázcích 5.5, 5.6 a 5.7 docházíme k závěru, že grafická vizualizace podporuje náš předpoklad, že náhodný vektor $(X, Y)^T$ popisující délku pažní kosti a průměr hlavice pažní kosti z levé strany u mužů z kmene



Obrázek 5.7: 3D-diagram (a) jádrového odhadu hustoty (vlevo); (b) hustoty dvourozměrného normálního rozložení (vpravo)

Tigara pochází z dvourozměrného normálního rozložení s parametry $\mu_1 = 296,34$, $\mu_2 = 43,90$, $\sigma_1^2 = 16,30^2$, $\sigma_2^2 = 3,21^2$ a $\rho = 0,85$. ★

Příklad 5.10. Řešený příklad

Za předpokladu, že náhodný vektor $(X, Y)^T$ popisující délku pažní kosti a průměr hlavičky pažní kosti z levé strany u mužů z kmene Tigara pochází z dvourozměrného normálního rozložení s parametry $\mu_1 = 296,34$, $\mu_2 = 43,90$, $\sigma_1^2 = 16,30^2$, $\sigma_2^2 = 3,21^2$ a $\rho = 0,85$, (a) určete marginální rozložení náhodné veličiny X , resp. Y ; (b) nakreslete graf křivky hustoty $f_X(x)$, resp. $f_Y(y)$ marginálního rozložení spolu s křivkou jádrového odhadu hustoty náhodné veličiny X , resp. Y .

Řešení příkladu 5.10

Za předpokladu, že náhodný vektor $(X, Y)^T$ pochází z dvourozměrného normálního rozložení s parametry $\mu_1 = 296,34$, $\mu_2 = 43,90$, $\sigma_1^2 = 16,30^2$, $\sigma_2^2 = 3,21^2$ a $\rho = 0,85$, můžeme z druhého bodu věty 5.2 odvodit, že náhodná veličina $X \sim N(296,34; 16,30^2)$, a náhodná veličina $Y \sim N(43,90; 3,21^2)$. Délka pažní kosti z levé strany u mužů z kmene Tigara se pohybuje okolo střední hodnoty 296,34 mm se směrodatnou odchylkou 16,30 mm. Průměr hlavičky pažní kosti z levé strany u mužů z kmene Tigara se pohybuje okolo střední hodnoty 43,90 mm se směrodatnou odchylkou 3,21 mm.

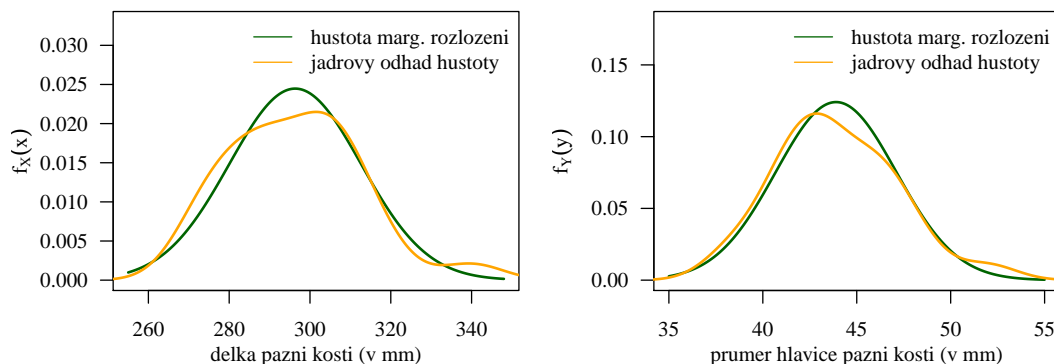
K vykreslení grafu křivky hustoty $f_X(x)$ marginálního rozložení náhodné veličiny X si nejprve vytvoříme posloupnost hodnot z intervalu se středem v bodě 296 mm, tj. například posloupnost v rozmezí 246 mm až 346 mm. Pro každý prvek této posloupnosti potom příkazem `dnorm()` vypočítáme hodnotu hustoty $f_X(x)$ rozložení $N(296,34; 16,30^2)$. Následně příkazem `plot()` s argumentem `type = 'l'` vykreslíme graf s křivkou hustoty $f_X(x)$ a příkazem `lines()` do něj doplníme křivku jádrového odhadu hustoty vypočítaného příkazem `density()`. Výsledný graf je zobrazen na obrázku 5.8 vlevo.

```
87 x <- seq(246, 346, length = 512)
88 fx <- dnorm(x, mL, sL)
89 plot(x, fx, type = 'l', lwd = 2, col = 'darkgreen', ylim = c(0, 0.033),
90       xlab = 'delka pazni kosti (v mm)', ylab = expression(f[X](x)), las = 1)
91 lines(density(humer.LM), col = 'orange', lwd = 2)
92 legend('topright', lwd = 1, col = c('darkgreen', 'orange'), legend = c('hustota marg.
    rozlozeni', 'jadrový odhad hustoty'), bty = 'n')
```

Analogicky vykreslíme graf hustoty $f_Y(y)$ marginálního rozložení a jádrového odhadu hustoty náhodné veličiny Y (viz obrázek 5.8 vpravo).

```

93 y <- seq(35, 55, length = 512)
94 fy <- dnorm(y, mHDL, sHDL)
95 plot(y, fy, type = 'l', lwd = 2, col = 'darkgreen', ylim = c(0, 0.18),
96       xlab = 'prumer hlavice pazni kosti (v mm)', ylab = expression(f[Y](y)),
97       las = 1)
98 lines(density(humer.HDLM), col = 'orange', lwd = 2)
99 legend('topright', lwd = 1, col = c('darkgreen', 'orange'), legend = c('hustota marg.
      rozlozeni', 'jadrovy odhad hustoty'), bty = 'n')
```



Obrázek 5.8: Křivka hustoty marginálního rozložení a jádrového odhadu hustoty (a) náhodné veličiny X (vlevo); (b) náhodné veličiny Y (vpravo)

★

Příklad 5.11. Neřešený příklad

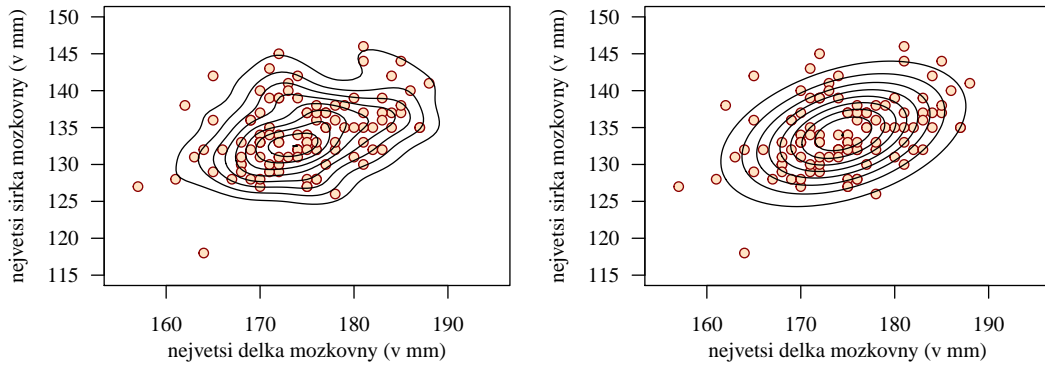
Načtete datový soubor 01-one-sample-mean-skull-mf.txt obsahující údaje o největší délce (skull.L) a šířce skull.B mozkovny mužů a žen ze starověké egyptské populace. Předpokládejme, že náhodný vektor $(X, Y)^T$ popisuje prostřednictvím náhodné veličiny X největší délku mozkovny a prostřednictvím náhodné veličiny Y největší šířku mozkovny žen starověké egyptské populace. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodného vektoru $(X, Y)^T$, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodného vektoru $(X, Y)^T$; (i) vykreslete tečkový diagram a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ nalezeného rozložení; (ii) vykreslete vrstevnicový diagram jádrového odhadu hustoty $k(x, y)$ a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty nalezeného rozložení $f(x, y)$; (iii) vykreslete 3D-diagram jádrového odhadu hustoty $k(x, y)$, resp. 3D-diagram hustoty nalezeného rozložení $f(x, y)$.

Výsledky: (a) $(X, Y)^T \sim N_2 \left((174, 53, 134, 15)^T, \begin{pmatrix} 6, 22^2 & 11, 12 \\ 11, 12 & 4, 70^2 \end{pmatrix} \right)$; (b-i) tečkový diagram superponovaný konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ dvourozměrného normálního rozložení viz obrázek 5.9 vlevo, resp. vpravo; (b-ii) vrstevnicový diagram superponovaný konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ viz obrázek 5.10 vlevo, resp. vpravo; (b-iii) 3D-diagram jádrového odhadu hustoty $k(x, y)$, resp. hustoty $f(x, y)$ viz obrázek 5.11 vlevo, resp. vpravo.

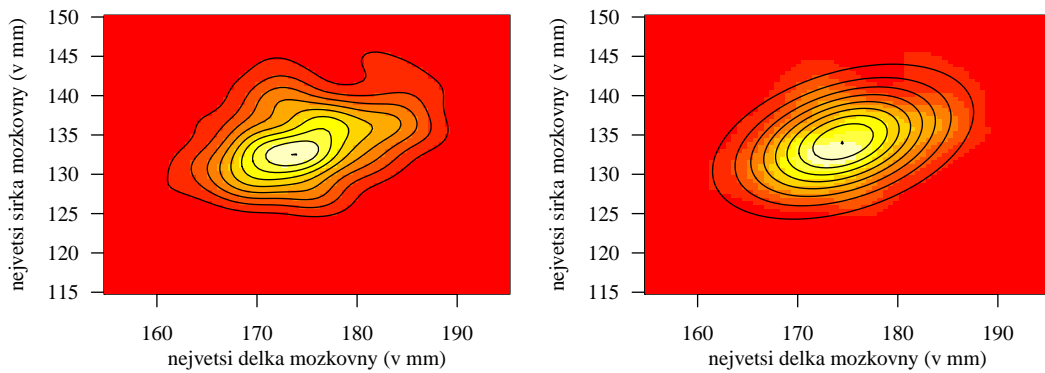
★

Příklad 5.12. Neřešený příklad

Za předpokladu, že náhodný vektor $(X, Y)^T$ popisující největší délku mozkovny a největší šířku mozkovny žen ze starověké egyptské populace pochází z dvourozměrného normálního rozložení s parametry $\mu_1 = 174, 53$, $\mu_2 = 134, 15$, $\sigma_1^2 = 6, 22^2$, $\sigma_2^2 = 4, 70^2$ a $\rho = 0, 38$, (a) určete marginální rozložení náhodné veličiny X , resp. Y ; (b) nakreslete graf křivky hustoty $f_X(x)$, resp. $f_Y(y)$ marginálního rozložení spolu s křivkou jádrového odhadu hustoty náhodné



Obrázek 5.9: Dvouzměrný tečkový diagram superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozložení (vpravo)

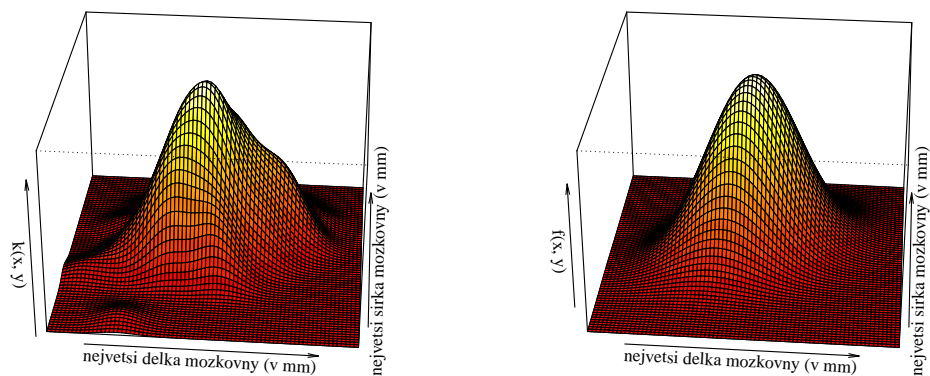


Obrázek 5.10: Vrstevnicový diagram jádrového odhadu hustoty superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozložení (vpravo)

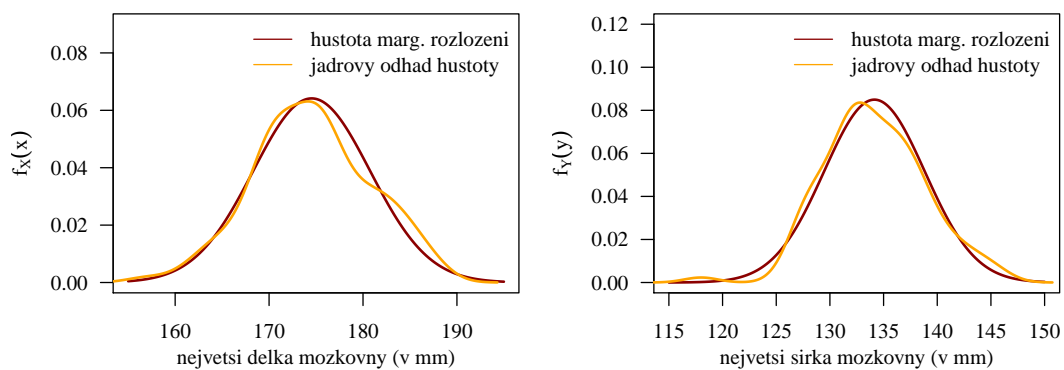
veličiny X , resp. Y .

Výsledky: (a) $X \sim N(174, 53; 6, 22^2)$; $Y \sim N(134, 15; 4, 70^2)$; (b) graf křivky hustoty $f_X(x)$, resp. $f_Y(y)$ marginálního rozložení a jádrového odhadu hustoty náhodné veličiny X , resp. Y viz obrázek 5.12 vlevo, resp. vpravo.

★



Obrázek 5.11: 3D-diagram (a) jádrového odhadu hustoty (vlevo); (b) hustoty dvourozměrného normálního rozložení (vpravo)



Obrázek 5.12: Křivka hustoty marginálního rozložení a jádrového odhadu hustoty (a) náhodné veličiny X (vlevo); (b) náhodné veličiny Y (vpravo)