5 Pravděpodobnostní modely pro spojité náhodné veličiny

5.1 Spojitá náhodná veličina

Spojitá náhodná veličina nabývá 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. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$) a s distribuční funkcí je spjata integrálním vztahem: $\forall x \in \mathbb{R}$: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$. 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. $\forall a, b \in \mathbb{R}$, $a < b$: $\Pr(a \leq X \leq b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \Pr(a < X < b) = F(b) - F(a) = \int_a^b f(x)\mathrm{d}x$.

5.1.1 Normální rozdělení

Normální rozdělení 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 odchylují od skutečné hodnoty μ zkoumané charakteristiky. Píšeme $X \sim N(\mu, \sigma^2)$. Parametr μ má význam střední hodnoty náhodné veličiny X, parametr σ^2, resp. σ představuje její rozptyl, resp. směrodatnou odchylku (viz kapitola 6).

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

Grafem hustoty normálního rozdělení je tzv. Gaussova křivka. Má zvonovitý tvar a je symetrická kolem osy procházející jejím vrcholem, který kulminuje nad hodnotou parametru μ.

Věta 5.1: Důležité vlastnosti normálního rozdělení:

- Jestliže $X \sim N(\mu, \sigma^2)$ a $Z = a + bX$, pak $Z \sim N(a + b\mu, b^2\sigma^2)$.
- Jestliže $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, $Z = \sum_{i=1}^{n} X_i$, pak $Z \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$.
- Jestliže $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, $M = \frac{1}{n}\sum_{i=1}^{n} X_i$, pak $M \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.
- Pravděpodobnost, že náhodná veličina $X \sim N(\mu, \sigma^2)$ se realizuje v nějakém intervalu $\langle a, b\rangle$, nemůžeme vypočítat jako $\Pr(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x$, protože jde o integrál z vyšší transcendentní funkce (funkce, kterou nelze vyjádřit prostřednictvím konečného počtu elementárních funkcí). 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 zvlášť užitečné následující pravděpodobnosti (společně tvoří pravidlo 3σ):

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

.

Poznámka: Uvedené tři rovnosti lze použít k jednoduchému ověření, zda se hodnoty v datovém souboru řídí normálním rozdělením. Z dat vypočítáme aritmetický průměr m a směrodatnou odchylku s. Pokud přibližně 68,27 % hodnot leží v intervalu $(m - s, m + s)$, 95,45 % hodnot leží v intervalu $(m - 2s, m + 2s)$ a 99,73 % hodnot leží v intervalu $(m - 3s, m + 3s)$, lze soudit, že data pocházejí z normálního rozdělení.

Vzorce pro výpočty pravděpodobností pomocí normálního rozdělení v softwaru R

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

5.1.2 Standardizované normální rozdělení

Pro $\mu = 0$, $\sigma^2 = 1$ se jedná o standardizované normální rozdělení, píšeme $U \sim N(0,1)$.
Hustota: $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.
Hodnoty distribuční funkce $F(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$ vypočítáme v softwaru R příkazem pnorm(u). Pro $u \geq 0$ lze hodnoty $F(u)$ nalézt také ve statistických tabulkách. Pro $u < 0$ se při použití tabulek 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
Načtěte datový soubor 13-two-samples-correlations-trunk.txt obsahující údaje o délce dolní končetiny v mm (lowex.L) a délce trupu v mm (tru.L) u mladých dospělých mužů a žen. Předpokládejme, že náhodná veličina X popisuje délku dolní končetiny v mm u žen. (a) Najděte rozdělení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a odhadněte parametry tohoto rozdělení; (b) ověřte, že nalezené rozdělení je vhodné na popis náhodné veličiny X; pro naměřená data vykreslete histogram a superponujte jej křivkou jádrového odhadu hustoty a křivkou hustoty nalezeného rozdělení. Obě křivky vzájemně porovnejte.

Řešení příkladu 5.1
Datový soubor načteme a vypíšeme prvních pět řádků tabulky. Z načteného souboru vybereme pouze údaje o délce dolní končetiny u žen a z výběru odstraníme řádky s chybějícími hodnotami. Nakonec příkazem length() zjistíme rozsah výběru a příkazem range() rozpětí naměřených hodnot.

```
data <- read.delim("13-two-samples-correlations-trunk.txt", sep = "\t")
head(data, n = 5)
lowex.LF <- na.omit(data[data$sex == "f", "lowex.L"])
n <- length(lowex.LF) # 100
range(lowex.LF) # 836; 1076
```

Datový soubor obsahuje údaje o délce dolní končetiny u 100 žen. Naměřené hodnoty se pohybují v rozmezí 836 mm až 1076 mm. Délka dolní končetiny u 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 rozdělení. O náhodné veličině X tedy předpokládáme, že pochází z normálního rozdělení s parametry μ a σ^2, kde μ je střední hodnota délky dolní končetiny u žen a σ^2 je rozptyl délky dolní končetiny u žen okolo střední hodnoty μ. Parametr μ odhadneme pomocí aritmetického průměru naměřených hodnot, tj.

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

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

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

Aritmetický průměr m vypočítáme příkazem mean(). Rozptyl s^2 budeme častěji zapisovat ve tvaru druhé mocniny směrodatné odchylky, jejíž hodnotu vypočítáme příkazem sd().

```
m <- mean(lowex.LF) # 940,5
s <- sd(lowex.LF) # 45,47116
```

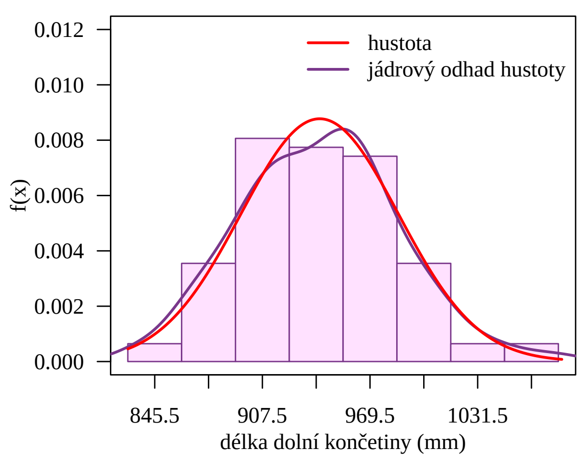
Odhad parametru střední hodnoty $m = 940{,}50$, odhad parametru rozptylu $s^2 = 45{,}47^2$. Délka dolní končetiny u žen se pohybuje okolo střední hodnoty 940,50 mm se směrodatnou odchylkou 45,47 mm. O náhodné veličině X předpokládáme, že pochází z normálního rozdělení s parametry $\mu = 940{,}50$ a $\sigma^2 = 45{,}47^2$, 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 rozdělení $N(940{,}50; 45{,}47^2)$. Naměřené hodnoty roztřídíme do osmi třídicích intervalů o ekvidistantní délce 31 mm stanovením hranic 830, 861, …, 1078 mm. Histogram vykreslíme 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.

```
r <- round(1 + 3.3 * log10(n)) # 8
# 1076 - 835 = 241; 248 / 8 = 31 -> seq(830, 1078, by = 31)
b.lowex.LF <- seq(from = 830, to = 1078, by = 31)
hist(lowex.LF, prob = T, breaks = b.lowex.LF, include.lowest = F, ylim = c(0, 0.012),
    axes = F, col = "thistle1", border = "mediumorchid4", xlab = "délka dolní končetiny
    (mm)", ylab = "f(x)", main = "")
box(bty = "o")
c.lowex.LF <- hist(lowex.LF, breaks = b.lowex.LF, plot = F)$mids
axis(side = 1, at = c.lowex.LF)
axis(side = 2, las = 1)
lines(density(lowex.LF), lwd = 2, col = "mediumorchid4")
```

Pro vykreslení křivky hustoty rozdělení $N(940,50; 45,47^2)$ je třeba nejprve příkazem seq() vytvořit posloupnost hodnot x pokrývající rozpětí naměřených délek dolní končetiny u žen. Pro každý prvek této posloupnosti potom příkazem dnorm() vypočítáme hodnotu hustoty rozdělení $N(940,50; 45,47^2)$. Vstupními argumenty příkazu budou posloupnost x (argument x), odhad parametru μ (argument mean) a odhad parametru σ (argument sd). Křivku hustoty vykreslíme příkazem lines(). Nakonec do grafu doplníme legendu popisující význam vykreslených křivek. Výsledný histogram je zobrazen na obrázku 5.1.

```
x <- seq(from = 830, to = 1080, length = 512)
fx <- dnorm(x = x, mean = m, sd = s)
lines(x, fx, col = "red", lwd = 2)
legend("topright", bty = "n", lwd = c(2, 2), col = c("red", "mediumorchid4"), legend =
    c("hustota", "jádrový odhad hustoty"))
```



Obrázek 5.1: Histogram pro délku dolní končetiny u žen superponovaný křivkou jádrového odhadu hustoty a křivkou hustoty normálního rozdělení

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

Poznámka: Grafická vizualizace je nástrojem sloužícím k subjektivnímu posouzení normálního rozdělení 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 v mm u žen pochází z rozdělení $N(940,50; 45,47^2)$, vypočítejte pravděpodobnost, že délka dolní končetiny u 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. K výpočtu pravděpodobností použijte nezaokrouhlené odhady parametrů μ a σ^2 vypočítané v příkladu 5.1.

Řešení příkladu 5.2

Pravděpodobnost, že délka dolní končetiny u žen bude menší než $962,50$ mm, tj. $\Pr(X < 962,50) = \Pr(X \leq 962,50)$ odpovídá hodnotě distribuční funkce $F(x)$ rozdělení $N(940,50; 45,47^2)$ v bodě $x = 962,50$. Tuto hodnotu vypočítáme příkazem pnorm(). Vstupními argumenty příkazu budou hodnota x (argument q), odhad parametru μ (argument mean) a odhad parametru σ (argument sd).

```
pnorm(q = 962.50, mean = m, sd = s) # 0,6857443
```

Pravděpodobnost, že délka dolní končetiny u 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 u žen bude nejvýše $974,13$ mm. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F(x)$ v bodě $x = 974,13$, vypočítáme příkazem pnorm().

```
1 - pnorm(q = 974.13, mean = m, sd = s) # 0,2297745
```

Pravděpodobnost, že délka dolní končetiny u 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 u žen bude nejvýše $950,00$ mm, odečteme pravděpodobnost, že délka dolní končetiny u 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 příkazem pnorm().

```
pnorm(q = 950.00, mean = m, sd = s) - pnorm(q = 930.00, mean = m, sd = s) # 0,1740559
```

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

Pravděpodobnost, že délka dolní končetiny u žen bude menší než $962,50$ mm, je $0,6857$ ($68,57$ %). Pravděpodobnost, že délka dolní končetiny u žen bude větší než $974,13$ mm, je $0,2298$ ($22,98$ %). Pravděpodobnost, že délka dolní končetiny u žen bude v rozmezí $930,00$ mm až $950,00$ mm, je $0,1741$ ($17,41$ %). Pravděpodobnost, že délka dolní končetiny u žen bude rovná $943,25$ mm, je $0,0000$ ($0,00$ %). ★

Příklad 5.3. Řešený příklad

Za předpokladu, že náhodná veličina X popisující délku dolní končetiny v mm u žen pochází z rozdělení $N(940,50; 45,47^2)$, (a) zjistěte rozdělení náhodné veličiny $M = \frac{1}{8}\sum_{i=1}^{8} X_i$ popisující průměrnou délku dolní končetiny v mm u osmi žen a odhadněte parametry tohoto rozdělení; (b) vypočítejte pravděpodobnost, že průměrná délka dolní končetiny u 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. K výpočtu odhadů parametrů rozdělení náhodné veličiny M použijte nezaokrouhlené odhady parametrů μ a σ^2 vypočítané v příkladu 5.1.

Řešení příkladu 5.3

Za předpokladu, že náhodná veličina X pochází z normálního rozdělení 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 rozdělení s parametry $\mu_M = \mu = 940,50$ a $\sigma_M^2 = \frac{45,47^2}{8} = 16,08^2$, tj. $M \sim N(940,50; 16,08^2)$.

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

Průměrná délka dolní končetiny u osmi žen se pohybuje okolo střední hodnoty $940,50$ mm se směrodatnou odchylkou $16,08$ mm.

Pravděpodobnost, že průměrná délka dolní končetiny u osmi žen bude menší než $962,50$ mm, tj. $\Pr(M < 962,50) = \Pr(M \leq 962,50)$ odpovídá hodnotě distribuční funkce $F_M(x)$ rozdělení $N(940,50; 16,08^2)$ v bodě $x = 962,50$. Tuto hodnotu vypočítáme příkazem pnorm().

```
pnorm(q = 962.50, mean = m, sd = sM) # 0,9144157
```

Pravděpodobnost, že průměrná délka dolní končetiny u 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 u osmi žen bude nejvýše $974,13$ mm. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F_M(x)$ v bodě $x = 974,13$, vypočítáme příkazem pnorm().

```
1 - pnorm(q = 974.13, mean = m, sd = sM) # 0,01822483
```

Pravděpodobnost, že průměrná délka dolní končetiny u 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 u osmi žen bude nejvýše $950,00$ mm, odečteme pravděpodobnost, že průměrná délka dolní končetiny u 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 příkazem pnorm().

```
pnorm(q = 950.00, mean = m, sd = sM) - pnorm(q = 930.00, mean = m, sd = sM) # 0,4658779
```

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

Pravděpodobnost, že průměrná délka dolní končetiny u osmi žen bude menší než $962,50$ mm, je $0,9144$ ($91,44$ %). Pravděpodobnost, že průměrná délka dolní končetiny u osmi žen bude větší než $974,13$ mm, je $0,0182$ ($1,82$ %). Pravděpodobnost, že průměrná délka dolní končetiny u osmi žen bude v rozmezí $930,00$ mm až $950,00$ mm, je $0,4659$ ($46,59$ %). Pravděpodobnost, že průměrná délka dolní končetiny u osmi žen bude rovná $943,25$ mm, je $0,0000$ ($0,00$ %).

★

Příklad 5.4. Řešený příklad

Předpokládejme, že náhodná veličina X popisující délku dolní končetiny v mm u žen pochází z rozdělení $N(940,50; 45,47^2)$ a náhodná veličina M popisující průměrnou délku dolní končetiny v mm u osmi žen pochází z rozdělení $N(940,50; 16,08^2)$. 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.

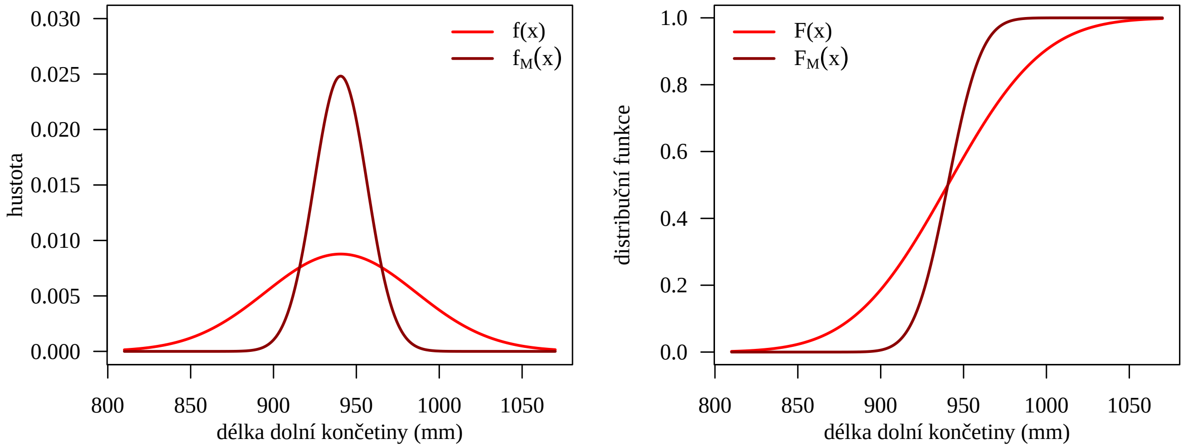
Řešení příkladu 5.4

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 mm až 1071 mm. Pro každý prvek této posloupnosti potom příkazem dnorm() vypočítáme hodnotu hustoty $f(x)$ rozdělení $N(940,50; 45,47^2)$, resp. hustoty $f_M(x)$ rozdělení $N(940,50; 16,08^2)$. 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.

```
x <- seq(from = 810, to = 1071, length = 512)
fx <- dnorm(x = x, mean = m, sd = s)
fM <- dnorm(x = x, mean = m, sd = sM)
plot(x, fx, type = "l", las = 1, ylim = c(0, 0.03), lwd = 2, col = "red", xlab = "délka
    dolní končetiny (mm)", ylab = "hustota")
lines(x, fM, lwd = 2, col = "darkred")
legend("topright", bty = "n", lwd = c(2, 2), col = c("red", "darkred"), legend =
    c("f(x)", expression(f[M](x))))
```

Nyní se zaměříme na vykreslení grafu s křivkami distribučních funkcí. Nejprve pro každý prvek posloupnosti z rozmezí 810 mm až 1071 mm vypočítáme pomocí funkce pnorm() hodnotu distribuční funkce $F(x)$ rozdělení $N(940,50; 45,47^2)$, resp. distribuční funkce $F_M(x)$ rozdělení $N(940,50; 16,08^2)$. 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.

```
Fx <- pnorm(q = x, mean = m, sd = s)
FM <- pnorm(q = x, mean = m, sd = sM)
plot(x, Fx, type = "l", las = 1, lwd = 2, col = "red", xlab = "délka dolní končetiny
    (mm)", ylab = "distribuční funkce")
lines(x, FM, lwd = 2, col = "darkred")
legend("topleft", bty = "n", lwd = c(2, 2), col = c("red", "darkred"), legend =
    c("F(x)", expression(F[M](x))))
```



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

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 hodnot blízkých nule, od určitého okamžiku pak začíná strmě růst až do vrcholu v bodě 940, 50 mm, a následně strmě klesá. Ve srovnání s křivkou hustoty $f_M(x)$ křivka hustoty $f(x)$ od začátku roste pozvolna až do svého vrcholu v bodě 940, 50 mm a od tohoto bodu pozvolna klesá. Tento průběh je důsledkem osmkrát většího rozptylu rozdělení náhodné veličiny X. Z obrázku 5.2 vpravo vidíme, že křivka distribuční funkce $F_M(x)$ nabývá na poměrně dlouhém úseku hodnot blízkých nule, následně začná 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 délky dolní končetiny u žen je menších než 940, 50 mm a stejně tak 50 % hodnot průměrné délky dolní končetiny u osmi žen je menších než 940, 50 mm.

★

Příklad 5.5. Neřešený příklad

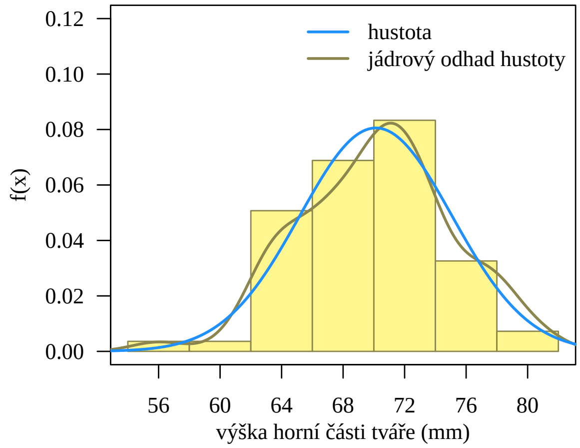
Načtěte datový soubor 15-anova-means-skull.txt obsahující údaje o výšce horní části tváře v mm (upface.H) u mužů bantuské, čínské, malajské, německé a peruánské populace. Předpokládejme, že náhodná veličina X popisuje výšku horní části tváře v mm u mužů malajské populace. (a) Najděte rozdělení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a odhadněte parametry tohoto rozdělení; (b) ověřte, že nalezené rozdělení je vhodné na popis náhodné veličiny X; pro naměřená data vykreslete histogram a superponujte jej křivkou jádrového odhadu hustoty a křivkou hustoty nalezeného rozdělení. 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 rozdělení $N(70,13;4,95^2)$ viz obrázek 5.3.

★

Příklad 5.6. Neřešený příklad

Za předpokladu, že náhodná veličina X popisující výšku horní části tváře v mm u mužů malajské populace pochází z rozdělení $N(70,13;4,95^2)$, vypočítejte pravděpodobnost, že výška horní části tváře u mužů malajské populace bude (a) větší než 70,75 mm; (b) menší než 66,15 mm; (c) rovná 82,30 mm; (d) v rozmezí 64,40 mm až 72,40 mm. K výpočtu pravděpodobností použijte nezaokrouhlené odhady parametrů μ a σ^2 vypočítané v příkladu 5.5.
Výsledky: (a) $\Pr(X > 70,75) = 0,4502$; (b) $\Pr(X < 66,15) = 0,2108$; (c) $\Pr(X = 82,30) = 0,0000$; (d) $\Pr(64,40 < X < 72,40) = 0,5530$.

★



Obrázek 5.3: Histogram pro výšku horní části tváře u mužů superponovaný křivkou jádrového odhadu hustoty a křivkou hustoty normálního rozdělení

Příklad 5.7. Neřešený příklad
Za předpokladu, že náhodná veličina X popisující výšku horní části tváře v mm u mužů malajské populace pochází z rozdělení $N(70,13;4,95^2)$, (a) zjistěte rozdělení náhodné veličiny $M = \frac{1}{5}\sum_{i=1}^{5} X_i$ popisující průměrnou výšku horní části tváře v mm u pěti mužů malajské populace a odhadněte parametry tohoto rozdělení; (b) vypočítejte pravděpodobnost, že průměrná výška horní části tváře u pěti mužů malajské populace bude (i) větší než 70,75 mm; (ii) menší než 66,15 mm; (iii) rovná 82,30 mm; (iv) v rozmezí 64,40 mm až 72,40 mm. K výpočtu odhadů parametrů rozdělení náhodné veličiny M použijte nezaokrouhlené odhady parametrů μ a σ^2 vypočítané v příkladu 5.5.
Výsledky: (a) $M \sim N(70,13;2,21^2)$; (b-i) $\Pr(M > 70,75) = 0,3898$; (b-ii) $\Pr(M < 66,15) = 0,0362$; (b-iii) $\Pr(M = 82,30) = 0,0000$; (b-iv) $\Pr(64,40 < M < 72,40) = 0,8424$.

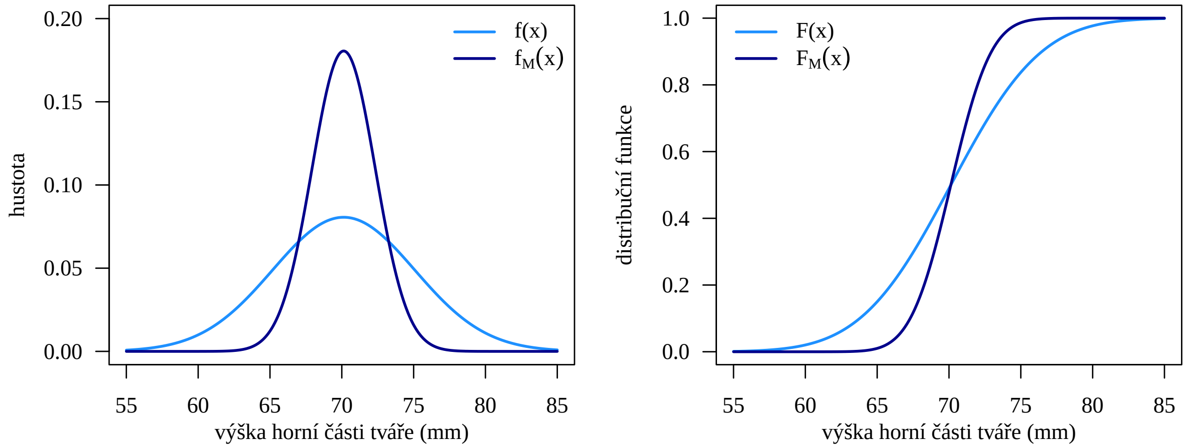
★

Příklad 5.8. Neřešený příklad
Předpokládejme, že náhodná veličina X popisující výšku horní části tváře v mm u mužů malajské populace pochází z rozdělení $N(70,13;4,95^2)$ a náhodná veličina M popisující průměrnou výšku horní části tváře v mm u pěti mužů malajské populace pochází z rozdělení $N(70,13;2,21^2)$. 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. ★

5.2 Spojitý dvourozměrný náhodný vektor

Jestliže obě složky dvourozměrného náhodného vektoru $(X,Y)^T$ jsou spojité 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 pravděpodobnosti 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:

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



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

5.2.1 Dvourozměrné normální rozdělení

Dvourozměrné normální rozdělení vzniká ve dvourozměrných situacích za podobných okolností jako normální rozdělení. 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 $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho \in \langle -1; 1 \rangle$ se jedná o standardizované dvourozměrné normální rozdělení. Píšeme $\mathbf{U} = (U_1, U_2)^T \sim N_2(\mathbf{0}, \mathbf{I})$, kde $\mathbf{0} = (0,0)^T$ je nulový vektor a $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ je jednotková matice druhého řádu.

Věta 5.2: Důležité vlastnosti dvourozměrného normálního rozdělení:

- 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 rozdělení 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 rozdělení je však nekorelovanost ekvivalentní se stochastickou nezávislostí veličin X a Y.

Příklad 5.9. Řešený příklad
Načtěte datový soubor 31-goldman-alaska-hfrt.csv obsahující údaje o délce pažní kosti z levé strany v mm (humerus.L) a vertikálním průměru (zkráceně též v. p.) hlavice pažní kosti z levé strany v mm (humerus.HDL) u mužů a žen tří aljašských populací. Předpokládejme, že máme náhodný vektor $(X,Y)^T$, kde náhodná veličina X popisuje délku pažní kosti z levé strany v mm a náhodná veličina Y popisuje vertikální průměr hlavice pažní kosti z levé strany v mm u mužů z kmene Tigara. (a) Najděte rozdělení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodného vektoru $(X,Y)^T$, a odhadněte parametry tohoto rozdělení; (b) ověřte, že nalezené rozdělení je vhodné na popis náhodného vektoru $(X,Y)^T$; (i) vykreslete dvourozměrný tečkový diagram a superponujte jej konturami jádrového odhadu hustoty $k(x,y)$, resp. konturami hustoty $f(x,y)$ nalezeného rozdělení; (ii) vykreslete

vrstevnicový diagram jádrového odhadu hustoty $k(x,y)$ a superponujte jej konturami jádrového odhadu hustoty $k(x,y)$, resp. konturami hustoty $f(x,y)$ nalezeného rozdělení; (iii) vykreslete 3D-diagram (trojrozměrný diagram) jádrového odhadu hustoty $k(x,y)$, resp. 3D-diagram hustoty $f(x,y)$ nalezeného rozdělení. K vykreslení vrstevnicových diagramů a 3D-diagramů použijte 10 barev z palety terrain.colors().

Řešení příkladu 5.9

Datový soubor načteme a vypíšeme prvních pět řádků a deset sloupců tabulky. Z načteného souboru vybereme údaje o délce pažní kosti a vertikálním průměru hlavice pažní kosti z levé strany u mužů z kmene Tigara a z výběru odstraníme řádky s chybějícími hodnotami. Z výběru separujeme naměřené délky pažní kosti a naměřené vertikální průměry hlavice pažní kosti, zjistíme rozsah datového souboru a rozpětí naměřených hodnot každého znaku.

```
data <- read.delim("31-goldman-alaska-hfrt.csv", sep = ";", dec = ".")
head(data, n = c(5, 10))
```

```
  loc pop sex humerus.L humerus.R humerus.HDL humerus.HDR humerus.ADL humerus.ADR femur.L
1 ala tig   m     308.5        NA       47.55          NA       22.00          NA     443
2 ala ipi   m     311.0       310       44.44       44.11       22.12       22.68     415
3 ala ipi   m     289.0       298       42.94       44.41       20.36       22.09     398
4 ala ipi   f     295.0       302       42.51       42.06       19.35       19.97     395
5 ala ipi   f     270.5       281       39.74       39.84       19.42       19.38      NA
```

```
data.M <- na.omit(data[data$sex == "m" & data$pop == "tig", c("humerus.L",
    "humerus.HDL")])
humerus.LM <- data.M$humerus.L
humerus.HDLM <- data.M$humerus.HDL
length(humerus.LM) # 25
range(humerus.LM) # 271,5; 340,0
range(humerus.HDLM) # 38,15; 52,00
```

Datový soubor obsahuje údaje o délce pažní kosti a vertikálním průměru hlavice pažní kosti z levé strany u 25 mužů z kmene Tigara. Naměřené hodnoty délky pažní kosti se pohybují v rozmezí $271,5$ mm až $340,00$ mm. Naměřené hodnoty vertikálního průměru hlavice pažní kosti se pohybují v rozmezí $38,15$ mm až $52,00$ mm.

Délka pažní kosti i vertikální 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 rozdělení. O náhodném vektoru $(X,Y)^T$ tedy předpokládáme, že pochází z dvourozměrného normálního rozdělení 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 vertikálního průměru hlavice pažní kosti a ρ je koeficient korelace. Parametr μ_1, resp. σ_1^2 odhadneme pomocí aritmetického průměru m_1, resp. rozptylu s_1^2 naměřených hodnot délky pažní kosti. Parametr μ_2, resp. σ_2^2 odhadneme pomocí aritmetického průměru m_2, resp. rozptylu s_2^2 naměřených hodnot vertikálního průměru hlavice pažní kosti. 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 zapisovat ve tvaru druhých mocnin směrodatných odchylek, jejichž hodnoty vypočítáme příkazem sd(). Pearsonův koeficient korelace r_{12} vypočítáme příkazem cor() s argumentem method = "pearson".

```
mL <- mean(humerus.LM) # 296,34
mHDL <- mean(humerus.HDLM) # 43,8984
sL <- sd(humerus.LM) # 16,30294
sHDL <- sd(humerus.HDLM) # 3,21493
r12 <- cor(humerus.LM, humerus.HDLM, method = "pearson") # 0,8523391
```

Odhady parametrů středních hodnot $m_1 = 296,34$, $m_2 = 43,90$, odhady parametrů rozptylů $s_1^2 = 16,30^2$, $s_2^2 = 3,21^2$, odhad parametru korelačního koeficientu $r_{12} = 0,85$. O náhodném vektoru $(X,Y)^T$ předpokládáme, že pochází z dvourozměrného normálního rozdělení 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 $f(x,y)$ rozdělení

$N_2\left((296,34;\, 43,90)^T, \begin{pmatrix} 16,30^2 & 44,67 \\ 44,67 & 3,21^2 \end{pmatrix}\right)$. Jádrový odhad hustoty $k(x,y)$ vypočítáme příkazem kde2d() z knihovny MASS. Vstupními argumenty příkazu budou vektory humerus.LM (argument x) a humerus.HDLM (argument y), argument n specifikující počet bodů mřížky, pro kterou bude jádrový odhad počítán (je-li n = c(75, 75), pak celkový počet bodů mřížky bude 75×75), a argument lim = c($\mathsf{x}_{\mathsf{min}}, \mathsf{x}_{\mathsf{max}}, \mathsf{y}_{\mathsf{min}}, \mathsf{y}_{\mathsf{max}}$) určující rozpětí hodnot z hlediska osy x a z hlediska osy y. Výstupem funkce bude posloupnost bodů osy x o délce 75 s hodnotami v rozpětí $\mathsf{x}_{\mathsf{min}}$ až $\mathsf{x}_{\mathsf{max}}$, posloupnost bodů osy y o délce 75 s hodnotami v rozpětí $\mathsf{y}_{\mathsf{min}}$ až $\mathsf{y}_{\mathsf{max}}$ a matice z, která má dimenzi 75×75 a obsahuje pro každý bod $[x_i, y_j]$, $i, j = 1, \ldots, 75$, hodnotu jádrového odhadu hustoty.

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

Hustotu dvourozměrného normálního rozdělení $f(x,y)$ vypočítáme pomocí funkce d2norm() implementované v skriptu AS1-sbirka-funkce.R. Syntax a výstupy funkce d2norm() jsou analogické jako u funkce kde2d().

```
source("AS1-sbirka-funkce.R")
fxy <- d2norm(x = humerus.LM, y = humerus.HDLM, n = c(75, 75), lim = c(255, 348, 35,
    55))
```

Dvourozměrný tečkový diagram vykreslíme příkazem plot(). Argumenty xlim a ylim nastavíme na stejné hodnoty, jaké jsme použili v argumentu lim příkazu kde2d(). Kontury jádrového odhadu hustoty $k(x,y)$ vykreslíme příkazem contour(). Prvními třemi vstupními argumenty příkazu budou výstupy funkce kde2d(), tj. posloupnost x, posloupnost y a matice z. Dále následuje argument add = T určující, že se kontury vykreslí do stávajícího tečkového diagramu, argument levels specifikující výšky kontur (příkazem seq() nadefinujeme posloupnost $l = 10$ výšek kontur v rozpětí od nuly po maximální hodnotu jádrového odhadu hustoty; tyto kontury od sebe oddělují $l - 1$ vrstev) a argument drawlabels = F potlačující u každé kontury vypsání popisku s výškou kontury. Výsledný graf je zobrazen na obrázku 5.5 vlevo.

```
l <- 10
plot(humerus.LM, humerus.HDLM, las = 1, pch = 21, col = "darkgreen", bg =
    "darkolivegreen1", xlim = c(255, 348), ylim = c(35, 55), xlab = "délka pažní kosti
    (mm)", ylab = "v. p. hlavice pažní kosti (mm)")
contour(kxy$x, kxy$y, kxy$z, add = T, levels = seq(0, max(kxy$z), length = l),
    drawlabels = F)
```

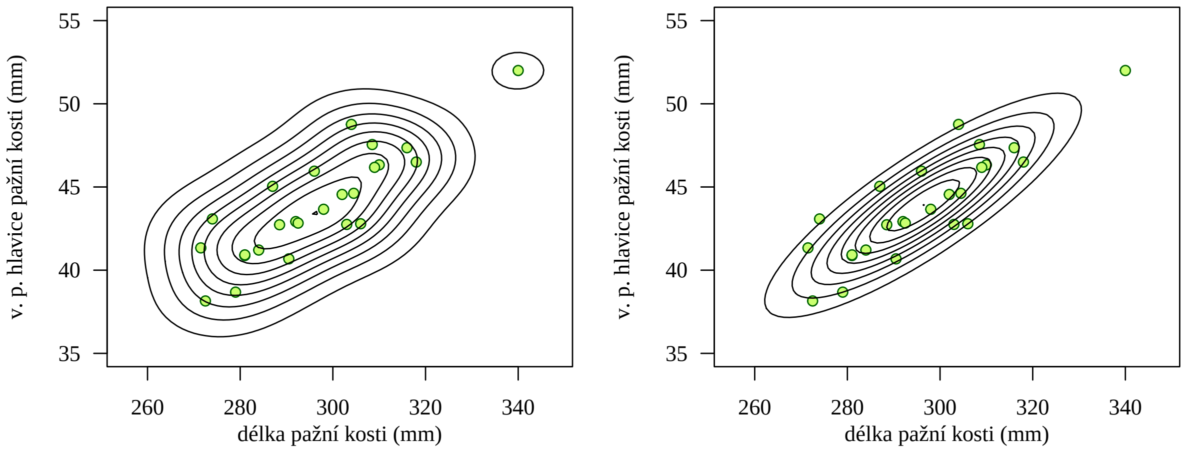
Analogickým způsobem vykreslíme tečkový diagram s konturami hustoty $f(x,y)$ dvourozměrného normální rozdělení (viz obrázek 5.5 vpravo).

```
plot(humerus.LM, humerus.HDLM, las = 1, pch = 21, col = "darkgreen", bg =
    "darkolivegreen1", xlim = c(255, 348), ylim = c(35, 55), xlab = "délka pažní kosti
    (mm)", ylab = "v. p. hlavice pažní kosti (mm)")
contour(fxy$x, fxy$y, fxy$z, add = T, levels = seq(0, max(fxy$z), length = l),
    drawlabels = F)
```

Vrstevnicový diagram jádrového odhadu hustoty $k(x,y)$ vykreslíme příkazem image(). Prvními třemi vstupními argumenty příkazu 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(), neboť l-tým odstínem je vždy automaticky přednastavená bílá barva. Volit můžeme i další palety barev, například paletu heat.colors(), topo.colors(), cm.colors(), hcl.colors(), rainbow(), apod. Dále argumentem breaks specifikujeme výšky oddělující od sebe jednotlivé vrstvy diagramu. Syntax argumentu breaks je analogická syntaxi argumentu levels funkce contour(). Nakonec v příkazu nastavíme otočení popisků měřítka osy y o 90° (argument las), popisek osy x (argument xlab) a popisek osy y (argument ylab). Kontury doplníme do vrstevnicového diagramu zvlášť příkazem contour(). Výsledný graf je zobrazen na obrázku 5.6 vlevo.

```
image(kxy$x, kxy$y, kxy$z, col = terrain.colors(l - 1), breaks = seq(0, max(kxy$z),
    length = l), las = 1, xlab = "délka pažní kosti (mm)", ylab = "v. p. hlavice pažní
    kosti (mm)")
contour(kxy$x, kxy$y, kxy$z, add = T, levels = seq(0, max(kxy$z), length = l),
    drawlabels = F)
```

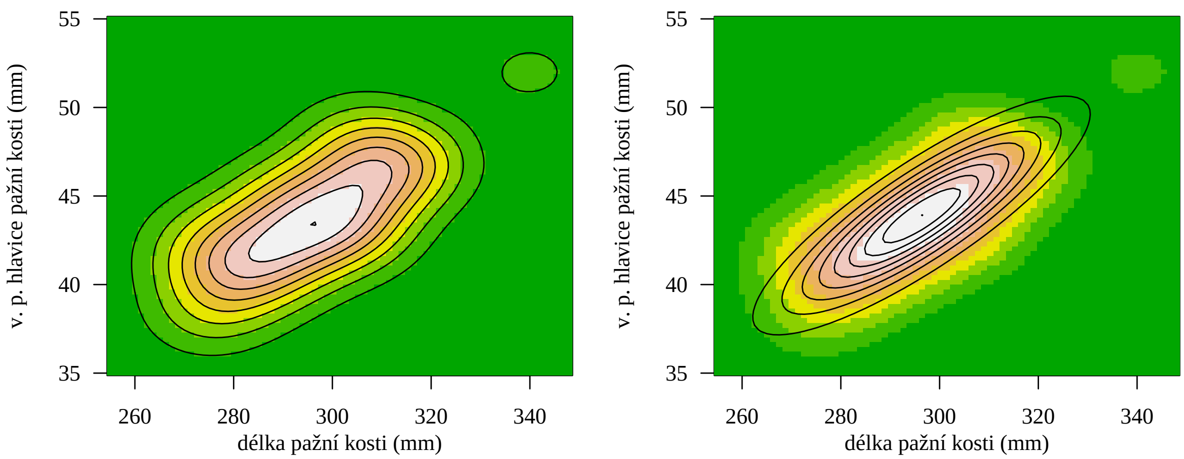
Analogickým způsobem vykreslíme vrstevnicový diagram jádrového odhadu hustoty $k(x,y)$ superponovaný konturami hustoty $f(x,y)$ dvourozměrného normálního rozdělení (viz obrázek 5.6 vpravo). Diagram slouží k porovnání



Obrázek 5.5: Dvourozměrný tečkový diagram superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozdělení (vpravo) pro délku pažní kosti a vertikální průměr hlavice pažní kosti z levé strany u mužů

tvaru dvourozměrných reálných dat s tvarem dvourozměrného normálního rozdělení s parametry odhadnutými na základě dat.

```
image(kxy$x, kxy$y, kxy$z, col = terrain.colors(l - 1), breaks = seq(0, max(kxy$z),
    length = l), las = 1, xlab = "délka pažní kosti (mm)", ylab = "v. p. hlavice pažní
    kosti (mm)")
contour(fxy$x, fxy$y, fxy$z, add = T, levels = seq(0, max(fxy$z), length = l),
    drawlabels = F)
```



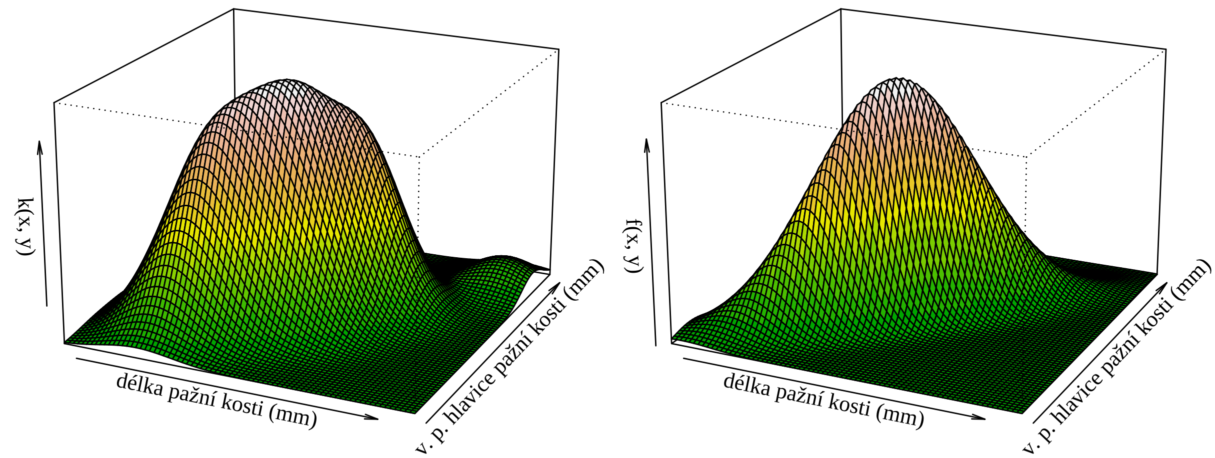
Obrázek 5.6: Vrstevnicový diagram jádrového odhadu hustoty superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozdělení (vpravo) pro délku pažní kosti a vertikální průměr hlavice pažní kosti z levé strany u mužů

3D-diagram jádrového odhadu hustoty $k(x, y)$ vykreslíme příkazem persp3D() z knihovny GA. První tři vstupní argumenty příkazu jsou opět výstupy funkce kde2d(). 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 ukazujícími směr růstu hodnot na osách (toto nastavení používáme pro přehlednější a elegantnější vzhled 3D-diagramu), argumentem col.palette vybereme paletu barev a argumentem border = "black" stanovíme černou barvu sítě 3D-diagramu. Nakonec specifikujeme popisek osy x, y a z argumenty xlab, ylab a zlab. Výsledný diagram je zobrazen na obrázku 5.7 vlevo.

```
GA::persp3D(kxy$x, kxy$y, kxy$z, phi = 20, theta = 25, ticktype = "simple", col.palette
    = terrain.colors, border = "black", xlab = "délka pažní kosti (mm)", ylab = "v. p.
    hlavice pažní kosti (mm)", zlab = "k(x, y)")
```

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

```
GA::persp3D(fxy$x, fxy$y, fxy$z, phi = 20, theta = 25, ticktype = "simple", col.palette
    = terrain.colors, border = "black", xlab = "délka pažní kosti (mm)", ylab = "v. p.
    hlavice pažní kosti (mm)", zlab = "f(x, y)")
```



Obrázek 5.7: 3D-diagram jádrového odhadu hustoty (vlevo); 3D-diagram hustoty dvourozměrného normálního rozdělení (vpravo) pro délku pažní kosti a vertikální průměr hlavice pažní kosti z levé strany u mužů

Porovnáním grafů zobrazujících jádrový odhad hustoty $k(x, y)$ a hustotu $f(x, y)$ dvourozměrného normálního rozdělení na obrázcích 5.5, 5.6 a 5.7 zjistíme, že grafická vizualizace podporuje náš závěr, že náhodný vektor $(X, Y)^T$ popisující délku pažní kosti z levé strany v mm a vertikální průměr hlavice pažní kosti z levé strany v mm u mužů z kmene Tigara pochází z dvourozměrného normálního rozdělení 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 z levé strany v mm a vertikální průměr hlavice pažní kosti z levé strany v mm u mužů z kmene Tigara pochází z dvourozměrného normálního rozdělení 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) zjistěte marginální rozdělení náhodné veličiny X, resp. Y; (b) nakreslete graf křivky hustoty $f_X(x)$, resp. $f_Y(y)$ marginálního rozdělení spolu s křivkou jádrového odhadu hustoty náhodné veličiny X, resp. Y.

Řešení příkladu 5.10

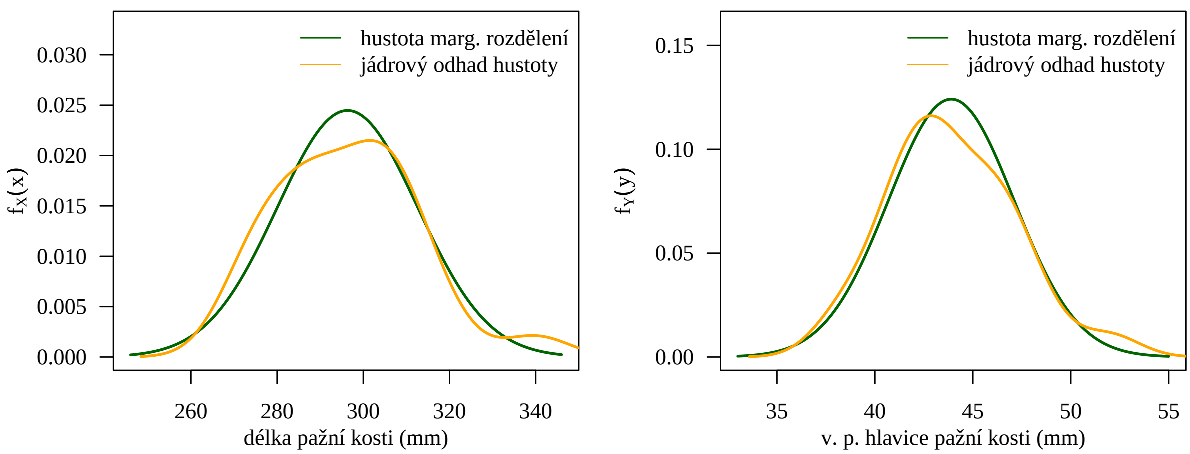
Za předpokladu, že náhodný vektor $(X, Y)^T$ pochází z dvourozměrného normálního rozdělení s parametry $\mu_1 = 296,34$, $\mu_2 = 43,90$, $\sigma_1^2 = 16,30^2$, $\sigma_2^2 = 3,21^2$ a $\rho = 0,85$, můžeme z druhého bodu věty 5.2 odvodit, že náhodná veličina $X \sim N(296,34; 16,30^2)$ a náhodná veličina $Y \sim N(43,90; 3,21^2)$. 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. Vertikální průměr hlavice 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 rozdělení 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 $f_X(x)$ rozdělení $N(296,34; 16,30^2)$. V příkazu použijeme nezaokrouhlené odhady parametrů μ_1 a σ_1^2 vypočítané v příkladu 5.9. 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(). Nakonec do grafu doplníme legendu rozlišující obě vykreslené křivky. Výsledný graf je zobrazen na obrázku 5.8 vlevo.

```
x <- seq(from = 246, to = 346, length = 512)
fx <- dnorm(x = x, mean = mL, sd = sL)
plot(x, fx, type = "l", las = 1, lwd = 2, col = "darkgreen", ylim = c(0, 0.033), xlab =
    "délka pažní kosti (mm)", ylab = expression(f[X](x)))
lines(density(humerus.LM), lwd = 2, col = "orange")
legend("topright", bty = "n", lwd = c(1, 1), col = c("darkgreen", "orange"), legend =
    c("hustota marg. rozdělení", "jádrový odhad hustoty"))
```

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

```
y <- seq(from = 33, to = 55, length = 512)
fy <- dnorm(x = y, mean = mHDL, sd = sHDL)
plot(y, fy, type = "l", las = 1, lwd = 2, col = "darkgreen",  ylim = c(0, 0.16), xlab =
    "v. p. hlavice pažní kosti (mm)", ylab = expression(f[Y](y)))
lines(density(humerus.HDLM), lwd = 2, col = "orange")
legend("topright", bty = "n", lwd = c(1, 1), col = c("darkgreen", "orange"), legend =
    c("hustota marg. rozdělení", "jádrový odhad hustoty"))
```



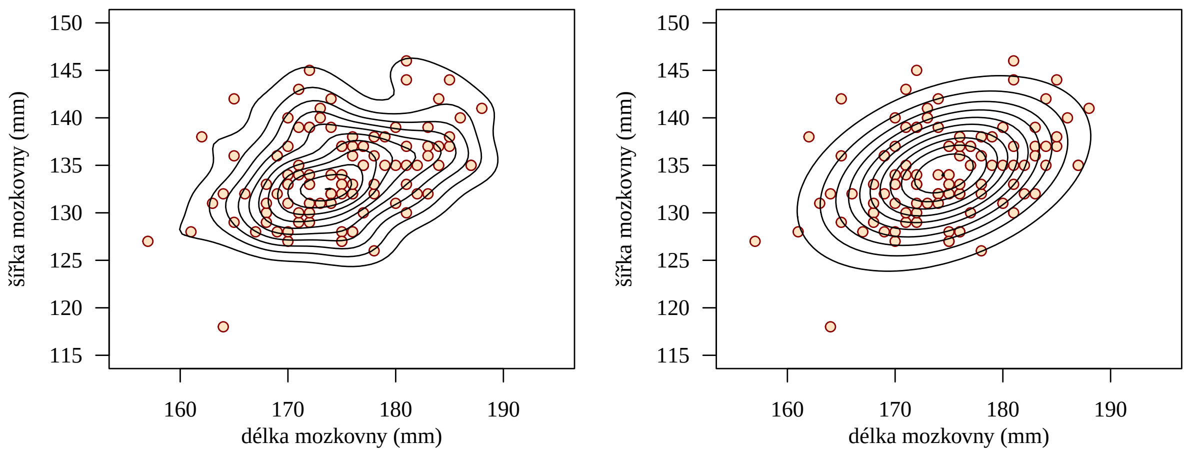
Obrázek 5.8: Křivka hustoty marginálního rozdělení a jádrového odhadu hustoty pro délku pažní kosti (vlevo) a pro vertikální průměr hlavice pažní kosti (vpravo) z levé strany u mužů

★

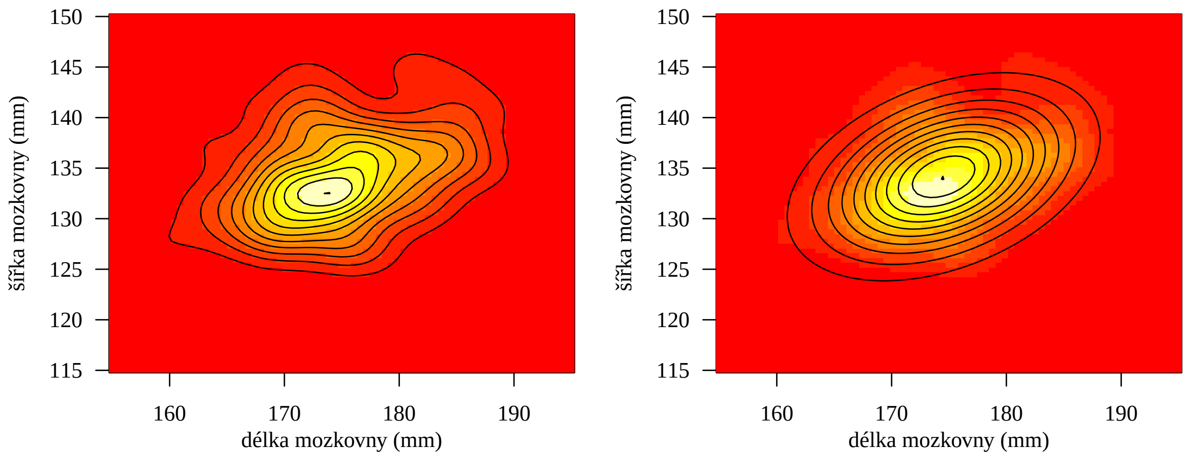
Příklad 5.11. Neřešený příklad

Načtěte datový soubor 01-one-sample-mean-skull-mf.txt obsahující údaje o délce mozkovny v mm (skull.L) a šířce mozkovny v mm (skull.B) u mužů a žen starověké egyptské populace. Předpokládejme, že máme náhodný vektor $(X, Y)^T$, kde náhodná veličina X popisuje délku mozkovny v mm a náhodná veličina Y popisuje šířku mozkovny v mm u žen starověké egyptské populace. (a) Najděte rozdělení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodného vektoru $(X, Y)^T$, a odhadněte parametry tohoto rozdělení; (b) ověřte, že nalezené rozdělení je vhodné na popis náhodného vektoru $(X, Y)^T$; (i) vykreslete dvourozměrný tečkový diagram a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ nalezeného rozdělení; (ii) vykreslete vrstevnicový diagram jádrového odhadu hustoty $k(x, y)$ a superponujte jej konturami jádrového odhadu hustoty $k(x, y)$, resp. konturami hustoty $f(x, y)$ nalezeného rozdělení; (iii) vykreslete 3D-diagram jádrového odhadu hustoty $k(x, y)$, resp. 3D-diagram hustoty $f(x, y)$ nalezeného rozdělení. K vykreslení vrstevnicových diagramů a 3D-diagramů použijte 12 barev z palety heat.colors().

Výsledky: (a) $(X,Y)^T \sim N_2\left((174,53;\ 134,15)^T, \begin{pmatrix} 6,22^2 & 11,12 \\ 11,12 & 4,70^2 \end{pmatrix}\right)$, $r_{12} = 0,38$; (b-i) dvourozměrný tečkový diagram superponovaný konturami jádrového odhadu hustoty $k(x,y)$, resp. konturami hustoty $f(x,y)$ dvourozměrného normálního rozdělení 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.



Obrázek 5.9: Dvourozměrný tečkový diagram superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozdělení (vpravo) pro délku mozkovny a šířku mozkovny u žen

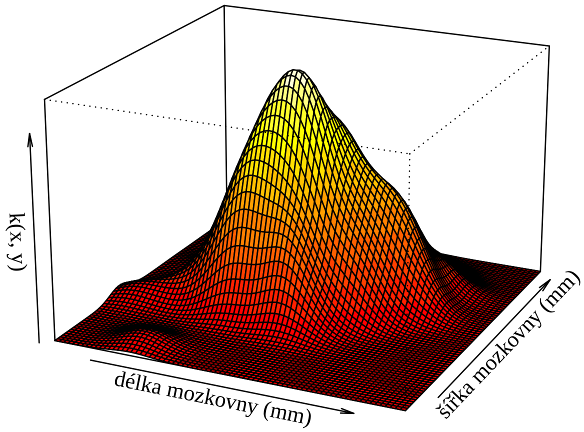


Obrázek 5.10: Vrstevnicový diagram jádrového odhadu hustoty superponovaný (a) konturami jádrového odhadu hustoty (vlevo); (b) konturami hustoty dvourozměrného normálního rozdělení (vpravo) pro délku mozkovny a šířku mozkovny u žen

★

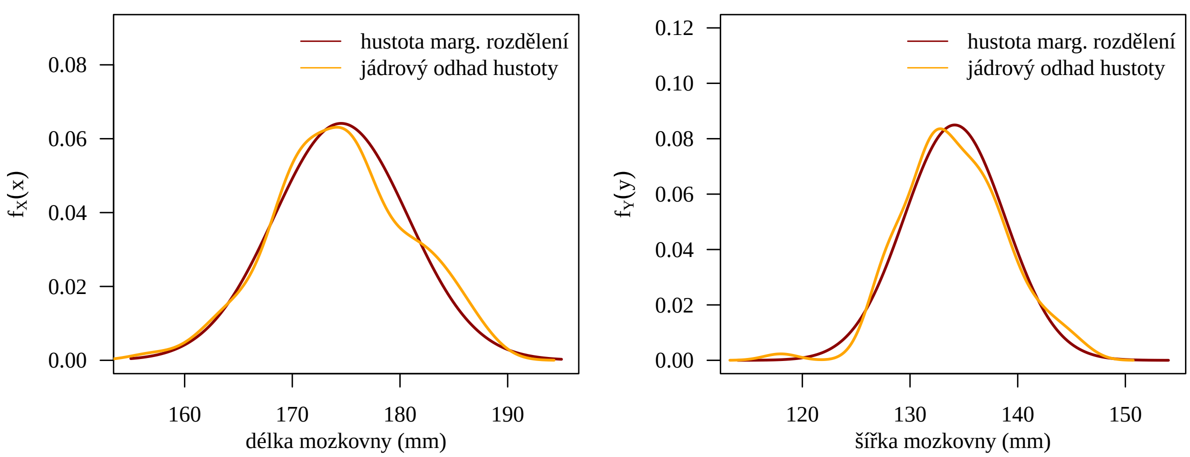
Příklad 5.12. Neřešený příklad

Za předpokladu, že náhodný vektor $(X,Y)^T$ popisující délku mozkovny v mm a šířku mozkovny v mm u žen starověké egyptské populace pochází z dvourozměrného normálního rozdělení 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) zjistěte marginální rozdělení náhodné veličiny X, resp. Y; (b) nakreslete graf křivky hustoty $f_X(x)$, resp. $f_Y(y)$ marginálního rozdělení spolu s křivkou jádrového odhadu hustoty náhodné veličiny X, resp. Y.



Obrázek 5.11: 3D-diagram jádrového odhadu hustoty (vlevo); 3D-diagram hustoty dvourozměrného normálního rozdělení (vpravo) pro délku mozkovny a šířku mozkovny u žen

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 rozdělení a jádrového odhadu hustoty náhodné veličiny X, resp. Y viz obrázek 5.12 vlevo, resp. vpravo.



Obrázek 5.12: Křivka hustoty marginálního rozdělení a jádrového odhadu hustoty pro délku mozkovny (vlevo) a pro šířku mozkovny (vpravo) u žen

★