

Zadání příkladů – Statistická inference II – 2017

Příklad 1. Směs dvou normálních rozdělání Nechť náhodná veličina X pochází ze směsi dvou normálních rozdělání $X \sim [pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2)]$. Potom marginální hustota náhodné veličiny X má tvar

$$f(x_i, \boldsymbol{\theta}) = \sum_{b_i \in \{0,1\}} f(x_i, b_i, \boldsymbol{\theta}) = f(x_i, 1, \boldsymbol{\theta}_1) + f(x_i, 0, \boldsymbol{\theta}_2),$$

kde

$$f(x_i, 1, \boldsymbol{\theta}_1) = \frac{p}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}\right)$$

je sdružená hustota za podmínky, že data pochází z první skupiny a

$$f(x_i, 0, \boldsymbol{\theta}_2) = \frac{1-p}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}\right)$$

je sdružená hustota za podmínky, že data pochází z druhé skupiny.

Logaritmická věrohodnostní funkce náhodné veličiny X má tvar

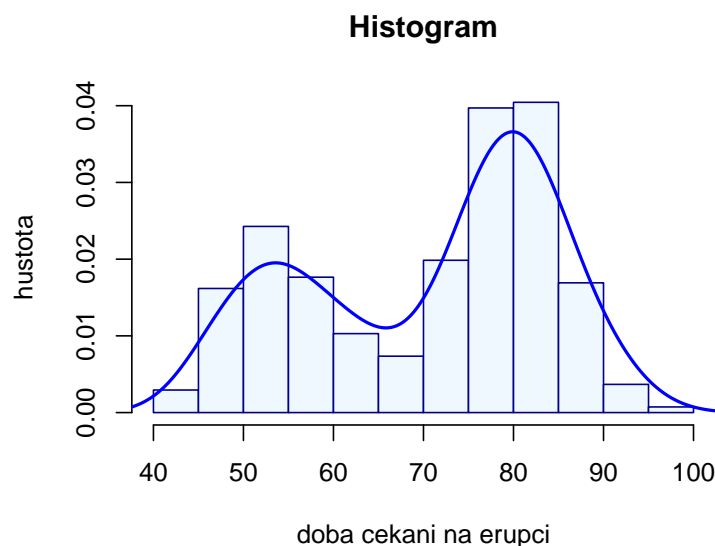
$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta}).$$

Příklad 2. Odhad parametrů směsi dvou normálních rozdělání

1. Načtete datový soubor `faithful` obsahující údaje o době čekání na erupci (`waiting`) a o době trvání erupce (`eruption`), přičemž se zaměříte na proměnnou `waiting`.
2. Nakreslete histogram doby čekání na erupci a superponujte jej křivkou jádrového odhadu.
3. Pomocí funkce `optim()` odhadněte parametry $p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ smíšeného rozdělání $[pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2)]$ náhodné proměnné `waiting`.
4. Pomocí funkce `optim()` nalezněte rozptyly odhadů parametrů $\hat{p}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2$.

Body 1.–3. aplikujte také na proměnnou `eruption`.

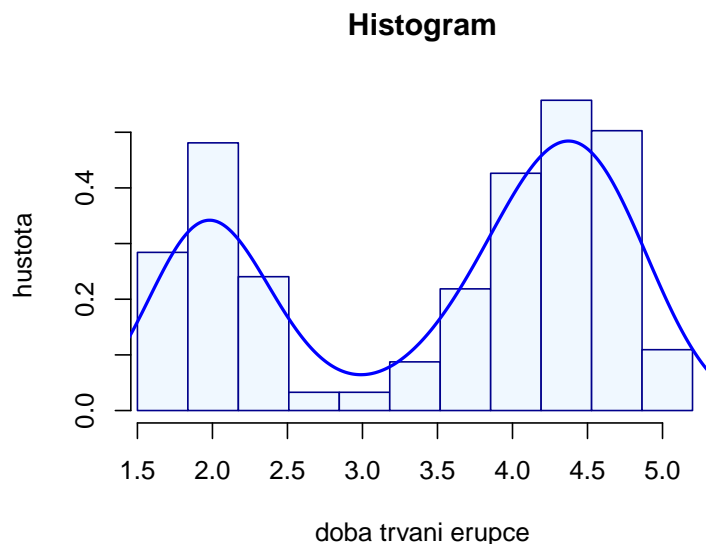
```
a) ##           p      mu1 sigma1      mu2 sigma2
## MLE 0.3609 54.6145 5.8698 80.0908 5.8682
## Var 0.0010 0.4893 0.2884 0.2547 0.1608
```



```

b) ##           p      mu1 sigma1      mu2 sigma2
## MLE 0.3482 2.0183 0.2356 4.2733 0.4370
## Var 0.0009 0.0007 0.0005 0.0012 0.0007

```



Příklad 3. Dvourozměrná Newton-Raphsonova metoda Nechť náhodná veličina X pochází z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , tj. $X \sim N(\mu, \sigma^2)$. Hustota náhodné veličiny X má tvar

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

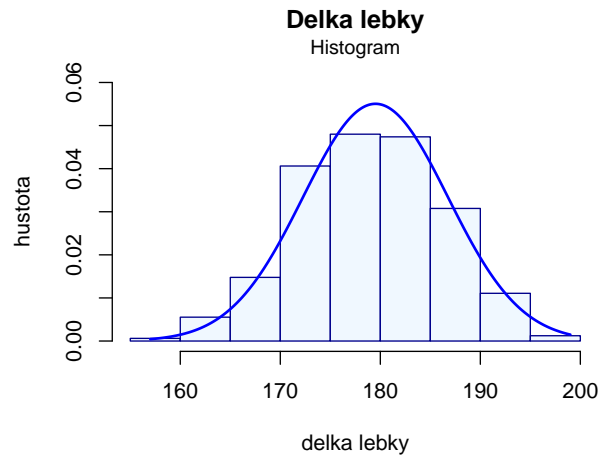
1. Odvoďte tvar věrohodnostní a logaritmické věrohodnostní funkce pro $N(\mu, \sigma^2)$.
2. Odvoďte tvar skóre funkce pro parametr μ a pro parametr σ (ne σ^2 !!!).
3. Odvoďte tvary druhých parciálních derivací logaritmické věrohodnostní funkce podle parametrů μ a σ (celkem 4).
4. Naprogramujte v R dvourozměrnou Newton-Raphsonovu metodu pro normální rozdělení $N(\mu, \sigma^2)$. Funkci pojmenujte `NMnorm`.

Naprogramovanou funkci nyní vyzkoušíme na reálných datech.

5. Načtěte datový soubor `one-sample-mean-skull-mf.txt`.
6. Vykreslete histogram proměnné délka lebky (`skull.L`).
7. Pomocí naprogramované funkce `NMnorm` odhadněte parametr μ a σ proměnné `skull.L`.
8. Odhady získané pomocí funkce `NMnorm` porovnejte s bodovými odhady parametrů μ a σ .
9. Body 6–8 aplikujte také na proměnnou šířka lebky (`skull.B`).

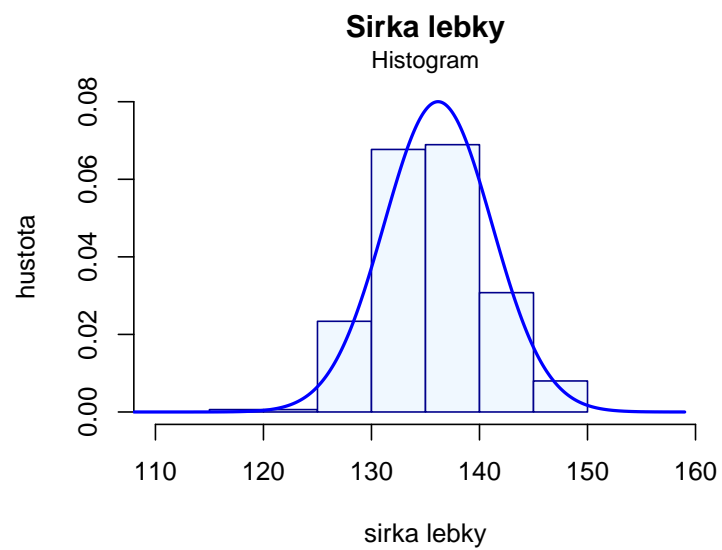
a) Délka lebky

```
##                mu sigma
## Newtonova metoda 179.5169 7.1390
## exaktni vypocet 179.5169 7.2249
```



b) Šírka lebky

```
##                mu sigma
## Newtonova metoda 136.1662 4.9247
## exaktni vypocet 136.1662 4.9708
```



Příklad 4. Broydenova metoda

1. Naprogramujte v R Broydenovu metodu (dvourozměrnou metodu sečen) pro normální rozdělení $N(\mu, \sigma^2)$. Funkci pojmenujte `BrMnorm()`.
 2. Načtěte datový soubor *one-sample-mean-skull-mf.txt*.
 3. Pomocí funkce `BrMnorm()` získejte odhady parametrů μ a σ délky lebky skull.L.
 4. Pomocí funkce `BrMnorm()` získejte odhady parametrů μ a σ šířky lebky skull.B.
3. Broydenova metoda aplikovaná na proměnnou skull.L

```
##                mu  sigma
## Broydenova metoda 179.5174 7.2333
## exaktní vypočet   179.5169 7.2249
```

4. Broydenova metoda aplikovaná na proměnnou skull.B

```
##                mu  sigma
## Broydenova metoda 136.1481 5.0360
## exaktní vypočet   136.1662 4.9708
```

Příklad 5. MC experiment pro Waldovy empirické intervaly spolehlivosti Necht

(a) $X \sim N(0, 1)$;

(b) $X \sim pN(0, 1) + (1 - p)N(0, 4)$, kde $p = 0.9$, tedy jde o směs dvou normálních rozdělání $X \sim N(0, 1)$ a $X \sim N(0, 4)$ v poměru 9 : 1.

Pro obě části (a) i (b) Vygenerujte $M = 100$ náhodných výběrů s rozsahem $n = 500$ a vypočítejte:

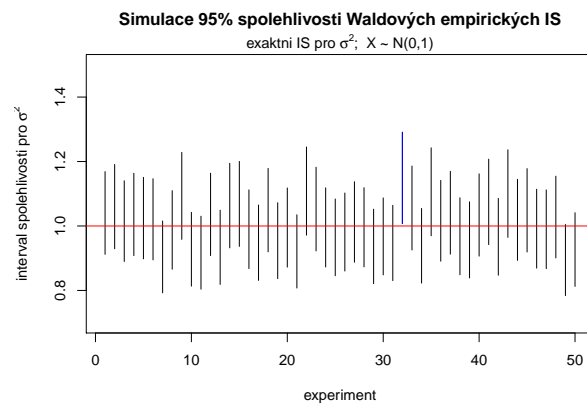
1. Waldovy exaktní empirické $100(1 - \alpha) \%$ IS pro rozptyl σ^2 , když μ neznáme.
2. Waldovy asymptotické empirické $100(1 - \alpha) \%$ IS pro rozptyl σ^2 , když μ neznáme.
3. Waldovy asymptotické empirické $100(1 - \alpha) \%$ IS pro směrodatnou odchylku σ , když μ neznáme.

Vždy spočítejte, kolik IS obsahuje rozptyl $\sigma^2 = 1$ (resp. směrodatnou odchylku $\sigma = 1$). Toto číslo podělené hodnotou M představuje simulovanou hladinu významnosti α .

a) $X \sim N(0, 1)$

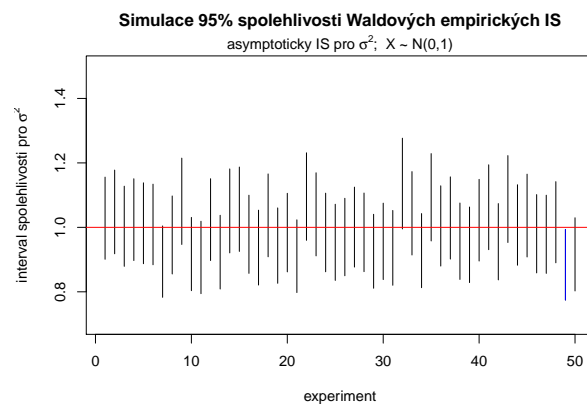
1. Waldovy exaktní empirické $100(1 - \alpha) \%$ IS pro rozptyl σ^2 , když μ neznáme.

```
##                               n
## simulovany pocet 49.0
## presny pocet    47.5
```



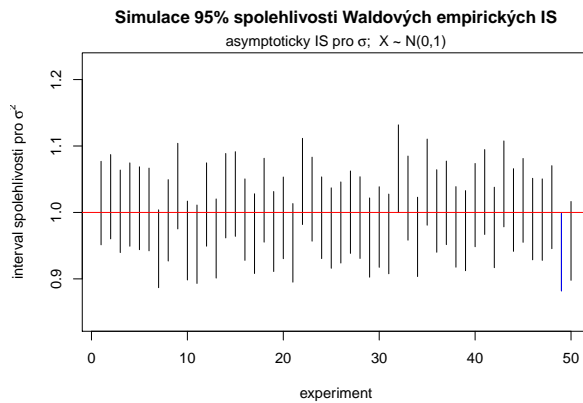
2. Waldovy asymptotické empirické $100(1 - \alpha) \%$ IS pro rozptyl σ^2 , když μ neznáme.

```
##                               n
## simulovany pocet 49.0
## presny pocet    47.5
```



3. Waldovy asymptotické empirické $100(1 - \alpha)\%$ IS pro směrodatnou odchylku σ , když μ neznáme.

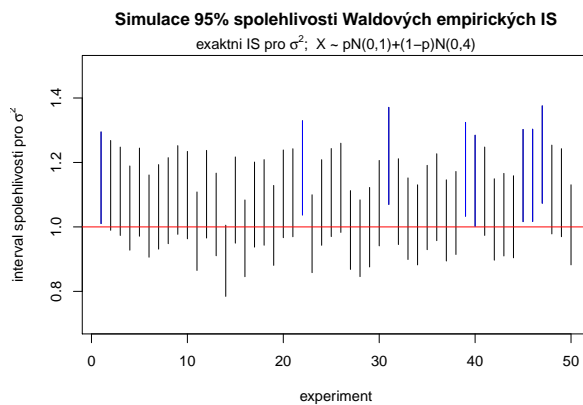
```
## n
## simulovany pocet 49.0
## presny pocet 47.5
```



b) $X \sim pN(0,1) + (1 - p)N(0,4)$

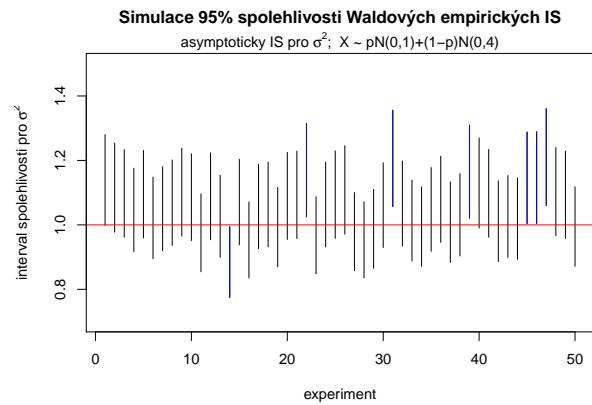
1. Waldovy exaktní empirické $100(1 - \alpha)\%$ IS pro rozptyl σ^2 , když μ neznáme.

```
## n
## simulovany pocet 42.0
## presny pocet 47.5
```



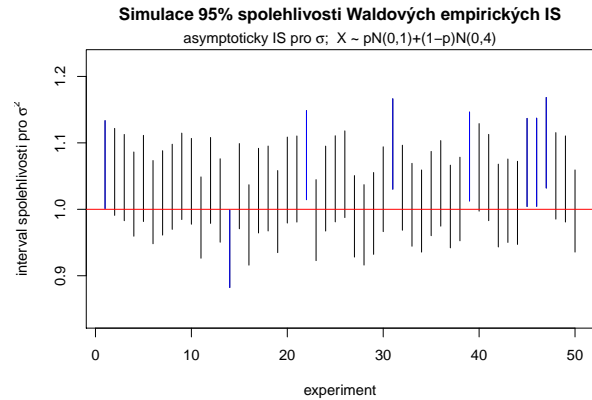
2. Waldovy asymptotické empirické $100(1 - \alpha)\%$ IS pro rozptyl σ^2 , když μ neznáme.

```
## n
## simulovany pocet 43.0
## presny pocet 47.5
```



3. Waldovy asymptotické empirické $100(1 - \alpha)\%$ IS pro směrodatnou odchylku σ , když μ neznáme.

```
##          n
## simulovany pocet 42.0
## presny pocet    47.5
```



Příklad 6. Simultánní oblasti spolehlivosti + elipsa spolehlivosti pro střední hodnotu a rozptyl (resp. směrodatnou odchylku)) Empirické $100(1 - \alpha)\%$ asymptotické intervaly spolehlivosti Waldova typu pro μ , σ^2 a σ jsou pro neznámé σ definovány následujícím způsobem:

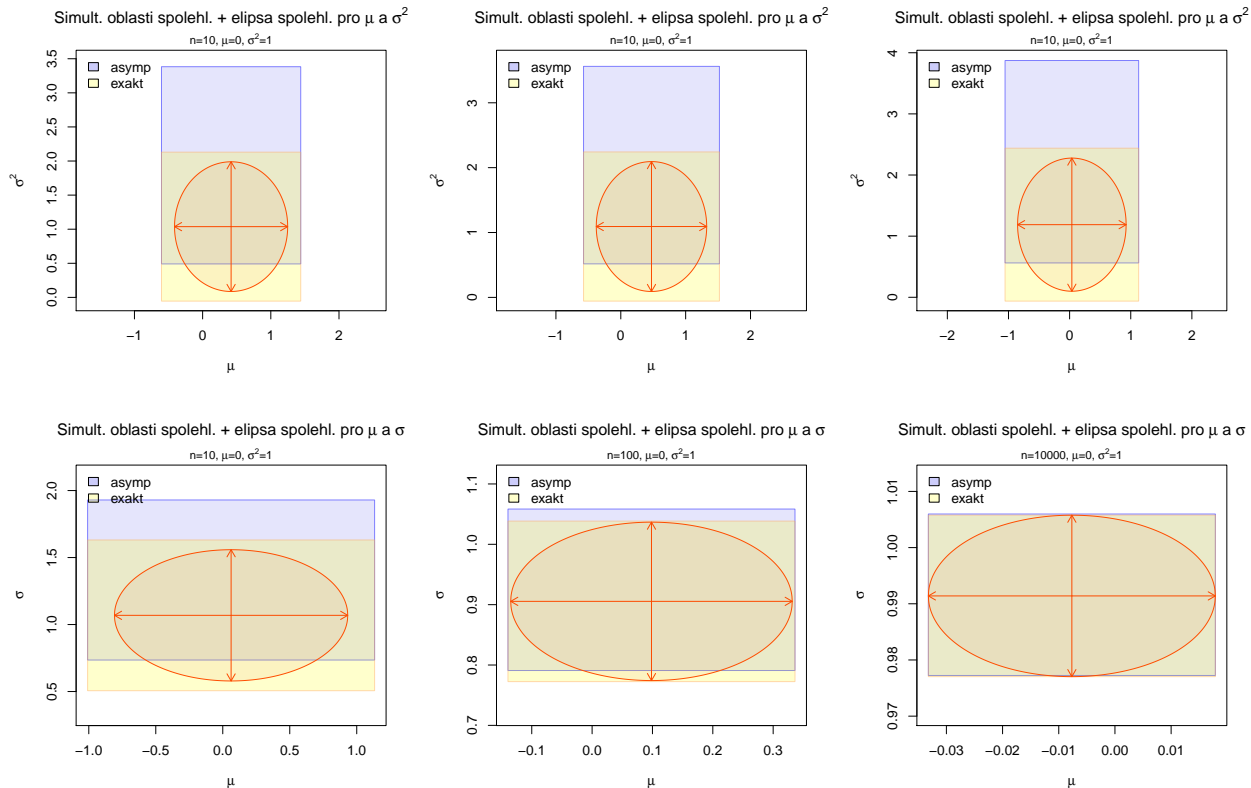
$$\Pr \left(\bar{x} - u_{1-\alpha/2} \sqrt{\hat{\sigma}^2/n} < \mu < \bar{x} - u_{\alpha/2} \sqrt{\hat{\sigma}^2/n} \right) = 1 - \alpha$$

$$\Pr \left(\hat{\sigma}^2 - u_{1-\alpha/2} \sqrt{2\hat{\sigma}^4/n} < \sigma^2 < \hat{\sigma}^2 - u_{\alpha/2} \sqrt{2\hat{\sigma}^4/n} \right) = 1 - \alpha$$

$$\Pr \left(\hat{\sigma} - u_{1-\alpha/2} \sqrt{\hat{\sigma}^2/2n} < \sigma < \hat{\sigma} - u_{\alpha/2} \sqrt{\hat{\sigma}^2/2n} \right) = 1 - \alpha$$

1. (a) Nakreslete simultánní množinu spolehlivosti pro $\theta = (\mu, \sigma^2)^T$ použitím exaktních intervalů spolehlivosti pro μ a pro σ^2 .
- (b) Nakreslete simultánní množinu spolehlivosti pro $\theta = (\mu, \sigma^2)^T$ použitím asymptotických intervalů spolehlivosti pro μ a pro σ^2 .
- (c) Do obrázku dokreslete $100(1-\alpha)\%$ elipsu spolehlivosti pro $\theta = (\mu, \sigma^2)^T$ použitím asymptotických intervalů spolehlivosti pro μ a pro σ^2 .
2. (a) Nakreslete simultánní množinu spolehlivosti pro $\theta = (\mu, \sigma)^T$ použitím exaktních intervalů spolehlivosti pro μ a pro σ .
- (b) Nakreslete simultánní množinu spolehlivosti pro $\theta = (\mu, \sigma)^T$ použitím asymptotických intervalů spolehlivosti pro μ a pro σ .
- (c) Do obrázku dokreslete $100(1-\alpha)\%$ elipsu spolehlivosti pro $\theta = (\mu, \sigma)^T$ použitím asymptotických intervalů spolehlivosti pro μ a pro σ .

Použijte (1) $n = 10$, (2) $n = 100$, (3) $n = 10000$. V (1), (2) a (3) zvolte $\mu = 0$ a $\sigma^2 = 1$ resp. $\sigma^2 = 4$. Koeficient spolehlivosti simultánní množiny zvolte zvolte $1 - \alpha = 0.95$.

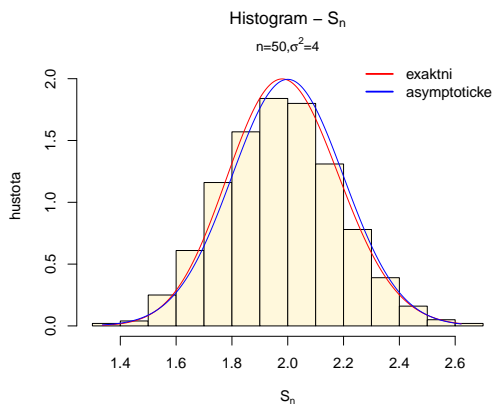
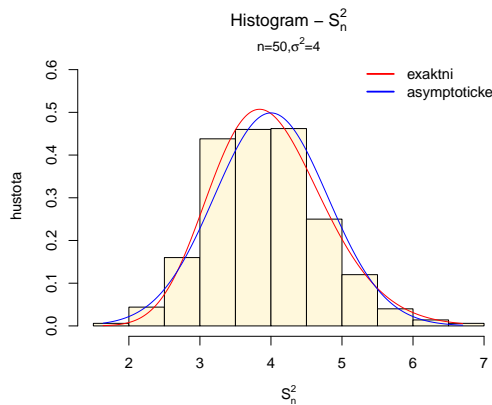
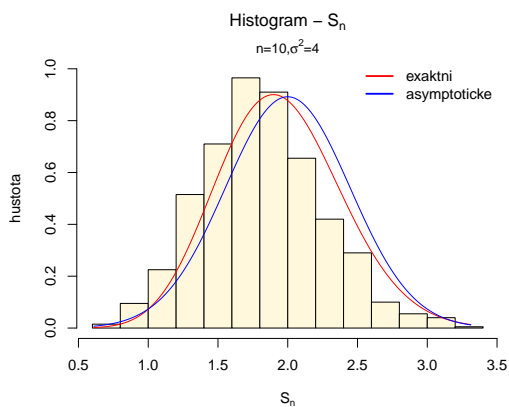
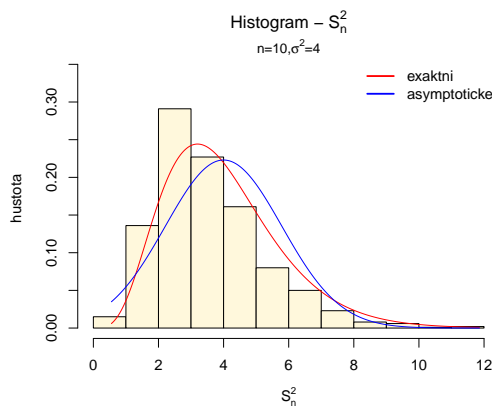


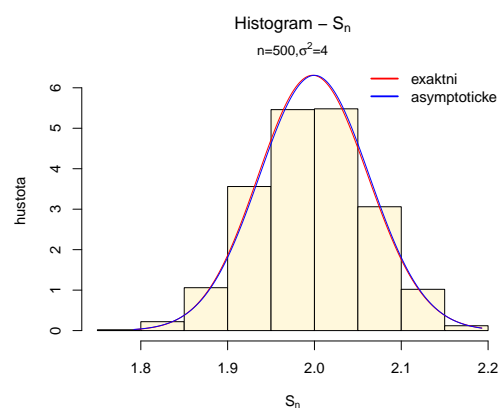
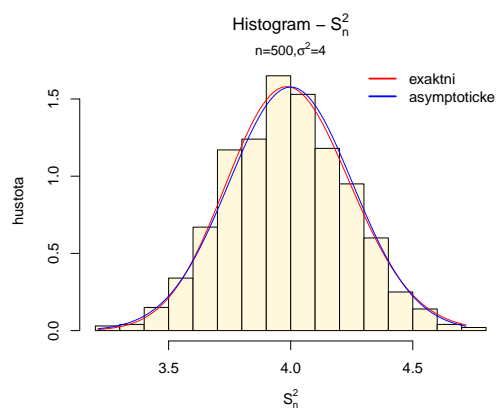
Příklad 7. Rozdělení výběrového rozptylu a výběrové směrodatné odchylky: simulační studie Necht $X \sim N(\mu, \sigma^2)$, potom

1. výběrový rozptyl $S_n^2 \sim N(\sigma^2, \frac{2\sigma^4}{n})$;
2. výběrová směrodatná odchylka $S_n \sim N(\sigma, \frac{\sigma^2}{2n})$.
3. testovací statistika $\frac{nS_n^2}{\sigma^2}$ pochází z χ^2 rozdělení o n stupních volnosti, tj. $\frac{nS_n^2}{\sigma^2} \sim \chi_n^2$.

Vygenerujte $M = 1000$ náhodných výběrů z normálního rozdělení $N(\mu, \sigma^2)$ o rozsahu n , kde $\mu = 0$ a $\sigma^2 = 4$, resp. $\sigma^2 = 1$. Použijte (i) $n = 10$, (ii) $n = 50$ a (iii) $n = 500$.

- (a) Pro každý náhodný výběr vypočítejte statistiku $S_{n,i}^2, i = 1, \dots, M$ a statistiky $S_{n,i}^2$ zobrazte pomocí histogramu. Histogram superponujte křivkami hustoty asymptotického a exaktního rozdělení statistiky S_n^2 .
- (b) Pro každý náhodný výběr vypočítejte statistiku $S_{n,i}, i = 1, \dots, M$ a statistiky $S_{n,i}$ zobrazte pomocí histogramu. Histogram superponujte křivkami hustoty asymptotického a exaktního rozdělení statistiky S_n .





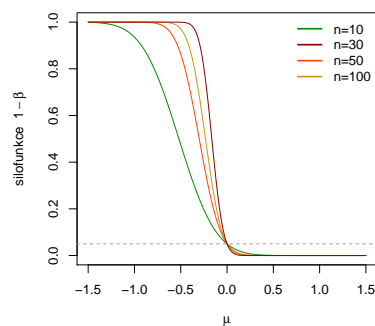
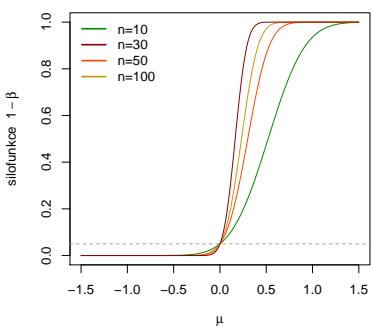
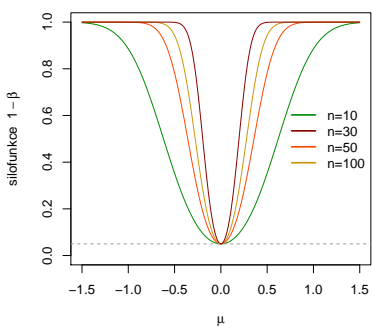
Příklad 8. P Předpokládejme, že $X \sim N(\mu, \sigma^2)$, kde σ^2 známe. Necht' $\theta = \mu$. Testujeme všechny tři typy hypotéz

a) $H_{01} : \mu = \mu_0$ oproti $H_{11} : \mu \neq \mu_0$ (oboustranná);

b) $H_{02} : \mu \leq \mu_0$ oproti $H_{12} : \mu > \mu_0$ (pravostranná);

c) $H_{03} : \mu \geq \mu_0$ oproti $H_{13} : \mu < \mu_0$ (levostranná).

1. Odvoďte tvary silofunkcí pro všechny tři typy hypotéz (a)–(c), t.j. tvary $\beta_{11}^*(\mu)$, $\beta_{12}^*(\mu)$ a $\beta_{13}^*(\mu)$.
2. Nakreslete silofunkce pro všechny tři typy hypotéz (a)–(c), kde $\mu_0 = 0$, a $\sigma^2 = 1$. Do jednoho obrázku zakreslete vždy tvary silofunkcí pro $n = 10$, $n = 30$, $n = 50$ a $n = 100$. Hladinu významnosti α zvolte 0.05. Hodnoty μ volte rozumně, např. v intervalu $\langle -1.5; 1.5 \rangle$.



Příklad 9. Rozdělení testovací statistiky pro test o střední hodnotě μ , když σ^2 neznáme

1. Necht' náhodný výběr X pochází z normálního rozdělení, t.j. $X \sim N(\mu, \sigma^2)$, kde $\mu = 600$ a $\sigma^2 = 100^2$. Rozsah náhodného výběru $n = 20$. Pomocí simulační studie v \mathbb{R} porovnejte rozdělení testovací statistiky pro test 'nepřesně zvolené' nulové hypotézy $H_0: \mu \leq 500$ (alternativní hypotéza $H_1: \mu > 500$), když rozptyl σ^2 neznáme, s rozdělením testovací statistiky nulové hypotézy $H_0: \mu \leq 600$ (alternativní hypotéza $H_1: \mu > 600$), opět když σ^2 neznáme.

Nasimulujte M pseudonáhodných výběrů, $M=1, \dots, 10\,000$ a pro každý vypočítejte realizaci testovací statistiky $t_{W,\lambda}^{(m)} = \frac{\bar{x}_m - \mu_0}{s_m} \sqrt{n}$ pro nulovou hypotézu $H_0: \mu \leq 500$ oproti $H_1: \mu > 500$. Histogram superponujte jednak křivkou hustoty necentrálního t -rozdělení s $n - 1$ stupni volnosti a parametrem necentrality λ ($\lambda = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$, kde μ_1 je vzata z alternativní hypotézy) a jednak křivkou hustoty centrálního studentova rozdělení. Obě křivky potom vzájemně okometricky porovnejte.

2. Necht' nyní X pochází ze směsi dvou normálních rozdělení, t.j. $X \sim [pN(\mu, 100^2) + (1 - p)N(\mu, 150^2)]$, kde $p = 0.9$ a $\mu = 600$. Proved'te simulační studii popsanou v bodě (1) pro tento náhodný výběr.

