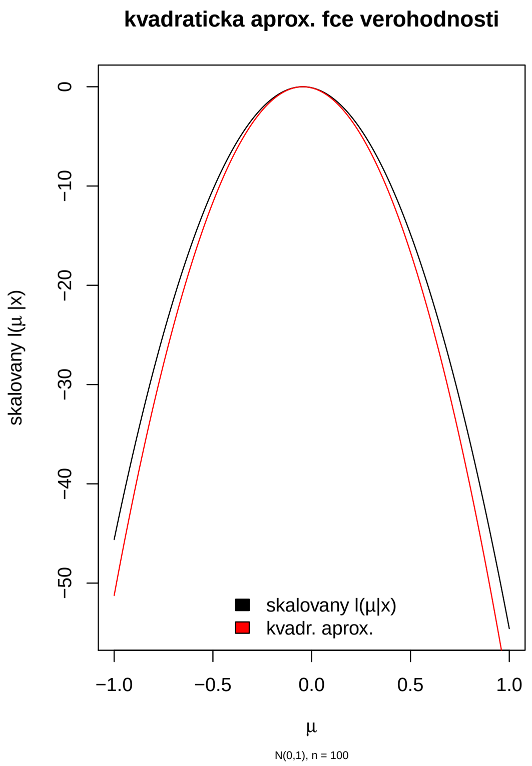
2 Statistická inference II - cvičení 16-03-01

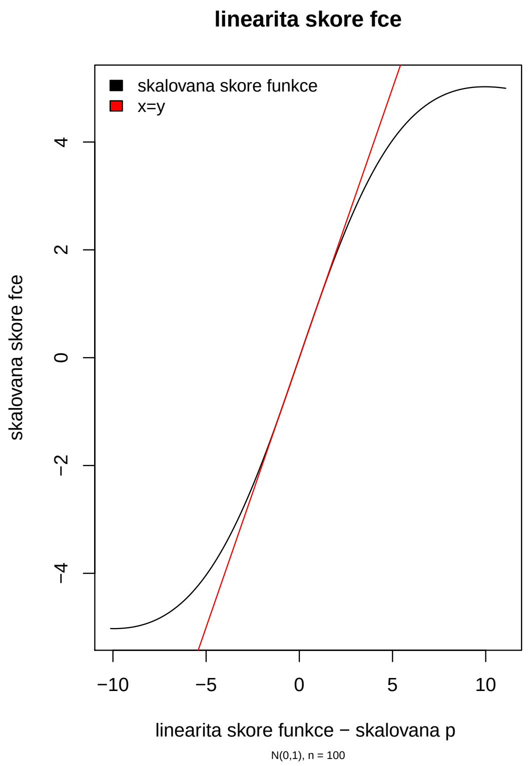
Řešení příkladů z domácí úlohy - SI I

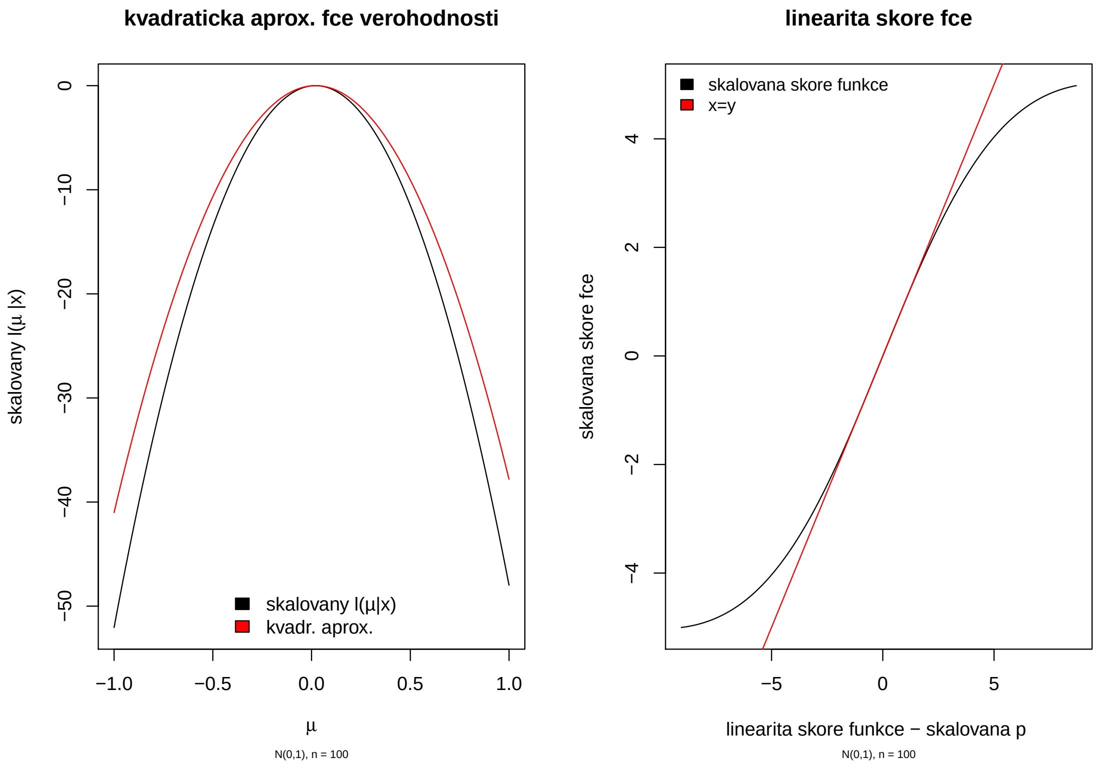
Příklad č.4 (kvadratická aproximace profilové funkce věrohodnosti)

(a) Nakreslete škálovaný logaritmus profilové funkce věrohodnosti normálního rozdělení pro μ. Na ose x bude μ a na ose y $\ln \mathcal{L}_P(\mu|\mathbf{x}) = l_P(\mu|\mathbf{x}) - l_P(\widehat{\mu}|\mathbf{x})$. Porovnejte $\ln \mathcal{L}_P(\mu|\mathbf{x})$ s kvadratickou aproximací vypočítanou pomocí Taylorova rozvoje $\ln \mathcal{L}_P(\mu|\mathbf{x}) = \ln(\frac{L_P(\mu|\mathbf{x})}{L_P(\widehat{\mu}|\mathbf{x})}) \approx -\frac{1}{2}\mathcal{I}(\widehat{\mu})(\mu - \widehat{\mu})^2$.

(b) Nechť skóre funkce $S(\mu) = \frac{\partial}{\partial \mu} \ln L_P(\mu|\mathbf{x})$. Vezmeme-li derivaci kvadratické aproximace uvedené výše, dostaneme $S(\mu) \approx -\mathcal{I}(\widehat{\mu})(\mu - \widehat{\mu})$ nebo $-\mathcal{I}^{-1/2}(\widehat{\mu})S(\mu) \approx \mathcal{I}^{1/2}(\widehat{\mu})(\mu - \widehat{\mu})$. Potom zobrazením pravé strany na ose x a levé strany na ose y dostaneme asymptoticky lineární funkci s jednotkovým sklonem. Asymptoticky také platí $\mathcal{I}^{1/2}(\overline{X})(\mu - \overline{X}) \stackrel{\mathcal{D}}{\sim} N(0,1)$. Je postačující mít rozsah osy x rovný $\langle -2, 2 \rangle$, protože funkce je asymptoticky (lokálně) lineární na tomto intervalu. Rozumně škálujte osu y. Zobrazte pro (a) $n = 10$, (b) $n = 100$ a (c) $n = 1000$. Použijte (1) $X \sim N(0,1)$ a (2) $X \sim (1-p)N(0,1) + pN(0,2)$, kde $p = 0.05$. Okomentujte rozdíly mezi (a), (b) a (c), stejně jako rozdíly mezi (1) a (2).



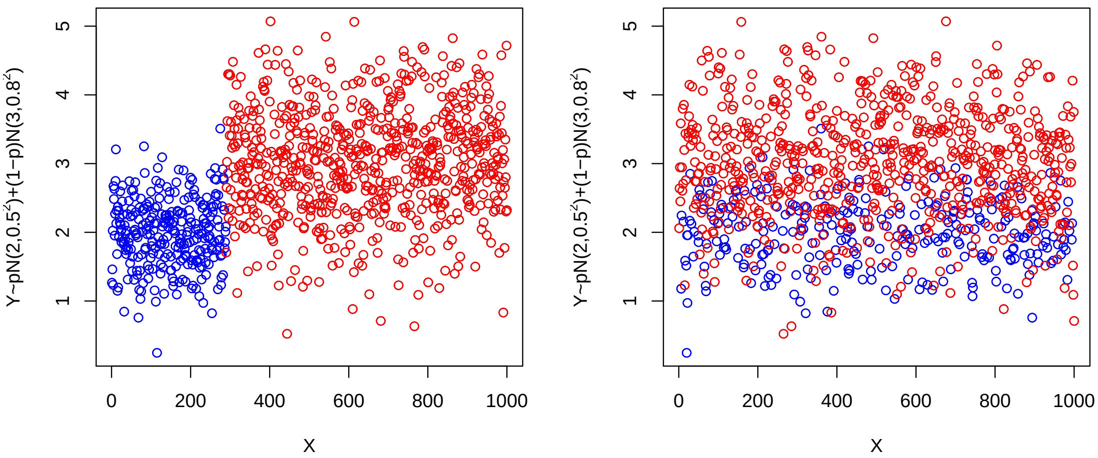



kvadraticka aprox. fce verohodnosti
skalovany l(μ |x)
0
−10
−20
−30
−40
−50
skalovany l(μ|x)
kvadr. aprox.
−1.0
−0.5
0.0
0.5
1.0
μ
N(0,1), n = 100
linearita skore fce
skalovana skore funkce
x=y
skalovana skore fce
4
2
0
−2
−4
−5
0
5
linearita skore funkce − skalovana p
N(0,1), n = 100

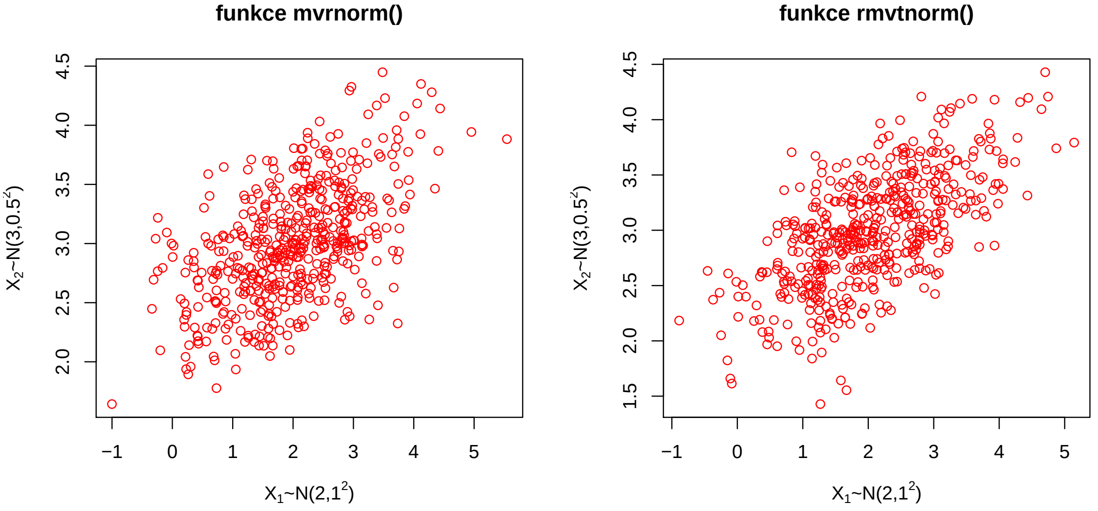
Statistická inference II – příklady

Příklad č.1 (Generování pseudonáhodných čísel)

1. Nechť $Y \sim pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma^2)$, kde $\mu_1 = 2$, $\mu_2 = 3$, $\sigma_1^2 = 0.5^2$ a $\sigma_2 = 0.8^2$. Vygenerujte $n = 1000$ pseudonáhodných čísel z tohoto smíšeného rozdělení.



2. Nechť $X_1 \sim N(\mu_1, \sigma_1^2)$ a $X_2 \sim N(\mu_2, \sigma_2^2)$. Potom náhodný vektor $(X_1, X_2)^T$ pochází z dvourozměrného normálního rozdělení s vektorem středních hodnot $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ a variační maticí $\boldsymbol{\Sigma}$
$$\boldsymbol{\Sigma} \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix} \tag{1}$$
Vygenerujte $n = 500$ pseudonáhodných čísel z dvourozměrného normálního rozdělení, pokud $\mu_1 = 2$, $\mu_2 = 3$, $\sigma_1^2 = 0.5^2$, $\sigma_2^2 = 1$, $\rho = 0.3$. K vygenerování použijte
   (a) funkci mvrnorm() z knihovny MASS;
   (b) funkci rmvnorm() z knihovny mvtnorm.



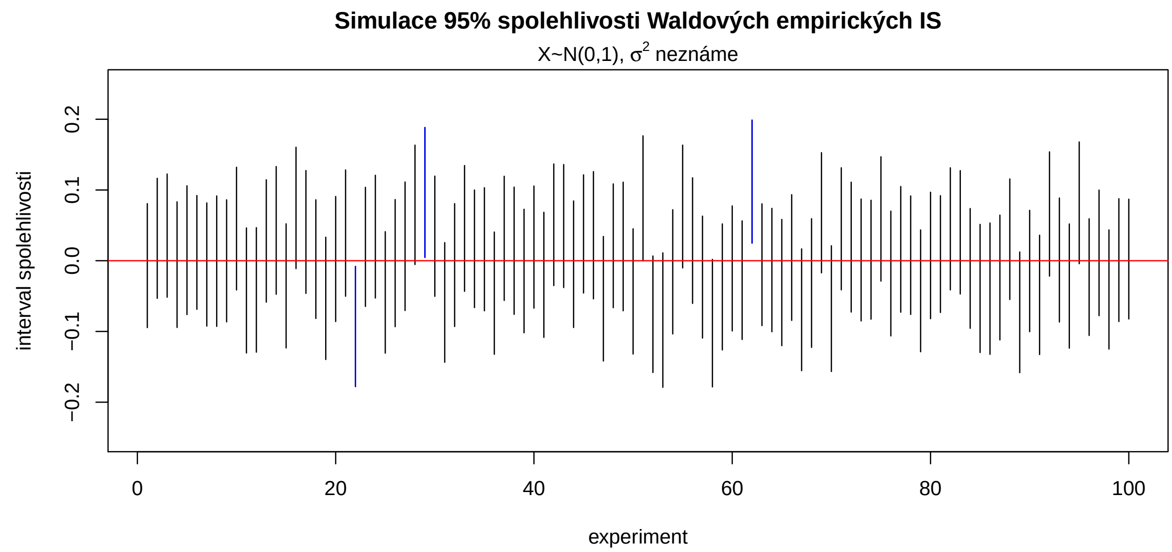
Příklad č.2 (MC experiment pro IS)
Nechť

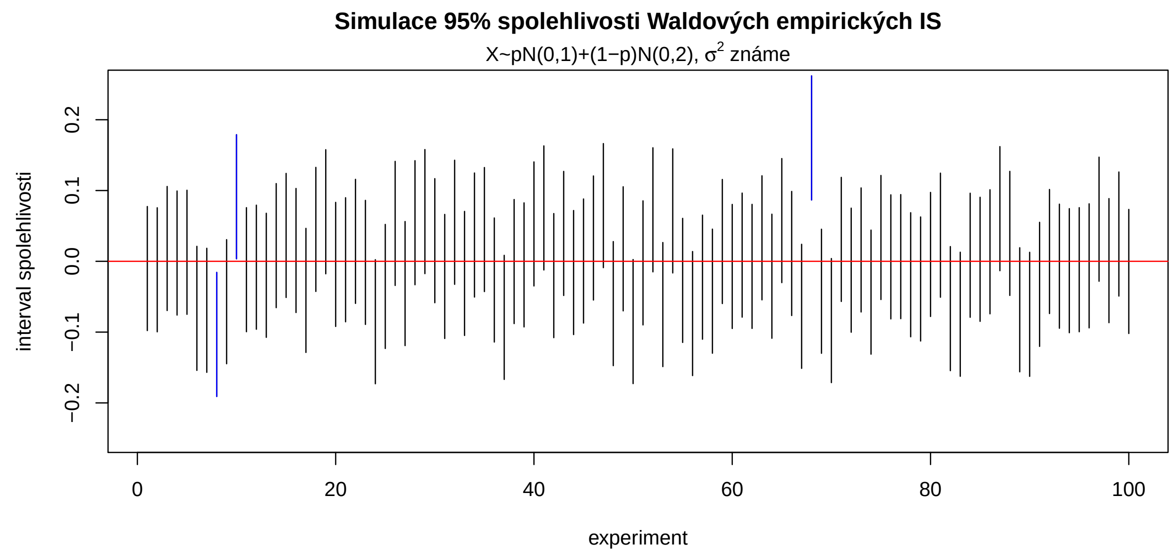
(a) $X \sim N(0,1)$;

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

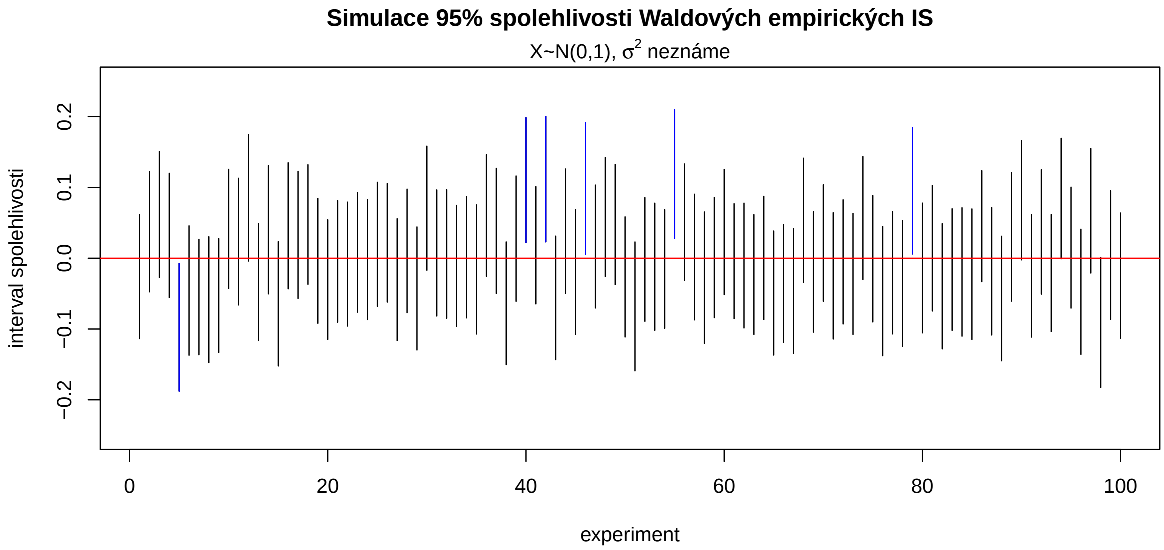
Pro obě části (a) i (b):

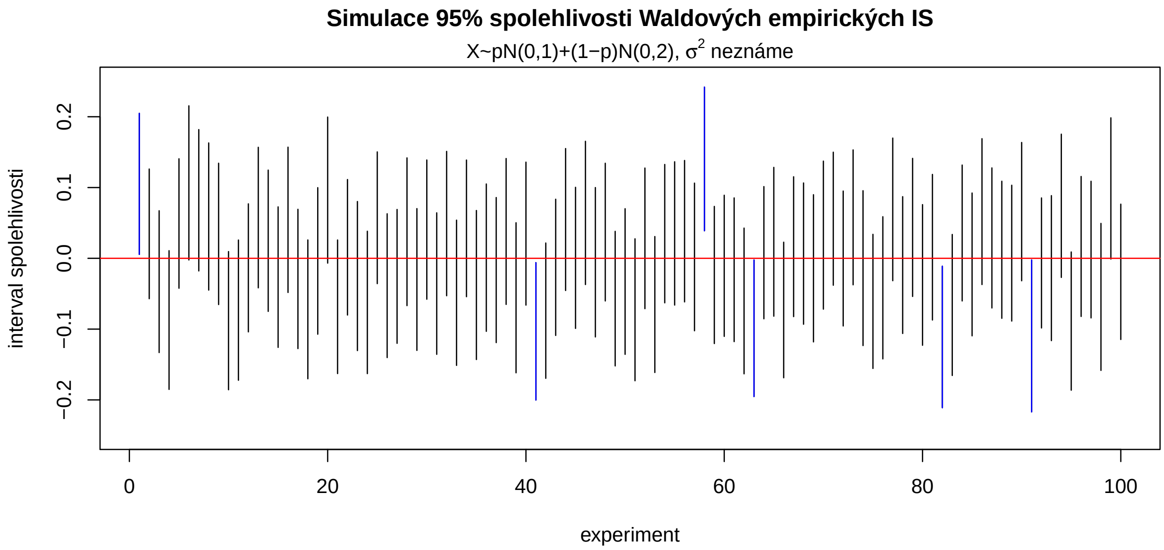
1. Vygenerujte $M = 100$ náhodných výběrů s rozsahem $n = 500$ a vypočítejte Waldovy $100(1-\alpha)\,\%$ empirické IS pro střední hodnotu μ, když σ^2 známe. Spočítejte, kolik IS obsahuje střední hodnotu $\mu = 0$. Toto číslo podělené hodnotou M představuje simulovanou hladinu významnosti α.





2. Vygenerujte $M = 100$ náhodných výběrů s rozsahem $n = 500$ a vypočítejte Waldovy $100(1-\alpha)\,\%$ empirické IS pro střední hodnotu μ, když σ^2 známe. Spočítejte, kolik IS obsahuje střední hodnotu $\mu = 0$. Toto číslo podělené hodnotou M představuje simulovanou hladinu významnosti α.





3. Postupy (1) a (2) zopakujte $N = 100\times$ a zjistěte, jaká je průměrná simulovaná hladina významnosti α v těchto stech pokusech.