

10. cvičení z PST

6. prosince 2017

10.1 (metoda momentů a max. věrohodnosti - spojité rozdělení)

Náhodná veličina X s oborem hodnot $\langle a, +\infty \rangle$ má hustotu

$$f_X(t) = \begin{cases} 0 & , t \in (-\infty, a), \\ e^{a-t} & , t \in \langle a, \infty \rangle, \end{cases}$$

kde $a \in \mathbb{R}$ je parametr. Pomocí metody maximální věrohodnosti i metody momentů odhadněte parametr a .

Úlohu vyřešte obecně pro realizaci $\mathbf{x} = (x_1, x_2, \dots, x_n)$ a také pro konkrétní realizaci

$$\mathbf{x} = (1, 2, 2, 2, 3, 3, 4)$$

rozsahu $n = 7$.

Řešení:

Funkce f_X je posunutá hustota exponenciálního rozdělení s parametrem $\tau = 1$. Je to tedy opět hustota, tj. je nezáporná a platí $\int_{\mathbb{R}} f_X(t) dt = 1$. Realizované výsledky musí spadat do oboru hodnot, tj.

$$x_i \in \langle a, +\infty \rangle$$

pro všechna $i = 1, \dots, n$ neboli musí platit, že

$$a \leq \min\{x_1, \dots, x_n\} .$$

Metoda maximální věrohodnosti:

Naším cílem je maximalizovat funkci

$$\Lambda(a) = \prod_{i=1}^n e^{a-x_i} = e^{na} \cdot e^{-\sum_i x_i} .$$

Tato funkce nabývá maxima pro největší přípustnou hodnotu parametru

$$\hat{a} = \min\{x_1, \dots, x_n\} .$$

Pro konkrétní zadání je to pak

$$\hat{a} = \min\{1, 2, 2, 2, 3, 3, 4\} = 1 .$$

Metoda momentů:

Porovnáme teoretickou střední hodnotu

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} t \cdot f_X(t) dt = \int_a^{\infty} t e^{a-t} dt = [-t e^{a-t}]_{t=a}^{\infty} + \int_a^{\infty} e^{a-t} dt = \\ &= a + [-e^{a-t}]_{t=a}^{\infty} = a + 1 \end{aligned}$$

a výběrový průměr \bar{x} . Odtud tak pro parametr \hat{a} dostaneme

$$\hat{a} = \bar{x} - 1 ,$$

POKUD je ovšem splněno, že $a \leq \min\{x_1, \dots, x_n\}$!

Pro konkrétní zadání je $\bar{x} = \frac{1+2+2+2+3+3+4}{7} = \frac{17}{7}$ a tedy $\hat{a} = \frac{17}{7} - 1 \doteq 1.43$, což ale **NENÍ** menší než $\min\{1, 2, 2, 2, 3, 3, 4\} = 1$. V tomto případě tedy metoda momentů **NEDÁVÁ** žádný odhad.

10.2 (metoda momentů - spojité rozdělení)

Náhodná veličina X má biexponenciální rozdělení s hustotou

$$f_X(t) = c \cdot e^{-a|t|}, \quad t \in \mathbb{R}$$

kde $a, c > 0$. Metodou momentů odhadněte parametry na základě realizace rozsahu $n = 10$, z níž jsme vypočítali realizaci výběrového průměru $\bar{x} = 2$ a realizaci výběrového rozptylu $s_{\bar{x}}^2 = 4$.

Řešení:

Aby funkce f_X byla hustotou, musí být splněno, že

$$1 = \int_{-\infty}^{\infty} f_X(t) dt = 2c \int_0^{\infty} e^{-at} dt = 2c \left[\frac{e^{-at}}{-a} \right]_{t=0}^{\infty} = \frac{2c}{a}$$

tedy $c = \frac{a}{2}$ a zbývá odhadnout parametr a .

Hustota je sudá funkce a proto střední hodnota $E(X)$ (stejně jako všechny liché momenty) bude nulová. Příslušné nevlastní integrály existují, protože klesající exponenciála převáží jakýkoliv polynom v nekonečnu.

Porovnáním teoretické střední hodnoty a výběrového průměru dostaneme $0 = E(X) = \bar{x} = 2$ což sice evidentně neplatí, ale nedává to žádné omezení na volbu parametrů a . Navíc těžko můžeme čekat, že se měřením přesně trefíme do teoretické hodnoty.

Musíme proto použít další moment:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} t^2 f_X(t) dt = \int_0^{\infty} t^2 (ae^{-at}) dt = \underbrace{[t^2 \cdot (-e^{-at})]_{t=0}^{\infty}}_{=0} + \int_0^{\infty} 2te^{-at} dt = \\ &= \underbrace{\left[2t \cdot \frac{e^{-at}}{-a} \right]_{t=0}^{\infty}}_{=0} + \frac{2}{a} \int_0^{\infty} e^{-at} dt = \frac{2}{a^2}, \end{aligned}$$

který chceme srovnat s druhým výběrovým momentem (nikoliv s výběrovým rozptylem!):

$$m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Ten se dá získat z hodnot $\bar{x} = 2$ a $s_{\bar{x}}^2 = 4$ podobným vztahem jako máme pro rozptyl, tj. $D(X) = E(X^2) - (E(X))^2$, jen si musíme dát pozor na to, že někde je $\frac{1}{n}$ a někde zase $\frac{1}{n-1}$. Máme tedy

$$\frac{n-1}{n} \cdot s_{\bar{x}}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = m_2 - (\bar{x})^2$$

takže

$$m_2 = \frac{n-1}{n} \cdot s_{\bar{x}}^2 + (\bar{x})^2 = \frac{9}{10} \cdot 4 + 2^2 = \frac{38}{5}.$$

Srovnáním dostaneme

$$\frac{2}{a^2} = E(X^2) = m_2 = \frac{38}{5}$$

a celkem tak máme

$$a = \sqrt{\frac{5}{19}} = \frac{\sqrt{95}}{19} \doteq 0.512989 \quad \text{a} \quad c = \frac{a}{2} = \frac{\sqrt{95}}{38} \doteq 0.256495 .$$

10.3 (metoda momentů a max. věrohodnosti - spojité rozdělení)

Předpokládáme, že náhodná veličina X má posunutě exponenciální rozdělení s oborem hodnot $\langle T, \infty \rangle$ a s hustotou

$$f_X(t) = \begin{cases} \frac{1}{\tau} \exp\left(-\frac{t-T}{\tau}\right), & t \geq T, \\ 0 & \text{jinak,} \end{cases}$$

kde $\tau > 0$. Z realizace $x = (2, 3, 8, 4, 10, 3, 5)$ odhadněte parametry T, τ .

Řešení:

Realizované výsledky musí spadat do oboru hodnot, tj. $x_i \in \langle T, +\infty \rangle$ pro všechna i , neboli musí platit, že

$$T \leq \min\{x_1, \dots, x_n\} = \min\{2, 3, 8, 4, 10, 3, 5\} = 2 .$$

Tímto máme omezení $T \in (-\infty, 2)$.

Metoda maximální věrohodnosti:

Logaritmicko-věrohodnostní funkce je

$$\begin{aligned} \lambda(T, \tau) = \ln \Lambda(T, \tau) &= \ln \left(\prod_{i=1}^n \frac{1}{\tau} \exp\left(-\frac{x_i - T}{\tau}\right) \right) = -n \ln \tau - \frac{1}{\tau} \sum_{i=1}^n x_i + \frac{1}{\tau} n T = \\ &= -n \left(\ln \tau + \frac{\bar{x} - T}{\tau} \right) \end{aligned}$$

Pro $T \in (-\infty, 2)$ je zřejmě vždy $\lambda(T, \tau) \leq \lambda(2, \tau)$, takže maximum stačí hledat pro $\hat{T} = \min_i x_i = 2$.

Pro parametr τ najdeme maximum pomocí derivace

$$0 = \frac{\partial \lambda}{\partial \tau}(2, \hat{\tau}) = -n \left(\frac{1}{\hat{\tau}} - \frac{\bar{x} - 2}{\hat{\tau}^2} \right) = n \cdot \frac{\bar{x} - 2 - \hat{\tau}}{\hat{\tau}^2}$$

tedy

$$\hat{\tau} = \bar{x} - 2 .$$

V našem případě máme $\bar{x} = 5$, takže maximálně věrohodné odhady parametrů jsou

$$\hat{T} = 2, \quad \hat{\tau} = 3 .$$

Metoda momentů:

Momenty pro X můžeme spočítat buď přímo, nebo prostě použijeme, že veličina $Y := X - T$ má exponenciální rozdělení $\text{Exp}(\tau)$. Pro $u \geq 0$ totiž máme

$$F_Y(u) = P(X - T \leq u) = P(X \leq u + T) = \int_T^{u+T} \frac{1}{\tau} \exp\left(-\frac{t-T}{\tau}\right) dt =$$

$$= \left[\frac{s+T=t}{ds=dt} \right] = \int_0^u \frac{1}{\tau} \exp\left(-\frac{s}{\tau}\right) ds .$$

Pro $Y \sim \text{Exp}(\tau)$ je $E(Y) = \tau$ a $D(Y) = \tau^2$. Tudíž

$$E(X) = E(Y + T) = E(Y) + T = \tau + T$$

a

$$E(X^2) = D(X) + (E(X))^2 = D(Y) + (E(Y) + T)^2 = \tau^2 + (\tau + T)^2 .$$

Výběrové momenty pak jsou

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{35}{7} = 5, \quad m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{227}{7} \doteq 32.429 ,$$

a z rovnic

$$\begin{aligned} \hat{\tau} + \hat{T} &= E(X) = \bar{x} = 5, \\ (\hat{\tau} + \hat{T})^2 + \hat{\tau}^2 &= E(X^2) = m_2 = \frac{227}{7}, \end{aligned}$$

dostaneme tak řešení

$$\begin{aligned} \hat{\tau} &= \sqrt{m_2 - \bar{x}^2} = \sqrt{\frac{227}{7} - 5^2} = \frac{2\sqrt{91}}{7} \doteq 2.7255 \\ \hat{T} &= \bar{x} - \hat{\tau} = 5 - \sqrt{\frac{52}{7}} \doteq 2.2745 \end{aligned}$$

kteří ovšem NEODPOVÍDÁ zadání, neboť $\hat{T} > x_1 = 2$, takže nalezený model nepřipouští pozorovanou hodnotu x_1 (ta by měla nulovou hustotu pravděpodobnosti).

10.4 (metoda momentů)

Datový soubor $\mathbf{x} = (-4, -3, -2, -1.5, 0.5, 1, 2.5, 3)$ je realizací náhodné veličiny X , která má spojité rovnoměrné rozdělení v intervalu $\langle -h, h \rangle$. Metodou momentů určete odhad parametru h (a ověřte, zda odhad odpovídá zadání).

Řešení:

Rozsah souboru je $n = 8$. Realizované výsledky musí spadat do oboru hodnot, tj. $|x_i| \leq h$ pro všechna i neboli musí platit, že

$$h \geq \max\{|x_1|, \dots, |x_n|\} = 4 .$$

Náhodná veličina X má hustotu pravděpodobnosti

$$f_X(x) = \begin{cases} \frac{1}{2h} & , \quad |x| \leq h, \\ 0 & , \quad |x| > h. \end{cases}$$

a tedy střední hodnotu $E(X)$ rovnou nule. Proto musíme použít další momenty (jen pro představu:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -\frac{3.5}{8} = -0.4375). \text{ Dostaneme}$$

$$E(X^2) = \int_{-h}^h \frac{x^2}{2h} dx = \left[\frac{x^3}{6h} \right]_{-h}^h = \frac{h^2}{3} .$$

Odhad druhého momentu je

$$m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{47.75}{8} = 5.96875.$$

Odhad parametru získáme jako řešení rovnice

$$\frac{\hat{h}^2}{3} = E(X^2) = m_2 = \frac{47.75}{8} \implies \hat{h} = \sqrt{17.90625} \doteq 4.2316.$$

Protože všechny hodnoty ze souboru leží v intervalu $\langle -h, h \rangle = \langle -4.2316, 4.2316 \rangle$, můžeme nalezenou hodnotu h tudíž považovat za hledaný odhad parametru rozdělení.

10.5 (intervalový odhad pomocí CLV)

Počet X ryb, které rybář uloví za den je popsán Poissonovým rozdělením s parametrem $\lambda = 3$. Na ryby jde $n = 100$ -krát za rok. Najděte symetrický interval, v němž se počet ulovených ryb za rok nachází s pravděpodobností 95%.

Řešení:

Pro $i = 1, \dots, n$ máme veličiny

$$X_i = \text{''počet ryb ulovených za } i\text{-tý den''}$$

které jsou za nezávislé a mají Poissonovo rozdělení s parametrem $\lambda = 3$.

Takže

$$E(X_i) = \lambda = 3 \quad \text{a} \quad D(X_i) = \lambda = 3.$$

Počet ryb ulovených za $n = 100$ dnů je

$$Y = \sum_{i=1}^n X_i.$$

Takže $E(Y) = n \cdot \lambda = 100 \cdot 3 = 300$ a (díky nezávislosti) je $D(Y) = n \cdot \lambda = 100 \cdot 3 = 300$. Odhadem z CLV má veličina Y přibližně normální rozdělení $N(300, 300) =: N(\mu, \sigma^2)$. Pro takovou veličinu má symetrický $1 - \alpha = 95\%$ intervalový odhad tvar

$$\begin{aligned} & \langle \mu - \sigma \cdot \Phi^{-1} \left(1 - \frac{\alpha}{2} \right), \mu + \sigma \cdot \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \rangle = \\ & = \langle 300 - 10\sqrt{3} \cdot \Phi^{-1}(0.975), 300 + 10\sqrt{3} \cdot \Phi^{-1}(0.975) \rangle \doteq \\ & \doteq \langle 300 - 33.948, 300 + 33.948 \rangle. \end{aligned}$$

10.6 (intervalový odhad pomocí CLV)

Veličina X , představující životnost žárovky, má exponenciální rozdělení s parametrem τ . Průměrná životnost $n = 64$ náhodně vybraných žárovek je $\bar{x} = 210$ hodin. Pomocí centrální limitní věty určete oboustranný symetrický 95% interval spolehlivosti pro parametr τ .

Řešení:

Pro veličinu X máme $E(X) = \tau$ a $D(X) = \tau^2$. Veličina průměrná životnost $n = 64$ žárovek je dána jako

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

Pro ni máme $E(\bar{X}) = \tau$ a $D(\bar{X}) = \frac{\tau^2}{n}$. Podle CLV má veličina \bar{X} přibližně normální rozdělení $N(\mu, \sigma^2) := N(\tau, \frac{\tau^2}{n})$. Oboustranný symetrický $1 - \alpha = 95\%$ odhad spolehlivosti pro realizaci \bar{x} veličiny \bar{X} je

$$\mu - \sigma \cdot \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \bar{x} \leq \mu + \sigma \cdot \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) ,$$

po dosažení to je

$$\tau - \frac{\tau}{\sqrt{n}} \cdot \Phi^{-1} (0.975) \leq \bar{x} \leq \tau + \frac{\tau}{\sqrt{n}} \cdot \Phi^{-1} (0.975)$$

a po úpravě dostáváme

$$\frac{\bar{x}}{1 + \frac{\Phi^{-1}(0.975)}{\sqrt{n}}} \leq \tau \leq \frac{\bar{x}}{1 - \frac{\Phi^{-1}(0.975)}{\sqrt{n}}}$$

Realizace veličiny \bar{X} je $\bar{x} = 210$ (hodin). Hledaný oboustranný symetrický 95% interval spolehlivosti (v hodinách) pro τ je tak přibližně

$$\left\langle \frac{210}{1 + \frac{1.96}{\sqrt{64}}}, \frac{210}{1 - \frac{1.96}{\sqrt{64}}} \right\rangle \doteq \langle 168.675, 278.146 \rangle .$$

10.7 (intervalový odhad pro střední hodnotu)

Soubor dat (75, 85, 58, 72, 70, 75) je náhodným výběrem z normálního rozdělení $N(\mu, \sigma^2)$. Stanovte oboustranný symetrický 95% interval spolehlivosti pro střední hodnotu μ .

Řešení:

K určení intervalového odhadu použijeme statistiku

$$T = \frac{\bar{X} - \mu}{S_X} \sqrt{n}$$

která má Studentovo rozdělení $t(n - 1)$, kde $n = 6$ je rozsah souboru. Oboustranný symetrický $1 - \alpha = 95\%$ interval spolehlivosti pro realizaci t veličiny T je

$$-q_{t(n-1)} \left(1 - \frac{\alpha}{2} \right) \leq \frac{\bar{x} - \mu}{s_x} \sqrt{n} \leq q_{t(n-1)} \left(1 - \frac{\alpha}{2} \right) .$$

Po úpravě máme

$$\bar{x} - \frac{s_x}{\sqrt{n}} q_{t(5)}(0.975) \leq \mu \leq \bar{x} + \frac{s_x}{\sqrt{n}} q_{t(5)}(0.975) .$$

Uurčíme si realizaci výběrového průměru \bar{x} a výběrového rozptylu s_x^2

$$\bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{435}{6} = 72.5$$

$$s_x^2 = \frac{1}{5} \left(\sum_{i=1}^6 x_i^2 - 6 \cdot (\bar{x})^2 \right) = \frac{385.5}{5} \doteq 77.1, \quad s_x \doteq 8.781 .$$

Z tabulek kvantilů Studentova rozdělení dostaneme $q_{t(5)}(0.975) \doteq 2.57$, a hledaný interval je tedy

$$63.29 \leq \mu \leq 81.71 .$$

10.8 (intervalový odhad pro rozptyl)

Soubor (70, 84, 89, 70, 74, 70) je náhodným výběrem z normálního rozdělení $N(\mu, \sigma^2)$. Určete oboustranný symetrický 95% interval spolehlivosti pro rozptyl σ^2 .

Řešení:

Interval spolehlivosti pro rozptyl určíme ze statistiky

$$T = \frac{(n-1) S_{\mathbf{X}}^2}{\sigma^2} ,$$

má rozdělení $\chi^2(n-1)$, kde $n = 6$ je rozsah výběru a $S_{\mathbf{X}}^2$ je výběrový rozptyl. Oboustranný symetrický $1 - \alpha = 95\%$ interval spolehlivosti pro realizaci t veličiny T je

$$q_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right) \leq \frac{(n-1) s_{\mathbf{x}}^2}{\sigma^2} \leq q_{\chi^2(n-1)}\left(1 - \frac{\alpha}{2}\right) .$$

Po úpravě máme

$$\frac{(n-1) s_{\mathbf{x}}^2}{q_{\chi^2(n-1)}\left(1 - \frac{\alpha}{2}\right)} \leq \sigma^2 \leq \frac{(n-1) s_{\mathbf{x}}^2}{q_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right)} .$$

Realizace výběrového průměru a rozptylu jsou

$$\bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{457}{6} \doteq 76.17$$

$$s_{\mathbf{x}}^2 = \frac{1}{5} \left(\sum_{i=1}^6 x_i^2 - 6 \cdot (\bar{x})^2 \right) \doteq \frac{341.787}{5} \doteq 68.358$$

a příslušné kvantily jsou

$$q_{\chi^2(5)}(0.025) = 0.831$$

$$q_{\chi^2(5)}(0.975) = 12.83 .$$

Po dosazení dostáváme hledaný interval

$$26.64 \doteq \frac{341.787}{12.83} \leq \sigma^2 \leq \frac{341.787}{0.831} \doteq 411.296 .$$

10.9 (intervalový odhad pro rozptyl)

Deset opakovaných měření obsahu alkoholu ve vzorku krve má průměrnou hodnotu $\bar{x} = 0.15$ promile (alkoholu v krvi) a směrodatnou odchylku $s_{\bar{x}} = 0.01$ promile (alkoholu v krvi). Jakou hodnotu σ_0 překročí chyba metody (t.j. směrodatná odchylka) s pravděpodobností nejvýše $\alpha = 1\%$? Uveďte použité předpoklady.

Řešení:

U veličiny

$Y = \text{“obsahu alkoholu v krvi (v promilích)”}$

budeme předpokládat normální rozdělení $N(\mu, \sigma^2)$. Naše veličina

$X = \text{“chyba měření obsahu alkoholu v krvi (v promilích)”}$

bude tedy určena jako $X = Y - \mu$ s rozdělením $N(0, \sigma^2)$. Jednotlivá měření považujeme za nezávislá. Hledáme horní $1 - \alpha = 99\%$ intervalový odhad pro parametr rozptylu σ^2 , který bude tvaru $(0, \sigma_0^2)$.

Pro statistiku

$$T = \frac{(n-1)S_{\mathbf{X}}^2}{\sigma^2}$$

s χ^2 -rozdělením s $n-1 = 9$ stupni volnosti máme, že s pravděpodobností $1 - \alpha = 99\%$ pro realizaci t veličiny T platí, že:

$$q_{\chi^2(n-1)}(\alpha) \leq t = \frac{(n-1)s_{\mathbf{x}}^2}{\sigma^2}$$

takže

$$\sigma^2 \leq \frac{(n-1)s_{\mathbf{x}}^2}{q_{\chi^2(n-1)}(\alpha)} =: \sigma_0^2$$

a hledaná hranice tak je

$$\sigma_0 = s_{\mathbf{x}} \cdot \sqrt{\frac{n-1}{q_{\chi^2(n-1)}(\alpha)}} = 0.01 \cdot \sqrt{\frac{9}{q_{\chi^2(9)}(0.01)}} \doteq \frac{0.03}{\sqrt{2.0879}} \doteq 0.02076.$$