

9. cvičení z PST

29. listopadu 2017

9.1 (CLV - odhad pravděpodobnosti)

Hodíme 420 krát pravidelnou šestistěnnou hrací kostkou a výsledky hodů sčítáme. Pomocí centrální limitní věty odhadněte pravděpodobnost, že součet bude ležet mezi čísly 1400 a 1550.

Řešení:

Pro $i = 1, \dots, n$ (kde $n = 420$) si zavedeme diskrétní veličiny

$X_i =$ "hodnota, která padne na kostce při i -tém hodu"

Velichiny X_i považujeme za nezávislé. Střední hodnota a rozptyl jsou

$$E(X_i) = 3.5$$

$$D(X_i) = \frac{35}{12} .$$

Součet hodnot se vyjádří jako

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

se střední hodnotou

$$E(X) = \sum_{i=1}^n E(X_i) = 420 \cdot 3.5 = 1470$$

a rozptylem

$$D(X) = \sum_{i=1}^n D(X_i) = 420 \cdot \frac{35}{12} = 1225 = (35)^2 .$$

Máme teď určit $P(1400 \leq X \leq 1550)$. Podle centrální limitní věty má veličina X přibližně normální rozdělení $N(1470, (35)^2)$.

Můžeme tak psát:

$$\begin{aligned} P(1400 \leq X \leq 1550) &\doteq \Phi\left(\frac{1550 - 1470}{35}\right) - \Phi\left(\frac{1400 - 1470}{35}\right) = \Phi\left(\frac{16}{7}\right) - \Phi(-2) \doteq \\ &\doteq \Phi(2.2857) - 1 + \Phi(2) \doteq 0.98886 + 0.97725 - 1 = 0.96611 . \end{aligned}$$

Postup se dá ještě trochu změnit (protože veličina X má za hodnoty pouze přirozená čísla):

$$\begin{aligned} P(1400 \leq X \leq 1550) &= P(1399 < X \leq 1550) \doteq \Phi\left(\frac{1550 - 1470}{35}\right) - \Phi\left(\frac{1399 - 1470}{35}\right) = \Phi\left(\frac{16}{7}\right) - \Phi\left(-\frac{71}{35}\right) \doteq \\ &\doteq \Phi(2.2857) - 1 + \Phi(2.02857) \doteq 0.98886 + 0.97875 - 1 = 0.96761 . \end{aligned}$$

Předchozí postup dal 96.611% a tento postup zase 96.761%, což je rozdíl 0.15%.

9.2 (CLV - odhad počtu)

Na oboru má studovat 600 studentů. Přibližně jen 2/3 z přijatých studentů se pak nakonec zapíše na studium. Kolik studentů by se mělo přijmout, aby počet zapsaných byl co největší, ale aby překročil 600 s pravděpodobností nejvýše 5%? Jaký pak bude průměrný počet zapsaných studentů?

Řešení:

Nechť $n \in \mathbb{N}$ je počet přijatých studentů. Pro $i = 1, \dots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \text{ } i\text{-tý přijatý student se zapíše na studium,} \\ 0 & , \text{ } i\text{-tý přijatý student se nezapíše na studium.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělením, kde $P(X_i = 1) = \frac{2}{3}$. Počet zapsaných studentů bude veličina

$$Y_n = \sum_{i=1}^n X_i ,$$

kteřá má binomické rozdělení $\text{Bi}\left(n, \frac{2}{3}\right)$. Zajímá nás teď největší n takové, že

$$P(Y_n > 600) \leq 0.05 .$$

Opět použijeme centrální limitní větu na normovanou veličinu

$$\text{norm}(Y_n) = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - \frac{2}{3}n}{\sqrt{\frac{2}{3} \cdot \frac{1}{3}n}} = \frac{3Y_n - 2n}{\sqrt{2n}} .$$

Pomocí úprav nerovností teď můžeme psát :

$$\begin{aligned} 0.05 &\geq P(Y_n > 600) = P\left(\frac{3Y_n - 2n}{\sqrt{2n}} > \frac{3 \cdot 600 - 2n}{\sqrt{2n}}\right) = \\ &= P\left(\text{norm}(Y_n) > \frac{1800 - 2n}{\sqrt{2n}}\right) = 1 - P\left(\text{norm}(Y_n) \leq \frac{1800 - 2n}{\sqrt{2n}}\right) = \\ &= 1 - F_{\text{norm}(Y_n)}\left(\frac{1800 - 2n}{\sqrt{2n}}\right) \doteq 1 - \Phi\left(\frac{1800 - 2n}{\sqrt{2n}}\right) \end{aligned}$$

tedy

$$\Phi\left(\frac{1800 - 2n}{\sqrt{2n}}\right) \geq 0.95$$

a

$$\left(\sqrt{2n}\right)^2 + \Phi^{-1}(0.95) \cdot \sqrt{2n} - 1800 \leq 0 .$$

Dostali jsme kvadratickou nerovnost a hledáme největší $n \in \mathbb{N}$, které ji splňuje. Z kořenů kvadratické rovnice si tedy vezmeme ten, co je více vpravo. Označme si ještě pro jednoduchost

$$b = \Phi^{-1}(0.95) \doteq 1.645 .$$

Dostaneme tak

$$\sqrt{2n} \leq \frac{-b + \sqrt{b^2 + 4 \cdot 1800}}{2}$$

a

$$n \leq \frac{(\sqrt{b^2 + 7200} - b)^2}{8} \doteq \frac{(\sqrt{1.645^2 + 7200} - 1.645)^2}{8} \doteq 865.774$$

Mělo by se tedy přijmout $n = 865$ studentů. Průměrný počet těch, co se zapíší tak bude

$$E(Y_n) = \frac{2}{3} \cdot n = \frac{2}{3} \cdot 865 \doteq 576,67 .$$

Opět můžeme udělat ještě trochu jiné řešení (díky tomu, že hodnoty veličiny Y_n jsou pouze přirozená čísla):

$$\begin{aligned} 0.05 &\geq P(Y_n > 600) = P(Y_n \geq 601) = P\left(\text{norm}(Y_n) \geq \frac{3 \cdot 601 - 2n}{\sqrt{2n}}\right) = \\ &= 1 - P\left(\text{norm}(Y_n) < \frac{1803 - 2n}{\sqrt{2n}}\right) = 1 - F_{\text{norm}(Y_n)}\left(\frac{1803 - 2n}{\sqrt{2n}}\right) \doteq \\ &\doteq 1 - \Phi\left(\frac{1803 - 2n}{\sqrt{2n}}\right) \end{aligned}$$

Obdobně dostaneme

$$n \leq \frac{(\sqrt{b^2 + 7212} - b)^2}{8} \doteq \frac{(\sqrt{1.645^2 + 7212} - 1.645)^2}{8} \doteq 867.245$$

Tentokrát tedy vyjde $n = 867$ studentů.

9.3 (CLV a Čebyševova nerovnost - spojité rovnoměrné rozdělení, odhad intervalu)

Chodíme náhodně k tramvaji, která jezdí po 6 minutách. Jezdíme dvakrát denně, 20 dní v měsíci. Jestliže označíme X_i náhodnou veličinu dobu čekání a T průměrnou dobu čekání za měsíc, pak vypočítejte a odhadněte pomocí Čebyševovy nerovnosti číslo tak $\varepsilon > 0$, aby $P(|T - 3| < \varepsilon) \geq 0.9$.

Řešení:

Příklad 9: <https://math.feld.cvut.cz/prucha/psi/wcv7.pdf>

9.4 (CLV - odhad intervalu)

Házíme 100-krát pravidelnou mincí a náhodná veličina X je počet rubů. Určete (přibližně) číslo ε tak, aby $P(|X - E(X)| < \varepsilon) = 0.9$.

Řešení:

Náhodná veličina X má binomické rozdělení $\text{Bi}(100, 0.5)$. Je tedy $E(X) = 100 \cdot 0.5 = 50$ a $D(X) = 100 \cdot 0.5 \cdot (1 - 0.5) = 25$. Z centrální limitní věty vyplývá, že můžeme předpokládat pro náhodnou veličinu X přibližně normální rozdělení $N(50, 25)$. Tedy $F_X(t) \doteq \Phi\left(\frac{t-50}{5}\right)$ pro $t \in \mathbb{R}$.

Číslo ε určíme z rovnice

$$\begin{aligned} 0.9 &= P(|X - E(X)| < \varepsilon) = P(50 - \varepsilon < X < 50 + \varepsilon) = F_X(50 + \varepsilon) - F_X(50 - \varepsilon) \doteq \\ &\doteq \Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right) = 2\Phi\left(\frac{\varepsilon}{5}\right) - 1 \\ \Rightarrow \Phi\left(\frac{\varepsilon}{5}\right) &= 0.95 \Rightarrow \frac{\varepsilon}{5} = \Phi^{-1}(0.95) \Rightarrow \varepsilon \doteq 5 \cdot 1.645 = 8.225. \end{aligned}$$

9.5 (metoda momentů a max. věrohodnosti)

Odhadněte parametr w geometrického rozdělení

$$p_i = w^i(1 - w), \quad i \in \mathbb{N}_0$$

na základě realizace s následujícími četnostmi výsledků:

hodnota	0	1	2	3
pozorovaná četnost	20	10	7	3

Použijte metodu momentů i metodu maximální věrohodnosti.

Řešení:

Součet pravděpodobností všech hodnot je 1. Pravděpodobnosti hodnot v pokusu musí být nenulové, protože dané hodnoty byly skutečně realizovány, takže

$$0 < w < 1 .$$

Metoda maximální věrohodnosti:

Hledáme hodnotu q , která maximalizuje funkci věrohodnosti

$$\begin{aligned} L(w) &= \prod_{i=1}^n P(X_i = x_i) = (1-w)^{20} (w(1-w))^{10} (w^2(1-w))^7 (w^3(1-w))^3 = \\ &= w^{33}(1-w)^{40} \end{aligned}$$

kde X_i jsou jednotlivé nezávislé veličiny (pokusy) a x_i naměřené hodnoty. Funkce L je nezáporná a spojitá na uzavřené množině $\langle 0, 1 \rangle$, takže zde nabývá maxima. To nemůže být v krajních bodech a proto je nabyto uvnitř dané množiny. To tedy odpovídá hledání maxima funkce

$$\ell(w) = \ln(L(w)) = 33 \cdot \ln(w) + 40 \cdot \ln(1-w)$$

na intervalu $(0, 1)$. Protože maximum existuje, musí pro něj platit

$$0 = \ell'(q) = \frac{33}{w} - \frac{40}{1-w}$$

neboli

$$w = \frac{33}{73} \doteq 0.45205$$

což vyhovuje zadání.

Metoda momentů:

Střední hodnota je

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} iw^i(1-w) = \sum_{i=1}^{\infty} iw^i - \sum_{i=1}^{\infty} iw^{i+1} = \sum_{i=1}^{\infty} iw^i - \sum_{i=2}^{\infty} (i-1)w^i = \\ &= w + \sum_{i=2}^{\infty} (i-i+1)w^i = \sum_{i=1}^{\infty} w^i = w \sum_{i=1}^{\infty} w^{i-1} = \frac{w}{1-w} \end{aligned}$$

její odhad z realizace je

$$\bar{x} = \frac{1}{n} \sum_i i n_i = \frac{1}{20+10+7+3} \cdot (0 \cdot 20 + 1 \cdot 10 + 2 \cdot 7 + 3 \cdot 3) = \frac{33}{40} .$$

Porovnáním dostaneme

$$\frac{w}{1-w} = E(X) = \bar{x} = \frac{33}{40}$$

což dává opět řešení

$$w = \frac{33}{73} \doteq 0.45205$$

jako v předchozí metodě.

Jak je snadno vidět, v případě geometrického rozdělení dostáváme pro jeho parametr w vždy stejné výsledky pro obě metody.

9.6 (metoda max. věrohodnosti)

Počet kazů X na tabulkách skla se řídí Poissonovým rozdělením. Bylo pozorováno

$X =$ počet kazů na dané tabulce	0	1	2	3	5
pozorovaná četnost	17	4	1	2	1

Metodou maximální věrohodnosti určete parametr λ tohoto Poissonova rozdělení.

Řešení:

Máme realizaci náhodného výběru (x_1, \dots, x_n) .

Pro náhodnou veličinu X s rozdělením Poiss(λ) je $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Věrohodnostní funkce je

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \left(\frac{\lambda^0}{0!} e^{-\lambda}\right)^{17} \left(\frac{\lambda^1}{1!} e^{-\lambda}\right)^4 \left(\frac{\lambda^2}{2!} e^{-\lambda}\right)^1 \left(\frac{\lambda^3}{3!} e^{-\lambda}\right)^2 \left(\frac{\lambda^5}{5!} e^{-\lambda}\right)^1$$

$$= \frac{\lambda^{17}}{1440} e^{-25\lambda},$$

logaritmicke-věrohodnostní funkce je

$$l(\lambda) = \ln L(\lambda) = 17 \ln \lambda - 25\lambda - \ln 1440$$

a z její derivace

$$\frac{d}{d\lambda} l(\lambda) = \frac{17}{\lambda} - 25.$$

získáme řešení

$$\frac{17}{\hat{\lambda}} - 25 = 0 \quad \implies \quad \hat{\lambda} = \frac{17}{25}.$$

9.7 (metoda momentů a max. věrohodnosti)

Náhodná veličina X nabývá hodnot s pravděpodobnostmi dle tabulky, kde c, q jsou reálné parametry rozdělení. Z četností hodnot v náhodném výběru, uvedených v tabulce, odhadněte parametry c a q .

hodnota i	1	2	3
pravděpodobnost $p_X(i)$	$c - q$	c	$c + q$
četnost n_i	8	10	5

Řešení:

Protože součet pravděpodobností všech hodnot je 1, musí být

$$1 = (c - q) + c + (c + q) = 3c$$

tedy $c = \frac{1}{3}$. Současně musí být pravděpodobnosti nezáporné, tj. $0 \leq c - q = \frac{1}{3} - q$ a $0 \leq c + q = \frac{1}{3} + q$, takže $|q| \leq \frac{1}{3}$. Zbývá tedy odhadnout parametr q .

Metoda maximální věrohodnosti:

Hledáme hodnotu q , která maximalizuje funkci věrohodnosti

$$L(q) = P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) = \left(\frac{1}{3} - q\right)^8 \cdot \left(\frac{1}{3}\right)^{10} \cdot \left(\frac{1}{3} + q\right)^5$$

kde X_i jsou jednotlivé nezávislé veličiny (pokusy) a x_i naměřené hodnoty. Funkce L je nezáporná a spojitá na uzavřené množině $\langle -\frac{1}{3}, \frac{1}{3} \rangle$, takže zde nabývá maxima. To nemůže být v krajních bodech a proto je nabyto uvnitř dané množiny. To odpovídá hledání maxima funkce

$$\ell(q) = \ln(L(q)) = 8 \cdot \ln\left(\frac{1}{3} - q\right) + 10 \cdot \ln\frac{1}{3} + 5 \cdot \ln\left(\frac{1}{3} + q\right)$$

na intervalu $(-\frac{1}{3}, \frac{1}{3})$. Protože maximum existuje, musí pro něj platit

$$0 = \ell'(q) = \frac{-8}{\frac{1}{3} - q} + \frac{5}{\frac{1}{3} + q}$$

Odhad parametru q je

$$q = -\frac{1}{13} \doteq -0.07692.$$

Odhady pravděpodobností hodnot 1, 2, 3 jsou tedy

$$p_X(1) = \frac{16}{39} \doteq 0.4103 \quad p_X(2) = \frac{1}{3} \doteq 0.3333 \quad p_X(3) = \frac{10}{39} \doteq 0.2564$$

což vyhovuje zadání.

Metoda momentů:

Střední hodnota je

$$E(X) = \left(\frac{1}{3} - q\right) + 2 \cdot \frac{1}{3} + 3 \cdot \left(\frac{1}{3} + q\right) = 2 + 2q$$

její odhad z realizace je

$$\bar{x} = \frac{1}{n} \sum_i i n_i = \frac{1}{8 + 10 + 5} \cdot (1 \cdot 8 + 2 \cdot 10 + 3 \cdot 5) = \frac{43}{23}.$$

Porovnáním dostaneme

$$2 + 2q = E(X) = \bar{x} = \frac{43}{23}$$

což odpovídá hodnotě

$$q = -\frac{3}{46} \doteq -0.06522.$$

Odhady pravděpodobností hodnot 1, 2, 3 jsou tedy

$$p_X(1) = \frac{55}{138} \doteq 0.3986 \quad p_X(2) = \frac{1}{3} \doteq 0.3333 \quad p_X(3) = \frac{37}{138} \doteq 0.2681$$

což opět vyhovuje zadání.

9.8 (metoda momentů a max. věrohodnosti)

V urně je mnoho hracích kostek, z nichž některé jsou správné, některé falešné. Na falešných padá šestka s pravděpodobností $1/2$, zbývající čísla mají stejnou pravděpodobnost. Opakovaně jsme vytáhli

kostku, hodili jí a vrátili ji zpět. Četnost výsledků udává tabulka:

hodnota	1	2	3	4	5	6
četnost	18	20	12	15	10	25

Odhadněte, kolik procent kostek je falešných.

Řešení:

Podíl falešných kostek označme $p \in \langle 0, 1 \rangle$. Naše náhodná veličina je

$X =$ “hodnota, která padne na dané kostce”

a můžeme ji vyjádřit jako směs $X = \text{Mix}_p(X_1, X_2)$ složenou z náhodných veličin

$X_1 =$ “hodnota, která padne na falešné kostce”

$X_1 :$ “množina falešných kostek” $\rightarrow \mathbb{R}$

$X_2 =$ “hodnota, která padne na správné kostce”

$X_2 :$ “množina správných kostek” $\rightarrow \mathbb{R}$.

Metoda momentů: Máme

$$E(X_1) = \frac{1}{10}(1 + \dots + 5) + \frac{1}{2} \cdot 6 = \frac{9}{2}$$

a

$$E(X_2) = \frac{1}{6}(1 + \dots + 6) = \frac{7}{2}$$

a z definice směsi tak dostaneme

$$E(X) = p \cdot E(X_1) + (1 - p) \cdot E(X_2) = p \cdot \frac{9}{2} + (1 - p) \cdot \frac{7}{2} = p + \frac{7}{2}.$$

Realizace výběrového průměru je

$$\bar{x} = \frac{18 \cdot 1 + 20 \cdot 2 + 12 \cdot 3 + 15 \cdot 4 + 10 \cdot 5 + 25 \cdot 6}{18 + 20 + 12 + 15 + 10 + 25} = \frac{354}{100} = 3.54.$$

Srovnáním dostaneme

$$\hat{p} + 3.5 = E(X) = \bar{x} = 3.54,$$

což dává $\hat{p} = 0.04 \in \langle 0, 1 \rangle$, a to vyhovuje zadání.

Metoda maximální věrohodnosti:

Z definice směsi máme pro její pravděpodobnostní funkci, že

$$p_X(i) = p \cdot p_{X_1}(i) + (1 - p) \cdot p_{X_2}(i) = \begin{cases} p \frac{1}{10} + (1 - p) \frac{1}{6} = \frac{5-2p}{30} & , i = 1, \dots, 5, \\ p \frac{1}{2} + (1 - p) \frac{1}{6} = \frac{1+2p}{6} & , i = 6. \end{cases}$$

Ve směsi rozdělení šestka padla 25×, ostatní čísla padla 75× (není třeba mezi nimi rozlišovat, protože mají stejnou pravděpodobnost). Tedy

$$L(p) = \left(\frac{5-2p}{30} \right)^{75} \cdot \left(\frac{1+2p}{6} \right)^{25},$$

$$\ell(p) = 75 \ln(5-2p) + 25 \ln(1+2p) - 75 \ln 30 - 25 \ln 6.$$

Maximum nastává pro \hat{p} takové, že

$$0 = \ell'(\hat{p}) = \frac{-150}{5 - 2\hat{p}} + \frac{50}{1 + 2\hat{p}} = 50 \cdot \frac{2 - 8\hat{p}}{(5 - 2\hat{p})(1 + 2\hat{p})},$$

$$\hat{p} = \frac{1}{4} \in \langle 0, 1 \rangle.$$

protože na intervalu $\langle 0, \frac{1}{4} \rangle$ je $\ell' > 0$ a na $(\frac{1}{4}, 1)$ je $\ell' < 0$. Tato hodnota je i v souladu s počátečními omezujícími podmínkami.

9.9 (metoda momentů a max. věrohodnosti)

Dvě diskrétní náhodné veličiny X, Y mají pravděpodobnostní funkce dané tabulkou. Odhadněte koeficient c směsi $Z = \text{Mix}_c(X, Y)$ z četností jejích realizací uvedených v tabulce.

hodnota	1	2	3	4
p_X	0.1	0.2	0.2	0.5
p_Y	0.5	0.2	0.2	0.1
četnost	30	20	15	35

Řešení:

Z definice směsi máme pro parametr c nutnou podmínku $0 \leq c \leq 1$.

Metoda momentů: Z definice směsi $Z = \text{Mix}_c(X, Y)$ dostaneme

$$E(Z) = c \cdot E(X) + (1 - c) \cdot E(Y).$$

Pro střední hodnoty X a Y máme

$$E(X) = 0.1 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.5 = 3.1$$

$$E(Y) = 0.5 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.1 = 1.9.$$

Takže dostaneme

$$E(Z) = c \cdot E(X) + (1 - c) \cdot E(Y) = 1.9 + 1.2 \cdot c.$$

Hodnota realizace výběrového průměru je

$$\bar{z} = \frac{30 + 2 \cdot 20 + 3 \cdot 15 + 4 \cdot 35}{100} = \frac{51}{20} = 2.55.$$

Jejich srovnáním dostáváme

$$1.9 + 1.2 \cdot c = E(Z) = \bar{z} = 2.55$$

takže výsledek je

$$c = \frac{13}{24} \doteq 0.5417.$$

Metoda maximální věrohodnosti: Z definice $Z = \text{Mix}_c(X, Y)$ pro pravděpodobnostní funkci dostaneme

$$p_Z = c \cdot p_X + (1 - c) \cdot p_Y$$

hodnota	1	2	3	4
p_Z	$0.5 - 0.4c$	0.2	0.2	$0.1 + 0.4c$

Pro funkci věrohodnosti pak máme

$$L(c) = (0.5 - 0.4 \cdot c)^{30} \cdot 0.2^{20+15} \cdot (0.1 + 0.4 \cdot c)^{35}$$

Funkce L je nezáporná a spojitá na uzavřené množině $\langle 0, 1 \rangle$, takže zde nabývá maxima. To odpovídá hledání maxima funkce

$$\ell(c) = \ln L(c) = 30 \cdot \ln(0.5 - 0.4c) + 35 \cdot \ln 0.2 + 35 \cdot \ln(0.1 + 0.4c) .$$

Derivace

$$0 = \ell'(\hat{c}) = -\frac{30 \cdot 0.4}{0.5 - 0.4\hat{c}} + \frac{35 \cdot 0.4}{0.1 + 0.4\hat{c}} = \frac{5.8 - 10.4\hat{c}}{(0.5 - 0.4\hat{c})(0.1 + 0.4\hat{c})}$$

je nulová ve stacionárním bodě

$$\hat{c} = \frac{29}{52} \doteq 0.5577 .$$

V intervalu $(0, \frac{29}{52})$ je ℓ' evidentně kladná (o znaménku rozhoduje jen výraz v čitateli, výraz ve jmenovateli je kladný) a v intervalu $(\frac{29}{52}, 1)$ je ℓ' zase záporná. Takže v bodě $\hat{c} = \frac{29}{52} \doteq 0.5577$ je skutečně věrohodnost maximální.