Bayesovské metody 4- Odhady

6. října 2020

Outline

Outline

Bodové odhady

Připomeňme:

$\boldsymbol{X} - (X_1, \ldots, X_n)^\top$ – hustota $r(\boldsymbol{x}|\boldsymbol{\theta})$ vzhledem k ν_n,

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ – parametr, $\boldsymbol{\theta} \in \Theta$

$\pi(\boldsymbol{\theta}|\boldsymbol{X})$– aposteriorní hustota

$q(\boldsymbol{\theta})$ – apriorní hustota

Úloha bodu odhadu parametru $\boldsymbol{\theta}$ lze formulovat jako statistický rozhodovací problém $(\Theta, \mathcal{D}, R)$, kde $\mathcal{D} = \Theta$

$\delta(\boldsymbol{X})$ – odhad parametru $\boldsymbol{\theta}$

$L(\boldsymbol{\theta}, \delta(\boldsymbol{X}))$ – ztrátová funkce

$R(\boldsymbol{\theta}, \delta) = E(L(\boldsymbol{\theta}, \delta(\boldsymbol{X}))|\boldsymbol{\theta})$ — riziková fce

$E\Big(R(\boldsymbol{\theta}, \delta)\Big)$ — bayesovská riziková fce

bayesovska rozhodovaci fce $\delta_*(\boldsymbol{x})$ minimalizuje

$$\int_\Theta L(\theta, \delta(\boldsymbol{x}))\pi(\theta|\boldsymbol{x})d\lambda(\theta)$$

vzhledem k $\delta(\boldsymbol{x})$

Odhady, jednorozměrný případ

Typicky:

$$L_{a,w}(\theta, \delta(\boldsymbol{x})) = w(\theta)|\theta - \delta(\boldsymbol{x})|^a, \quad \theta \in \Theta, \quad \delta \in \Delta,$$

$w(\theta)$ – nezáporná měřitelná funkce, často hustota popř. $w(\theta) \equiv 1$

$a > 0$ – nejčastěji $a = 1$ nebo 2

Bayesovska rozhodovaci fce (odhad) $\delta_{a,w}(\boldsymbol{x})$ minimalizuje

$$\int_\Theta |\theta - \delta(\boldsymbol{x})|^a w(\theta)\pi(\theta|\boldsymbol{x}) d\lambda(\theta)\Big) = \int_\Theta |\theta - \delta|^a \pi_1(\theta|\boldsymbol{x}) d\lambda(\theta) \times \frac{r_1(\boldsymbol{x})}{r(\boldsymbol{x})}$$

$$q_1(\theta) = \frac{q(\theta)w(\theta)}{\int_\Theta q(\theta)w(\theta)d\theta}, \quad \pi_1(\theta|\boldsymbol{x}) = \frac{r(\boldsymbol{x}|\theta)q_1(\theta)}{r_1(\boldsymbol{x})},$$

$$r_1(\boldsymbol{x}) = \int_\Theta r(\boldsymbol{x}|\theta)q_1(\theta)d\lambda(\theta)$$

$q_1,\ \pi_1,\ r_1$ — nové hustoty, vhodné pro výpočet

pro $a = 2$ je $E(|\theta - \delta(\boldsymbol{X})|^2|\boldsymbol{X} = \boldsymbol{x})$ minimum vzhledem k $\delta(\boldsymbol{X})$ je dosaženo pro $\delta(\boldsymbol{X}) = E(\theta|\boldsymbol{X})$

poznámka – $\min_a E(Y - a)^2$ je dosaženo pro $a = EY$

Řešení pro w obecné a $a = 2$:

$$\delta_{2,w}(\mathbf{x}) = \frac{\int_\Theta \theta \pi_1(\theta|\mathbf{x}) d\lambda(\theta)}{\int_\Theta \pi_1(\theta|\mathbf{x}) d\lambda(\theta)} = E_1\Big(\theta|\mathbf{x}\Big)$$

$\rho^*_{w,2}(\mathbf{x}) = E_1\big(var_1(\theta|\mathbf{x})\big)$ – bayesovské riziko
předpoklady: $0 < \int_\Theta \theta^2 \pi_1(\theta|\mathbf{x}) < \infty$

Řešení pro w obecné a $a = 1$ je $\delta_{1,w}(\mathbf{x})$ – median aposteriorního rozdělení π_1

Příklad 1, pokračování. Problém určení IQ sledovaného dítěte.
Výsledek testu X má rozdělení $N(\theta, 100)$, θ má rozd. $N(100, 225)$
Aposteriorní rozdělení θ při daném X je normální

$$N\Big(X\frac{225}{225+100} + 100\frac{100}{225+100}, 69,23\Big)$$

Pro ztratovou funkci $L(\theta, \delta(x)) = (\theta - \delta(x))^2 \exp\{\frac{1}{2r}(\theta - 100)^2\}$, $r > 225$
dáváme větší váhu hodnotám θ dále od 100

Pak apriorni rozdeleni $q_1(\theta)$ je $N\Big(100, \Big(\frac{1}{225} - \frac{1}{r}\Big)^{-1}\Big)$
a aposteriorni $\pi_1(\theta|x)$ je
$N\Big(x\frac{225r}{225r+100(r-225)} + 100\frac{100(r-225)}{225r+100(r-225)}, \frac{100\cdot 225r}{225r+100(r-225)}\Big)$

a dostaneme bayesovskou rozhodovací fci, tj. odhad θ:

$$\delta_{2,w}(x) = x\frac{225r}{225r + 100(r-225)} + 100\frac{100(r-225)}{225r + 100(r-225)}$$

porovnáme-li ji s odhadem θ pro $w(\theta) = 1$: $\delta_{1,2}(x)$

pak pro $x < 100$ je $\delta_{2,w}(x) < \delta_{1,2}(x)$ a pro $x > 100$ je $\delta_{2,w}(x) > \delta_{1,2}(x)$

Příklad Úkolem dvou fyziků je odhadnout jistou fyzikální konstatntu θ.

1. fyzik vyjádří představu (apriorní rozdělení) parametru θ – $N(900, 400)$

2. fyzik vyjádří představu (apriorní rozdělení) parametr θ – $N(800, 6400)$

výsledek pokusu X – $N(\theta, 1600)$

1. fyzik aposteriorní rozdělení – $N(\frac{x+3600}{5}, 320)$, $\quad N(\frac{\overline{x}_n+3600/n}{1+4/n}, \frac{1600}{n+4})$

2 fyzik aposteriorní rozdělení – $N(\frac{4x+800}{5}, 1750)$, $\quad N(\frac{4\overline{x}_n+800/n}{4+1/n}, \frac{6400}{4n+1})$

opakuje se pokus nezávisle n krát

Příklad Empirické bayesovské postupy. X má Poissonovo rozdělení s parametrem $\lambda > 0$, z minulosti $Y_1, \dots, Y_N$ nezávislé, Y_i má Poissovovo rozdělení s par. λ_i, $\lambda_1, \dots, \lambda_N$ nezávislé stejně rozdělené s hustotou $q(.)$

Nejprve předpokládejme že λ_i má rozdělení gama $(\alpha, \beta), \alpha > 0, \beta > 0$

Pak pro nepodmíněnou střední hodnotu a nepodmíněný rozptyl náhodné veličiny Y_i

$$EY_i = E\lambda_i = \beta/\alpha, \quad var\{Y_i\} = E\lambda_i + E\Big(\lambda_i - \beta/\alpha\Big)^2 = (\beta/\alpha)(\alpha+1)$$

$$\widehat{\alpha} = \frac{\overline{Y}}{\widehat{\sigma}^2 - \overline{Y}}, \quad \widehat{\beta} = \frac{\overline{Y}^2}{\widehat{\sigma}^2 - \overline{Y}}, \quad je - li \quad \widehat{\sigma}^2 - \overline{Y} > 0$$

kde $\widehat{\sigma}^2 = \sum_i (Y_i - \overline{Y})^2/(N-1)$

Odhad λ při kvadratické ztrátové fci je $\widehat{\lambda} = \frac{\widehat{\beta}+x}{1+\widehat{\alpha}},$, $\quad je - li \quad \widehat{\sigma}^2 - \overline{Y} > 0$

Jiný postup– pro obecnou apriorní hustotu $q(.)$ při kvadratické ztrátové funkci má odhad θ tvar

$$\widetilde{\lambda} = \frac{\int_0^\infty \lambda r(x|\lambda) q(\lambda) d\lambda}{\int_0^\infty \lambda r(x|\lambda) q(\lambda) d\lambda} = \frac{(x+1)r(x+1)}{r(x)}$$

$r(x) = P(X = x)$, což lze odhadnout $\widehat{r}(x) = (pocet\ Y_i = x)/N$ a tedy máme nový odhad

$$\widetilde{\widetilde{\lambda}} = \frac{(x+1)\widehat{r}(x+1)}{\widehat{r}(x)}$$

vhodne jen pro velka N, jinak znacne nestabilni

Odhady, vícerozměrný případ

Ztrátová funkce

$$L_A(\boldsymbol{\theta}, \boldsymbol{\delta}(\boldsymbol{X})) = \left(\boldsymbol{\theta} - \boldsymbol{\delta}(\boldsymbol{X})\right)^\top \boldsymbol{A}\left(\boldsymbol{\theta} - \boldsymbol{\delta}(\boldsymbol{X})\right),$$

$\boldsymbol{A}$ – symetrická pozitivně semidefinitní matice typu $k \times k$
$\boldsymbol{\delta}(\boldsymbol{X}) = (\delta(\boldsymbol{X}), \dots, \delta(\boldsymbol{X}))^\top$ odhad parametru $\boldsymbol{\theta}$

Jsou-li prvky varianční matice $var\{\boldsymbol{X}|\boldsymbol{\theta}\}$ konečné, bayesovská optimální rozhodovací funkce $\boldsymbol{\delta}^*_A(\boldsymbol{X})$ splňuje

$$\left(E(\boldsymbol{\theta}|\boldsymbol{X}) - \boldsymbol{\delta}^*_A(\boldsymbol{X})\right)^\top \boldsymbol{A}\left(E(\boldsymbol{\theta}|\boldsymbol{X}) - \boldsymbol{\delta}^*_A(\boldsymbol{X})\right) = 0$$

a bayesovské riziko je

$$\rho^*_A(q) = E\{tr\{\boldsymbol{A}\, var(\boldsymbol{\theta})\}\} \cdot \int_\Theta r(\boldsymbol{X}|\boldsymbol{\theta}) q(\boldsymbol{\theta} d\lambda(\boldsymbol{\theta})$$

O platnosti se lze přesvědčit vhodnou úpravou

$$\int_\Theta r(\boldsymbol{X}|\boldsymbol{\theta}) q(\boldsymbol{\theta}) \left(\boldsymbol{\theta} - \boldsymbol{\delta}(\boldsymbol{X})\right)^\top \boldsymbol{A}\left(\boldsymbol{\theta} - \boldsymbol{\delta}(\boldsymbol{X})\right) d\lambda(\boldsymbol{\theta})$$

Priklad Jsou-li X_i, $i = 1, \ldots, n$ nezávislé a X_i rozdělení $N(\mu, \sigma^2)$, pak konjugované rozdělení parametrů μ, σ^{-2} je normální-gama rozdělení:
μ podmíněno σ^{-2} je $N(a, \sigma^2/r)$,
σ^{-2} má gama rozdělení (c, d),
parametry (a, r, c, d), $a \in \mathcal{R}_1$, $r > 0$, $2d = 1, 2, \ldots, d > 0$

aposteriorní rozdělení je normální-gama $\mu^*, r + n, c^*, d + n/2)$

$$\mu^* = \frac{ra + n\bar{x}_n}{n + r}, \quad c^* = c + \frac{1}{2}\sum_i (x_i - \bar{x}_n)^2 + \frac{rn(\bar{x}_n - a)^2}{2(r + n}, \quad \bar{x}_n = \sum_i x_i/n.$$

Při ztrátové funkci $L_A(.,.)$ s $\boldsymbol{A}$ lib.sym., pozitivně definit. jsou odhady μ a σ^{-2}:

$$\widehat{\mu} = \frac{ra + \sum_i X_i}{r + n}, \quad \widehat{\sigma^{-2}} = \frac{d + n/2}{c + \frac{1}{2}\sum_i (X_i - \overline{X}_n)r + \frac{rn(\overline{X}_n - a)^2}{2(r+n)}}$$

Pro $r = 0$, $c = 0$, $d = -1/2$ dostaneme klasické odhady.

$\frac{\mu - \widehat{\mu}}{(d+n/2)r/c^*}$ má $t-$rozdělení o $2(d + n/2)$ stupních volnosti
maximálně věrohodný odhad parametru σ^{-2} je založený na aposteriorní

$$\frac{d + n/2 - 1}{c + \frac{1}{2}\sum_i (X_i - \overline{X}_n)r + \frac{rn(\overline{X}_n - a)^2}{2(r+n)}}$$

Věrohodnostní množiny

Věrohodnostní množiny jsou bayesovskou analogií konfidenčních množin (oblastí) či intervalů při klasickém přístupu. Pozor na nejednotnou českou terminologii!

Konfidenční oblast pro parametr $\boldsymbol{\theta}$ s koeficientem spolehlivosti $1-\alpha$ – je definovaná jako borelovská množina $D_\alpha(\boldsymbol{X}) \subsetneqq \Theta$ s vlastností

$$P(\boldsymbol{\theta} \in D_\alpha(\boldsymbol{X})|\boldsymbol{\theta}) = 1-\alpha, \quad pro \quad vs. \quad \in \Theta,$$

t.j. $D_\alpha(\boldsymbol{X})$ pokrývá skutečnou hodnotu parametru $\boldsymbol{\theta}$ s pravděpodobností $1-\alpha$.

Při bayesovském přístupu – $100(1-\alpha)\%$ věrohodnostní množina (credible region v angličtině) parametru $\boldsymbol{\theta}$ je definovaná ako borelovská množina $C_\alpha(\boldsymbol{X}) \subsetneqq \Theta$ s vlastností

$$P(\boldsymbol{\theta} \in C_\alpha(\boldsymbol{X})|\boldsymbol{X}) = \int_{C_\alpha(\boldsymbol{X})} \pi(\boldsymbol{\theta}|\boldsymbol{X}) d\lambda\boldsymbol{\theta} = 1-\alpha,$$

pravděpodobnost, že $\boldsymbol{\theta}$ náleží do $C_\alpha(\boldsymbol{X})$ je $1-\alpha$; $1-\alpha$ se nazývá věrohodnost.

Snaha zkonstruovat $C_\alpha(\boldsymbol{X})$ s co nejmenším obsahem:

$C^*_\alpha(\boldsymbol{X}) = \{\boldsymbol{\theta};\, \pi(\boldsymbol{\theta}|\boldsymbol{X}) \geq k_\alpha\}$

s vlastností $\int_{C^*_\alpha(\boldsymbol{X})} \pi(\boldsymbol{\theta}|\boldsymbol{X} d\lambda(\boldsymbol{\theta}) = 1-\alpha, \quad \int_{C^*_\alpha(\boldsymbol{X})} d\lambda(\boldsymbol{\theta}) \leq \int_{C_\alpha(\boldsymbol{X})} d\lambda(\boldsymbol{\theta})$

Příklad 1, pokrač. Chceme zkonstruovat 95% verohodný interval pro kvocient inteligence θ. Aposteriorní rozdělení je $N(x\frac{225}{325} + 100\frac{100}{325}, 69,23)$.
Tedy nejkratší 95% verohodný interval pro θ je

$$(x\frac{225}{325} + 100\frac{100}{325} - 16,3,\ x\frac{225}{325} + 100\frac{100}{325} + 16,3)$$

Klasický 95% interval spolehlivosti je $(x - 19,6,\ x + 19,6)$.

Příklad $X_1, \ldots, X_n$ je náh. výběr z $N(\mu, \sigma^2)$, $\mu, \sigma^{-2} > 0$ neznáme. Apriorní rozdělení parametrů μ, σ^{-2} je gama-normální (a, r, c, d)
Pak aposteriorní rozdělení μ, σ^{-2} je gama-normální (μ^*, r^*, c^*, d^*) a aposteriorní rozdělení $(\mu - \mu^*)(d^*, r/c^*)^{1/2}$ je t-rozdělení s $2d^*$ stupni volnosti, kde

$$\mu^* = \bar{x}_n \frac{n}{n+r} + a\frac{r}{n+r},$$
$$c^* = c + \frac{1}{2}\sum_i (x_i - \bar{x}_n)^2 + \frac{rn(\bar{x}_n - a)^2}{2(r+n}$$
$$d^* = d + n/2$$

nejkratší 95% věrohodný interval pro μ má tvar

$$\left(\mu^* - t_{1-\alpha/2}(2d^*)\left(c^*(d^*r)^{-1}\right)^{1/2},\ \mu^* + t_{1-\alpha/2}(2d^*)\left(c^*(d^*r)^{-1}\right)^{1/2}\right)$$

$t_{1-\alpha/2}(2d^*)$ je $100(1-\alpha/2)\%$ kvantil t- rozdělen o $2d*$, interval je nejkratší

marginální aposteriorní rozděleni σ^{-2} je gama rozdělení s parametry (c^*, d^*), které je jednovrcholové, ale ne symetrické, maximálně věrohodný odhad zal.na aposteriorním rozdělení je $\widehat{\sigma}^{-2} = (d^* - 1)/c^*$
podobně jako při klasickém přístupu se neužívá nejkratší maximalně věrohodný interval pro σ^{-2}

Pro $r \to 0,\ c \to 0,\ d \to -1/2$ dostaneme $c^* = \frac{1}{2}S_n^2$, $d^* = (n-1)/2$ a 95% věrohodný interval pro μ má tvar

$$(\overline{X}_n - t_{1-\alpha/2}(n-1)S_n/\sqrt{n}, \overline{X}_n + t_{1-\alpha/2}(n-1)S_n/\sqrt{n})$$

$S_n^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X}_n)^2$
a 95% věrohodný interval pro σ^2 má tvar

$$(S_n^2(n-1)/\chi^2_{1-\alpha/2}, S_n^2(n-1)/\chi^2_{\alpha/2})$$

$\chi^2_{\alpha/2}(n-1)$ -kvantil χ^2-rozdělení o $n-1$ stupních volnosti
Tedy bayesovske věrohodnostní intervaly spolehlivosti a klasické intervalové odhady se shodují.