Bayesovské metody 5- Testy

12. října 2020

Outline

Outline

Testování hypotéz

Připomeňme:
$\boldsymbol{X} - (X_1, \dots, X_n)^\top$ – hustota $r(\boldsymbol{x}|\boldsymbol{\theta})$ vzhledem k ν_n,
$\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$ – parametr, $\boldsymbol{\theta} \in \Theta$
$\pi(\boldsymbol{\theta}|\boldsymbol{X})$– aposteriorní hustota
$q(\boldsymbol{\theta})$ – apriorní hustota

$\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0 \quad \textit{versus} \quad \mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1$

$\Theta_0 \cap \Theta_1 = \emptyset, \quad \Theta_0 \cup \Theta_1 \subseteqq \Theta$

$d_i = \{\textit{plati} \quad \mathcal{H}_i\}, \quad i = 0, 1, \quad \mathcal{D} = \{d_0, d_1\}$

Typicky:

$$
\begin{aligned}
L(\boldsymbol{\theta}, d_i) &= 0, \quad \boldsymbol{\theta} \in \Theta_i, \quad i = 0, 1 \\
L(\boldsymbol{\theta}, d_i) &\geq 0, \quad \boldsymbol{\theta} \notin \Theta_i, \quad i = 0, 1
\end{aligned}
$$

Obvyklé typy ztrátových funkcí

$$L_*(\boldsymbol{\theta}, d_i) = 0, \quad \boldsymbol{\theta} \in \Theta_i, \quad i = 0, 1$$
$$L_*(\boldsymbol{\theta}, d_i) = a_i, \quad \boldsymbol{\theta} \notin \Theta_i, \quad i = 0, 1$$

$$L_{**}(\boldsymbol{\theta}, d_i) = 0, \quad \boldsymbol{\theta} \in \Theta_i, \quad i = 0, 1$$
$$L_{**}(\boldsymbol{\theta}, d_i) = K_i\, d(\boldsymbol{\theta}, \Theta_i), \quad \boldsymbol{\theta} \notin \Theta_i, \quad i = 0, 1$$

kde $d(\boldsymbol{\theta}, \Theta_i)$ je vzdálenost $\boldsymbol{\theta}$ od Θ_i a $K_i > 0$ konstanty

Riziková funkce;

$$\begin{aligned} R_*(\boldsymbol{\theta}, \delta) &= a_0 P(\delta(\boldsymbol{X}) = d_0 | \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_1 \\ &= a_1 P(\delta(\boldsymbol{X}) = d_1 | \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_0 \end{aligned}$$

Bayesovská riziková funkce

$$\begin{aligned} \rho_*(q, \delta) = a_0 E\Big[\int_{\Theta_1} P(\delta(\boldsymbol{X}) = d_0 | \boldsymbol{\theta}) q(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) \Big] \\ + a_1 E\Big[\int_{\Theta_0} P(\delta(\boldsymbol{X}) = d_1 | \boldsymbol{\theta}), q(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) \Big] \end{aligned}$$

Jednoduché vlastnosti pro lib. $\delta \in \Delta$

$$P(\delta(\boldsymbol{X}) = d_0|\boldsymbol{\theta}) + P(\delta(\boldsymbol{X}) = d_1|\boldsymbol{\theta}) = 1, \quad pro \quad vs. \quad \boldsymbol{\theta} \in \Theta$$

$$\int_{\Theta_1} L_*(\boldsymbol{\theta}, d_0)\pi(\boldsymbol{\theta}|\boldsymbol{X})d\lambda(\boldsymbol{\theta}) = a_0 P(\boldsymbol{\theta} \in \Theta_1|\boldsymbol{X})$$

$$\int_{\Theta_0} L_*(\boldsymbol{\theta}, d_1)\pi(\boldsymbol{\theta}|\boldsymbol{X})d\lambda(\boldsymbol{\theta}) = a_1 P(\boldsymbol{\theta} \in \Theta_0|\boldsymbol{X})$$

Bayesovská rozhodovací funkce δ_*

$$\begin{aligned}\delta_*(\boldsymbol{X}) =& d_0 \quad je-li \quad P(\boldsymbol{\theta} \in \Theta_1|\boldsymbol{X}) < \frac{a_1}{a_0 + a_1},\\ =& d_1 \quad je-li \quad P(\boldsymbol{\theta} \in \Theta_1|\boldsymbol{X}) > \frac{a_1}{a_0 + a_1},\\ =& lib. \quad je-li \quad P(\boldsymbol{\theta} \in \Theta_1|\boldsymbol{X}) = \frac{a_1}{a_0 + a_1}\end{aligned}$$

Příklad, skripta str. 72 Doba čekání na autobus na zastávce má rovnoměrné rozdělení na $(0, \theta), \theta > 0$, t.j. $r(x|\theta) = \theta^{-1} I\{x \in (0, \theta)\}$.

Testujeme $\mathcal{H}_0 : \theta \leq d$ proti $\mathcal{H}_1 : \theta > d$, kde $d > 0$ je dáno.

Apriorní konjugované rozdělení θ – Paretovo rozdělení s parametry (a, x_0), tj. $q(\theta|a, x_0) = (a/x_0)(x_0/\theta)^{a+1} I\{\theta > x_0\}, \quad E\theta = a x_0 (a-1)^{-1}$

Aposteriorní rozdělení θ – Paretovo rozdělení s parametry $(a + n, \max_{i=0,1,\ldots n} x_i)$, tedy

$$P(\theta \leq d|\boldsymbol{X}) = \int_0^d (n+a)\Big(\max_{0\leq i\leq n} X_i\Big)^{a+n} \theta^{-a-1-n} I\{\theta > \max_{0\leq i\leq n} X_i\} d\theta$$

$$= 1 - \Big(\max_{0\leq i\leq n} X_i/d\Big)^{a+n}, \quad d > \max_{0\leq i\leq n} X_i,$$

$$= 0 \quad , \quad d < \max_{0\leq i\leq n} X_i.$$

Při $L_*(\theta, \delta(\boldsymbol{X})$ s $a_1 = a_0 > 0$ je bayes. rozhodovací funkce

$$\delta_*(\boldsymbol{X}) = d_1 \quad pro \quad P(\theta \leq d)|\boldsymbol{X}) < P(\theta > d)|\boldsymbol{X})$$
$$\delta_*(\boldsymbol{X}) = d_0 \quad pro \quad P(\theta \leq d)|\boldsymbol{X}) > P(\theta > d)|\boldsymbol{X})$$
$$= lib., \quad pro \quad P(\theta \leq d)|\boldsymbol{X}) = P(\theta > d)|\boldsymbol{X})$$

Prosim dopocitat pro hodnoty ze skript str. 74, nahore, tedy

$$d = 15,\ x_0 = 5,\ a = 3, \quad \max(x_1, ..., x_5) = 14$$

Jake bude rozhodnuti pri $L_*(.,.)$, $a_0 = a_1$?

Bayesovské metody 5 - Testy

Outline

Testy

Uvažujme nyní $L_{**}(\theta, d)$ pro

$$\mathcal{H}_0 : \theta \leq \theta_0, \quad \textit{versus} \quad \mathcal{H}_1 : \theta > \theta_0$$

$$\begin{aligned} L_{**}(\theta, d_0) &= 0, \quad \theta \leq \theta_0 \\ &= \theta - \theta_0, \quad \theta > \theta_0 \\ L_{**}(\theta, d_1) &= \theta_0 - \theta, \quad \theta \leq \theta_0 \\ &= 0, \quad \theta > \theta_0 \end{aligned}$$

Bayesovská rozhodovací funkce

$$\begin{aligned} \delta_{**}(\boldsymbol{X}) &= d_0 \quad E(\theta|\boldsymbol{X}) \leq \theta_0 \\ &= d_1 \quad E(\theta|\boldsymbol{X}) > \theta_0 \end{aligned}$$

Odvození

$$E\Big(L_{**}(\theta, \delta(\boldsymbol{X}))|\boldsymbol{X}\Big) = \int_{\theta 0}^{\infty} (\theta - \theta_0)\pi(\theta|\boldsymbol{X})d\lambda(\theta), \quad \delta(\boldsymbol{X}) = d_0$$

$$E\Big(L_{**}(\theta, \delta(\boldsymbol{X}))|\boldsymbol{X}\Big) = \int_{0}^{\theta_0} (\theta_0 - \theta)\pi(\theta|\boldsymbol{X})d\lambda(\theta), \quad \delta(\boldsymbol{X}) = d_1$$

Je-li

$$\int_{\theta 0}^{\infty} (\theta - \theta_0)\pi(\theta|\boldsymbol{X})d\lambda(\theta) \leq \int_{0}^{\theta_0} (\theta_0 - \theta)\pi(\theta|\boldsymbol{X})d\lambda(\theta)$$

pak bayesovské rozhodnutí je d_0.

Testy při $\lambda(\Theta_0) = 0$

Nutné modifikovat buď hypotézy nebo apriorní rozdělení $q(\boldsymbol{\theta})$, popř. obojí.

$\lambda(\Theta_0) = 0$ implikuje $P(\boldsymbol{\theta} \in \Theta_0) = 0$

(i) Hypotézy

$$\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0 \quad \textit{versus} \quad \mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1 = \Theta - \Theta_0$$

modifikujeme

$$\mathcal{H}_0^* : \boldsymbol{\theta} \in \quad \textit{versus} \quad \mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1^* = \Theta - \Theta{*}_0$$

$0 < P(\boldsymbol{\theta} \in \Theta_0^*) < 1, \quad \Theta_0 \subset \Theta_0^*$

(ii) modifikujeme původní apriorní rozdělení $q(.)$ následovně:

$$P(\boldsymbol{\theta} \in \Theta_0) = a \in (0,1), \quad P(\boldsymbol{\theta} \in B) = (1-a)\int_B q(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}), \quad B \subseteq \Theta_0,$$

Příklad 1 Test IQ dítěte $\mathcal{H}_0 : \theta = 100$ versus $\mathcal{H}_1 : \theta \neq 100$
X má $N(\theta, 100)$, θ má $N(100, 225)$
$P(\theta = 100) = 0$
(i) změníme hypotézy $\mathcal{H}_0^* : \theta \in (100 - a, 100 + a)$ versus
$\mathcal{H}_1^* : |\theta - 100| > a \neq (100 - a, 100 + a)$, volíme $a > 0$

(ii) změníme apriorní rozdělení zvolíme $q \in (0,1)$ a $P(\theta = 100) = q$ a pro $B \subseteq \mathcal{R} - \{100\}$

$P(\theta \in B) = (1-q)\int_B q(\theta) d\lambda(\theta)$ volíme $q \in (0,1)$
Pak

$$P(\theta = 100|X) = \frac{qr(x|\theta = 100)}{qr(x|\theta = 100) + (1-q)\int_B r(x|\theta)q(\theta)d\lambda(\theta)}$$

$$P(\theta \in B|X) = \frac{(1-q)\int_B r(x|\theta = 100) + \int_B r(x|\theta)q(\theta)d\lambda(\theta)}{qr(x|\theta = 100) + (1-q)\int_B r(x|\theta)q(\theta)d\lambda(\theta)}$$

Lindleyův paradox – při nevhodné volbě q dostaneme podivné závěry = detaily ve skriptech na str.78 a též v různých knihách a článcích

Příklad 1, pokrač. IQ dítěte, úloha diskriminace mezi hypotézami:

$$\mathcal{H}_1 : \theta \leq 90, \quad \mathcal{H}_2 : 90 < \theta < 110, \quad \mathcal{H}_3 : \theta \geq 110$$

Ztrátová funkce:

$$\begin{aligned}
L(\theta, d_1) &= 0 \qquad \theta \leq 90 \\
&= \theta - 90, \quad 90 < \theta < 110 \\
&= 2(\theta - 90) \quad \theta \geq 110 \\
L(\theta, d_2) &= 90 - \theta \qquad \theta \leq 90 \\
&= 0, \qquad 90 < \theta < 110 \\
&= \theta - 90 \qquad \theta \geq 110 \\
L(\theta, d_3) &= 2(110 - \theta) \qquad \theta \leq 90 \\
&= 110 - 90, \quad 90 < \theta < 110 \\
&= 0 \qquad \theta \geq 110
\end{aligned}$$

Přímý výpočet pro lib. rozhodovací fci $\delta(x)$

$$E(L(\theta, \delta(X)|X = x) = \int_{90}^{110} (\theta - 90)\pi(\theta|x)d\theta + 2\int_{110} (\theta - 90)\pi(\theta|x)d\theta, \quad \delta(x) = d_1$$

$$E(L(\theta, \delta(X)|X = x) = \int_{-\infty}^{90} (90 - \theta)\pi(\theta|x)d\theta + \int_{110} (\theta - 90)\pi(\theta|x)d\theta, \quad \delta(x) = d_2$$

$$E(L(\theta, \delta(X)|X = x) = 2\int_{-\infty}^{90} (110 - \theta)\pi(\theta|x)d\theta + 2\int_{90}^{110} (110 - \theta)\pi(\theta|x)d\theta, \quad \delta(x) = d_3$$

Bayesovské metody
5 - Testy

Outline

Testy

pro $x = 115$ je

$$E(L(\theta, \delta(X)|X = 115) = 34,32 \quad \delta(115) = d_1,$$
$$- = 3,55 \quad \delta(115) = d_2,$$
$$= 3,24 \quad \delta(115) = d_3$$

Tedy bayesovské rozhodnutí je d_3.

$$P(\theta \leq 90|x = 115) = 0,005 \quad P(90 \leq \theta \leq 110|x = 115) = 0,475$$
$$P(\theta \geq 110|x = 115) = 0,520$$

Tím končí úvodní kapitoly do tzv. bayesovské statistiky. Snaha byla vysvětlit základní principy bayesovského uvažování a na příkladech ukázat možnosti využití.
V kapitolách o odhadech a testech se výklad soustředil na konstrukce založené a statistických rozhodovacích funkcích. Samozřejmě lze aplikovat např. metodu maximální věrohodnosti ci metodu podílem věrohodností založená na aposteriorních rozděleních.