Bernstein-von-Mises

13. prosince 2020

Outline

Bernstein-von-Mises

Outline

Opakování z předchozího

Outline

Bernstein-von-Mises

Outline

Opakování z předchozího

Obecná Bayesova věta

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ – náhodný vektor s hustotou $q(\theta)$ vzhl. k σ - konečné míře λ na $(\Theta, \mathcal{B}(\Theta))$, Θ - neprázdná borelovská podmnožina $\mathcal{R}^k$,
$\mathcal{B}(\Theta)$ – borelovské podmnožiny Θ

$\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ – náhodný vektor s podmíněnou hustotou $r(\boldsymbol{x}|\boldsymbol{\theta})$ při danném $\boldsymbol{\theta}$ vzhl. k σ -konečné míře ν_n na $(\mathcal{R}^n, \mathcal{B}^n)$, tj.

$$P(\boldsymbol{\theta} \in B, \boldsymbol{X} \in C) = \int_B \left(\int_C r(\boldsymbol{x}|\boldsymbol{\theta}) d\nu_n(\boldsymbol{x}) \right) q(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta})$$

pro B, C lib. měř. množiny

Bayesova věta pro podmíněnou hustotu $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ náhodného vektoru $\boldsymbol{\theta}$ při daném $\boldsymbol{x}$ platí

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{x}) &= \frac{q(\boldsymbol{\theta}) r(\boldsymbol{x}|\boldsymbol{\theta})}{\int_\Theta q(\boldsymbol{\theta}) r(\boldsymbol{x}|\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta})}, \quad \int_\Theta q(\boldsymbol{\theta}) r(\boldsymbol{x}|\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) > 0 \\ &= 0, \qquad \textit{jinak}\end{aligned}$$

$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto q(\boldsymbol{\theta}) r(\boldsymbol{x}|\boldsymbol{\theta})$ – zkrácené značení

$r(\boldsymbol{x}) = \int_\Theta q(\boldsymbol{\theta}) r(\boldsymbol{x}|\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta})$ – marginální hustota $\boldsymbol{X}$

Odhady , testy i konfindenční množiny se konstruují na základě aposterioního rozdělení, např. maximální aposteriorní věrohodnost

Konjugované systémy rozdělení

Konstrukci vyložíme pro případ: $(X_1, \ldots, X_n)$ má hustotu $\prod_{i=1}^n r(x_i|\boldsymbol{\theta})$ vzhledem k σ-konečné míře $\prod_{i=1}^n \nu(x_i)$, $\boldsymbol{\theta} \in \Theta \neq \emptyset$, $\Theta \in \mathcal{B}_k$, (ν je typicky Lebesgueova míra nebo čítací míra.)

Systém Q apriorních hustot $q(\boldsymbol{\theta})$ nazveme systémem konjugovaným s hustotami $\{r(x|\boldsymbol{\theta}), \theta \in \Theta\}$, jestliže pro vš. dost velká n a při libovolných hodnotách $\boldsymbol{X} = \boldsymbol{x}$, které splňují

$$0 < \int_\Theta \prod_{i=1}^n r(\boldsymbol{x}_i|\theta) q(\theta) d\lambda(\theta) < \infty$$

patří aposteriorní hustota do systému Q.

Jsou-li $X_1, \ldots, X_n$ i.i.d. s rozdělení $N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ známá, konjugované rozdělení parametru μ je normální a $N(a, b^2)$, $a \in \mathcal{R}$,, pak aposteriorní rozdělení je

$$N\Big(\frac{1}{n}\sum_i X_i \frac{b^2}{b^2 + \sigma_0^2/n} + a\frac{\sigma_0^2/n}{b^2 + \sigma_0^2/n}, \frac{b^2\sigma_0^2/n}{b^2 + \sigma_0^2/n}\Big).$$

2. Bernstein-von-Mises – limitní věty

poprve J. Doob (1949) pro spec. pripad, zobezneno mnoha autory

Motivační příklad
$X_1, \ldots, X_n$ – i.i.d. (podmíněně) , $N(\theta, \sigma_0^2)$, $\quad \sigma_0^2 > 0$ známé
θ ma $N(a, b^2)$, $a \in \mathcal{R}^1$, $b^2 > 0$ -hyperparametry

aposteriorní rozdělení veličiny $\sqrt{n}(\theta - \overline{X}_n)$ je normální
$N\Big(\frac{\sigma_0^2/n}{\sigma_0^2/n+b^2}\sqrt{n}(a - \overline{X}_n),\ \frac{\sigma_0^2 b^2}{\sigma_0^2/n+b^2}\Big)$

Je-li θ_0 skutečná hodnota parametru θ, pak pro $n \to \infty$

$$\overline{X}_n \to \theta_0,\ a.s. \quad \frac{\sigma_0^2 b^2}{\sigma_0^2/n + b^2} \to \sigma_0^2$$

a také platí pro $n \to \infty$

$$\sup_x \left|P\Big(\sqrt{n}(\theta - \overline{X}_n)/\sigma_0 \le x | \boldsymbol{X}_n\Big) - \Phi(x)\right| \to 0, \quad a.s.,$$

$\Phi(x)$ – distribuční funkce $N(0,1)$

Poznámka Limitní rozdělení nezávisí na hyperparametrech a, b^2.

Příklad 1.-připomenutí Úkolem dvou fyziků je odhadnout jistou fyzikální konstatntu θ.

1. fyzik vyjádří představu (apriorní rozdělení) parametru $\theta - N(900, 400)$
2. fyzik vyjádří představu (apriorní rozdělení) parametr $\theta - N(800, 6400)$

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

V knize J.Anděl: Matematická statistika, vydaná 1985 nakladatelstvím SNTL, kapitola XVI,je možné nalézt tvrzeni i s důkazem ve formě ε, δ.

Zobecnění

$X_1, \dots, X_n$ jsou nezávislé stejně rozdělené s hustotou $r(x|\theta)$ hl.k μ ,$\theta \in \Theta$ je parametr $q(\theta)$ – apriorní hustota

Předpokládejme:
(a) Θ je otevřená podmnožina $\mathcal{R}$,

(b)$\int \frac{\partial}{\partial \theta} r(x|\theta) d\mu(x) = 0$, $\forall \theta \in \Theta$, a jakasi hladkost parcialni derivace $r(x|\theta)$ vzhl. k *theta* integralu

(c) Fisherova informace $0 < J(\theta) < \infty$

(d) je-li skutečná hodnota parametru θ_0, pak $q(\theta_0) > 0$ a $q(\theta)$ je spojita v $(\theta_0 - \epsilon, \theta_0 + \epsilon)$ pro nějaké $\epsilon > 0$.

Je-li $\widehat{\theta}$ striktně konsistentní max. věrohodnostní odhad parametru θ, pak pro $n \to \infty$

$$P(\sqrt{J(\theta_0) n}(\theta - \widehat{\theta}) \leq x | \mathbf{X}) \to^{a.s.} \Phi(x)$$

pro vš. x, kde θ_0 je skutečná hodnota parametru θ

Tato věta se často nazývá "asymptotic normality of the posterior"

Je-li $J(\theta)$ v okolí bodu θ_0 tvrzení platí nahradíme-li $J(\theta_0)$ veličinou $J(\widehat{\theta})$.

Existuje mnoho různých zobecnění

Mírné zpřesnění tvrzení jen pro spec. případ

$X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené s rozdělením $N(\theta, 1)$, θ - parametr
$q(\theta)$ – apriorní hustota —dostatečně hladká

Pak platí:

$$P(\sqrt{n}(\theta - \overline{X}_n) \leq x|\boldsymbol{X}) = \Phi(x) - \frac{q'(\overline{X}_n)}{q'(\overline{X}_n)} \frac{\phi(x)}{\sqrt{n}} - \frac{q''(\overline{X}_n)}{q'(\overline{X}_n)} \frac{x\phi(x)}{n} + O(n^{-3/2})$$

stejnoměrně v x, $q'(.)$, $q''(.)$ jsou derivace , $\phi(x)$ je hustota $N(0,1)$

Za určitých podmínek platí: pro θ_0 -skut.hodnota par., $\widehat{\theta}$– max. verhod.odhad

$$E(\theta|\boldsymbol{X}) = \widehat{\theta} + \frac{1}{n} \frac{1}{\sigma(\overline{X}_n)} \left(\frac{q'(\overline{X}_n)}{q(\overline{X}_n)} + a_n(\overline{X}_n) \right) + o_P(n^{-2})$$

$\sigma(\overline{X}_n)$ a $a(\overline{X}_n)$ jsou funkce nezavisející na $q(.)$.

Bernstein-von-Mises

Outline

Opakování z předchozího

Bernstein-von-Mises

Outline

Opakování z předchozího

Ještě pár poznámek – Za určitých podmínek (dost komplikovaných) platí:

θ_0 -skut. hodnota parametru
$\widehat{\theta}$– max. věrohodný odhad (klasický)
$\widehat{\theta}^*$ – medián par. θ aposteriorního rozdělení $\pi(\theta|\boldsymbol{X})$

$$(i)\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \to^{P_{\theta_0}} 0, \quad \forall\theta,$$
$$\sqrt{n}(\widehat{\theta}^* - \theta) \to^{\mathcal{D}} N(0, J^{-1}(\theta)),$$

$$(ii)E\left[\sqrt{n}(|\widehat{\theta} - \theta| - |\widehat{\theta}^* - \widehat{\theta}|)|\boldsymbol{X}\right] = o(1)$$
$$E\left[\sqrt{n}(|\theta - \widehat{\theta}| - |\theta|)|\boldsymbol{X}\right] = \min_d E\left[\sqrt{n}(|\theta - \widehat{\theta}| - |d - \theta|)|\boldsymbol{X}\right] + o_P(1)$$

(i) pro nebayesovský přístup (ii) pro bayesovský přístup

Výše uvedené výsledky byly zobecněny na vetsinu "rozumných"parametrických konečně-dimenzionálních modelů. Pro případ dimenze pozorovani rostoucích zároveň s počtem pozorování a u neparametrických modelů je situace jiná. Tím se nebudeme zabývat.

Tato část přednášky byla připravena na základě kapitoly 20 v knize
A. Das Gupta: Asymptotic Theory of Statistics and Probability, Springer 2008.

2. Steinův efekt; James-Steinův odhad

$\boldsymbol{X}$ – $N_p(\boldsymbol{\theta}, \boldsymbol{I}_p)$, $\boldsymbol{I}_p$ – jednotková matice $p \times p$

úloha: odhad parametru $\boldsymbol{\theta}$

$\widehat{\boldsymbol{\theta}} = \boldsymbol{X}$ – LSE, NNLO, max. věrohodný odhad

$L(\boldsymbol{\theta}, \boldsymbol{\delta}(\boldsymbol{X})) = ||\boldsymbol{\theta} - \boldsymbol{\delta}(\boldsymbol{X})||^2$ – kvadratická ztrátová funkce

$$EL(\boldsymbol{\theta}, \widehat{\theta}(\boldsymbol{X})) = p, \qquad E\boldsymbol{X} = \boldsymbol{\theta}$$

Pro $p > 2$ existuje odhad $\boldsymbol{\delta}^{JS}(\boldsymbol{X})$ s vlastností

$$E\Big(L(\boldsymbol{\theta}, \boldsymbol{\delta}^{JS}(\boldsymbol{X}))|\boldsymbol{\theta}\Big) < E\Big(L(\boldsymbol{\theta}, \widehat{(}\boldsymbol{X}))|\boldsymbol{\theta}\Big)$$

James a Stein (1961) navrhli:

$$\boldsymbol{\delta}^{JS}(\boldsymbol{X}) = \left(1 - \frac{p-2}{||X||^2}\right)\boldsymbol{X},$$

$$\boldsymbol{\delta}^{JS+}(\boldsymbol{X}) = \left(1 - \frac{p-2}{||X||^2}\right)_+\boldsymbol{X}$$

$x_+ = \max(x, 0)$

Bernstein-von-Mises

Outline

Opakování z předchozího

James- Stein odhad jako empirický bayesovský odhad

$X_i - -N(\theta_i, 1)$, $i = 1, \ldots, p$ nezávislé (podmíněně)
$\theta_i - -N(0, \tau^2)$, $\quad \tau^2 > 0$ nezávislé
Rozdělení $\boldsymbol{X}$ při daném τ^2 – $N(_p(\mathbf{0}, (1+\tau^2)\boldsymbol{I})$

τ^2 odhadneme met. max. věrohodnosti, t.j.

$$\max_{\tau^2 > 0} \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi(1+\tau^2)}} \exp\{-\frac{\sum_i X_i^2}{2(1+\tau^2)}\}$$

dostaneme

$$\begin{aligned} \widehat{\tau}^2 =& \frac{1}{p}\sum_{i=1}^{p} X_i^2 - 1, \quad \sum_{i=1}^{p} X_i^2 > p, \\ =& 0, \quad jinak \end{aligned}$$

aposteriorní rozdělení $\pi(\boldsymbol{\theta}|\boldsymbol{X})$ je $N_p(\boldsymbol{X}\frac{\tau^2}{1+\tau^2}, \frac{\tau^2+1}{\tau^2})$
Tedy

$$\delta_B(\boldsymbol{X}) = \boldsymbol{X}\frac{\tau^2}{1+\tau^2}$$

a po nahrazeni $\widehat{\tau}^2$ za τ^2 dostaneme

$$\widehat{\delta}_B(\boldsymbol{X}) = \boldsymbol{X}\left(1 - \frac{p}{\sum_i X_i^2}\right)_+$$