Bayesovské metody,2

28. září 2020

Outline

Outline

Outline

Bayesovské metody, 2

Outline

Opakování z předchozího

Apriorní rozdělení

Obecná Bayesova věta

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$ – 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

$\boldsymbol{X} = (X_1, \dots, X_n)^\top$ – 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

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

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

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

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

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

Outline

Základní problémy

• Volba apriorního rozdělení

• konstrukce odhadů, testů, konfidenčních množin (založeny na aposteriorních rozděleních)

Volby apriorního rozdělení

Velice důležité a často obtížné

• známe, vyplyne z kontextu, popř. známe z minulosti z obdobných situací-viz příklad s IQ dítěte

• konjugované rozdělení – volíme, tak aby apriorní i aposteriorní rozdělení bylo téhož typu – nejpoužívanější

• princip neučitosti – stejná váha všem možným hodnotám parametrů, Jeffreysova věta

• empirické bayesovské metody – využijí se pozorování získaných za obdobné situace v minulosti

• často se volí tak, aby odhady nebayesovké i bayesovské byly stejné

• Existují i další metody, některé později

———————-

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

Bayesovské metody, 2

Outline

Opakování z předchozího

Apriorní rozdělení

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.

Systémů typicky existuje mnoho, ale je vhodné nalézt menší systémy.

Jedna možná konstrukce založená na postačujících statistikách. Tu si uvedeme.

Předpokládejme, že existuje n_0 takové, že pro vš. $n_0 \le n$ existují nezáporné funkce g_n a h_n takové, že:

$$\prod_{i=1}^{n} r(x_i|\boldsymbol{\theta}) = g_n(\boldsymbol{T}_n(\boldsymbol{x});\, \boldsymbol{\theta}) h_n(\boldsymbol{x})$$

že $\boldsymbol{T}_n$ je r-rozměrná postačující statistika, r nezávisí na n
Ozn. $S_n = \{\boldsymbol{t};\ \boldsymbol{t} = \boldsymbol{T}_n(\boldsymbol{X}),\, 0 < \int_\Theta g_n(\boldsymbol{t};\, \boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) < \infty\}$

Definujme

$$f_{n,\boldsymbol{t}}(\boldsymbol{\theta}) = g_n(\boldsymbol{t};\, \boldsymbol{\theta})\Big(\int_\Theta g_n(\boldsymbol{t};\, \boldsymbol{\theta}) d\lambda(\boldsymbol{\theta})\Big)^{-1}$$

Systém hustot $\{f_{n,\boldsymbol{t}}(\boldsymbol{\theta}),\, \boldsymbol{t} \in S_n, \, , n \ge n_0\}$ je konjugovaný se systémem $\{r(x|\boldsymbol{\theta}),\, \boldsymbol{\theta} \in \Theta\}$.

Náznak odvození: Stači ukázat, že pro aposteriorní rozdělení $\pi(\boldsymbol{\theta}|\boldsymbol{X})$ prislušné apriornímu rozděleni $f_{m,\boldsymbol{t}}(\boldsymbol{\theta})$, $\boldsymbol{t} \in S_m$, $m \ge n_0$, platí

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto f_{m,\boldsymbol{t}}(\boldsymbol{\theta}) \prod_{i=1}^{n} r(x_i|\boldsymbol{\theta}) \propto g_m(\boldsymbol{t};\, \boldsymbol{\theta}) g_n(\boldsymbol{T}_n(\boldsymbol{x});\, \boldsymbol{\theta})$$

Stačí si uvědomit, že existují náhodné veličiny $Y_1, \dots, Y_m$, že $(X_1, \dots, X_n, Y_1, \dots, Y_m)$, podmíněnou hustotu

$$\prod_{i=1}^{n} r(x_i|\boldsymbol{\theta}) \prod_{j=1}^{m} r(y_j|\boldsymbol{\theta}) \propto g_m(\boldsymbol{T}_m(\boldsymbol{y});\, \boldsymbol{\theta}) g_n(\boldsymbol{T}_n(\boldsymbol{x});\, \boldsymbol{\theta})$$

Příklady některých systémů konjugovaných rozdělení

Předpokládáme – $X_1, \ldots, X_n$ – iid , X_i

• má-li X_i, binomické rozdělení s parametry m, p, $p \in (0,1)$ je parametr zájmu, pak platí

$$r(x|p) = P(X_i = x|p) = \binom{m}{x} p^x (1-p)^{m-x}, \quad x = 0, \ldots, m$$

konjugované rozdělení parametru p je beta rozdělení

$$q(\theta) \propto p^{\alpha-1}(1-p)^{\beta-1}, \quad p \in (0,1)), \quad \alpha > 0, \beta > 0$$

Je-li apriorní rozdělení beta(α, β) pak aposteriorní rozdělení je beta$(\alpha + \sum_i X_i, \beta + mn - \sum_i X_i)$

• má-li X_i, Poissonovo rozdělení s parametrem $\lambda > 0$, platí

$$r(x|p) = P(X_i = x|p) = \exp\{-\lambda\} \frac{\lambda^x}{x!}, \quad x = 0, \ldots,$$

konjugované rozdělení parametru λ je rozdělení

$$q(\lambda) \propto \lambda^{\alpha-1} \exp\{-\lambda\beta\}, \quad \lambda > 0, \quad \alpha > 0, \beta > 0$$

Je-li apriorní rozdělení gamma (α, β) pak aposteriorní rozdělení je gamma $(\alpha + \sum_i X_i, \beta + n)$

• má-li X_i rozdělení $N(\mu, \sigma_0^2)$, variance σ_0^2 známá, μ je parametr zájmu, hustota $(X_1, \dots, X_n)$ má tvar

$$r(\boldsymbol{x}|\mu) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\left\{-\frac{\sum_i (x_i-\mu)^2}{\sigma_0^2}\right\}$$

konjugované rozdělení parametru μ je $N(a, b^2)$, $a \in \mathcal{R}$, $b^2 > 0$

Je-li apriorní rozdělení $N(a, b^2)$, pak aposteriorní rozdělení je

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

• má-li X_i rozdělení $N(\mu_0, \sigma^2)$, μ_0 známá, konjugované rozdělení parametru σ^{-2} je gama rozdělení $t, m/2$, $t > 0, m = 1, 2....$

Je-li apriorní rozdělení gama $t, m/2$, $t > 0, m = 1, 2, \dots$ pak aposteriorní rozdělení je gama $t + \sum_i (X_i - \mu_0)^2/2, (m+n)/2$

- má-li X_i rozdělení $N(\mu, \sigma^2)$, konjugované rozdělení pro μ, σ^{-2} je normální-gama rozdělení: μ podmíněno σ^{-2} je $N(a, \sigma^2/r)$, σ^{-2} má gama rozdělení (c, d), tedy parametry tohoto rozdění jsou (a, r, c, d), $a \in \mathcal{R}_1$, $r > 0$, $2d = 1, 2, \ldots, d > 0$

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

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

Otázka je apriorní a aposteriorní rozdělení μ – lze spočítat jako marginální rozdělení

Ve skritech jsou další příklady rozdělen i– exponenciální, rovnoměrné a různá mnohorozměrná.

Výše uvedená metoda dává jen systém apriorních rozdělení, která závisí na parametru - **hyperparametry**. Toto je problém a řeší se různě.

Princip neurčitosti

• apriorní rozdělení rovnoměrné na množině možných hodnot parametru, jen u rozptylu se obvykle volí jinak,

$$q_0(\theta) = 1, \quad \theta \in \Theta$$

Někdy $\int_\Theta q_0(\theta) d\lambda(\theta) = \infty$ pak mluvíme o nevlastní hustotě – lze použít opatrně

Má řadu nevýhod, nejpodstatnější souvisí se změnou parametrizace.

Např. jsou-li $X_1, \ldots, X_n$ i.i.d. s Poissonovým, rozdělením s parametrem $\theta > 0$ a volíme aprirorní rozdělení $q_0(\theta) = 1$ pro $\theta > 0$, jinak 0, pak aposteriorní je gama s parametry $(n, \sum_i X_i + 1)$.
Při parametrizaci $\lambda = \sqrt{\theta}$, toto apriorní rozdělení odpovídá $q(\lambda) = \lambda$, $\lambda > 0$. Tedy není již rovnoměrné.

Jeffreysova hustota

Jeffreys navrhl apriorní hustotu, která která "nezávisí na parametrizaci".

Regulární systém hustot tvaru $\{r(\boldsymbol{x}|\boldsymbol{\theta});\ \boldsymbol{\theta} \in \Theta\}$ splňuje:

- Θ neprázdná otevřená množina v $\mathcal{R}_k$.
- $M = \{\boldsymbol{x};\ r(\boldsymbol{x}|\boldsymbol{\theta}) > 0\}$ nezávisí na $\boldsymbol{\theta}$).
- Pro skoro vš. $\boldsymbol{x} \in M$ (vzhledem k σ- konečné míře ν_n) existuje konečná parciální derivace $r_i'(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial r(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta_i}}$, $i = 1, \ldots, k$.
- Pro každé i a pro vš. $\boldsymbol{\theta} \in \Theta$ platí $\int_M r_i'(\boldsymbol{x}|\boldsymbol{\theta}) d\nu_n(\boldsymbol{x}) = 0$.
- Pro každou dvojici (i, j) existuje konečný integrál

$$J_{i,j}(\boldsymbol{\theta}) = \int_M \frac{r_i'(\boldsymbol{x}|\boldsymbol{\theta}) r_j'(\boldsymbol{x}|\boldsymbol{\theta})}{r^2(\boldsymbol{x}|\boldsymbol{\theta})} r(\boldsymbol{x}|\boldsymbol{\theta}) d\nu_n(\boldsymbol{x})$$

- Matice $\boldsymbol{J}(\boldsymbol{\theta}) = \Big(J_{i,j}(\boldsymbol{\theta})\Big)_{i,j=1,\ldots,k}$ je pozitivně definitní pro každé $\boldsymbol{\theta}$. Fisherova informační matice.

Jeffreysova věta $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ s podmíněnou hustotou $r(\boldsymbol{x}|\boldsymbol{\theta})$ splňujíci podmínky regularity. Předpokládejme

$$c = \int_\Theta r(\boldsymbol{x}|\boldsymbol{\theta})\Big(\det \boldsymbol{J}(\boldsymbol{\theta})\Big)^{1/2} d\boldsymbol{\theta} \in (0, \infty)$$

H – regulární prosté zobrazení Θ na $\Theta^* \in \mathcal{B}_k$. Ozn. $\boldsymbol{\eta} = H(\boldsymbol{\theta})$, $r^*(\boldsymbol{x}|\boldsymbol{\eta}) = r(\boldsymbol{x}|H^{-1}\boldsymbol{\eta}))$. Pak $\{r^*(\boldsymbol{x}|\boldsymbol{\eta}), \boldsymbol{\eta} \in \Theta^*\}$ je regulární systém a platí pro lib. $B \in \mathcal{B}_k$

$$\int_B c^{-1} r(\boldsymbol{x}|\boldsymbol{\theta})\Big(\det \boldsymbol{J}(\boldsymbol{\theta})\Big)^{1/2} d\boldsymbol{\theta} = \int_{H(B)} c^{-1} r^*(\boldsymbol{x}|\boldsymbol{\eta})\Big(\det \boldsymbol{J}^*(\boldsymbol{\eta})\Big)^{1/2} d\boldsymbol{\eta}$$

kde $J^*(\boldsymbol{\eta})$ je Fisherova inf. matice a $\boldsymbol{D} = \left(\frac{\partial \eta_{\boldsymbol{u}}}{\partial \theta_{\boldsymbol{i}}}\right)_{i,\boldsymbol{u}=1,\dots k}$

Platí $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{D}^\top J^*(\boldsymbol{\eta})\boldsymbol{D}$.

———

Tvrzení věty říká: $P\Big(\boldsymbol{\theta} \in B|\boldsymbol{x}|\Big) = P\Big(\boldsymbol{\eta} \in H(B)|\boldsymbol{x}\Big)$

Volba apriorního rozdělení - Jeffreysova hustota: $q(\boldsymbol{\theta}) = \Big(det \boldsymbol{J}(\boldsymbol{\theta})\Big)^{1/2}$

Tady je výsledek pro některá rozdělení:

pro binomické rozdělení s parametry (m, p), apriorní rozd. je beta $(1/2, 1/2)$
pro Poissonovo rozdělení s θ apriorní rozd. je gama $(0, 1/2)$

Empirické bayesovské metody

Jsou k dipozici obdobná data z minulosti – $Y_1, \ldots, Y_N$, podmíněně nezávislé, Y_i s hustotou $r(y|\boldsymbol{\theta}_i)$, $i = 1, \ldots, N$, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ jsou nezávislé stejně rozdělené s hustotou $q(\boldsymbol{\theta})$

Pak nepodmíněné rozdělení Y_i je $r_{\boldsymbol{q}}(y) = \int_\Theta r(y|\boldsymbol{\theta})q(\boldsymbol{\theta})d\lambda(\boldsymbol{\theta})$, kde $q(\boldsymbol{\theta}) \in \{q(\boldsymbol{\theta}; \boldsymbol{\alpha}), \boldsymbol{\alpha} \in \Upsilon\}$ – $\boldsymbol{\alpha}$ – parametr

Tedy $Y_1, \ldots, Y_n$ jsou nezávislé stejně rozdělené s hustotou závisející na $\boldsymbol{\alpha}$.

$\boldsymbol{\alpha}$ na základě $Y_1, \ldots, Y_n$ odhadneme některou klasickou metodou, ozn. $\widehat{\boldsymbol{\alpha}}$ a pak použijeme jako apriorní rozdělení $q(\boldsymbol{\theta}; \widehat{\boldsymbol{\alpha}})$. Často užijeme momentovou metodou.

Příklad Y_i má (podmíněné) rozděleni $N(\theta, \sigma_0^2)$, σ_0^2 je známé, θ_i $N(\mu_{\boldsymbol{q}}, \sigma_{\boldsymbol{q}}^2)$, $\boldsymbol{\alpha} = (\mu_{\boldsymbol{q}}, \sigma_{\boldsymbol{q}}^2)$. Nepodmíněné rozdělení Y_i je $N(\mu_{\boldsymbol{q}}, \sigma^2)$, $\sigma^2 = \sigma_0^2 + \sigma_{\boldsymbol{q}}^2$
Použijeme odhady

$$\widehat{\mu}_{\boldsymbol{q}} = \frac{1}{N}\sum_{i=1}^N Y_i, \quad \widehat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^N (Y_i - \overline{Y}_N)^2, \quad \overline{Y}_N = \frac{1}{N}\sum_{i=1}^N Y_i.$$

Za apriorní rozdělení vezmeme $N(\widehat{\mu}_{\boldsymbol{q}}, \widehat{\sigma}_{\boldsymbol{q}}^2)$, kde $\widehat{\sigma}_{\boldsymbol{q}}^2 = max(\widehat{\sigma}^2 - \sigma_0^2, 0)$

Bayesovské metody, 2

Outline

Opakování z předchozího

Apriorní rozdělení

. Existuje řada zobecnění:

Konvexni směs konjugovaných rozdělení

Hierarchický bayesovský model

$$q(\boldsymbol{\theta}_0) = \int_{\Theta_1 \times \Theta_2} q_1(\boldsymbol{\theta}_0|\boldsymbol{\theta}_1) q_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) q_3(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \, d\boldsymbol{\theta}_2$$

$q_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)$ – podmíněná hustota $\boldsymbol{\theta}$ při daném $\boldsymbol{\theta}_2$

$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ - hyperparametry 1. resp.2. úrovně

$\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ – parametry apriorního rozdělení