NEPARAMETRICKÉ BAYESOVSKÉ ODHADY V KOZIOLOVĚ-GREENOVĚ MODELU NÁHODNÉHO CENZOROVÁNÍ

Michal Friesl

Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

•Princip •Příklady •V K.-G. modelu

Bayesovský přístup

Data

- pozorování X s hodnotami v $(\mathcal{X}, \mathcal{A})$
- X s rozdělením Q

Parametricky

- např. $Q = Q_{\mu,\sigma^2} = \mathrm{N}(\mu, \sigma^2)$
- μ, σ^2 parametry $\rightarrow$ náhodné veličiny
- apriorní rozdělení na $\mathbf{R}^2$

„Neparametricky“

- Q libovolné rozdělení na $(\mathcal{X}, \mathcal{A})$
- parametry $(Q(A), A \in \mathcal{A}) \rightarrow$ náhodný proces
- apriorní rozdělení na prostoru pravděpodobnostních měr

Dirichletův proces

$Q \sim \mathcal{D}(\alpha)$

- $\alpha = n_0 Q_0$ konečná míra na $(\mathcal{X}, \mathcal{A})$
- $(Q(A_1), \ldots, Q(A_k)) \sim \mathrm{D}(\alpha(A_1), \ldots, \alpha(A_k))$

Interpretace α

- $\mathrm{E}\, Q(A) = Q_0(A)$
- $\mathrm{var}\, Q(A) = Q_0(A)(1 - Q_0(A))/(n_0 + 1)$

Vlastnosti

- pokrývá širokou třídu rozdělení
- Q s pravděpodobností 1 diskrétní
- např. $Q(A) = \sum p_n \delta_{Y_n}(A)$,
  kde $Y_n \sim Q_0$, $p_n = \theta_n \prod_{j<n}(1 - \theta_j)$, $\theta_j \sim \mathrm{B}(1, n_0)$

Uplatnění Dirichletova procesu

Pozorování

- $X_1, \ldots, X_n$ náhodný výběr z Q
- $Q \sim \mathcal{D}(\alpha)$

Aposteriorní rozdělení

- $(Q \mid X_1, \ldots, X_n) \sim \mathcal{D}(\alpha + \sum \delta_{X_i})$

Reálné veličiny

- $\widehat{F}(x) = \mathrm{E}(Q(-\infty, x\rangle \mid X_1, \ldots, X_n) = \dfrac{n_0 F_0(x) + n F_n(x)}{n_0 + n}$
- $\widehat{\mathrm{E}\, Q} = \dfrac{n_0 \,\mathrm{E}\, Q_0 + n\bar{X}}{n_0 + n}$
- při $n_0 \to 0$ je $\widehat{F} \to F_n$, $\widehat{\mathrm{E}\, Q} \to \bar{X}$

Zprava neutrální procesy

Kromě Q

- distribuční funkce $F(t) = Q(-\infty, t\rangle$
- $F(t_1), \dfrac{F(t_2) - F(t_1)}{1 - F(t_1)}, \dots, \dfrac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$ nezávislé

Anebo

- kumulativní intenzita $\Lambda(t) = -\ln(1 - F(t))$
- $S(t) = \mathrm{P}(X > t) = \mathrm{e}^{-\Lambda(t)}$, $\mathrm{P}(X > t \mid X > s) = e^{-\Lambda(s,t\rangle}$
- Λ proces s nezávislými přírůstky

Vlastnosti

- skoro jistě diskrétní rozdělení
- aposteriorní opět zprava neutrální
- v časech pozorování vždy skoky

Gama proces

$\Lambda \sim \mathcal{G}(n_0, \Lambda_0)$

- $n_0 > 0$, Λ_0 libovolná kumulativní intenzita
- nezávislé přírůstky (zprava neutrální)
- $\Lambda(s,t\rangle = \Lambda(t) - \Lambda(s) \sim \mathrm{G}(n_0, n_0\Lambda_0(s,t\rangle))$

Parametry

- $\mathrm{E}\,\Lambda(s,t\rangle = \Lambda_0(s,t\rangle$
- $\mathrm{var}\,\Lambda(s,t\rangle = \Lambda_0(s,t\rangle/n_0$

Aposteriorně

- mezi pozorováními gama
- v časech pozorování skoky
- rozdělení nových skoků nezávisí na čase

Gama proces a odhady

Pozorování

- $X_1, \ldots, X_n$ náhodný výběr z Q
- $Q \sim \mathcal{G}(n_0, \Lambda_0)$

Pro jednoduchost

- Λ_0 spojitá, pozorování různá

$$1 - \widehat{F}(t) = \Big(\frac{n_0 + N_t}{n_0 + N_t + 1}\Big)^{n_0 \Lambda_0(t)} \prod_{i, X_i \leqq t} E_{X_i}$$

kde $N_t = \#\{i, X_i > t\}$ a

$$E_x = \Big(\frac{n_0 + N_x + 1}{n_0 + N_x}\Big)^{n_0 \Lambda_0(x)} \cdot \frac{\ln((n_0 + N_x + 2)/(n_0 + N_x + 1))}{\ln((n_0 + N_x + 1)/(n_0 + N_x))}$$

Cenzorování

Náhodné cenzorování

- některá pozorování předčasně ukončena
- kromě X_i ještě časový cenzor Y_i
- X_i a Y_i nezávislé

Pozorujeme

- $Z_i = \min(X_i, Y_i)$
- $\delta_i = I_{X_i \leqq Y_i}$

Při $Z_i = t$ pro zkoumané X_i

- skutečné pozorování $X_i = t$ $(\delta = 1)$
- částečná informace $X_i > t$ $(\delta = 0)$

Vliv cenzorování na odhady

Dirichletův proces

- aposteriorní není Dirichletův
- $\widehat{S}(t) = \dfrac{n_0 S_0(t) + N_t}{n_0 + n} \prod_{y \leqq t, y = Y_i = Z_i} \dfrac{n_0 S_0(y-) + N_{y-} - u_y}{n_0 S_0(y-) + N_y}$
- při $n_0 \to 0$ Kaplanův-Meierův odhad $\prod_{x \leqq t, x = X_i = Z_i} (1 - \dfrac{u_x}{N_{x-}})$

Zprava neutrální proces

- aposteriorní znovu zprava neutrální
- u $\mathcal{G}(n_0, \Lambda_0)$ v odhadu součin jen přes necenzorovaná
- i cenzorování $X_i \geqq Y_i$

Koziolův-Greenův model

Model náhodného cenzorování

- $X \sim F \leftrightarrow \Lambda$, $Y \sim G \leftrightarrow \Lambda_G$ nezávislé
- pozorujeme $Z = \min(X, Y)$, $\delta = I_{X \leqq Y}$
- navíc $1 - G = (1 - F)^\gamma$ pro nějaké $\gamma > 0$
- tj. $\Lambda_G = \gamma \Lambda$

Ve spojitém případě

- Z a δ nezávislé
- $\mathrm{P}(\delta = 1) = 1/(1 + \gamma)$

Oproti předchozímu

- doba X a cenzor Y mají společný parametr
- i v $Y = t$ nebo $Y \geqq t$ obsažena informace o Λ

Předpoklady

Apriorní rozdělení

- $\Lambda \sim \mathcal{G}(n_0, \Lambda_0)$
- γ nezávislé s Λ, s hustotou $\pi(\gamma)$

Pozorování

- data $= (Z_1, \delta_1), \ldots, (Z_n, \delta_n)$
- značíme $N_t = \#\{i, Z_i > t\}$

Pro jednoduchost zápisu

- Λ_0 spojitá
- pozorování Z_i navzájem různá

Po jednom pozorování

- Např. $Z = t, \delta = 0$, tj. $Y = t, X > t$

			věroh.		
A	$\Lambda(A)$	aprior.	$Y = t$	$X > t$	apost.
$(., s)$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
(s, t)	λ	$\mathrm{G}(n_0, a)$	$\mathrm{e}^{-\gamma\lambda}$	$\mathrm{e}^{-\lambda}$	$\mathrm{G}(n_0 + 1 + \gamma, a)$
$\{t\}$	μ	$\mathrm{G}(n_0, 0)$	$1 - \mathrm{e}^{-\gamma\mu}$	$\mathrm{e}^{-\mu}$	$\mathrm{G}_0(n_0 + 1, n_0 + 1 + \gamma)$
(t, u)	ν	$\mathrm{G}(n_0, b)$	1	1	$\mathrm{G}(n_0, b)$
$(u, .)$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

- $\mathrm{G}(n_0, a)$ s hustotou $\dfrac{n_0^a}{\Gamma(a)} \lambda^{a-1} \mathrm{e}^{-n_0\lambda}$, $\lambda > 0$
- $\mathrm{G}_0(n, m)$, $0 < n < m$, s hustotou $\dfrac{\mathrm{e}^{-n\mu} - \mathrm{e}^{-m\mu}}{-\mu \ln(n/m)}$, $\mu > 0$

Aposteriorní rozdělení

Λ při daném γ a datech

- nezávislé přírůstky
- mezi časy pozorování

$$\Lambda(s,t) \sim \mathrm{G}(n_0 + N_s(1+\gamma), n_0\Lambda_0(s,t))$$

- v čase s jedním pozorováním $Z_i = t$ $(\Lambda_0\{t\} = 0)$

$$\Lambda\{t\} = \begin{cases} \mathrm{G}_0(N_{Z_i}(1+\gamma), N_{Z_i}(1+\gamma)+1), & \delta_i = 1, \\ \mathrm{G}_0(N_{Z_i}(1+\gamma)+1, N_{Z_i}(1+\gamma)+1+\gamma), & \delta_i = 0, \end{cases}$$

A hustota γ ...

Aposteriorní rozdělení (pokrač.)

. . . a hustota γ

- $$\pi(\gamma \mid \text{data}) \propto \Big(\prod_{j=1}^{n} C_j(\gamma)\, D_j(\gamma)\Big)\pi(\gamma), \quad \gamma > 0.$$

  kde

$$C_j(\gamma) = \Big(\frac{n_0}{n_0 + N_{Z_{j-1}}(1+\gamma)}\Big)^{n_0 \Lambda_0(Z_{j-1}, Z_j)}$$

$$D_j(\gamma) = \begin{cases} \ln \dfrac{n_0 + N_{Z_j}(1+\gamma)}{n_0 + N_{Z_j}(1+\gamma) + 1}, & \delta_j = 1 \\ \ln \dfrac{n_0 + N_{Z_j}(1+\gamma) + 1}{n_0 + N_{Z_j}(1+\gamma) + 1 + \gamma} & \delta_j = 0. \end{cases}$$

Odhad

Pro $t \in (Z_{(i-1)}, Z_{(i)})$

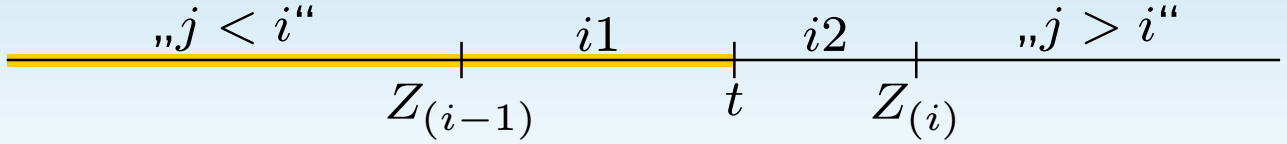
$$\widehat{S}(t) = \mathrm{E}(\mathrm{e}^{-\Lambda(-\infty,t\rangle}\,|\text{data}) = \frac{\int_0^\infty K_1^+(\gamma)K_2(\gamma)\pi(\gamma)\,\mathrm{d}\gamma}{\int_0^\infty K_1(\gamma)K_2(\gamma)\pi(\gamma)\,\mathrm{d}\gamma}$$

kde

$$K_1(\gamma) = \Big(\prod_{„j<i“} C_j(\gamma)D_j(\gamma)\Big)C_{i1}(\gamma)$$

$$K_2(\gamma) = C_{i2}(\gamma)D_i(\gamma)\Big(\prod_{„j>i“} C_j(\gamma)D_j(\gamma)\Big)$$

$$K_1^+(\gamma) = \Big(\prod_{„j<i“} C_j^+(\gamma)D_j^+(\gamma)\Big)C_{i1}^+(\gamma)$$



Uvolnění předpokladů

Skoky v apriorní intenzitě Λ_0

- při pozorováních v bodech skoků
- $\mathrm{G}_0 \to g_1 - g_2$, $D(\gamma) \to$ rozdíly $C(\gamma)$

Shody mezi časy pozorování

- složitější zápis, kombinace cenz. a necenz.
- místo G_0 např. $\mathrm{e}^{-\mu}(1-\mathrm{e}^{-\mu\gamma})(1-e^{-\mu})$ [1 cenz., 1 necenz.]
- místo $D_i(\gamma)$ např. $\ln \frac{(n_0+1)(n_0+2+\gamma)}{(n_0+1+\gamma)(n_0+2)}$

Jiné apriorní rozdělení

- $1-\exp(-\Lambda) \sim \mathcal{D}(\alpha)$
- v normovacích konstantách $(\Lambda \mid \gamma)$ rozdíly digama funkcí

NEPARAMETRICKÉ BAYESOVSKÉ ODHADY V KOZIOLOVĚ-GREENOVĚ MODELU NÁHODNÉHO CENZOROVÁNÍ

Michal Friesl

Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

•http://home.zcu.cz/~friesl/Archiv/Rob04.html