Bayesiánská analýza

IV. Normální lineární regresní model s jinou apriorní hustotou

Obsah tématu

Úvod

- Přirozeně konjugovaná apriorní hustota pro NLRM – omezující.
- Nezávislost apriorních omezení – neexistují analytické výsledky $\rightarrow$ potřeba posteriorních simulátorů.
- Gibbsův vzorkovač a Importance sampling.
- Věrohodnostní funkce se nemění! (stále NLRM)

Obsah tématu

Obsah tématu

Apriorní hustota

- Rozdělení vektoru β není pdmíněno přesností chyby, h.

$$p(\beta, h) = p(\beta)p(h)$$

- $p(\beta)$ z normálního rozdělení a $p(h)$ z rozdělení gama:

$$p(\beta) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\underline{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \underline{\beta})' \underline{V}^{-1}(\beta - \underline{\beta})\right],$$

$$p(h) = c_G^{-1} h^{\frac{\underline{\nu}-2}{2}} \exp\left(-\frac{h\underline{\nu}}{2\underline{s}^{-2}}\right),$$

kde c_G je integrační konstanta pro funkci gama rozdělení.

- $\underline{V}$ – apriorní kovarianční matice vektoru parametrů β.

Obsah tématu

Sdružená hustota

- Sdružená hustota nemá tvar známé hustoty:

$$p(\beta, h|y) \propto \left\{ \exp\left[-\frac{1}{2}\left\{ h(y - X\beta)'(y - X\beta) + (\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta}) \right\}\right]\right\} \times h^{\frac{N+\underline{\nu}-2}{2}} \exp\left[-\frac{h\underline{\nu}}{2\underline{s}^{-2}} \right].$$

- Podmíněné hustoty mají známou formu.
- $p(\beta|y, h) = \frac{p(\beta,h|y)}{p(h|y)}$.
- Z $p(\beta, h|y)$ můžeme vyvodit informaci o $p(\beta|y, h)$ pokud fixujeme h.

Sdružená hustota – úpravy

$$
\begin{aligned}
&h(y - X\beta)'(y - X\beta) + (\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta}) \\
&= (\beta - \overline{\beta})'\overline{V}^{-1}(\beta - \overline{\beta}) + Q,
\end{aligned}
$$

kde

$$
\begin{aligned}
\overline{V} &= (\underline{V}^{-1} + hX'X)^{-1}, \\
\overline{\beta} &= \overline{V}(\underline{V}^{-1}\underline{\beta} + hX'y), \\
Q &= hy'y + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - \overline{\beta}'\overline{V}^{-1}\overline{\beta}.
\end{aligned}
$$

Podmíněná hustota pro β

- Vypuštění členů nezahrnující vektor β (včetně výrazu Q):

$$p(\beta|y,h) \propto \exp\left[-\frac{1}{2}(\beta-\overline{\beta})'\overline{V}^{-1}(\beta-\overline{\beta})\right].$$

- Jádrová hustota vícerozměrného normálního rozdělení.

$$\beta|y,h \sim N(\overline{\beta},\overline{V})$$

Podmíněná hustota pro h

- Z posteriorní hustoty jako funkce h.

$$p(h|y,\beta) \propto h^{\frac{N+\underline{\nu}-2}{2}} \exp\left[-\frac{h}{2}\left\{(y-X\beta)'(y-X\beta)\right\}\right].$$

- Jádrová hustota gama rozdělení.

$$h|y,\beta \sim G(\overline{s}^{-2},\overline{\nu}),$$

kde

$$\overline{\nu} = N + \underline{\nu},$$
$$\overline{s}^2 = \frac{(y-X\beta)'(y-X\beta)+\underline{\nu}\underline{s}^2}{\overline{\nu}}.$$

Obsah tématu

Motivace

- $p(\beta, h|y) \neq p(\beta|y, h)p(h|y, \beta) \Rightarrow$ podmíněné hustoty neříkají vše o $p(\beta, h|y)$.
- Posteriorní simulátor – Gibbsův vzorkovač.
- Náhodné výběry z podmíněných hustot $\rightarrow$ množina náhodných vzorků z odpovídající sdružené hustoty.

Východiska

- p-rozměrný vektor parametrů θ, věrohodnostní funkce $p(y|\theta)$, apriorní hustota $p(\theta)$ a posteriorní hustota $p(\theta|y)$.
- Rozdělení vektoru parametrů θ do několika (B) bloků, tedy $\theta = (\theta'_{(1)}, \theta'_{(2)}, \ldots, \theta'_{(B)})'$, kde $\theta_{(j)}$ je skalár nebo vektor pro $j = 1, 2, \ldots, B$.
- V LRM: obvyklý počet bloků $B = 2$ (první blok $\theta_{(1)} = \beta$ a druhý blok $\theta_{(2)} = h$).
- *Plně podmíněná množina posteriorních rozdělení* (jsme z nich schopni generovat výběry):

$$p(\theta_{(1)}|y, \theta_{(2)}, \ldots, \theta_{(B)}), p(\theta_{(2)}|y, \theta_{(1)}, \theta_{(3)}, \ldots, \theta_{(B)}), \ldots$$
$$\ldots, p(\theta_{(B-1)}|y, \theta_{(1)}, \ldots, \theta_{B-2}, \theta_{(B)}), p(\theta_{(B)}|y, \theta_{(1)}, \ldots, \theta_{(B-1)}).$$

Postup (příklad)

- Sekvence vzorků $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)} \rightarrow$ Monte Carlo integrace pro získání $E[g(\theta)|y]$.
- Pro $B = 2$ mějme počáteční náhodný výběr z $p(\theta_{(2)}|y)$: $\theta_{(2)}^{(0)}$.
- $p(\theta|y) = p(\theta_{(1)}|y, \theta_{(2)})p(\theta_{(2)}|y) \Rightarrow$ výběr z $p(\theta_{(1)}|y, \theta_{(2)}^{(0)})$ je řádným výběrem $\theta_{(1)}$ z $p(\theta|y) \rightarrow \theta_{(1)}^{(1)}$.
- $p(\theta|y) = p(\theta_{(2)}|y, \theta_{(1)})p(\theta_{(1)}|y) \Rightarrow$ náhodný výběr z $p(\theta_{(2)}|y, \theta_{(1)}^{(1)})$ je platným výběrem $\theta_{(2)}$ z $p(\theta|y)$.
- $\theta^{(1)} = (\theta_{(1)}^{(1)\prime}, \theta_{(2)}^{(1)\prime})'$ je řádným výběrem z $p(\theta|y)$.
- Postup do nekonečna.
- Pokud najdeme $\theta_{(2)}^{(0)} \Rightarrow$ sekvenční výběr z $\theta_{(1)}$ podmíněný předchozím výběrem $\theta_{(2)}$ a výběr $\theta_{(2)}$ podmíněný takto získaným $\theta_{(1)}$ dává řadu náhodných výběrů (vzorků) z posteriorního rozdělení = Gibbsův vzorkovač.

Počáteční podmínky

- Problémem s $\theta_{(2)}^{(0)}$.
- Pokud bychom uměli získávat náhodné výběry z $p(\theta_{(2)}|y) \rightarrow$ přímé využítí spolu s $p(\theta_{(1)}|\theta_{(2)}, y)$ v rámci MC integrace.
- Splnění tzv. slabých podmínek = počáteční výběr $\theta_{(2)}^{(0)}$ žádnou roli $\rightarrow$ Gibbsův vzorkovač konverguje k sekvenci výběrů z $p(\theta|y)$.
- Obvyklá volba $\theta_{(2)}^{(0)} \rightarrow S$ replikací $\rightarrow$ prvních S_0 vzorků odstraníme (tzv. *burn-in replications*) $\Rightarrow S_1$ vzorků použijeme dále $(S_0 + S_1 = S)$.
- Příklad nesplnění: posteriorní hustota definována ve dvou různých oblastech, které nejsou vzájemně propojeny $\rightarrow$ Gibbsův vzorkovač poskytne výběry jen z jedné z těchto oblastí (do druhé oblasti se nebude schopen dostat).
- Není případ normálního-gama rozdělení.

Shrnutí algoritmu

1. Zvolíme počáteční hodnotu vektoru parametrů, $\theta^{(0)}$.
   Pro $s = 1, \dots, S$:
2. Provedeme náhodný výběr $\theta_{(1)}^{(s)}$ z $p(\theta_{(1)}|y, \theta_{(2)}^{(s-1)}, \theta_{(3)}^{(s-1)}, \dots, \theta_{(B)}^{(s-1)})$.
3. Provedeme náhodný výběr, $\theta_{(2)}^{(s)}$ z $p(\theta_{(2)}|y, \theta_{(1)}^{(s)}, \theta_{(3)}^{(s-1)}, \dots, \theta_{(B)}^{(s-1)})$.
   ...
4. Provedeme náhodný výběr $\theta_{(B)}^{(s)}$ z $p(\theta_{(B)}|y, \theta_{(1)}^{(s)}, \theta_{(2)}^{(s)}, \dots, \theta_{(B-1)}^{(s)})$.

Monte Carlo integrace

- S výběrů, $\theta^{(s)}$ pro $s = 1, \ldots, S$.
- Prvních S_0 výběrů vyhodíme (eliminace efektu počáteční volby $\theta^{(0)}$).
- S_1 výběrů zprůměrujeme $\rightarrow$ požadované posteriorní charakteristiky.
- Funkce parametrů $g(\cdot)$ a

$$\widehat{g}_{S_1} = \frac{1}{S_1} \sum_{s=S_0+1}^{S} g(\theta^{(s)}).$$

- $\widehat{g}_{S_1}$ konverguje ke střední hodnotě $E[g(\theta)|y]$ pro S_1 jdoucí k nekonečnu.

Problémy

- Jakákoliv volba bloků (ve většině případů se nabízí sama).
- Centrální limitní věta – přibližné určení chyby aproximace.
- Dva problémy:
    1. Potřeba ověřit, že volba $\theta^{(0)}$ nemá vliv na získané výsledky.
    2. Sekvence výběrů není i.i.d.; vektory $\theta^{(s)}$ a $\theta^{(s-1)}$ nejsou vzájemně nezávislé, neboť $\theta^{(s)}_{(j)}$ závisí na $\theta^{(s-1)}_{(l)}$ pro $j = 1, \ldots, B-1$ a $l > j$.
- Prakticky nutné pro dosažení požadované úrovně přesnosti vygenerovat mnohem více výběrů.

Obsah tématu

Numerická standardní chyba

- Využití centrální limitní věty – problém závislosti výběrů.
- Centrální limitní věta – přibližné určení chyby aproximace.

$$\sqrt{S_1}\{\widehat{g}_{S_1} - E[g(\theta)|y]\} \rightarrow N(0, \sigma_g^2)$$

pro S_1 jdoucí k nekonečnu.
- σ_g^2 má mnohem složitější formu; v literatuře zatím nebyl publikován dostatečně ověřený způsob jejího odhadu.
- Intuice: σ_g^2 by měla zohledňovat skutečnost, že $\theta^{(s)}$ pro $s = 1, \ldots, S$ je vzájemně korelovaná řada.

NSE – Geweke (1992)

- Z teorie časových řad.

$$\widehat{\sigma}_g^2 = \frac{S(0)}{S_1}$$

- Oprávnění tohoto odhadu spíše neformální $\times$ v praxi se osvědčuje.
- $S(0)$ je spektrální hustota řady $\theta^{(s)}$ pro $s = S_0 + 1, \ldots, S$ vyhodnocená v 0.
- NSE: $\frac{\widehat{\sigma}_g}{\sqrt{S_1}}$.

Geweke (1992) – pokračování

- Z ponechaných vzorků: první sadu S_A výběrů, prostřední S_B a poslední S_C.
- V praxi osvědčeno: $S_A = 0.1S_1$, $S_B = 0.5S_1$ a $S_C = 0.4S_1$.
- Pro potřeby diagnostiky nepoužíváme prostřední množinu vzorků S_B.
- Nechť $\widehat{g}_{S_A}$ a $\widehat{g}_{S_C}$ jsou odhady $E[g(\theta)|y]$ za použití prvních S_A vzorků (po vyhození S_0 výběrů) a S_C vzorků.
- Definujme $\frac{\widehat{\sigma}_A}{\sqrt{S_A}}$ a $\frac{\widehat{\sigma}_C}{\sqrt{S_C}}$ jako NSE odhadů.
- Centrální limitní věta:

$$CD \rightarrow N(0,1)$$

- Konvergenční diagnostika CD:

$$CD = \frac{\widehat{g}_{S_A} - \widehat{g}_{S_C}}{\frac{\widehat{\sigma}_A}{\sqrt{S_A}} + \frac{\widehat{\sigma}_C}{\sqrt{S_C}}},$$

- Porovnání s kritickými hodnotami standardizovaného normálního rozdělení.

Problémy

- Gibbsův vzorkovač – prochází posteriorní rozdělení.
- Bimodální posteriorní hustota – diagnostika selhává.
- Počáteční výběr $\theta^{(0)}$ je extrémě vzdálen $+$ míra korelace ve vzorcích vysoká (většinou CD diagnostika neselže).

Gelman a Rubin (1992) – motivace

- Problém neodeznění efektu počátečního výběru $\theta^{(0)}$.
- Praxe: použít různá $\theta^{(0)}$.
- $\theta^{(0,i)}$ pro $i = 1, \dots, m$ označuje m počátečních hodnot z různých oblastí parametrického prostoru (*overdispersed starting values*).
- $\theta^{(s,i)}$ pro $s = 1, \dots, S$ označuje S výběrů navzorkovaných Gibbsovým vzorkovačem z i-té počáteční hodnoty a $\widehat{g}_{S_i}^{(i)}$ označuje odpovídající odhad $E[g(\theta)|y]$.
- Pokud efekt počátečních podmínek odezněl, měly by být sekvence podobné $\Rightarrow$ porovnání rozptylů v rámci a mezi sekvencemi.

Gelman a Rubin (1992) – pokračování

- Vnitřní rozptyl sekvence:

$$s_i^2 = \frac{1}{S_1 - 1} \sum_{s=S_0+1}^{S} \left[g(\theta^{(s,i)}) - \widehat{g}_{S_i}^{(i)} \right]^2 .$$

- Průměrný rozptyl vnitřních rozptylů sekvencí:

$$W = \frac{1}{m} \sum_{i=1}^{m} s_i^2 .$$

- Mezisekvenční rozptyl:

$$B = \frac{S_1}{m-1} \sum_{i=1}^{m} (\widehat{g}_{S_i}^{(i)} - \widehat{g})^2 ,$$

kde

$$\widehat{g} = \frac{1}{m} \sum_{i=1}^{m} \widehat{g}_{S_i}^{(i)} .$$

Gelman a Rubin (1992) – statistika

- W je odhadem $var[g(\theta)|y]$.
- Další odhad rozptylu, $var[g(\theta)|y]$:

$$\widehat{var[g(\theta)}|y] = \frac{S_1 - 1}{S_1} W + \frac{1}{S_1} B.$$

- MCMC kovergenční diagnostika má podobu:

$$\widehat{R} = \frac{\widehat{var[g(\theta)}|y]}{W}.$$

- Hodnoty R by měly být větší než jedna; hodnoty blízké jedné indikují úspěšnou konvergenci.
- $\sqrt{\widehat{R}}$ označována jako *estimated potential scale reduction* (mez toho, jak vzdálené mohou být odhady směrodatné odchylky $g(\theta)$ díky nedostatečné konvergenci).

Obsah tématu

Savage-Dickey – motivace

- Není obecně uplatnitelná.
- Bayesův faktor pro vnořené modely (v případě určitých apriorních hustot).
- Neomezenáí verze modelu M_2 s vektorem parametrů $\theta = (\omega', \psi')'$.
- Věrohodnostní funkce a apriorní hustota: $p(y|\omega, \psi, M_2)$ a $p(\omega, \psi, M_2)$.
- Omezená verze modelu M_1 má $\omega = \omega_0$, kde ω_0 je vektor konstant (parametry vektoru ψ neomezené pro oba modely).
- Věrohodnostní funkce a apriorní hustota: $p(y|\psi, M_1)$ a $p(\psi|M_1)$ ($\omega = \omega_0$ v $M_1 \Rightarrow$ není třeba specifikovat apriorní hustotu pro tento vektor parametrů).

Savage-Dickey – definice

- Předpokládáme

$$p(\psi|\omega = \omega_0, M_2) = p(\psi, M_1).$$

- Potom BF_{12}, Bayesův faktor porovnávající M_1 a M_2 má podobu:

$$BF_{12} = \frac{p(\omega = \omega_0|y, M_2)}{p(\omega = \omega_0|M_2)}.$$

- $p(\omega = \omega_0|y, M_2)$ a $p(\omega = \omega_0|M_2)$ jsou neomezená apriorní a posteriorní hustota pravděpodobnosti vyhodnocená v bodě ω_0.
- Mnohdy $p(\psi|M_2) = p(\psi|M_1)$ (předchozí podmínka slabší).
- Nezatěžujeme se posteriorní analýzou M_1; není potřeba přímý výpočet marginální věrohodnosti.

Savage-Dickey – ilustrace

- M_1: omezení $\beta = \beta_0$ ($R\beta = r$ je jednoduchým rozšířením).
- Bayesův faktor:

$$BF_{12} = \frac{p(\beta = \beta_0|y, M_2)}{p(\beta = \beta_0|M_2)}.$$

- Jmenovatel lze snadno spočítat (marginální apriorní hustota pro β je normální):

$$p(\beta = \beta_0|M_2) = \frac{1}{(2\pi)^{\frac{k}{2}}}|\underline{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta_0 - \underline{\beta})'\underline{V}^{-1}(\beta_0 - \underline{\beta})\right]$$

Savage-Dickey – ilustrace (pokračování)

- Hustota $p(\beta|y, h, M_2)$ odpovídá normálnímu rozdělení $\times$ neznáme podobu $p(\beta|y, M_2)$.
- $p(\beta = \beta_0|y, M_2)$ lze snadno odhadnout.
- Gibbsovým vzorkovačem získáme $\beta^{(s)}$ a $h^{(s)}$ pro $s = S_0 + 1, \ldots, S$; zprůměrováním $p(\beta = \beta_0|y, h^{(s)}, M_2)$ přes všechny výběry $h^{(s)}$ získáme odhad $p(\beta = \beta_0|y, M_2)$.

$$\frac{1}{S_1} \sum_{s=S_0+1}^{S} p(\beta = \beta_0|y, h^{(s)}, M_2) \to p(\beta = \beta_0|y, M_2)$$

pro S_1 jdoucí k nekonečnu.

- Platí:

$$p(\beta = \beta_0|y, h^{(s)}, M_2) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\overline{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta_0 - \overline{\beta})'\overline{V}^{-1}(\beta_0 - \overline{\beta})\right].$$

Savage-Dickey – ilustrace (dokončení)

- Zákony pravděpodobnosti implikují

$$p(\beta = \beta_0|y, M_2) = \int p(\beta = \beta_0|y, h, M_2)p(h|y.M_2)dh.$$

- $p(\beta = \beta_0|y, h, M_2)$ v sobě neobsahuje nadále β (jen vektor konstant β_0).
- Lze psát:

$$p(\beta = \beta_0|y, M_2) = \int g(h)p(h|y)dh = E[g(h)|y],$$

  kde $g(h) = p(\beta = \beta_0|y, h, M_2)$.
- Posteriorní simulátory použitelné právě pro výpočet charakteristik jako je $E[g(h)|y]$.

Obsah tématu

Predikční hustota

- NLRM: $y^* = X^*\beta + \epsilon^*$.

$$p(y^*|y) = \iint p(y^*|y,\beta,h)p(\beta,h|y)d\beta dh$$

- ϵ^* je nezávislé na $\epsilon \Rightarrow p(y^*|y,\beta,h) = p(y^*|\beta,h)$:

$$p(y^*|\beta,h) = \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \exp\left[-\frac{h}{2}(y^* - X^*\beta)'(y^* - X^*\beta)\right].$$

- Zajímá nás $E[g(y^*)|y] = \int g(y^*)p(y^*|y)dy^*$.
- $y^{*(s)}$ pro $s = 1,\dots,S$ jakožto výběry z $p(y^*|y) \Rightarrow$ $\widehat{g}_Y = \frac{1}{S}\sum_{s=1}^{S} g(y^{*(s)})$ konverguje k $E[g(y^*)|y]$.
- Sumace od $S_0 + 1$ do S (vyhození prvních vzorků).

Predikční hustota a Gibbsův vzorkovač

- Obecný postup pto získání vzorků y^*.
- Pro každé $\beta^{(s)}$ a $h^{(s)}$, vezmeme výběr $y^{*(s)}$ z predikční hustoty $p(y^*|y, \beta^{(s)}, h^{(s)})$ (normální rozdělení).
- Máme tak $\beta^{(s)}$, $h^{(s)}$ a $y^{*(s)}$ pro $s = 1, \ldots, S \rightarrow$ zákony pravděpodobnosti říkají $p(\beta, h, y^*|y) = p(y^*|y, \beta, h)p(\beta, h|y)$.
- Strategie výběru nejdříve z posteriorní hustoty a potom z $p(y^*|y, \beta, h) \rightarrow$ výběr z $p(\beta, h, y^*|y)$.
- Obecné pravidlo: pokud máme výběry ze sdružené hustoty pravděpodobnosti $p(\theta, y^*|y)$, potom samostatné výběry θ jsou výběry z marginálního rozdělení $p(\theta|y)$ a samotné výběry y^* jsou výběrem z $p(y^*|y)$.

Obsah tématu

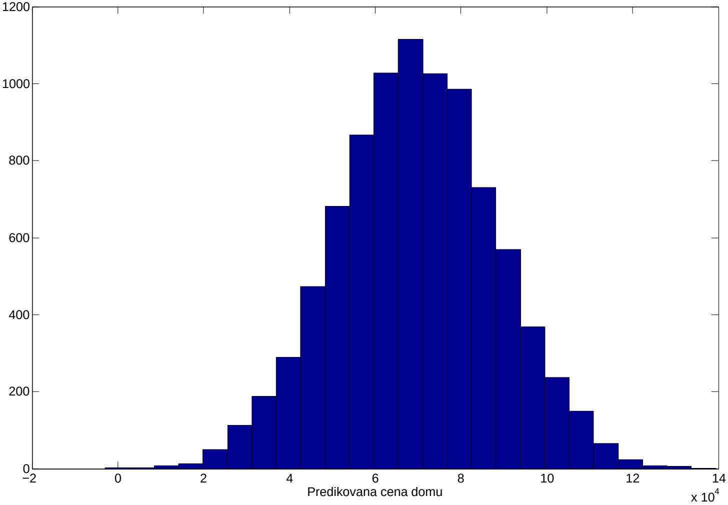
Příklad – ceny domů ve Windsoru

- Analogie předchozího příkladu.
- Nezávislé apriorní hustoty pro β (normální) a přesnost chyby, h (gama rozdělení).

Posteriorní výsledky pro parametr β

	Prior	Posterior	NSE	Gewekeho CD	P. podíl šancí pro $\beta_j = 0$
β_1	0 (10000)	-4119.01 (3251.44)	34.273	-0.065	1.39
β_2	10 (5)	5.45 (0.36)	0.004	-1.474	0.00
β_3	5000 (2500)	3228.83 (1080.46)	11.389	1.073	0.18
β_4	10000 (5000)	16136.64 (1605.11)	16.919	0.144	0.00
β_5	10 (5)	7685.55 (987.20)	10.406	-0.597	0.00

Predikční hustota pro dům o rozloze 5000 čtverečních stop.



Obsah tématu

Omezení ve tvaru nerovnosti

- $\beta \in A$, kde A je příslušná relevantní oblast parametrického prostoru.
- Bayesiánské řešení: zakomponovat omezení do apriorní hustoty.
- Posteriorní analýza: Importance Sampling.

Obsah tématu

Apriorní hustota

- $\beta \in A \Leftrightarrow$ oblast parametrického prostoru mimo A má apriorní váhy hodnoty 0.
- Pro ilustraci principu: přirozeně konjugovaná apriorní hustota.

$$p(\beta, h) \propto f_{NG}(\beta, h | \underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}) 1(\beta \in A)$$

- Indikační funkce $1(\beta \in A)$ nabývá hodnot 1, pokud $\beta \in A$ a 0 jinak.
- Marginální hustota pro β z omezeného t-rozdělení:

$$p(\beta) \propto f_t(\beta | \underline{\beta}, \underline{s}^2 \underline{V}, \underline{\nu}) 1(\beta \in A).$$

- Neinformativní varianta:

$$p(\beta, h) \propto \frac{1}{h} 1(\beta \in A).$$

Obsah tématu

Posteriorná hustota

- Odvození analogické z předchozích přednášek (navíc jen omezení skrze indikační funkce):

$$p(\beta|y) \propto f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu})1(\beta \in A).$$

- Posteriorní hyperparametry: vztahy pro informativní nebo neinformativní apriorn hustotu.
- Pokud předpokládáme nezávislou normální-gama apriorní hustotou $\rightarrow$ vztah pro $p(\beta|y, h)$ násoben $1(\beta \in A)$.

Obsah tématu

Úvod

- Pro některé volby A existují analytické výsledky.
- Vektor parametrů θ, věrohodnostní funkce $p(y|\theta)$, apriorní hustota $p(\theta)$ a posteriorní hustota $p(\theta|y)$.
- *Importance function*, $q(\theta)$: jsme schopni z ní získat náhodné výběry $\theta^{(s)}$ pro $s = 1, \ldots, S$.

$$\widehat{g}_S = \frac{1}{S} \sum_{s=1}^{S} g(\theta^{(s)})$$

nebude konvergovat k $E[g(\theta)|y]$ pro $S \to \infty$.

Importance sampling

- Vážené průměrování.
- Nechť $\theta^{(s)}$ pro $s = 1, \dots, S$ je náhodný výběr z $q(\theta)$ a definujme

$$\widehat{g}_S = \frac{\sum_{s=1}^{S} w(\theta^{(s)}) g(\theta^{(s)})}{\sum_{s=1}^{S} w(\theta^{(s)})}$$

  kde

$$w(\theta^{(s)}) = \frac{p(\theta = \theta^{(s)}|y)}{q(\theta = \theta^{(s)})}$$

- $\widehat{g}_S$ koverguje k $E[g(\theta)|y]$ pro S jdoucí k nekonečnu (při splnění slabých podmínek).
- Podmínka zejména, že $q(\theta)$ pokrývá $p(\theta|y)$ a $E[g(\theta|y)]$ existuje.

Importance sampling – pokračování

- Váhy jen v čitateli i jmenovateli $\rightarrow$ potřeba vyhodnotit pouze jádrové hustoty $p(\theta|y)$ a $q(\theta)$.
- Pokud $p^*(\theta|y) \propto p(\theta|y)$ a $q^*(\theta) \propto q(\theta)$, lze využít

$$w(\theta^{(s)}) = \frac{p^*(\theta = \theta^{(s)}|y)}{q^*(\theta = \theta^{(s)})}.$$

Importance sampling – problémy

- $q(\theta)$ dobře neaproximuje $p(\theta|y)$ ($w(\theta^{(s)})$ je nula pro téměř každý výběr) $\rightarrow$ potřeba enormně velkého S pro dostatečně přesný odhad $E[g(\theta)|y]$.
- Mnohdy raději Gibbsův vzorkovač.
- θ vícedimenzionální: obtížné nalezení vhodné importance funkce.

Importance sampling – příklad

- NLRM s omezením ve formě nerovností:

$$q(\beta) = f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}).$$

- Výpočet vah:

$$w(\beta^{(s)}) = 1(\beta^{(s)} \in A).$$

- Váhy rovny jedné ($\beta^{(s)} \in A$) nebo nule ($\beta^{(s)} \notin A$).
- Numerická standardní chyba:

$$\sqrt{S}\{\widehat{g}_S - E[g(\theta)|y]\} \to N(0, \sigma_g^2)$$

- σ_g^2 je možno konzistentně odhadnout jako

$$\widehat{\sigma}_g^2 = \frac{\frac{1}{S}\sum_{s=1}^{S}\left[w(\theta^{(s)})\left\{g(\theta^{(s)}) - \widehat{g}_S\right\}\right]^2}{\left[\frac{1}{S}\sum_{s=1}^{S} w(\theta^{(s)})\right]^2}.$$

- NSE $= \frac{\widehat{\sigma}_g}{\sqrt{S}}$.

Obsah tématu

Porovnání modelů – úvod

- M_1: NLRM s přirozeně konjugovaným priorem s omezeními ve tvaru nerovnosti ($\beta \in A$).
- M_2: tentýž model, přičemž předchozí restrikce neplatí ($\beta \notin A$).
- Např. ekonomická teorie může implikovat $\beta \in A \Rightarrow p(M_1|y)$ označuje pravděpodobnost, že ekonomická teorie je v souladu s daty.
- Neomezený NLRM s přirozeně konjugovaným priorem pro výpočet $p(M_1|y) = p(\beta \in A|y)$ a $p(M_2|y) = 1 - p(M_1|y)$.
- Analytické řešení pro lineární omezení, alternativně importance sampling pro neomezený model, $p(\beta \in A|y) = E[g(\theta)|y]$, kde $g(\theta) = 1(\beta \in A)$.
- Náhodné výběry u neomezené posteriorní hustoty $(f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}))$ a spočítání podílu těch, které splňují $\beta \in A \Rightarrow$ odhad $p(\beta \in A|y)$.
- Provedení importance sampling a zachováním informace o tom, kolik vzorků jsme nechali a kolik vyhodili (přiřadili jim nulovou váhu) $\rightarrow$ základ pro výpočet $p(M_1|y)$ a $p(M_2|y)$.

Vnořené modely

- Savageho-Dickeyeho poměr hustot: M_2 je NLRM s přirozeně konjugovaným priorem a nerovnostními omezeními a M_1 je stejný jako M_2 s tou výjimkou, že $\beta = \beta_0$.
- Předpokládáme v obou modelech tutéž apriorní hustotu pro h:

$$BF_{12} = \frac{p(\beta = \beta_0 | y, M_2)}{p(\beta = \beta_0 | M_2)}.$$

- Známe jen jádrové hustoty × formálně:

$$p(\beta) = \underline{c} f_t(\beta | \underline{\beta}, \underline{s}^2 \underline{V}, \underline{\nu}) 1(\beta \in A), p(\beta | y) \quad = \overline{c} f_t(\beta | \overline{\beta}, \overline{s}^2 \overline{V}, \overline{\nu}) 1(\beta \in A).$$

Vnořené modely – dokončení

$$BF_{12} = \frac{\overline{c} f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu})}{\underline{c} f_t(\beta|\underline{\beta}, \underline{s}^2\underline{V}, \underline{\nu})}.$$

- Vyhodnocení dvou hustot v $\beta = \beta_0$ a výpočet konstant $\underline{c}$ a $\overline{c}$.
- Někdy snadno z tabulek, jinak předchozí metoda výpočtu $p(M_1|y)$ (omezení $\beta \in A$ platí).
- $\overline{c} = \frac{1}{p(M_1|y)}$, protože

$$\overline{c} = \frac{1}{\int f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}) 1(\beta \in A) d\beta}$$

  a $p(M_1|y) = \int f_t(\beta|\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}) 1(\beta \in A) d\beta$.
- Výpočet $\underline{c}$ analogicky (využití importance sampling v rámci apriorní hustoty pravděpodobnosti).

Obsah tématu

Predikční hustota

- Standardní postup získání výběrů y^*.
- Výběry z importance function je třeba převážit.
- $\theta^{(s)}$, náhodný výběr z importance function a $y^{*(s)}$, náhodný výběr z $p(y^*|y,\theta^{(s)})$ pro $s = 1, \ldots, S$, potom

$$\widehat{g}_Y = \frac{\sum_{s=1}^{S} w(\theta^{(s)}) g(y^{*(s)})}{\sum_{s=1}^{S} w(\theta^{(s)})}$$

  konverguje k $E[g(y^*)|y]$ pro S jdoucí k nekonečnu.
- Váha $w(\theta^{(s)})$ vychází z definice importance sampling.
- Výpočet predikčních charakteristik všude tam, kde je prováděno importance sampling.

Obsah tématu

Příklad – ceny domů ve Windsoru

- Apriorní hyperparametry stejné jako v předchozích příkladech.
- Omezení: $\beta_2 > 5$, $\beta_3 > 2500$, $\beta_4 > 5000$ a $\beta_5 > 5000$.

Posteriorní výsledky pro parametr β

	Stř. hodnota	Směrodatná odchylka	NSE	P. podíl šancí pro $\beta_j = \underline{\beta}_j$
β_1	−5658.15	3011.44	41.245	1.20
β_2	5.50	0.30	0.004	0.00
β_3	3571.50	777.15	10.644	0.49
β_4	16638.59	1671.19	22.889	0.00
β_5	7454.92	925.41	12.675	0.22