Bayesiánská analýza

V. Nelineární regresní model

Obsah tématu

Úvod

- Pátá kapitola z Koop (2003) resp. učebního textu.
- Doposud:

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i.$$

- Cobb-Douglasova produkční funkce:

$$y = \alpha_1 x_1^{\beta_2} \cdot \ldots \cdot x_k^{\beta_k}.$$

- Logaritmování

$$\ln(y_i) = \beta_1 + \beta_2 \ln(x_{i2}) + \ldots + \beta_k \ln(x_{ik}) + \epsilon_i.$$

CES funkce

- CES produkční funkce:

$$y_i = \left(\sum_{j=1}^{k} \gamma_j x_{ij}^{\gamma_{k+1}} \right)^{\frac{1}{\gamma_{k+1}}}.$$

- Model:

$$y_i = \left(\sum_{j=1}^{k} \gamma_j x_{ij}^{\gamma_{k+1}} \right)^{\frac{1}{\gamma_{k+1}}} + \epsilon_i.$$

Obecně

- Předpoklady:
  1. $\epsilon \sim N(0_N, h^{-1}I_N)$.
  2. Všechny prvky matice X jsou pevná čísla (tj. nenáhodné proměnné) nebo náhodné veličiny nezávislé se všemi prvky vektoru ϵ + $p(X|\lambda)$, kde λ je vektor parametrů, který neobsahuje žádný z ostatních parametrů modelu.
- Obecně pracujeme s modelem:

$$y = f(X, \gamma) + \epsilon.$$

Obsah tématu

Věrohodnostní funkce

- Z normality náhodných složek:

$$p(y|\gamma, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp\left[-\frac{h}{2}\{y - f(X,\gamma)\}'\{y - f(X,\gamma)\} \right] \right\}.$$

Apriorní hustota

- Obecně $p(\gamma, h)$.
- Neinformativní varianta:

$$p(\gamma, h) = \frac{1}{h}.$$

- Uniformní rozdělení pro γ a $\ln(h)$.

Posteriorní hustota

- Součin věrohodnostní funkce a apriorní hustoty:

$$p(\gamma, h|y) \propto p(\gamma, h)\frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\left\{\exp\left[-\frac{h}{2}\{y - f(X,\gamma)\}'\{y - f(X,\gamma)\}\right]\right\}.$$

- Marginální hustota pro γ (při neinformativní apriorní hustotě):

$$p(\gamma|y) \propto \left[\{y - f(X,\gamma)\}'\{y - f(X,\gamma)\}\right]^{-\frac{N}{2}}.$$

Obsah tématu

Obecný algoritmus

1. Zvolíme počáteční hodnotu, $\theta^{(0)}$.
2. Vygenerujeme kandidátský výběr θ^* ze zvolené kandidátské hustoty $q(\theta^{(s-1)}; \theta)$.
3. Spočítáme akceptační pravděpodobnost (*acceptance probability*), $\alpha(\theta^{(s-1)}, \theta^*)$.
4. Přiřadíme $\theta^{(s)} = \theta^*$ s pravděpodobností $\alpha(\theta^{(s-1)}, \theta^*)$ a $\theta^{(s)} = \theta^{(s-1)}$ s pravděpodobností $1 - \alpha(\theta^{(s-1)}, \theta^*)$.
5. Opakujeme Krok 1, 2 a 3 celkem S krát.
6. Spočítáme průměr S výběrů $g(\theta^{(1)}), \ldots, g(\theta^{(S)})$.

Akceptační pravděpodobnost

1. Princip: procházet hlavně oblast vysoké hustoty (ale nejen ji) + tendence vracet se do této oblasti.
2. Více zamítat kandidáty z oblasti nízké hustoty.
3. Podoba akceptační pravděpodobnosti:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min\left[\frac{p(\theta = \theta^*|y)q(\theta^*; \theta = \theta^{(s-1)})}{p(\theta = \theta^{(s-1)}|y)q(\theta^{(s-1)}; \theta = \theta^*)}, 1\right]$$

4. Kandidátská hustota: q (obecně nemusí být symetrická).
5. Potřeba dobré volby kandidátské hustoty!

Obsah tématu

Princip

- Kandidátské hustoty nezávislé na výběrech.
- $q(\theta^{(s-1)}; \theta) = q^*(\theta)$.
- Užitečný v případech, kdy existuje vhodná aproximace posteriorní hustoty (kandidátská hustota).
- Zjednodušení kandidátské hustoty:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min\left[\frac{p(\theta = \theta^*|y)q^*(\theta = \theta^{(s-1)})}{p(\theta = \theta^{(s-1)}|y)q^*(\theta = \theta^*)}, 1\right].$$

Souvislost s importance sampling

- Pokud definujeme váhy:

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

- $q(\theta^{(s-1)}; \theta) = q^*(\theta)$.
- Akceptační pravděpodobnost:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min \left[\frac{w(\theta^*)}{w(\theta^{(s-1)})}, 1 \right].$$

Volba kandidátské hustoty

- Není obecné pravidlo.
- Obvykle využití „klasické" metody maximální věrohodnosti (maximal likelihood).
- Klasická ekonometrie: ML estimátor asymptoticky normální s kovarianční maticí
$$var(\widehat{\theta}_{ML}) = I(\theta)^{-1}.$$
- Informační matice:
$$I(\theta) = -E\left[\frac{\partial^2 \ln(p(y|\theta))}{\partial\theta\partial\theta'}\right].$$
- Záporný inverzní Hessián: $\widehat{var(\widehat{\theta}_{ML})}$.

Volba kandidátské hustoty – pokračování

- Pokud velikost vzorku dostatečně velká + relativně neinformativní apriorní hustota $\Rightarrow$ posteriorní hustota aproximativně normální se střední hodnotou $\widehat{\theta}_{ML}$ a kovarianční maticí aproximativně $\widehat{var(\widehat{\theta}_{ML})}$.
- Někdy maximalizace posteriorní hustoty $\rightarrow \widehat{\theta}_{\max}$ a kovarianční maticí aproximativně $\widehat{var(\widehat{\theta}_{\max})}$ (lepší v případě informativního prioru).
- $q^*(\theta) = f_N(\theta|\widehat{\theta}_{ML}, \widehat{var(\widehat{\theta}_{ML})})$ dobrá $\times$ obvyklejší
  $q^*(\theta) = f_t(\theta|\widehat{\theta}_{ML}, \widehat{var(\widehat{\theta}_{ML})}, \nu)$ (důležitost konců hustoty; nejméně tak tlusté jako posteriorní hustota).

Další otázky

- Konvergenční diagnostiky.
- Problémy: multimodální rozdělení nebo posterior jen v omezené oblasti (např. gama rozdělení).

Obsah tématu

Princip

- Pokud nenalezneme dobrou aproximaci posteriorní hustoty.
- Kandidátské výběry dle:

$$\theta^* = \theta^{(s-1)} + z.$$

- z: increment random variable.
- θ^* a $\theta(s-1)$ vstupují do vztahu symetricky $\Rightarrow$ $q(\theta^*; \theta = \theta^{(s-1)}) = q(\theta^{(s-1)}; \theta = \theta^*)$ (Metropolis algoritmus).
- Akceptační pravděpodobnost:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min\left[\frac{p(\theta = \theta^*|y)}{p(\theta = \theta^{(s-1)}|y)}, 1\right].$$

Volba z

- Obvykle vícerozměrné normální rozdělení.
- $\theta^{(s-1)}$ je střední hodnota.
- Potřeba kovarianční matice Σ!

$$q(\theta^{(s-1)};\theta) = f_N(\theta|\theta^{(s-1)},\Sigma).$$

- Problém s akceptační pravděpodobností (v průměru).
- Není obecné pravidlo, z praxe pro dimenzi blížící se k nekonečnu, 0.23.
- Jiné pravidlo: zhruba 0.5.
- Použití konvergenčních diagnostik!

Volba Σ

- Snadné experimenty pro skalární Σ.
- Jinak obvyklé $\Sigma = c\Omega$, kde c skalár a Ω odhad posteriorní kovarianční matice $\rightarrow$ experimenty s různými c.
- Nalezení Ω jako odhadu $var(\theta|y)$:
  - První způsob: začít s $\Sigma = cI_p$ a najít c pro které neextrémní akceptační pravděpodobnost (0.000001 nebo 0.99999) $\rightarrow$ hrubý odhad Ω a volba $\Sigma = c\Omega$ a hledání nové hodnoty c (nalézání lepších Ω a následně Σ).
  - Alternativně: $\Omega = \widehat{var(\widehat{\Omega}_{ML})}$.

Obsah tématu

Princip

- Kombinace Gibbsova vzorkovače a M-H algoritmu.
- V nelineárním regresním modelu volba neinformativní nebo nezávislé gama apriorní hustoty pro h implikue $p(h|y, \gamma)$ jako gama hustotu.
- $p(\gamma|y, h)$ není známá hustota $\Rightarrow$ M-H algoritmus.
- Následně $p(h|y, \gamma)$ jako gama rozdělení $\rightarrow$ platné výběry $\gamma^{(s)}$ a $h^{(s)}$ pro $s = 1, \ldots, S$.
- Obecně: systém plně podmíněných posteriorních hustota, kdy z některých výběry snadné, z jiných M-H algoritmem.

Obsah tématu

Princip

- Analýza kvality modelu, alternativa k posteriornímu podílu šancí.
- Rozlišení mezi aktuálně pozorovanými daty y a pozorovatelnými daty y° generovanými modelem $\rightarrow$ y° náhodný vektor rozměru $N \times 1$ s funkcí hustoty pravděpodobnosti $p(y^\circ|\theta)$.
- Funkce $g(\cdot)\rightarrow$ $p(g(y^\circ)|y)$ jako souhrn veškeré informace z modelu o $g(y^\circ)$ po shlédnutí dat.
- Snadno $g(y)$ pro pozorovaná data $\rightarrow$ pokud $g(y)$ je z odlehlého konce $p(g(y^\circ)|y)$, potom model nemůže dobře vysvětlovat $g(y)$, tedy $g(y)$ není tím typem charakteristiky dat, které lze přijatelně generovat modelem.
- Formální získání pravděpodobnosti okrajových oblastí způsobem podobným p-hodnotě z klasické statistiky.
- Posteriorní predikční p-hodnota je pravděpodobnost modelu generovat datovou sadu, která bude mít extrémnější vlastnosti než ta, kterou skutečně pozorujeme (jednostranná nebo oboustranná p-hodnota).

Výpočet

- $p(g(y^\circ)|y)$ možno spočítat použitím simulačních metod:

$$p(g(y^\circ)|y) = \int p(g(y^\circ)|\theta, y)p(\theta|y)d\theta = \int p(g(y^\circ)|\theta)p(\theta|y)d\theta.$$

- Díky podmíněnosti vektorem parametrů θ, aktuální data nepřinášejí žádnou dodatečnou informaci o y°.
- Posteriorní simulátor poskytuje výběry z $p(\theta|y)$ a jsme tak schopni simulovat výběry z $p(g(y^\circ)|\theta)$ pro daný posteriorní výběr parametrů θ.

Využití

- Posteriorní predikční p-hodnotu lze využít dvěma způsoby:
  1. Jako měřítko souladu modelu s daty, tedy jaká je pravděpodobnost, že data byly generována dle tohoto modelu.
  2. Pro porovná různých modelů, tedy jestliže jeden model poskytuje posteriorní predikční p-hodnoty výrazně nižší než druhý model, potom je to důkaz v neprospěch tohoto druhého modelu.

Příklad pro test normality – úvod

- Příklad z úvodu (pro $i = 1, \ldots, N$):

$$y_i^\circ = f(X_i, \gamma) + \epsilon_i.$$

- Z předpokladů na náhodnou složku:

$$p(y^\circ|\gamma, h) = f_N(y^\circ|f(X, \gamma), h^{-1} I_N).$$

- Zjednodušení pro neinformativní apriorní hustotou (h lze vyintegrovat):

$$p(y^\circ|\gamma) = f_t(y^\circ|f(X, \gamma), \bar{s}^2 I_N, N),$$

kde

$$\bar{s}^2 = \frac{[y - f(X, \gamma)]'[y - f(X, \gamma)]}{N}.$$

- Nalezení příslušného percentilu, který vytváří $g(y)$ v rámci hustoty $p(g(y^\circ)|y)$.

Příklad pro test normality – definice

- Možnosti volby $g(\cdot)$: analýza reziduí z hlediska vlastností (splnění předpokladů).
- V bayesovském kontextu jsou chyby ϵ_i pro $i = 1, \ldots, N$ dány jako

$$\epsilon_i = y_i - f(X, \gamma)$$

- Předpoklad normality = nulovost

$$Skew = \frac{\sqrt{N} \sum_{i=1}^{N} \epsilon_i^3}{\left[\sum_{i=1}^{N} \epsilon_i^2\right]^{\frac{3}{2}}} \quad Kurt = \frac{N \sum_{i=1}^{N} \epsilon_i^4}{\left[\sum_{i=1}^{N} \epsilon_i^2\right]^2} - 3.$$

- Nepozorujeme $\epsilon_i \rightarrow$ na základě posteriorního simulátoru:

$$E[Skew|y] = E\left\{ \left. \frac{\sqrt{N} \sum_{i=1}^{N} [y_i - f(X_i, \gamma)]^3}{\left[\sum_{i=1}^{N} [y_i - f(X_i, \gamma)]^2\right]^{\frac{3}{2}}} \right| y \right\}.$$

Příklad – praktický postup

1. Vezmeme výběr $\gamma^{(s)}$ využitím posteriorního simulátoru.
2. Vygenerujeme reprezentativní datovou sadu y° z $p(y^{\circ}|\gamma^{(s)})$ užitím odpovídajících vztahů.
3. Definujeme $\epsilon_i^{(s)} = y_i - f(X_i, \gamma^{(s)})$ pro $i = 1, \ldots, N$ a vyhodnotíme hodnotu šikmosti v tomto bodě, abychom získali $Skew^{(s)}$.
4. Definujeme $\epsilon_i^{\circ(s)} = y_i^{\circ(s)} - f(X_i, \gamma^{(s)})$ pro $i = 1, \ldots, N$ a vyhodnotíme hodnotu špičatosti v tomto bodě, abychom získali $Skew^{\circ(s)}$.
5. Opakujeme Krok 1, 2, 3 a 4 celkem S krát.
6. Spočítáme průměr S výběrů $Skew^{(1)}, \ldots, Skew^{(S)}$ pro odhad $E[Skew|y]$.
7. Spočítáme podíl S vzorků $Skew^{\circ(1)}, \ldots, Skew^{\circ(S)}$, které jsou menší než odhad $E[Skew|y]$. Odhad p-hodnoty, pokud výsledek menší než 0.5 (jinak jedna mínus).

Obsah tématu

Princip

- Pro porovnání nevnořených modelů, či vnořených modelů, kdy nelze snadno využít Savage-Dickey density ratio.
- Obecnější metoda vyhodnocení posteriorního podílu šancí.
- Porovnání různých funkčních vztahů $f(\cdot)$.
- Princip: inverzi marginální věrohodnosti pro model M_i, který závisí na vektoru parametrů θ, lze zapsat jako $E[g(\theta)|y, M_i]$ pro specifickou volbu $g(\cdot)$.

Metoda Gelfanda a Deye

- Nechť $p(\theta|M_i)$ označuje apriorní hustotu, $p(y|\theta, M_i)$ věrohodnostní funkci a $p(\theta|y, M_i)$ posteriorní hustotu pro model M_i definovaný v oblasti Θ. Pokud $f(\theta)$ je funkce hustoty pravděpodobnosti definovaná na oblasti obsahující Θ, potom

$$E\left[\frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)}\middle| y, M_i\right] = \frac{1}{p(y|M_i)}$$

Důkaz

$$E\left[\frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)}\middle| y, M_i\right]$$

$$\begin{aligned}
&= \int \frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)} p(\theta|y,M_i)d\theta \\
&= \int \frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)} \frac{p(\theta|M_i)p(y|\theta,M_i)}{p(y|M_i)} d\theta \\
&= \frac{1}{p(y|M_i)} \int f(\theta)d\theta \\
&= \frac{1}{p(y|M_i)}.
\end{aligned}$$

Problémy

- Na první pohled lze pro jakoukoliv p.d.f $f(\theta)$ nastavit:

$$g(\theta) = \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)}.$$

- Následné využití výsledků posteriorního simulátoru k odhadu $E[g(\theta)|y, M_i]$.
- Pro úspěch potřeba obezřetné volby $f(\theta)$.
- Geweke (1999): $\frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)}$ musí být shora omezena, tedy musí nabývat konečných hodnot pro jakoukoliv volbu θ.

Geweke (1999) – úvod

- Strategie definování $f(\theta)$ jako funkce normální hustoty pravděpodobnosti s ohraničenými okraji.
- Důvod: obtížné ověřit, zda-li $\frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)}$ je konečné na okrajích hustoty normálního rozdělení $\rightarrow$ odříznutím je $f(\theta)$ pro tyto potenciálně problematické oblasti nulová.
- Formálně: $\widehat{\theta}$ a $\widehat{\Sigma}$ jsou odhady $E(\theta|y,M_i)$ a $var(\theta|y,M_i)$ z posteriorní simulace.
- Pro určitou oblast pravděpodobnosti $p \in (0,1)$ označuje $\widehat{\Theta}$ oblast definičního oboru funkce $f(\theta)$:

$$\widehat{\Theta} = \{\theta : (\widehat{\theta} - \theta)' \widehat{\Sigma}^{-1} (\widehat{\theta} - \theta) \leq \chi^2_{1-p}(k)\}.$$

- $\chi^2_{1-p}(k)$ je $(1-p)$ procentní kvantil rozdělení chí-kvadrát s k stupni volnosti (k je počet prvků vektoru θ).

Geweke (1999) – dokončení

- Geweke doporučuje $f(\theta)$ jako funkci vícerozměrné normální hustoty pravděpodobnosti omezené do oblasti $\widehat{\Theta}$:

$$f(\theta) = \frac{1}{p(2\pi)^{\frac{k}{2}}} |\widehat{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\widehat{\theta}-\theta)'\widehat{\Sigma}^{-1}(\widehat{\theta}-\theta)\right] 1(\theta \in \widehat{\Theta}).$$

- 1() je indikační funkce + očekává se, že nejlepších výsledků dosahujem při volbě nízkých hodnot p (např. $p = 0.01$) (při odhadu marginální věrohodnosti zahrnujeme mnohem více vzorků).
- Dodatečný náklad zkoušení různých hodnot p je velmi malý.
- Standarní způsob spočítání NSE – využití pro vyhodnocení přesnosti odhadu marginální věrohodnosti (např. balík BACC).
- Metoda pro jakýkoliv model: nutný posteriorní simulátor a známé $p(\theta|M_i)$ a $p(y|\theta, M_i)$ (netriviální, někdy známe jen jádrové hustoty).
- Gewekova implementace jen v případě $\widehat{\Theta} \in \Theta$ (pokud ne, nabízeny další postupy).

Obsah tématu

Princip

- Využijeme výběry z $p(\gamma|y)$, které nám poskytuje M-H algoritmus.
- Tvorba vzorků z podmíněné predikční hustoty $p(y^*|y, \gamma^{(s)})$.
- Lze ověřit:
$$p(y^*|y, \gamma) = f_t(y^*|f(X^*, \gamma), \bar{s}^2 I_T, N).$$
- Výběry z y^* podmíněné vektorem γ lze tedy získat velmi snadno.

Obsah tématu

Příklad

- CES produkční funkce.