

Bayesiánská analýza

II. Normální lineární regresní model s přirozeně konjugovanou apriorní hustotou (jediná vysvětlující proměnná)

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

- Koop (2003) - kapitola 2
- Normální lineární regresní model s jedinou vysvětlující proměnnou (není třeba maticové algebry).
- Přirozeně konjugovaný prior \Rightarrow analytické výsledky.
- Význam před rozvojem výpočetní techniky.
- Ukázka propojení klasické a bayesovské ekonometrie.

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

Lineární regresní model

- y_i – pozorovaná data vysvětlované proměnné ($i = 1, \dots, N$).
- x_i – pozorovaná data vysvětlující proměnné ($i = 1, \dots, N$).
- LRM s jedinou vysvětlující proměnnou a bez úrovně konstanty (pro zjednodušení):

$$y_i = \beta x_i + \epsilon_i$$

- ϵ_i – náhodná složka (chyba); β – parametr.
- Vektorově pro $y = (y_1, \dots, y_N)'$, $x = (x_1, \dots, x_N)'$ a $\epsilon = (\epsilon_1, \dots, \epsilon_N)'$:

$$y = \beta x + \epsilon$$

Předpoklady

- Předpoklady o ϵ_i a x_i determinují podobu věrohodnostní funkce.
- Klasické předpoklady (později uvolněny):
 - 1 ϵ_i je i.i.d. $N(0, \sigma^2)$ (independent and identically distributed).
 - 2 x_i je fixní proměnná (tj. nenáhodná veličina), popř. náhodná s omezeními.

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce**
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

Věrohodnostní funkce I

- Věrohodnostní funkce = funkce sdružené hustoty pravděpodobnosti pro všechna data podmíněná neznámými parametry:

$$p(y|\beta, \sigma^2).$$

- Získáme z definovaných předpokladů.
 - $p(y_i|\beta, \sigma^2)$ má normální rozdělení,
 - $E(y_i|\beta, \sigma^2) = \beta x_i$,
 - $\text{var}(y_i|\beta, \sigma^2) = \sigma^2$.

Věrohodnostní funkce II

- Věrohodnostní funkce:

$$p(y|\beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \beta x_i)^2 \right].$$

Věrohodnostní funkce III

- Další vyjádření věrohodnostní funkce:

$$\sum_{i=1}^N (y_i - \beta x_i)^2 = \nu s^2 + (\beta - \hat{\beta})^2 \sum_{i=1}^N x_i^2,$$

- kdy

$$\nu = N - 1$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$s^2 = \frac{\sum (y_i - \hat{\beta} x_i)^2}{\nu}$$

- $\hat{\beta}$, s^2 a $\nu = N - 1$ OLS odhady pro β , rozptyl reziduí a počet stupňů volnosti.

Věrohodnostní funkce IV

- Obvykle využíváme přesnost chyby, h , místo variability:

$$h = \frac{1}{\sigma^2}$$

- Věrohodnostní funkce:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{1}{2}} \exp \left[-\frac{h}{2} (\beta - \hat{\beta})^2 \sum_{i=1}^N (x_i)^2 \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp \left[-\frac{h\nu}{2s^{-2}} \right] \right\}.$$

- První člen je podobný jádrové hustotě (kernelu) normální hustoty pro β ; druhý člen je podobný rozdělení gama pro h .

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota**
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

Konjugovaná apriorní hustota

- Konjugovaná apriorní hustota – interpretovatelná jako dodatečná data.
- Jestliže je \mathcal{F} třída výběrových rozdělení $p(y|\theta)$ a \mathcal{P} je třída apriorních rozdělení pro θ , potom třída \mathcal{P} je konjugovaná pro \mathcal{F} jestliže

$$p(\theta|y) \in \mathcal{P} \quad \text{pro všechna } p(\cdot|\theta) \in \mathcal{F} \quad \text{a} \quad p(\cdot) \in \mathcal{P}.$$

- Vágní definice, pokud \mathcal{P} je třída všech rozdělení, potom \mathcal{P} je vždy konjugovaná bez ohledu na použitou třídu výběrových rozdělení.
- Zájem o třídy **přirozeně konjugovaných apriorních hustot**, kdy \mathcal{P} je množina všech hustot majících stejnou funkční podobu jako věrohodnostní funkce.

Prior I

- Apriorní hustota – libovolná forma, obvykle určitá třída priorů, které lze snadno interpretovat a (nebo) zjednodušují výpočet.
- Apriorní hustota pro β a h : $p(\beta, h)$.
- Pravidla pravděpodobnosti: $p(\beta, h) = p(\beta|h)p(h)$.
- Forma věrohodnostní funkce – konjugovaný prior jako normální rozdělení pro $\beta|h$ a gama rozdělení pro h .

Prior II

- Normální-gama rozdělení:

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V})$$

$$h \sim G(\underline{s}^{-2}, \underline{\nu})$$

- Přirozeně konjugovaný prior pro β a h :

$$\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu})$$

- Volba hodnoty apriorních hyperparametrů: $\underline{\beta}$, \underline{V} , \underline{s}^{-2} a $\underline{\nu}$ (zohlednění apriorní informace).

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota**
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

Posterior I

- Posteriorní hustota je proporcionální součinu věrohodnostní funkce a apriorní hustoty.
- Normální-gama rozdělení (apriorní hustota je tedy skutečně přirozeně konjugovaná):

$$\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu}),$$

přičemž platí

$$\begin{aligned}\bar{V} &= \frac{1}{\underline{V}^{-1} + \sum x_i^2}, \\ \bar{\beta} &= \bar{V}(\underline{V}^{-1}\underline{\beta} + \hat{\beta} \sum x_i^2), \\ \bar{\nu} &= \underline{\nu} + N.\end{aligned}$$

- \bar{s}^{-2} je implicitně definováno následovně:

$$\bar{\nu}\bar{s}^2 = \underline{\nu}\underline{s}^2 + \nu s^2 + \frac{(\hat{\beta} - \underline{\beta})^2}{\underline{V} + \frac{1}{\sum x_i^2}}.$$

Vlastnosti posteriorní hustoty I

- $p(\beta|y, h)$ má normální rozdělení.
- Chceme $p(\beta|y)$.
- Marginální posteriorní rozdělení pro β je t -rozdělení:

$$\beta|y \sim t(\bar{\beta}, \bar{s}^2 \bar{V}^2, \bar{\nu}).$$

- Z definice t -rozdělení vyplývá:

$$E(\beta|y) = \bar{\beta},$$
$$\text{var}(\beta|y) = \frac{\bar{\nu} \bar{s}^2}{\bar{\nu} - 2} \bar{V}.$$

Vlastnosti posteriorní hustoty II

- Bayesiánský bodový odhad $\bar{\beta}$ je vážený průměr OLS odhadu a apriorní střední hodnoty $\underline{\beta}$.
- Váhy jsou proporcionální k $\sum x_i^2$ a \underline{V}^{-1} .
- \underline{V}^{-1} = důvěra v apriorní a datovou informaci (v klasické ekonometrii je $(\sum x_i^2)^{-1}$ proporcionální k variabilitě $\hat{\beta}$).
- Alternativní intuice: jednoduchý případ $x_i = 1$ pro $i = 1, \dots, N \Rightarrow \sum x_i^2 = N$ a váha přiřazená k $\hat{\beta}$ je velikost vzorku (rozumné měřítko pro objem informací v datech).

Vlastnosti posteriorní hustoty III

- Rozptyl OLS estimátoru $s^2(\sum x_i^2)^{-1}$ (využití např. k testování statistické významnosti parametru, kdy t -statistika pro testování $\beta = 0$ je $\frac{\hat{\beta}}{\sqrt{s^2(\sum x_i^2)^{-1}}}$).
- Bayesiánská analogie – posteriorní rozptyl β (podobná podobu \times zahrnuje datovou i nedatovou informaci).
- Přirozeně konjugovaná apriorní hustota = prior vyplývá z fiktivní datové množiny (podobná role $\underline{\nu}$ a N).
- $\underline{\nu}$: apriorní velikost vzorku.

Vlastnosti posteriorní hustoty IV

- Fiktivní datová množina \rightarrow lze využít při specifikaci $\underline{\beta}$, \underline{V} , \underline{s}^{-2} a $\underline{\nu}$.
- Citlivostní analýza prioru – empirické výsledky lze prezentovat za využití různých priorů.
- Síla datové informace může i nemusí rozhodnout (nutnost zdůvodnění volby apriorních hyperparametrů).

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota**
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace

Neinformativní prior I

- Neinformativní prior z nastavení $\underline{\nu} = 0$ a $\underline{V}^{-1} = 0$ (tj. $\underline{V} \rightarrow \infty$).
- V tomto případě: $\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$, kde

$$\begin{aligned}\bar{V} &= \frac{1}{\sum x_i^2}, \\ \bar{\beta} &= \hat{\beta}, \\ \bar{\nu} &= N, \\ \bar{\nu} \bar{s}^2 &= \nu s^2.\end{aligned}$$

Neinformativní prior II

- Jedná se čistě o OLS odhad.
- Rozdílná interpretace:
 - bayesovský přístup – β je náhodná veličina;
 - klasický přístup – $\hat{\beta}$ je náhodná veličina.
- Tato apriorní „hustota“ neintegruje na hodnotu jedna \rightarrow tzv. nepravý prior (improper prior).
- Problémy s použitím.

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů**
- 7 Predikce
- 8 Empirická ilustrace

Porovnání modelů

- Dva modely M_1 a M_2 .
- M_j pro $j = 1, 2$ mají různé vysvětlující proměnné:

$$y_i = \beta_j x_{ji} + \epsilon_{ji}.$$

- Přirozeně konjugovaná normální-gama apriorní hustota:

$$\beta_j, h_j | M_j \sim NG(\underline{\beta}_j, \underline{V}_j, \underline{s}_j^{-2}, \underline{\nu}_j).$$

- Posteriorní hustota:

$$\beta_j, h_j | y, M_j \sim NG(\bar{\beta}_j, \bar{V}_j, \bar{s}_j^{-2}, \bar{\nu}_j).$$

Posterioční podíl šancí I

$$PO_{12} = \frac{p(y|M_1)p(M_1)}{p(y|M_2)p(M_2)}$$

- Apriorní pravděpodobnosti modelů $p(M_j)$ (před konfrontací s daty); neinformativní volba $p(M_1) = p(M_2) = \frac{1}{2}$.
- Marginální věrohodnost $p(y|M_j)$:

$$p(y|M_j) = \iint p(y|\beta_j, h_j)p(\beta_j, h_j)d\beta_j dh_j.$$

- Lze pro náš případ spočítat analyticky:

$$p(y|M_j) = c_j \left(\frac{\bar{V}_j}{\underline{V}_j} \right)^{\frac{1}{2}} (\bar{\nu}_j \bar{s}_j^2)^{-\frac{\bar{\nu}_j}{2}}.$$

Posteriorní podíl šancí II

- c_j konstanta nezávislá na datech

$$c_j = \frac{\Gamma\left(\frac{\bar{\nu}_j}{2}\right) (\underline{\nu}_j \bar{s}_j^2)^{\frac{\nu_j}{2}}}{\Gamma\left(\frac{\nu_j}{2}\right) \pi^{\frac{N}{2}}}.$$

- Posteriorní podíl šancí:

$$PO_{12} = \frac{c_1 \left(\frac{\bar{V}_1}{\underline{V}_1}\right)^{\frac{1}{2}} (\bar{\nu}_1 \bar{s}_1^2)^{-\frac{\bar{\nu}_1}{2}} p(M_1)}{c_2 \left(\frac{\bar{V}_2}{\underline{V}_2}\right)^{\frac{1}{2}} (\bar{\nu}_2 \bar{s}_2^2)^{-\frac{\bar{\nu}_2}{2}} p(M_2)}.$$

Faktory ovlivňující bayesovské porovnání modelů

- 1 Čím větší je apriorní podíl šancí $\frac{p(M_1)}{p(M_2)}$, tím vyšší je podpora pro M_1 .
- 2 $\bar{\nu}_j \bar{s}_j^2$ obsahuje člen $\nu_j s_j^2$, což je součet čtverců chyb (měřítko souladu modelu s daty).
- 3 Pokud všechny ostatní charakteristiky jsou stejné: posteriorní podíl šancí upřednostní model s větší koherencí mezi apriorní a datovou informací ($(\hat{\beta}_j - \underline{\beta}_j)^2$ vstupuje do $\bar{\nu}_j \bar{s}_j^2$).
- 4 $\left(\frac{\bar{V}_j}{\underline{V}_j}\right)$ je podíl posteriorní a apriorní variancí \Rightarrow pokud vše ostatní zůstane stejné, upřednostní model s větší apriorní informací (menší apriorní variancí), relativně k posteriorní informaci.
- 5 Použitím neinformativního prioru není marginální věrohodnost definována (stějně tak posteriorní podíl šancí).

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce**
- 8 Empirická ilustrace

Predikční hustota

- Chceme předpovědět y^* v bodě x^* :

$$y^* = \beta x^* + \epsilon^*.$$

- y^* není pozorováno.

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

- Lze ukázat:

$$y^*|y \sim t(\bar{\beta}x^*, \bar{s}^2\{1 + \bar{V}x^{*2}\}, \bar{\nu}).$$

- Bodová předpověď a s ní spojená nejistota (např. směrodatná odchylka predikce).

Bayesovské průměrování modelů

- Předpokládáme, že známe $p(M_j|y)$, pro $j = 1, 2$
- Bayesovské průměrování modelů v sobě zahrnuje průměrování přes všechny modely.
- Zákony pravděpodobnosti implikují:

$$p(y^*|y) = p(y^*|y, M_1)p(M_1|y) + p(y^*|y, M_2)p(M_2|y).$$

- Nezaměřovat na jediný model a pracovat např. s $p(y^*|y, M_1) \times$ průměrování výsledků přes oba (všechny) modely s váhami danými posteriorními hustotou pravděpodobnosti.

Obsah tématu

- 1 Lineární regresní model
- 2 Věrohodnostní funkce
- 3 Apriorní hustota
- 4 Posteriorní hustota
- 5 Neinformativní apriorní hustota
- 6 Porovnání modelů
- 7 Predikce
- 8 Empirická ilustrace**

Umělá data

- Umělá data pro ilustraci s $N = 50$, $\beta = 2$ a $h = 1$.
- Neinformativní prior a informativní prior s $\underline{\beta} = 1.5$, $\underline{V} = 0.25$, $\nu = 10$ a $\underline{s}^{-2} = 1$.
- „Originální“ data obsahem *data_NLRM1.mat* + skript *priklad_NLRM1.m*.

Apriorní a posteriorní charakteristiky parametru β

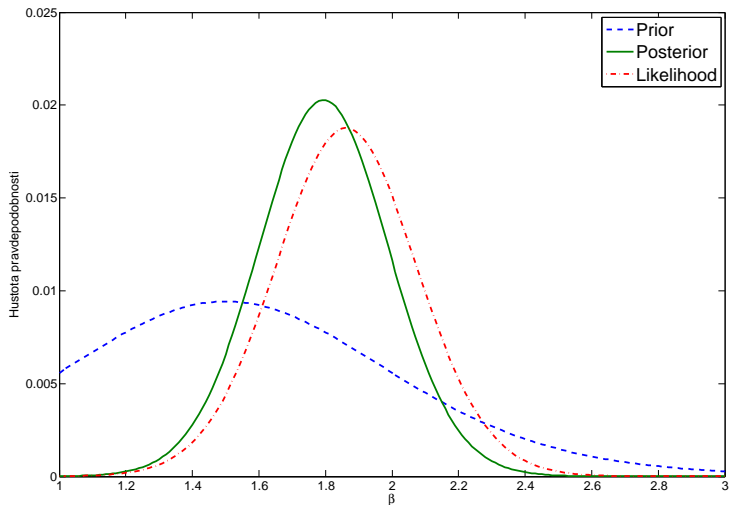
	Prior	Posterior	
	Informativní	(Inf. prior)	(Neinf. prior)
Stř. hodnota	1.500	1.794	1.862
Sm. odchylka [*]	0.559	0.199	0.216

* Jedná se o směrodatnou odchylku příslušného t -rozdělení, apriorní či posteriorní V tedy neudává přímo hodnotu rozptylu.

Apriorní a posteriorní charakteristiky parametru h

	Prior	Posterior	
	Informativní	(Inf. prior)	(Neinf. prior)
Stř. hodnota	1.000	1.215	1.283
Sm. odchylka [*]	0.447	0.222	0.257

* Jedná se o směrodatnou odchylku odpovídajícího gama rozdělení.

Marginální apriorní a posteriorní hustoty pro parametr β .

Apriorní a posteriorní charakteristiky pro model jen s úroňovou konstantou

	β		h	
	Prior	Posterior	Prior	Posterior
Stř. hodnota	1.500	0.990	1.000	0.917
Sm. odchylka	0.559	0.145	0.447	0.167

Další výsledky

- Posteriorní podíl šancí je 7346 $\Rightarrow p(M_1|y) = 0.9999$ a $p(M_2|y) = 0.0001$.
- Predikce pro $x^* = 0.5$:
 - při využití informativního prioru je $y^*|y \sim t(0.897, 0.833, 60)$,
 - V případě neinformativní apriorní hustoty je $y^*|y \sim t(0.931, 0.7915, 50)$.