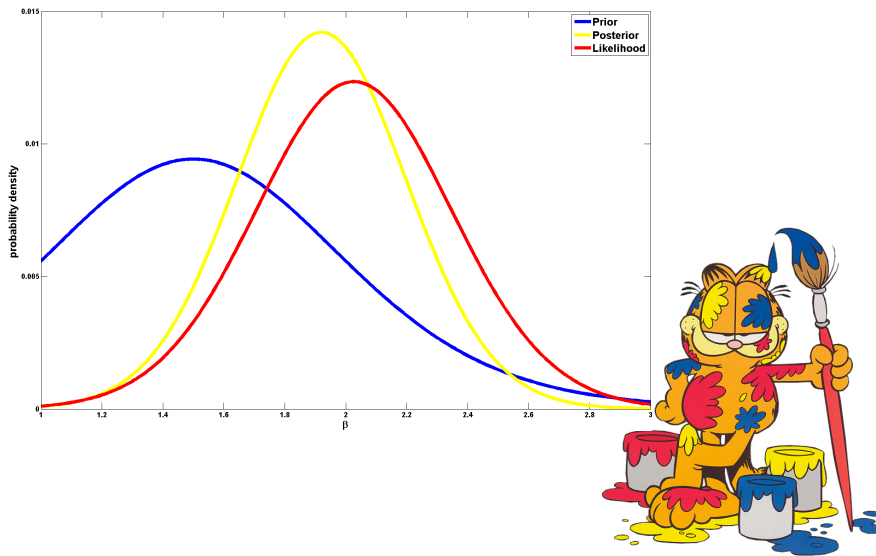


MASARYKOVA UNIVERZITA  
EKONOMICKO-SPRÁVNÍ FAKULTA



---

## Bayesiánská analýza

---

z anglického originálu pro potěšení sebe i druhých  
přeložil, opravil, doplnil, v  $\text{\TeX}$ u vysázel  
a v Matlabu funkční aplikace připravil

DANIEL NĚMEC

podzim 2011



# Obsah

<b>Předmluva</b>	<b>xiii</b>
<b>1 Základní principy a pojmy Bayesiánské ekonometrie</b>	<b>1</b>
1.1 Bayesiánská teorie	1
1.2 Bayesiánské výpočetní postupy	5
1.3 Software pro Bayesiánskou analýzu	8
1.4 Shrnutí	10
<b>2 NLRM s přirozeně konjugovaným priorem – jediný regresor</b>	<b>11</b>
2.1 Úvod	11
2.2 Věrohodnostní funkce	11
2.3 Apriorní hustota	14
2.4 Posteriorní hustota	15
2.5 Porovnání modelů	22
2.6 Predikční hustota	26
2.7 Empirická ilustrace	28
2.8 Shrnutí	31
<b>3 NLRM s přirozeně konjugovaným priorem - více regresorů</b>	<b>33</b>
3.1 LRM v maticovém vyjádření	33
3.2 Věrohodnostní funkce	34
3.3 Apriorní hustota	35
3.4 Posteriorní hustota	35
3.5 Porovnání modelů	39
3.5.1 Porovnání modelů s omezením ve tvaru nerovnosti	40
3.5.2 Omezení ve tvaru rovnosti	41
3.5.3 Intervaly nejvyšší posteriorní hustoty (HPDI)	44
3.6 Predikční hustota	46
3.7 Monte Carlo integrace	46
3.8 Empirická ilustrace	48
3.9 Shrnutí	55

<b>4</b>	<b>Normální lineární regresní model s jinými priory</b>	<b>57</b>
4.1	NLRM s nezávislou normální-gama apriorní hustotou . . . . .	58
4.1.1	Apriorní hustota . . . . .	58
4.1.2	Posteriorní hustota . . . . .	58
4.1.3	Gibbsův vzorkovač . . . . .	60
4.1.4	Markov Chain Monte Carlo diagnostiky . . . . .	62
4.1.5	Porovnání modelů: Savage-Dickey density ratio . . . . .	65
4.1.6	Předpověď . . . . .	67
4.1.7	Empirická ilustrace . . . . .	68
4.2	NLRM s omezeními ve tvaru nerovnosti . . . . .	72
4.2.1	Apriorní hustota . . . . .	72
4.2.2	Posteriorní hustota . . . . .	73
4.2.3	Importance sampling . . . . .	73
4.2.4	Porovnání modelů . . . . .	75
4.2.5	Předpověď . . . . .	76
4.2.6	Empirická ilustrace . . . . .	77
4.3	Shrnutí . . . . .	79
<b>5</b>	<b>Nelineární regresní model</b>	<b>81</b>
5.1	Úvod . . . . .	81
5.2	Věrohodnostní funkce . . . . .	83
5.3	Apriorní hustota . . . . .	83
5.4	Posteriorní hustota . . . . .	83
5.5	Metropolis-Hastings algoritmus . . . . .	84
5.5.1	Independence Chain M-H algoritmus . . . . .	86
5.5.2	Random Walk Chain M-H algoritmus . . . . .	88
5.5.3	Metropolis-within-Gibbs . . . . .	90
5.6	Měřítko kvality modelu: Posteriorní predikční p-hodnota . . . . .	91
5.7	Porovnání modelů: Metoda Gelfanda-Deye . . . . .	94
5.8	Predikce . . . . .	96
5.9	Empirická ilustrace . . . . .	97
5.10	Shrnutí . . . . .	102
<b>6</b>	<b>Lineární regresní model s obecnou kovarianční maticí chyb</b>	<b>105</b>
6.1	Úvod . . . . .	105
6.2	Model s obecnou kovarianční maticí . . . . .	106
6.2.1	Věrohodnostní funkce . . . . .	107
6.2.2	Apriorní hustota . . . . .	107
6.2.3	Posteriorní hustota . . . . .	108
6.3	Heteroskedasticita ve známé podobě . . . . .	109
6.3.1	Empirická ilustrace . . . . .	110
6.4	Heteroskedasticita v neznámé podobě . . . . .	112
6.4.1	Bayesovský výpočet . . . . .	114
6.4.2	Empirická ilustrace . . . . .	116
6.5	Autokorelace náhodných složek . . . . .	116
6.5.1	Bayesovský výpočet . . . . .	118

6.5.2	Empirická ilustrace . . . . .	121
6.6	Modely zdánlivě nesouvisejících regresí . . . . .	121
6.6.1	Apriorní hustota . . . . .	122
6.6.2	Bayesovský výpočet . . . . .	123
6.6.3	Empirická ilustrace . . . . .	124
6.7	Shrnutí . . . . .	124
<b>7</b>	<b>Lineární regresní model s panelovými daty</b>	<b>125</b>
7.1	Úvod . . . . .	125
7.2	Souhrnný model (pooled model) . . . . .	126
7.3	Modely individuálních vlivů (individual effects models) . . . . .	127
7.3.1	Věrohodnostní funkce . . . . .	128
7.3.2	Apriorní hustota . . . . .	128
7.3.3	Bayesovský výpočet . . . . .	130
7.4	Model náhodných koeficientů . . . . .	132
7.4.1	Věrohodnostní funkce . . . . .	132
7.4.2	Hierarchická apriorní hustota . . . . .	133
7.4.3	Bayesovský výpočet . . . . .	133
7.5	Porovnání modelů: Chibova metoda . . . . .	134
7.6	Empirická ilustrace . . . . .	135
7.7	Analýza efektivity a model stochastických hranic . . . . .	136
7.7.1	Úvod do modelu stochastických hranic . . . . .	136
7.7.2	Věrohodnostní funkce . . . . .	137
7.7.3	Hierarchická apriorní hustota pro model stochastických hranic . . . . .	138
7.7.4	Bayesovský výpočet . . . . .	139
7.7.5	Empirická ilustrace . . . . .	140
7.8	Rozšíření . . . . .	140
7.9	Shrnutí . . . . .	140
<b>8</b>	<b>Úvod do časových řad</b>	<b>141</b>
8.1	Úvod . . . . .	141
8.2	Local level model . . . . .	142
8.2.1	Věrohodnostní funkce a apriorní hustota . . . . .	144
8.2.2	Posteriorní hustota . . . . .	146
8.2.3	Empirické bayesiánské metody . . . . .	147
8.2.4	Empirická ilustrace: Local level model . . . . .	150
8.3	Obecný stavový model . . . . .	150
8.3.1	Bayesovský výpočet pro stavový model . . . . .	151
8.3.2	Empirická ilustrace . . . . .	156
8.4	Rozšíření . . . . .	156
8.5	Shrnutí . . . . .	156

<b>9</b>	<b>Modely kvalitativních a omezených vysvětlovaných proměnných</b>	<b>157</b>
9.1	Úvod	157
9.2	Jednorozměrné modely	159
9.3	Tobit model	160
9.3.1	Empirická ilustrace: tobit model	161
9.4	Probit model	161
9.4.1	Empirická ilustrace: probit model	163
9.5	Uspořádaný probit model	164
9.6	Multinomiální probit model	166
9.6.1	Empirická ilustrace: Multinomiální probit model	171
9.7	Rozšíření probit modelů	171
9.8	Další rozšíření	172
9.9	Shrnutí	174
<b>10</b>	<b>Flexibilní modely</b>	<b>175</b>
10.1	Bayesovská neparametrická a semiparametrická regrese	176
10.1.1	Přehled	176
10.1.2	Parciální lineární model	177
10.1.3	Aditivní verze parciálního lineárního modelu	182
10.2	Kompozice (mixture) normálních modelů	182
10.2.1	Přehled	182
10.2.2	Věrohodnostní funkce	182
10.2.3	Apriorní hustota	182
10.2.4	Bayesovský výpočet	182
10.2.5	Porovnání modelů: Informační kritéria	182
10.2.6	Empirická ilustrace: Kompozice normálních modelů	182
10.3	Rozšíření a alternativní přístupy	182
10.4	Souhrn	182
<b>11</b>	<b>Bayesiánské průměrování modelů</b>	<b>183</b>
<b>12</b>	<b>Další modely, metody a otázky bayesiánské ekonometrie</b>	<b>185</b>
12.1	Úvod	185
12.2	Další metody	186
12.3	Další otázky	188
12.3.1	Identifikace	188
12.4	Další modely	188
12.4.1	Modely časových řad	188
12.4.2	Endogenita, výběr vzorku a další otázky	188
12.4.3	Modely s nestandardními vysvětlovanými proměnnými	188
12.4.4	Strukturální modely	188
12.4.5	Bayesovské neparametrické metody	188
<b>A</b>	<b>Úvod do maticové algebry</b>	<b>189</b>

---

<b>B Úvod do pravděpodobnosti a matematické statistiky</b>	<b>195</b>
B.1 Základy pravděpodobnosti . . . . .	195
B.2 Běžná rozdělení pravděpodobnosti . . . . .	199
B.3 Úvod do teorie výběru . . . . .	201
B.4 Další užitečné teoremy . . . . .	202
<b>C Užitečné funkce v Matlabu</b>	<b>203</b>
C.1 Základní příkazy Matlabu . . . . .	203
C.2 Generátory náhodných čísel . . . . .	203
C.3 Ekonometrický toolbox . . . . .	204
<b>Literatura</b>	<b>206</b>





# Seznam tabulek

1.1	Přehled software pro bayesiánskou ekonometrii. . . . .	9
2.1	Apriorní a posteriorní charakteristiky parametru $\beta$ . . . . .	28
2.2	Apriorní a posteriorní charakteristiky parametru $h$ . . . . .	29
2.3	Apriorní a posteriorní charakteristiky pro model jen s úrovníovou konstantou . . . . .	31
3.1	Apriorní a posteriorní střední hodnoty pro $\beta$ (směrodatné odchyly v závorkách) . . . . .	51
3.2	Apriorní a posteriorní střední hodnoty pro $h$ (směrodatné odchyly v závorkách) . . . . .	51
3.3	Porovnání modelů zahrnující parametr $\beta$ . . . . .	52
3.4	Posteriorní výsledky pro parametr $\beta_2$ spočítané alternativním způsobem . . . . .	55
4.1	Apriorní a posteriorní střední hodnoty pro parametr $\beta$ (směrodatné odchyly v závorkách) . . . . .	70
4.2	Posteriorní výsledky pro parametr $\beta$ . . . . .	78
5.1	Posteriorní výsledky na základě dvou M-H algoritmů . . . . .	98
6.1	Posteriorní výsledky pro $\beta$ , $h$ a $\alpha$ . . . . .	110



# Seznam obrázků

2.1	Marginální apriorní a posteriorní hustoty pro parametr $\beta$ . . . . .	29
3.1	Predikční hustota ceny domu (přirozeně konjugovaná apriorní hustota). . . . .	54
4.1	Predikční hustota ceny domu (Gibbsův vzorkovač). . . . .	71
5.1	Posteriorní predikční hustota šikmosti. . . . .	100
5.2	Posteriorní predikční hustota špičatosti. . . . .	101



# Předmluva

Tato práce je podkladovým materiálem k předmětu Bayesiánská analýza, ve svých počátcích vyučovaným na Ekonomicko-správní fakultě pod názvem Regresní analýza. Obsahově i strukturou vychází z vynikající knihy Garyho Koopa – Bayesian econometrics [19] a jedná se tak o jakýsi neoficiální překlad.

Verze pro rok 2011 by měla dále rozšířit a opravit původní text upravovaný od roku 2007, a to v několika ohledech. Prvním z nich je doplnění kapitoly věnované flexibilním modelům, bayesovskému průměrování modelů a dalším vybraným problémům bayesovské ekonometrie. Doplněny budou rovněž části věnované empirickým příkladům, v první fázi minimálně v rozsahu, jaký nabízí Gary Koop.

Postupně bych rád rozšířil a doplnil jednotlivé kapitoly o další pasáže, které v Koopově knize najít nelze, případně jsou tam uváděny v redukované podobě. Jedná se zejména o vybraná témata z neméně zajímavé učebnice Tonyho Lancastera [21]. Dalším příkladem by mohla být specifická problematika konvergenčních diagnostik.

Protože je bayesiánská ekonometrie nerozlučitelně spjata s využitím počítačů, rád bych rozšířil část přílohy věnované popisu užitečných funkcí a toolboxů pro Matlab. Práce je to však náročná zejména z časového hlediska, přesto pevně věřím, že se mi alespoň většinu z plánovaných aktivit podaří uskutečnit.



# Kapitola 1

## Základní principy a pojmy Bayesiánské ekonometrie

Ekonometrii lze definovat nejlépe jako samostatnou vědní disciplínu, která v sobě propojuje a rozšiřuje zejména poznatky ekonomické teorie, matematické ekonomie, ekonomické statistiky a matematické statistiky. Ekonometrie dodává ekonomické teorii empirický rozměr, přičemž obvykle využívá matematickou formulaci ekonomického problému, což je hlavní náplní matematické ekonomie. Každá empirická analýza je založena na reálných naměřených datech, které poskytuje ekonomická statistika. Role ekonometra v tomto případě nespočívá jen v pasivním přejímání statistických dat a ukazatelů, ale i v jejich aktivní analýze a v jejich pečlivém výběru pro účely praktického modelování. Příkladem může být ekonomický pojem “úroková míra” (použitý v rámci formulace nějaké hypotézy vycházející z ekonomické teorie), pro který lze nalézt desítky různých statistických ukazatelů. Poznatky matematické statistiky jsou následně základem ekonometrických technik a nástrojů, které jsou obvykle vytvářeny a voleny s ohledem na specifický charakter ekonomických dat a modelů.

V tomto textu se budeme zabývat jedním z možných přístupů k ekonometrické analýze, a to bayesovským přístupem. Tento přístup je hojně využíván i v jiných vědních disciplínách jakými jsou biologie či medicína. Ačkoliv se tedy budeme zabývat bayesiánskou<sup>1</sup> ekonometrií, prezentované principy (a mnohdy i modely) mají univerzální platnost.

### 1.1 Bayesiánská teorie

Bayesiánská ekonometrie (či obecněji analýza) je založena na několika jednoduchých zákonech pravděpodobnosti. Díky tomu je bayesovský přístup charakterizován vysokou mírou univerzálnosti. Ať už budeme chtít odhadovat parametry nějakého modelu, porovnávat různě specifikované modely nebo snad vytvářet

---

<sup>1</sup>V dalším textu je užíváno zcela rovnocenně přídavné jméno „bayesovský“ i „bayesiánský“.

předpovědi, všechny tyto aktivity v sobě budou zahrnovat jednotnou aplikaci několika zákonů pravděpodobnosti.

Pro ilustraci bayesiánského principu předpokládejme, že máme dvě spojité náhodné veličiny  $A$  a  $B$ . Jedno ze základních pravidel pravděpodobnosti nám říká, že

$$p(A, B) = p(A|B)p(B),$$

kde  $p(A, B)$  je simultánní či sdružená hustota pravděpodobnosti,  $p(A|B)$  je podmíněná hustota pravděpodobnosti náhodné veličiny  $A$  za předpokladu realizace  $B$  a  $p(B)$  je marginální hustota pravděpodobnosti. Toto pravidlo samozřejmě platí i pro diskrétní náhodné veličiny, v jejichž případě stačí hustotu pravděpodobnosti nahradit pravděpodobnostní funkcí. Logicky samozřejmě platí i vztah

$$p(A, B) = p(B|A)p(A).$$

Bayesovo pravidlo, které je alfou a omegou bayesiánské ekonometrie, vyplývá z rovnosti výše uvedených výrazů pro sdruženou hustotu pravděpodobnosti  $p(A, B)$ :

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}. \quad (1.1)$$

Pro ekonometrii a ekonomii vůbec je charakteristická práce s modely. Typickým zástupcem ekonometrického modelu je regresní model, který můžeme chápat jako určitou formu nějakého obecnějšího ekonomického modelu. V rámci něho nás zajímají koeficienty resp. parametry tohoto modelu, které obvykle chceme odhadnout. Označme si tedy písmenem  $y$  vektor či matici dat a písmenem  $\theta$  vektor či matici parametrů modelu, kterým se snažíme vysvětlit chování veličiny  $y$ . Chceme se tedy dozvědět něco o parametrech  $\theta$  na základě dostupných dat  $y$ . Bayesiánská ekonometrie pro tento účel využívá právě Bayesova pravidla. Stačí, pokud ve vztahu (1.1) nahradíme náhodnou veličinu  $B$  parametry  $\theta$  a náhodnou veličinu  $A$  získanými daty obsaženými ve vektoru (či matici)  $y$ . Získáme tak vztah:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \quad (1.2)$$

Právě podmíněná hustota pravděpodobnosti  $p(\theta|y)$  je v centru zájmu bayesiánců, neboť v sobě obsahuje odpověď na otázku: „Co můžeme říci o parametrech  $\theta$ , při nám známých datech?“. Chápání parametrů  $\theta$  jako náhodné veličiny<sup>2</sup> odlišuje bayesiánský přístup od přístupu klasické ekonometrie (v anglické literatuře je přesnější označení *frequentist econometrics*). Bayesiánská ekonometrie je založena na subjektivním chápání pravděpodobnosti. V tomto kontextu tedy budeme brát jako fakt, že tento přístup nám umožňuje poznat něco neznámého (např. regresní koeficienty) za předpokladu, že něco známe (např. data), přičemž nejlepší souhrnnou charakteristikou takového přístupu k poznání je koncept podmíněné pravděpodobnosti.

<sup>2</sup>Všimněme si, že pro korektnost předchozího vztahu zde logicky musí dojít k nahrazení jedné náhodných veličin náhodnými veličinami druhými.



Pokud nás zajímá skutečně jen to, dozvědět se něco o parametrech  $\theta$ , můžeme v rovnici (1.2) ignorovat člen  $p(y)$ , neboť ten v sobě  $\theta$  nezahrnuje. Můžeme tedy psát:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (1.3)$$

Podmíněná hustota pravděpodobnosti  $p(\theta|y)$  se nazývá posteriorní hustota (či posteriorní hustota pravděpodobnosti – *posterior density*), funkce hustoty pravděpodobnosti (v angličtině *propability density function* – *p.d.f*)  $p(y|\theta)$  je tzv. věrohodnostní funkce (*likelihood function*) a výraz  $p(\theta)$  je tzv. apriorní hustota (*prior density*). Tento vztah se obvykle čte tak, že „posteriorní hustota je proporcionalní součinu věrohodnostní funkce a apriorní hustoty.“

Apriorní hustota či zkráceně „prior“  $p(\theta)$  v sobě obsahuje informaci o parametrech  $\theta$  nezávislou na datech. Jinými slovy v sobě obsahuje vše, co o  $\theta$  víme před tím, než se podíváme na data. Apriorní hustota je jednou z kontroverzních věcí bayesovské ekonometrie. V dalších kapitolách bude probírána otázka tzv. informativních a neinformativních priorů (apriorních hustot). Existují i bayesiánské empirické metody, které využívají informaci obsaženou v datech pro stanovení apriorní hustoty, čímž ovšem poněkud narušují základní premisu bayesiánství. Nicméně, vzhledem k úspěchům v praxi je tento druh metod čím dál populárnější a používanější.

Věrohodnostní funkce  $p(y|\theta)$  je podmíněnou hustotou pravděpodobnosti dat. V tomto kontextu je chápána jako proces, kterým jsou data generována. Např. v lineárním regresním modelu často předpokládáme, že náhodná složka má normální rozdělení, což dále implikuje i to, že  $p(y|\theta)$  bude mít podobu odpovídající normálnímu rozdělení, jehož konkrétní podoba je závislá na hodnotě regresních koeficientů a variabilitě náhodné složky.

Posteriorní hustota či zkráceně „posterior“  $p(\theta|y)$  v sobě zahrnuje informaci o tom, co vše víme o parametrech  $\theta$ , když jsme vzali v úvahu data. Rovnici (1.3) je tedy možno chápat jako „updating rule“, tedy jakési pravidlo, které nám říká, jak aktualizovat naši apriorní informaci o parametrech  $\theta$  na základě její konfrontace s daty. Posteriorní hustota je tak kombinací datové i nedatové informace.

V řadě aplikací nás může zajímat porovnání různých modelů, což lze uplatnit v rámci testování alternativních ekonomických hypotéz. Model je formálně definován věrohodnostní funkcí a apriorní hustotou. Předpokládejme, že máme  $m$  různých modelů  $M_i$  pro  $i = 1, \dots, m$ , které se snaží vysvětlit chování  $y$ .  $M_i$  závisí samozřejmě na parametrech  $\theta^i$ . Posteriorní hustotu tak můžeme zapsat jako

$$p(\theta^i|y, M_i) = \frac{p(y|\theta^i, M_i)p(\theta^i|M_i)}{p(y|M_i)}, \quad (1.4)$$

kdy indexy  $i$  využíváme proto, abychom odlišili prior, věrohodnostní funkci a posterior pro každý z jednotlivých modelů.

Logika bayesovské ekonometrie nás vede k užití Bayesova pravidla pro odvození pravděpodobnostní výpovědi o tom co neznáme (např. zda-li model je korektní či ne) za předpokladu platnosti toho, co známe (např. data). Tím je myšleno to, že můžeme použít posteriorní pravděpodobnost modelu k vyjádření

míry podpory náležející modelu  $M_i$ . Stačí tedy opět dosadit ve vztahu (1.1)  $B = M_i$  a  $A = y$ , abychom získali:

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)}. \quad (1.5)$$

V této rovnici označuje výraz  $p(M_i)$  apriorní pravděpodobnost modelu, což je měřítko toho, jakou pravděpodobnost tomuto modelu přisuzujeme bez znalosti dat. Výraz  $p(y|M_i)$  je tzv. marginální věrohodnost (marginal likelihood), kterou můžeme získat ze vztahu (1.4), pokud integrujeme obě její strany podle  $\theta^i$  a pokud využijeme známou skutečnost, že  $\int p(\theta^i|y, M_i)d\theta^i = 1$ . Po úpravě tak dostáváme:

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i|M_i)d\theta^i. \quad (1.6)$$

V dalších kapitolách uvidíme, jak prakticky marginální věrohodnost spočítáme.

Výraz ve jmenovateli rovnice (1.5) není snadné přímo spočítat, proto obvykle porovnáváme dva modely,  $i$  a  $j$ , pomocí tzv. posteriorního podílu šancí (posterior odds ratio),  $PO_{ij}$ , což je vzájemný podíl posteriorních pravděpodobností porovnávaných modelů:

$$PO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_j)p(M_j)}. \quad (1.7)$$

Mnohdy je možné spočítat posteriorní podíl přímo. Pokud tento podíl spočítáme pro každou dvojici modelů a budeme-li předpokládat, že naše množina modelů je úplná (tj.  $p(M_1|y) + p(M_2|y) + \dots + p(M_m|y) = 1$ ), lze tímto spočítat i samotnou posteriorní pravděpodobnosti modelů. Např. pokud máme dva modely ( $m = 2$ ), pak lze využít rovnice

$$p(M_1|y) + p(M_2|y) = 1$$

a

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

abychom získali

$$p(M_1|y) = \frac{PO_{12}}{1 + PO_{12}}$$

a

$$p(M_2|y) = 1 - p(M_1|y).$$

Obvykle přisuzujeme všem modelům stejnou apriorní váhu. V takovémto případě nám přejde posteriorní podíl šancí na podíl mezních věrohodností modelů, jenž je označován jako **Bayesův faktor**,  $BF_{ij}$ :

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)}. \quad (1.8)$$

Poslední otázkou, které budeme na tomto místě věnovat pozornost, je problematika předpovědí. Zajímá nás tedy, jak na základě pozorovaných dat  $y$

můžeme předpovídat budoucí nepozorované hodnoty  $y^*$ . Bayesovský přístup vede k tomu, že je třeba uvažovat o neurčitosti  $y^*$  v kontextu podmíněné pravděpodobnosti. Předpověď je tak založena na tzv. predikční hustotě pravděpodobnosti (*predictive density*),  $p(y^*|y)$ , resp. v případě více alternativních modelů na predikční hustotě  $p(y^*|y, M_i)$ . Tuto predikční hustotu můžeme zapsat do běžnějšího vyjádření s využitím několika jednoduchých pravidel pro hustoty pravděpodobnosti. Jedním z nich je to, že marginální hustotu pravděpodobnosti lze získat ze sdružené hustoty pravděpodobnosti prostou integrací:

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

Člen uvnitř integrálu lze dále rozepsat, čímž získáme obvyklé vyjádření predikční hustoty za pomoci posteriorní hustoty pravděpodobnosti:

$$p(y^*|y) = \int p(y^*|y, \theta)p(\theta|y) d\theta. \quad (1.9)$$

Tímto jsme si představili základní teoretické pojmy a principy bayesovské ekonometrie. Pokud jednou přijmeme, že neznámé charakteristiky (např.  $\theta$ ,  $M_i$  a  $y^*$ ) jsou náhodné veličiny, dostáváme se k vnitřně logickému a nadále již nekontroverznímu bayesovskému přístupu, který využívá matematicky ověřená pravidla pravděpodobnosti pro statistickou analýzu zkoumaných charakteristik. V rámci každého modelu je třeba mít na paměti, že bayesiánská ekonometrie vyžaduje výběr apriorní hustoty a věrohodnostní funkce, které jsou dále využity k formulaci posteriorní hustoty (1.3). Zdůrazníme tedy, že posteriorní hustota je základem analýzy neznámých parametrů modelu. Pro porovnání více modelů je možno použít posteriorní pravděpodobnosti modelu (1.5), posteriorní podíl šancí (1.7) nebo Bayesův faktor (1.8), k jejichž výpočtu je obvykle třeba spočítat marginální věrohodnosti modelů (1.6). Předpověď je pak založena na predikční hustotě pravděpodobnosti obvykle získávané z (1.9). Ač se to může zdát k neuvěření, těchto několik rovnic je možno využít ke statistické analýze jakýchkoliv představitelných aplikací.

## 1.2 Bayesiánské výpočetní postupy

Na ekonometrické půdě byl bayesiánský přístup přes svou teoretickou a konceptuální eleganci dlouho mimo hlavní pozornost. Důvody pro to byly v zásadě dvojího druhu: problematika apriorní informace a problém výpočetní. První problém se tak týká víceméně filozofické debaty užití "subjektivní" apriorní informace v rámci "objektivní" ekonomické vědy. Argumentem zde nicméně je to, že v ekonometrii jako takové v podstatě celý proces tvorby modelů v sobě zahrnuje širokou škálu nedatových informací, ať již jde např. o rozhodnutí jaký typ modelu použít, jaké proměnné zahrnout, jaká zvolit kritéria pro porovnání modelů či odhad parametrů apod. Z tohoto ohledu je bayesiánský přístup zcela čistý a rigorózní v tom, jak s takovýmto typem nedatové informace pracovat. Navíc, pokud je k dispozici apriorní informace, měla by být užita, neboť větší množství

informací je vždy preferováno před menším. Poslední obranou je i skutečnost, že bayesiánství “vyvinulo” pro celou řadu modelů neinformativní priory, což dává člověku volnost v rozhodnutí, zda-li informativní apriorní hustotu využít nebo nevyužít.

Problém výpočetní náročnosti tak hrál historicky mnohem důležitější roli a byl překonán až v průběhu posledních dvaceti let, což je spojeno s rozvojem výpočetní techniky. Praktické využití bayesiánských metod totiž vyžaduje aktivní zapojení počítačů. Třeba takové rovnice pro porovnání modelů a predikci v sobě přímo či nepřímo obsahují integrály. Jen v ojedinělých případech existuje jejich analytické řešení. Obvykle tak pro vyhodnocení těchto integrálů potřebujeme počítač, s čímž souvisí rozvoj celé řady výpočetních algoritmů. Samotná rovnice definující posteriorní hustotu vyhodnocení integrálů nevyžaduje, nicméně prezentace informací o parametrech si již nějaké to počítání vynucuje. Hustota  $p(\theta|y)$  shrnuje vše, co můžeme říct o parametrech  $\theta$  po shlédnutí dat. Prezentovat veškerou informaci o  $p(\theta|y)$  je možné pouze v ojedinělých případech, kdy se jedná o posteriorní hustotu v jednoduché formě nebo vektor  $\theta$  je jednorozměrný. V tomto případě si můžeme např. vykreslit posteriorní hustoty pravděpodobnosti. Obecně nám však jde o prezentaci různých numerických souhrnů informace obsažené v posteriorní hustotě. A zde se opět dostáváme k integrování. Je tak např. žádoucí prezentovat bodový odhad parametru  $\theta$ . Bayesovský přístup řeší volbu takového odhadu v rámci teorie rozhodování (decision theory). Pro naše potřeby je dostačující vědět, že vcelku intuitivní bodové odhady jako průměr (střední hodnota), medián či modus posteriorní hustoty je možno odvodit právě v kontextu této teorie rozhodování.

Předpokládejme, že chceme jako bodový odhad využít posteriorní střední hodnotu (posterior mean) a předpokládejme dále, že  $\theta$  je  $k$ -prvkový vektor,  $\theta = (\theta_1, \dots, \theta_k)'$ . Posteriorní střední hodnota je tedy definována jako:

$$E(\theta_i|y) = \int \theta_i p(\theta|y) d\theta. \quad (1.10)$$

Až na několik výjimek nelze tento integrál řešit analyticky. Vhodné je rovněž prezentovat míru stupně neurčitosti spojenou s bodovým odhadem. Obvyklým měřítkem je v tomto případě posteriorní směrodatná odchylka, tedy odmocnina posteriorního rozptylu, který lze vypočítat jako

$$\text{var}(\theta_i|y) = E(\theta_i^2|y) - \{E(\theta_i|y)\}^2,$$

což vyžaduje vyhodnocení integrálu v (1.10) v podobě

$$E(\theta_i^2|y) = \int \theta_i^2 p(\theta|y) d\theta.$$

Dle kontextu řešeného problému je možné prezentovat jiné vlastnosti posteriorního rozdělení, např. zda-li některý z parametrů je kladný. To znamená opět vyhodnocení intergálu, tentokrát ve formě:

$$p(\theta_i \geq 0|y) = \int_0^\infty p(\theta|y) d\theta.$$

Všechny tyto posteriorní výpočty lze shrnout do jednotného vztahu:

$$E(g(\theta)|y) = \int g(\theta)p(\theta|y)d\theta, \quad (1.11)$$

kde  $g(\theta)$  je funkce, která nás zajímá. Např.  $g(\theta) = \theta_i$ , pokud počítáme posteriorní střední hodnotu pro  $\theta_i$ . Podobně bude  $g(\theta) = 1(\theta_i \geq 0)$  v případě, kdy počítáme pravděpodobnost, že  $\theta_i$  je kladné. Zde vystupuje  $1(A)$  jako indikační funkce, která je rovna jedné pokud podmínka  $A$  platí a nule v ostatních případech. Predikční hustota spadá také do tohoto obecného zápisu, neboť můžeme položit  $g(\theta) = p(y^*|y, \theta)$ . Jedněmi z výjimek, které nespádají pod tento obecný zápis jsou marginální věrohodnost a kvantily posteriorní hustoty.

V rámci dalšího výkladu bude střední hodnota  $E[g(\theta)|y]$  v rámci všech diskutovaných modelů existovat, podobně jako funkce  $g(\cdot)$ . Pro některé modely však existovat nemusí. Příkladem je Cauchyho rozdělení, což je  $t$ -rozdělení s jedním stupněm volnosti, pro které střední hodnota neexistuje. Při využití bayesiánských metod v rámci nového modelu tedy může být důležité existenci střední hodnoty  $E[g(\theta)|y]$  ověřit. V případě, kdy  $p(\theta|y)$  odpovídá hustotě pravděpodobnosti, však budou vždy existovat kvantily této hustoty. Nejsme-li se jistí existencí  $E[g(\theta)|y]$ , je možné vždy prezentovat posteriorní informaci na kvantilovém základě (např. medián nebo mezikvartilové rozpětí).

V ojedinělých případech je možné výraz (1.11) vyhodnotit analyticky. Obecně je však potřeba využít výpočetní techniku. V moderní bayesiánské ekonometrii je převažujícím přístupem posteriorní simulace. Existuje celá řada simulátorů a všechny jsou aplikací či rozšířením zákona velkých čísel a centrální limitní věty. Jedním z důsledků zákona velkých čísel je:

**Teorém 1.1** (Monte Carlo integrace). Nechť  $\theta^{(s)}$  pro  $s = 1, \dots, S$  je náhodný výběr (vzorek) z  $p(\theta|y)$  a definujme

$$\hat{g}_S = \frac{1}{S} \sum_{s=1}^S g(\theta^{(s)}). \quad (1.12)$$

Potom  $\hat{g}_S$  konverguje k  $E[g(\theta)|y]$  pro  $S$  jdoucí k nekonečnu.

V praxi je tím myšleno to, že pokud počítač dokáže generovat náhodný výběr z posteriorního rozdělení, můžeme aproximovat  $E[g(\theta)|y]$  jednoduchým průměrováním funkce, která nás zajímá, a to vyhodnocené v rámci tohoto náhodného výběru (vzorku). Tvorba náhodných výběrů či vzorkování z posteriorní hustoty je právě nazývána posteriorní simulací, kdy  $\theta^{(s)}$  je příslušný výběr (draw) či replikace (replication). Výše uvedený teorém je v podstatě nejjednodušším posteriorním simulátorem (tzn. generujeme náhodné výběry z posteriorního rozdělení), přičemž Monte Carlo integrace je způsob jak aproximovat střední hodnotu  $E[g(\theta)|y]$ .

Chyba aproximace by byla nulová pouze v případě, kdy by  $S$  bylo nekonečno. Můžeme si tak zvolit libovolné  $S$ , kdy samozřejmě s růstem  $S$  se zvyšují výpočetní nároky kladené na počítač. Je mnoho způsobů jak měřit chybu aproximace spojenou s konkrétní hodnotou  $S$ . Některé z nich jsou založeny na rozšíření

centrální limitní věty. Pro případ Monte Carlo integrace centrální limitní věta implikuje následující teorém:

**Teorém 1.2** (Numerická standardní chyba). Při využití definice a postupu z Teorému 1.1 platí

$$\sqrt{S} \{\hat{g}_S - E[g(\theta)|y]\} \rightarrow N(0, \sigma_g^2), \quad (1.13)$$

pro  $S$  jdoucí k nekonečnu, přičemž  $\sigma_g^2 = \text{var}[g(\theta)|y]$ .

Teorém 1.2 lze využít pro odhad chyby aproximace při Monte Carlo integraci s využitím vlastností normálního rozdělení. Skutečnost, že standardizované normální rozdělení má 95 % své hustoty uvnitř intervalu definovaném 1.96 násobkem směrodatné odchylky od střední hodnoty, nás vede tedy k tomu, že:

$$\Pr \left[ -1.96 \frac{\sigma_g}{\sqrt{S}} \leq \hat{g}_S - E[g(\theta|y)] \leq 1.96 \frac{\sigma_g}{\sqrt{S}} \right] = 0.95.$$

Volbou  $S$  lze zajistit, aby chyba odhadu  $\hat{g}_S - E[g(\theta|y)]$  bylo dostatečně malá, a to s vysokou pravděpodobností. V praxi je  $\sigma_g$  neznámé, ovšem Monte Carlo integrace nám umožňuje její aproximaci. Výraz  $\frac{\sigma_g}{\sqrt{S}}$  se označuje jako numerická standardní chyba a lze ji interpretovat jako míru chyby aproximace. Teorém 1.2 tedy implikuje, že máme-li např. vzorek velikosti  $S = 10000$ , pak numerická standardní chyba je 1 % posteriorní směrodatné odchylky.

Problémem je, že málokdy je možno Monte Carlo integraci tak snadno provést. Existují algoritmy pro náhodný výběr z mnoha obvyklých rozdělení (normální, chí-kvadrát), i když přesněji řečeno, tyto počítačové algoritmy nejsou z formálního výběru náhodné generátory, ale spíše pseudo-náhodné. V řadě modelů však posteriorní hustoty nemají žádnou z obvyklých a známých forem. V takovýchto případech je nutné využít jiných posteriorních simulátorů, jejichž vytvoření je úkolem náročnějším. K tomu se ale v dalších kapitolách samozřejmě vrátíme.

### 1.3 Software pro Bayesiánskou analýzu

Existuje řada počítačových programů a balíčků, které jsou užitečné pro bayesiánskou analýzu v rámci určité třídy modelů. Bayesiánská ekonometrie je však výpočetně mnohem náročnější než klasická ekonometrie. Pro nástroje a techniky klasické ekonometrie existuje celá řada programů, ve kterých uživatel může jednoduše klikat na odpovídající ikonky či nabídky a provést tak požadovaný ekonometrický postup (odhad parametrů modelu, diagnostický test apod.). Lze namítnout, že tato výhoda je i nevýhodou v tom smyslu, že podporuje ekonometry v používání jednoduše jen těch technik, které jsou v daném programu obsaženy. To může mnohdy svádět k jednoduché prezentaci výsledků odhadů, testových statistik a diagnostických testů, které nám program vrací, ale bez ohledu na to, jestli se jedná o adekvátní použití pro aplikaci či

problém, který řešíme. Bayesiánská analýza nás naopak nutí přemýšlet v intencích modelů (tzn. věrohodnostní funkce a apriorní hustoty), které jsou vhodné pro empirické otázky, které potřebujeme řešit. Nepřeborná řada možných věrohodnostních funkcí a apriorních hustot tak ztěžuje vytvoření bayesovského programu s širokým použitím. Z tohoto důvodu si bayesovští ekonometři vytvářejí vlastní programy a skripty v maticových programovacích jazycích jako je např. MATLAB, Gauss nebo Ox. Tento postup není zase tak obtížný jak by se mohlo zdát. Psát si vlastní program je navíc velmi dobrým způsobem jak se donutit plně porozumět podstatě daných ekonometrických postupů. V našem případě budeme řešit příklady s využitím MATLABU, který je pravděpodobně nejběžněji používaným počítačovým jazykem pro bayesovskou ekonometrii a statistiku. Pravdou je, že se jedná o komerční software, nicméně volně dostupnou variantou tohoto produktu je Octave (používající stejný syntax) a rozšířeným programovacím jazykem podobným Matlabu (s podobným syntaxem) je pak program R a jeho balíčky. Přehled programů a nástavbových balíčků nabízí tabulka 1.1.

**Tabulka 1.1:** Přehled software pro bayesiánskou ekonometrii.

Typ - Název	Výrobce	Poznámka
<i>Open source</i>		
Octave	University of Wisconsin	„analogie“ Matlabu
R	The R Project	různé knihovny
BUGS/WinBUGS	The BUGS Project	
<i>Komerční</i>		
GAUSS	Aptech Systems, Inc.	
Matlab	MathWorks, Inc.	
<i>Balíčky</i>		
BACC	John Geweke	i pro Matlab
DYNARE	M. Juillard	makromodelování
Econometrics Toolbox	James P. LeSage	Matlab, Octave

Programy Octave, R a Matlab jsou v podstatě systémy s vlastním programovacím jazykem (Octave a Matlab ho mají totožný) navržené pro složité numerické výpočty. Disponují celou řadou implementovaných či externě stahitelných toolboxů či pluginů (což jsou balíčky již hotových a naprogramovaných funkcí pro různé účely použití). Matlab tak má k dispozici implementovaný statistický toolbox (funkce pro základní regresní techniky a statistiku), ekonometrický toolbox (pro práci s časovými řadami) a optimalizační toolbox (nástroje pro optimalizaci). Zcela volně je pro Matlab dostupný tzv. **Econometrics Toolbox**, z části naprogramován a spravován Jamesem P. LeSagem [22], který obsahuje velké množství funkcí a nástrojů pro odhad široké škály ekonometrických modelů. Výhodou tohoto typu programů je však obrovská flexibilita, kdy je možné naprogramovat si funkce pro řešení i těch nejsložitějších problémů.

Zajímavým nástrojem pro makroekonomické modelování (využívající i klasické metody odhadu) je Juillardův *Dynare* [17].

Ale ani člověk, který si nechce zdokonalit své dovednosti v programování nemusí zoufat. Existují balíčky z naprogramovanými funkcemi pro standardní třídy modelů. Příkladem může být program BUGS, což je akronym pro Bayesian Inference Using Gibbs Sampling), využívající techniku Gibbsova vzorkovače (s níž se blíže seznámíme). Jiným relevantním balíčkem je BACC (Bayesian Analysis, Computation and Communication), který opět pokrývá velkou spoustu obvykle používaných modelů. Jeho snadné využití je pomocí dynamické knihovny pro Matlab (lze jej tak využívat prostřednictvím Matlabu). V rámci bayesovské analýzy se uplatní i nástroje a funkce LeSageho ekonometrického toolboxu, zmiňovaného výše. Existuje spousta dalších programků a balíčků pro bayesovskou analýzu, ty jsou však obvykle více směřovány na statistiky než ekonometri. Carlin a Louis [2] ve své příloze nabízejí přehled těchto programů a balíčků.

## 1.4 Shrnutí

V této kapitole jsme si představili obecně všechny základní pojmy a principy bayesiánské ekonometrie. Velkou výhodou bayesovského přístupu je to, že jsme schopni vložit veškerou obecnou teorii do jedné kapitoly, a to s využitím základních principů pravděpodobnosti. Základními stavebními kameny bayesovského přístupu jsou *věrohodnostní funkce (likelihood)* a *apriorní hustota pravděpodobnosti (prior)*. Jejich součinem definujeme *posteriorní hustotu pravděpodobnosti (posterior)*, která tvoří základ analýzy neznámých parametrů modelu. Různé modely lze porovnat s využitím *posteriorních pravděpodobností modelu (posterior model probabilities)*, které vyžadují výpočet *marginálních věrohodností (marginal likelihood)*. Předpověď je založena na *predikční hustotě pravděpodobnosti (predictive density)*. V naprosté většině případů nejsme schopni pracovat s těmito základními stavebními bloky v analytických tvarech. Proto zde na scénu přichází bayesovský způsob výpočtu, jehož dominantní metodou je *posteriorní simulace*.

Následující kapitoly se věnují konkrétním typům modelů a ukazuje se v nich, jak doposud představené abstraktní koncepty jsou zosobněny v konkrétním kontextu. Logika bayesiánské ekonometrie představená v této kapitole se promítá i do struktury následujících kapitol. Kapitoly tak obvykle začínají věrohodnostní funkcí a apriorní hustotou. Následně je odvozena posteriorní hustota spolu s výpočetními metodami posteriorní analýzy a porovnání modelů. V tomto smyslu uvažování (tedy věrohodnostní funkce, apriorní hustota, posteriorní hustota a výpočet) je dobré přemýšlet i v rámci práce na vlastních empirických projektech řešících problémy (nejen) ekonomie.



## Kapitola 2

# Normální lineární regresní model s přirozeně konjugovaným priorem a jedinou vysvětlující proměnnou

### 2.1 Úvod

Regresní model patří mezi základní stavební kameny ekonometrie. Tento model předpokládá závislost mezi proměnnou  $y$  (označovanou jako závisle proměnná či vysvětlovaná proměnná) a  $k$  vysvětlujícími proměnnými, označenými jako  $x_1, \dots, x_k$ , ve formě:

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon,$$

kde  $\epsilon$  představuje náhodnou složku resp. regresní chybu a  $x_1$  je obvykle roven jedné, pokud uvažujeme model s tzv. úrovnovou konstantou  $\beta_1$ . Je tedy zřejmé, že očekáváme lineární závislost mezi vysvětlovanou proměnnou a vysvětlujícími proměnnými, proto hovoříme o lineárním regresním modelu (LRM).

Pro maximální zjednodušení je v této kapitole základní koncept a motivace regresního modelu prezentována na příkladu s jedinou vysvětlující proměnnou.

### 2.2 Věrohodnostní funkce

Věrohodnostní funkci můžeme intuitivně chápat jako předpis či stochastické pravidlo, podle kterého jsou generována data. Nechť  $y_i$  a  $x_i$  označují postupně

naměřená (pozorovaná) data vysvětlované a vysvětlující proměnné pro jednotlivá pozorování  $i$ , kdy  $i = 1, \dots, N$ . Máme tedy celkem  $N$  pozorování. Jednotlivá pozorování mohou být brána jako pozorování v čase, pozorování týkající se jednotlivých subjektů (např. firem, států) apod. Pro zjednodušení tedy předpokládáme lineární regresní model ve tvaru:

$$y_i = \beta x_i + \epsilon_i, \quad (2.1)$$

kde  $\epsilon_i$  je náhodná chyba. Náhodná chyba je obvykle zosobněním chyb v měření (v pozorování), možné chyby specifikace modelu (obsahuje tedy nevysvětlenou složku modelu), popř. v moderní dynamické makroekonomii (pracující se systémem interdependentních, tedy vzájemně propojených rovnic) je ekvivalentem stochastických šoků (šoky v technologii, pokud máme rovnici produkční funkce, šoky v poptávce v rovnici poptávky apod.).

Předpoklady kladené na  $\epsilon_i$  a  $x_i$  určují tvar věrohodnostní funkce. Standardní předpoklady (které budou v dalších kapitolách postupně uvolněny) jsou následující:

1.  $\epsilon_i$  má normální rozdělení s nulovou střední hodnotou a rozptylem  $\sigma^2$ , přičemž  $\epsilon_i$  a  $\epsilon_j$  jsou na sobě nezávislé pro  $j \neq i$ . Zkráceně lze zapsat  $\epsilon_i \sim N(0, \sigma^2)$ . V anglickém textu je možno nalézt značení, že  $\epsilon_i$  je i.i.d.  $N(0, \sigma^2)$ , kde i.i.d. je zkráceně „independent and identically distributed“, tedy „nezávisle a stejně rozděleny“.
2.  $x_i$  jsou buď pevně daná čísla (tj. nenáhodné veličiny) nebo pokud se jedná o náhodné veličiny, pak musí být nezávislé vzhledem k  $\epsilon_i$ . Jejich funkce hustoty pravděpodobnosti tak může být označena jako  $p(x_i|\lambda)$ , kde  $\lambda$  je vektor parametrů, které nezahrnují  $\beta$  nebo  $\sigma^2$ .

Předpoklad nenáhodnosti vysvětlujících proměnných nemusí být v ekonomických aplikacích vždy rozumný, nicméně nezávislost jejich případných rozdělení na náhodné složce a parametrech  $\beta$  a  $\sigma^2$  již rozumná je. Lze tak hovořit o tom, že  $x$  je exogenní proměnná (ať náhodná, nebo deterministická). Požadavek normality náhodných složek vysvětluje to, proč hovoříme o normálním lineárním regresním modelu (NLRM).

Věrohodnostní funkce je definována jako funkce sdružené hustoty pravděpodobnosti pro všechna data podmíněná neznámými parametry (viz (1.3)). Pro stručnost značení definujeme vektor pozorované vysvětlované proměnné  $y = (y_1, y_2, \dots, y_N)'$  a vysvětlující proměnné  $x = (x_1, x_2, \dots, x_N)'$ . Věrohodnostní funkce má tedy logicky tvar podmíněné hustoty pravděpodobnosti  $p(y, x|\beta, \sigma^2, \lambda)$ . Druhý předpoklad (o nezávislosti vysvětlující proměnné na parametrech a rozptylu náhodné složky) umožňuje zapsat věrohodnostní funkci následujícím způsobem:

$$p(y, x|\beta, \sigma^2, \lambda) = p(y|x, \beta, \sigma^2)p(x|\lambda)$$

Pokud nás pravděpodobnostní rozdělení  $x$  nezajímá, můžeme pracovat s věrohodnostní funkcí  $p(y|x, \beta, \sigma^2)$ . Pro jednoduchost značení nebude vektor vysvětlujících proměnných  $x$  ve věrohodnostní funkci dále uváděn. Je však třeba mít

na paměti, že regresní model (ať už je s ním pracováno bayesovsky či klasicky) v sobě implicitně zahrnuje práci s rozdělením  $y$  podmíněným  $x$ , a ne sdruženým rozdělením těchto dvou náhodných vektorů.

Předpoklady kladené na náhodnou složku umožňují precizní formulaci věrohodnostní funkce:

- $p(y_i|\beta, \sigma^2)$  je normální,
- $E(y_i|\beta, \sigma^2) = E(\beta x_i + \epsilon_i|\beta, \sigma^2) = \beta x_i$ ,
- $var(y_i|\beta, \sigma^2) = var(\beta x_i + \epsilon_i|\beta, \sigma^2) = \sigma^2$ .

Střední hodnota  $y_i$  vychází z faktu, že střední hodnota náhodné složky je nulová. Rozptyl je ovlivněn pouze rozptylem náhodné složky, protože uvažujeme podmíněnou střední hodnotu, kdy  $\beta_i x_i$  je třeba brát jako pevné hodnoty (jejichž rozptyl je tudíž nulový).

Z definice normálního rozdělení plyne analytický tvar věrohodnostní funkce:

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

Protože pro  $i \neq j$  jsou  $\epsilon_i$  a  $\epsilon_j$  vzájemně nezávislé, znamená to, že  $y_i$  a  $y_j$  jsou nezávislé, tedy sdružená hustota všech pozorování je

$$p(y|\beta, \sigma^2) = \prod_{i=1}^N p(y_i|\beta, \sigma^2)$$

a věrohodnostní funkce je tak dána jako:

$$p(y|\beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \beta x_i)^2\right] \quad (2.2)$$

Pro další odvozování je vhodné přepsat věrohodnostní funkci trochu odlišným způsobem. Lze snadno ukázat platnost následujícího vztahu<sup>3</sup>:

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

kde

$$\nu = N - 1, \quad (2.3)$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}, \quad (2.4)$$

$$s^2 = \frac{\sum (y_i - \hat{\beta} x_i)^2}{\nu}. \quad (2.5)$$

<sup>3</sup>Stačí zapsat  $\sum (y_i - \beta x_i)^2 = \sum \{(y_i - \hat{\beta} x_i) - (\beta - \hat{\beta}) x_i\}^2$ .

V rámci klasické ekonometrie jsou  $\hat{\beta}$ ,  $s^2$  a  $\nu$  odhady získané metodou nejmenších čtverců (OLS - Ordinary Least Squares) postupně pro  $\beta$ , rozptyl reziduí a stupně volnosti. Jsou rovněž tzv. postačujícími statistikami pro (2.2). Mnohdy je jednodušší pracovat místo s rozptylem chyb raději s přesností chyby (error precision). Přesnost chyby je definována jednoduše jako:  $h = \frac{1}{\sigma^2}$ . Věrohodnostní funkci lze tedy přepsat do podoby:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{1}{2}} \exp \left[ -\frac{h}{2} (\beta - \hat{\beta})^2 \sum_{i=1}^N (x_i)^2 \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp \left[ -\frac{h\nu}{2s^2} \right] \right\} \quad (2.6)$$

Je dobré si všimnout, že první člen ve složených závorkách je podobný jádrové hustotě (kernelu) normální hustoty pro  $\beta$  a druhý člen je podobný rozdělení gama pro  $h$ . Vzhledem k tomu, že pojem jádrová hustota je v rámci bayesiánské ekonometrie hojně využíván, nebude od věci si blíže definovat, co to vlastně je.

**Definice 2.1** (Jádrová hustota – kernel). Hustota pravděpodobnosti náhodné veličiny  $X$  má obvykle podobu  $kg(x)$ , kde  $k$  je číselná konstanta, jejíž jedinou rolí je zajistit, aby integrál  $kg(x)$  přes všechna  $x$  byl roven jedné (a jednalo se tak o skutečnou funkci hustoty pravděpodobnosti). Zbylá část výrazu,  $g(x)$ , která zahrnuje  $x$ , se nazývá jádrem (jádrovou hustotou, kernelem) této funkce.

Jako příklad si vezměme třídu hustot pravděpodobnosti beta funkcí. V tomto případě je příslušným jádrem hustoty funkce  $x^{a-1}(1-x)^{b-1}$ , kdy  $k$  je podílem gama funkcí, jež lze pro různé typy rozdělení nalézt v Příloze B. Co je obsahem jádra závisí na tom, co považujeme za proměnné dané hustoty pravděpodobnosti. Pokud nás např. zajímá  $x$ , kernelem funkce normální hustoty pravděpodobnosti  $N(\mu, \sigma^2)$  je funkce  $\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$  a integrační konstantou je zde  $\frac{1}{\sqrt{2\pi}\sigma}$ . Na druhé straně, pokud uvažujeme normální hustotu pravděpodobnosti pro dané  $x$  jako funkci parametrů  $\mu$  a  $\sigma^2$ , je kernelem výraz  $\frac{1}{\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$ . V obou případech nás nezajímá člen  $\frac{1}{\sqrt{2\pi}}$ .

Jak již bylo řečeno, význam  $k$  spočívá v tom, aby hustota pravděpodobnosti integrovala do jedné. Pokud známe jádrovou hustotu, lze tuto konstantu nalézt integrací této hustoty, ovšem málokdy nás tato konstanta zajímá. Protože dokážeme rozpoznat třídu rozdělení podle jejich kernelu, je vcelku obvyklé zanedbávat tyto konstanty při praktických manipulacích s hustotami pravděpodobnosti. V případě Bayesova teorému je to více než zřejmé, neboť se prakticky využívá bez integrační konstanty  $p(y)$ :

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

tedy posteriorní hustota je proporcionální součinu věrohodnostní funkce a apriorní hustoty.

## 2.3 Apriorní hustota

Apriorní hustota pravděpodobnosti (zkráceně prior) v sobě obsahuje informaci, kterou má výzkumník před tím, než se podívá na data. Obvykle se vybírají

určité třídy priorů, které lze snadno interpretovat a zjednodušují výpočet. Přírozně konjugovaný prior má obě tyto přednosti. **Konjugované** apriorní rozdělení je takové, které po kombinaci s věrohodnostní funkcí dává posteriorní hustotu, která spadá do téže třídy rozdělení. **Přírozně konjugované** apriorní rozdělení pak má tu dodatečnou vlastnost, že má dokonce stejnou funkční formu jako věrohodnostní funkce. Důsledkem toho je, že apriorní informace může být interpretována stejným způsobem jako informace obsažená ve věrohodnostní funkci. Jinými slovy lze tento prior interpretovat tak, že pochází z fiktivního datového souboru, který je generován stejným procesem jako reálná data.

V jednoduchém lineárním regresním modelu je třeba specifikovat apriorní (sdruženou) hustotu pro parametry  $\beta$  a  $h$ , označovanou  $p(\beta, h)$ . Již ze zápisu této hustoty je zřejmé, že nezávisí na datech  $y$ . Obvykle se tato hustota zapisuje pomocí podmíněné a marginální hustoty pravděpodobnosti jako  $p(\beta, h) = p(\beta|h)p(h)$ . V takovémto případě pak definujeme apriorní hustotu pro  $\beta|h$  a pro  $h$ . Tvar věrohodnostní funkce ve vztahu (2.6) nasvědčuje tomu, že přírozně konjugovaný prior bude zahrnovat normální rozdělení pro  $\beta|h$  a gama rozdělení pro  $h$ . Tento typ rozdělení, který je součinem rozdělení gama a (podmíněného) normálního rozdělení, je označován jako normální-gama rozdělení. V souladu se značením uvedeným v Příloze B platí, že pokud definujeme:

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V}) \quad h \sim G(\underline{s}^{-2}, \underline{\nu}),$$

získáváme přírozně konjugovanou apriorní hustotu pro  $\beta$  a  $h$  jako

$$\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}). \quad (2.7)$$

Stačí pak vybrat konkrétní hodnoty tzv. apriorních hyperparametrů  $\underline{\beta}$ ,  $\underline{V}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$ . V dalším značení tedy budeme užívat podtržení pro parametry apriorní hustoty a čárku nahoře pro parametry posteriorní hustoty.

## 2.4 Posteriorní hustota

Posteriorní hustota v sobě shrnuje apriorní informaci a datovou informaci o neznámých parametrech. Je proporcionalní součinu věrohodnostní funkce (2.2) a apriorní hustotě (2.7). Lze ukázat, že posteriorní hustota má rovněž normální-gama rozdělení, což samozřejmě odpovídá tomu, že apriorní hustota je skutečně přírozně konjugovaná. To si nyní dokážeme. Posteriorní hustota pravděpodobnosti je tedy proporcionalní součinu apriorní hustoty a věrohodnostní funkce. Z definice normálního-gama rozdělení pro apriorní hustotu pravděpodobnosti a drobnou úpravou věrohodnostní funkce (2.6) spočívající v roznásobení mocnin se společným základem (kdy základ opíšeme a exponenty sečteme) můžeme

psát:

$$\begin{aligned} p(\beta, h|y) &\propto \left\{ h^{\frac{1}{2}} \exp \left[ -\frac{h}{2} (\beta - \underline{\beta})^2 \cdot \underline{V}^{-1} \right] \right\} \cdot \left\{ h^{\frac{\nu-2}{2}} \exp \left[ -\frac{h\nu s^2}{2} \right] \right\} \\ &\times \left\{ h^{\frac{N}{2}} \exp \left[ -\frac{h}{2} \left\{ \nu s^2 + (\beta - \hat{\beta})^2 \sum x_i^2 \right\} \right] \right\} \\ &= h^{\frac{1+\nu-2+N}{2}} \exp \left[ -\frac{h}{2} \left( \underline{\nu} s^2 + \nu s^2 + (\beta - \underline{\beta})^2 \underline{V}^{-1} + (\beta - \hat{\beta})^2 \sum x_i^2 \right) \right]. \end{aligned}$$

Dále můžeme označit  $\bar{\nu} = \underline{\nu} + N$ . Parametr  $\beta$  se nachází pouze v části předchozího exponentu, konkrétně v

$$(\beta - \underline{\beta})^2 \underline{V}^{-1} + (\beta - \hat{\beta})^2 \sum x_i^2,$$

a ten snadno zapíšeme jako

$$(\hat{\beta} - \underline{\beta})^2 \sum x_i^2 \bar{V} \underline{V}^{-1} + (\beta - \bar{\beta})^2 \bar{V}^{-1},$$

kde

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

Ukažme si nyní, že tomu tak opravdu je. V první fázi umocníme obsah závorek a sdružíme členy obsahující parametr  $\beta$ , čímž dostaneme

$$\beta^2 (\underline{V}^{-1} + \sum x_i^2) - 2\beta (\underline{\beta} \underline{V}^{-1} + \hat{\beta} \sum x_i^2) + \underline{\beta}^2 \underline{V}^{-1} + \hat{\beta}^2 \sum x_i^2.$$

Využitím vztahů pro  $\bar{V}$  a  $\bar{\beta}$  lze psát

$$\beta^2 \bar{V}^{-1} - 2\beta \bar{\beta} \bar{V}^{-1} + \underline{\beta}^2 \underline{V}^{-1} + \hat{\beta}^2 \sum x_i^2,$$

kdy po přidání členu  $(\bar{\beta}^2 \bar{V}^{-1} - \bar{\beta}^2 \bar{V}^{-1})$  získáme

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

Dále se můžeme soustředit na druhou část výrazu (po znaménku mínus) a dosadit původní výraz pro  $\bar{\beta}$ , tedy

$$\begin{aligned} & -\bar{\beta}^2 \bar{V}^{-1} + \underline{\beta}^2 \underline{V}^{-1} + \hat{\beta}^2 \sum x_i^2 = -\bar{V} (\underline{V}^{-1} \underline{\beta} + \hat{\beta} \sum x_i^2)^2 + \underline{\beta}^2 \underline{V}^{-1} + \hat{\beta}^2 \sum x_i^2 \\ &= -\bar{V} \left[ (\underline{V}^{-1})^2 \underline{\beta}^2 + 2\underline{\beta} \hat{\beta} \sum x_i^2 \underline{V}^{-1} + \hat{\beta}^2 (\sum x_i^2)^2 \right] + \underline{\beta}^2 \underline{V}^{-1} + \hat{\beta}^2 \sum x_i^2 \\ &= \bar{V} \left[ \underline{\beta}^2 \underline{V}^{-1} \bar{V}^{-1} + \hat{\beta}^2 \sum x_i^2 \bar{V}^{-1} - (\underline{V}^{-1})^2 \underline{\beta}^2 - 2\underline{\beta} \hat{\beta} \sum x_i^2 \underline{V}^{-1} - \hat{\beta}^2 (\sum x_i^2)^2 \right] \\ &= \bar{V} \left[ \underline{\beta}^2 \underline{V}^{-1} \underline{V}^{-1} + \underline{\beta}^2 \underline{V}^{-1} \sum x_i^2 + \hat{\beta}^2 \sum x_i^2 \underline{V}^{-1} + \hat{\beta}^2 (\sum x_i^2)^2 - \right. \\ &\quad \left. - (\underline{V}^{-1})^2 \underline{\beta}^2 - 2\underline{\beta} \hat{\beta} \sum x_i^2 \underline{V}^{-1} - \hat{\beta}^2 (\sum x_i^2)^2 \right] \\ &= \bar{V} \left( \underline{\beta}^2 \underline{V}^{-1} \sum x_i^2 + \hat{\beta}^2 \sum x_i^2 \underline{V}^{-1} - 2\underline{\beta} \hat{\beta} \sum x_i^2 \underline{V}^{-1} \right) \\ &= \bar{V} (\hat{\beta} - \underline{\beta})^2 \sum x_i^2 \underline{V}^{-1}. \end{aligned}$$

Parametr  $\beta$  se po této úpravě nachází již jen ve členu  $(\beta - \bar{\beta})^2 \bar{V}^{-1}$ . Díky tomu je jádrová hustota pro  $\beta|y, h$  dána jako

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

což je jádrová hustota normální hustoty pravděpodobnosti a platí tedy, že

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

Marginální posteriorní hustotu pro  $h$  lze odvodit na základě skutečnosti, že

$$p(h|y) = \int p(\beta, h|y) d\beta = \int p(h|y) p(\beta|y, h) d\beta.$$

Protože každá hustota pravděpodobnosti integrovaná přes všechny meze je rovna jedné, lze tímto vyintegrovat část sdružené posteriorní hustoty pravděpodobnosti zahrnující funkci normální hustoty pravděpodobnosti pro  $\beta$ , tedy  $p(\beta|y, h) \propto h^{\frac{1}{2}} \exp \left[ -\frac{h}{2} (\beta - \bar{\beta})^2 \bar{V}^{-1} \right]$ . Zůstává nám tedy

$$\begin{aligned} p(h|y) &\propto h^{\frac{\bar{\nu}-2}{2}} \exp \left[ -\frac{h}{2} \left\{ \underline{\nu s}^2 + \nu s^2 + (\hat{\beta} - \underline{\beta})^2 \sum x_i^2 \bar{V} \underline{V}^{-1} \right\} \right] \\ &= h^{\frac{\bar{\nu}-2}{2}} \exp \left[ -\frac{h}{2} \bar{\nu s}^2 \right], \end{aligned}$$

kde

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

Výraz pro  $p(h|y)$  je jádrová hustota gama rozdělení, z čehož vyplývá, že

$$p(h|y) \sim G(\bar{s}^{-2}, \bar{\nu}).$$

Protože rozdělení  $\beta|y, h$  je normální a  $h|y$  je gama, je zřejmé, že sdružená hustota rozdělení pro parametry  $\beta$  a  $h$  odpovídá normálnímu-gama rozdělení, v našem značení tedy  $NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ . Apriorní i posteriorní hustota jsou obě z normálního-gama rozdělení, čímž je ověřena konjugovanost apriorní hustoty.

Shrňme si tedy naše poznatky o posteriorní hustotě:

$$\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu}), \quad (2.8)$$

kde

$$\bar{V} = \frac{1}{\underline{V}^{-1} + \sum x_i^2}, \quad (2.9)$$

$$\bar{\beta} = \bar{V}(\underline{V}^{-1} \underline{\beta} + \hat{\beta} \sum x_i^2), \quad (2.10)$$

$$\bar{\nu} = \underline{\nu} + N \quad (2.11)$$

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

$$\bar{\nu}\bar{s}^2 = \underline{\nu}s^2 + \nu s^2 + \frac{(\hat{\beta} - \beta)^2}{\underline{V} + \frac{1}{\sum x_i^2}}, \quad (2.12)$$

což je zcela ekvivalentnímu přechozímu výrazu pro  $\bar{\nu}\bar{s}^2$ , protože

$$\sum x_i^2 \bar{V} \underline{V}^{-1} = \sum x_i^2 \frac{1}{\underline{V}^{-1} + \sum x_i^2} \underline{V}^{-1} = \frac{\sum x_i^2}{1 + \sum x_i^2 \underline{V}} = \frac{1}{\frac{1}{\sum x_i^2} + \underline{V}}.$$

V regresním modelu nás nejvíce zajímají koeficienty  $\beta$  vyjadřující marginální vliv příslušné vysvětlující proměnné na proměnnou vysvětlovanou. Posteriorní střední hodnota  $E(\beta|y)$  je hojně používaným bodovým odhadem a posteriorní rozptyl  $var(\beta|y)$  je obvykle využívanou mírou neurčitosti spojenou s bodovým odhadem. Posteriorní střední hodnotu lze spočítat jako:

$$E(\beta|y) = \iint \beta p(\beta, h|y) dh d\beta = \int \beta p(\beta|y) d\beta.$$

Pro její výpočet nás tedy zajímá marginální posteriorní hustota  $p(\beta|y)$ . V tomto případě ji lze spočítat analyticky s využitím vlastností normálního-gama rozdělení. Využijeme-li skutečnosti, že  $p(\beta, h|y) = \int p(\beta, h|y) dh$ , potom marginální posteriorní hustota pro  $\beta$  odpovídá Studentovu  $t$ -rozdělení:

$$\beta|y \sim t(\bar{\beta}, \bar{s}^2 \bar{V}, \bar{\nu}), \quad (2.13)$$

z čehož vyplývá:

$$E(\beta|y) = \bar{\beta}, \quad (2.14)$$

$$var(\beta|y) = \frac{\bar{\nu}\bar{s}^2}{\bar{\nu} - 2} \bar{V}. \quad (2.15)$$

Důkaz toho, že marginální hustota pravděpodobnosti skutečně odpovídá  $t$ -rozdělení je relativně snadný. Sdružená posteriorní hustota parametrů  $\beta$  a  $h$ ,  $p(\beta, h|y) \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ , má analytické vyjádření

$$\frac{1}{(2\pi)^{\frac{1}{2}}} h^{\frac{1}{2}} \bar{V}^{-\frac{1}{2}} \exp\left[-\frac{h}{2} \bar{V}^{-1} (\beta - \bar{\beta})^2\right] \left[\left(\frac{2\bar{s}^{-2}}{\bar{\nu}}\right)^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{\bar{\nu}}{2}\right)\right]^{-1} h^{\frac{\bar{\nu}-2}{2}} \exp\left(-\frac{h\bar{\nu}}{2\bar{s}^{-2}}\right).$$

Marginální hustotu získáme integrací přes všechna  $h$ . Členy neobsahující  $h$  můžeme vytknout před integrál a vzhledem k tomu, že přesnost chyby  $h$  nemůže nabývat záporných hodnot, můžeme psát

$$p(\beta|y) = \frac{\bar{V}^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \left[\left(\frac{2\bar{s}^{-2}}{\bar{\nu}}\right)^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{\bar{\nu}}{2}\right)\right]^{-1} \times \int_0^\infty h^{\frac{\bar{\nu}-1}{2}} \exp\left[-\frac{h}{2} \left(\bar{V}^{-1} (\beta - \bar{\beta})^2 + \frac{\bar{\nu}}{\bar{s}^{-2}}\right)\right] dh.$$



Pro výpočet výše uvedeného integrálu lze s úspěchem využít následující vztah odpovídající definici funkce gama:

$$\int_0^{\infty} x^r e^{-ax} dx = \frac{\Gamma(r+1)}{a^{r+1}} \quad (a > 0, r > -1),$$

čímž dostáváme

$$p(\beta|y) = \frac{\bar{V}^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \left[ \left( \frac{2\bar{s}^{-2}}{\bar{\nu}} \right)^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{\bar{\nu}}{2}\right) \right]^{-1} \Gamma\left(\frac{\bar{\nu}+1}{2}\right) \\ \times \left[ \frac{1}{2} \left( \bar{V}^{-1}(\beta - \bar{\beta})^2 + \frac{\bar{\nu}}{\bar{s}^{-2}} \right) \right]^{-\frac{\bar{\nu}+1}{2}}.$$

Po uspořádání jednotlivých členů jsme schopni získat vyjádření Studentova  $t$ -rozdělení  $t(\bar{\beta}, \bar{s}^2 \bar{V}, \bar{\nu})$ :

$$p(\beta|y) = \frac{\bar{\nu}^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{\bar{\nu}+1}{2}\right)}{(2\pi)^{\frac{1}{2}} \Gamma\left(\frac{\bar{\nu}}{2}\right)} (\bar{s}^2 \bar{V})^{-\frac{1}{2}} \left[ \bar{\nu} + (\bar{s}^2 \bar{V})^{-1} (\beta - \bar{\beta})^2 \right]^{-\frac{\bar{\nu}+1}{2}}.$$

Pokud jde o přesnost chyby  $h$ , pak z normálního-gama rozdělení plyne:

$$h|y \sim G(\bar{s}^{-2}, \bar{\nu}) \quad (2.16)$$

$$E(h|y) = \bar{s}^{-2} \quad (2.17)$$

$$var(h|y) = \frac{2\bar{s}^{-4}}{\bar{\nu}}. \quad (2.18)$$

Rovnice (2.8)-(2.18) ukazují, jak bayesiánský přístup kombinuje apriorní a datovou informaci. Tento model je jedním a snad i jediným z případů, u kterého není třeba provádět posteriorní simulaci a jsme schopni odvodit analytické řešení.

Klasická ekonometrie často využívá k odhadu parametru  $\beta$  estimátor metody nejmenších čtverců  $\hat{\beta}$ . Bayesovský bodový odhad  $\bar{\beta}$  je v tomto případě váženým průměrem OLS estimátoru a apriorní střední hodnoty  $\underline{\beta}$ . Váhy jsou proporcionální k výrazu  $\sum x_i^2$  a  $\underline{V}^{-1}$ . Poslední z těchto výrazů vyjadřuje velikost důvěry ve zvolený prior. Pokud zadáme vysokou variabilitu našemu prioru, vyjadřujeme tím velkou míru nejistoty o tom jaká asi bude hodnota  $\beta$ . Výsledkem tak je, že  $\underline{V}^{-1}$  je malé a jen malá váha je přiřazena apriorní střední hodnotě  $\underline{\beta}$ , tedy fakticky našemu apriornímu odhadu. Podobnou roli hraje i člen  $\sum x_i^2$ , snad jen s tím rozdílem, že v sobě nese datovou informaci. Volně řečeno, zohledňuje míru důvěry v to, že data v sobě obsahují informaci o nejlepším možném odhadu parametru  $\beta$ , kterým je OLS estimátor  $\hat{\beta}$ . V klasické ekonometrii je  $(\sum x_i^2)^{-1}$  proporcionální k variabilitě odhadu  $\hat{\beta}$ , k čemuž se ještě vrátíme. Nicméně, před tím si uveďme intuitivní příklad. Mějme jednoduchý případ, kdy  $x_i = 1$  pro  $i = 1, \dots, N$ . V takovémto případě je  $\sum x_i^2 = N$  a váha přiřazená k OLS odhadu  $\hat{\beta}$  je jednoduše velikost vzorku, což je rozumné měřítko pro objem informací v datech (a klasická ekonometrie to tak skutečně chápe).

V rámci výpočtu posteriorní střední hodnoty parametru jsou váhy přiřazeny přímo úměrně přesnosti (tj. převrácené hodnotě rozptylů) apriorní střední hodnoty a OLS odhadu tohoto parametru. V klasické ekonometrii je rozptyl OLS estimátoru našeho regresního modelu  $s^2(\sum x_i^2)^{-1}$ , což se využívá např. k testování statistické významnosti parametru (tzv.  $t$ -statistika pro testování  $\beta = 0$  je  $\frac{\hat{\beta}}{\sqrt{s^2(\sum x_i^2)^{-1}}}$ ). Bayesiánskou analogií je posteriorní rozptyl  $\beta$  daný v rovnici (2.15). Neformálně tak můžeme interpretovat část posteriorního rozptylu parametr (2.9) tak, že posteriorní přesnost je průměrem apriorní přesnosti ( $V^{-1}$ ) a datové přesnosti ( $\sum x_i^2$ ). Podobně můžeme vztah (2.12) intuitivně interpretovat tak, že posteriorní součet čtverců chyb ( $\overline{\nu s^2}$ ) je roven sumě apriorního součtu čtverců chyb ( $\nu s^2$ ), součtu čtverců chyb OLS estimátoru ( $\nu s^2$ ) a členu jenž měří míru souladu mezi apriorní a datovou informací.

Z výše uvedených rovnic tedy vypývá, že bayesiánský přístup kombinuje datovou a apriorní informaci. Přirozeně konjugovaná apriorní hustota navíc implikuje, že prior je možné interpretovat tak, že vyplývá z fiktivní datové množiny. Výrazy  $\underline{\nu}$  a  $N$  hrají ve vztazích (2.11) a (2.12) podobnou roli, tudíž  $\underline{\nu}$  lze chápat jako apriorní velikost vzorku.

Na první pohled se zdá, že klasický a Bayesiánský odhad mají mnoho společného. Klasický přístup odhaduje  $\hat{\beta}$  a jeho rozptyl  $s^2(\sum x_i^2)^{-1}$  a odhadem  $\sigma^2$  je  $s^2$ . Bayesiánský přístup rovněž počítá posteriorní střední hodnotu a rozptyl parametru  $\beta$ , tj.  $\bar{\beta}$  a  $\frac{\overline{\nu s^2}}{\overline{\nu s^2} - 2} \bar{V}$ , a odhaduje  $h = \sigma^{-2}$  pomocí posteriorní střední hodnoty  $\bar{s}^2$ . Existují zde však dva zásadní rozdíly: za prvé, bayesiánské vztahy vždy kombinují datovou a apriorní informaci, a za druhé, bayesiánci interpretují parametr (resp. obecně parametry)  $\beta$  jakožto náhodnou veličinu. Klasický přístup chápe pouze odhad  $\hat{\beta}$  jakožto náhodnou veličinu.

Použití přirozeně konjugovaného prioru nám implikuje to, že apriorní informace vstupuje do výpočtu stejným způsobem jako datová informace. To může pomoci se specifikací prioru. Při volbě  $\underline{\beta}$ ,  $\underline{V}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$  může pomáhat znalost toho, že  $\underline{\beta}$  je ekvivalentní OLS odhadu z imaginární datové sady o  $\underline{\nu}$  pozorování s imaginární částí rozptylu parametru  $\sum x_i^2$  rovnou  $\underline{V}^{-1}$  a stejně tak imaginárním rozptylem náhodné složky  $s^2$  daným jako  $\underline{s}^2$ . Ekonometrie je však vědou, s jejichž výstupy a výsledky přichází do styku široká veřejnost. V mnoha případech jsou čtenáři schopni posoudit a ztotožnit se s tím, co by mohlo být rozumným priorem (např. na základě ekonomické teorie, která může napovědět, jakých hodnot by parametry mohly nabývat). Ovšem v případech, kdy je možno přistoupit k problému se zcela rozdílnými priory, je možno bayesovskou analýzu spoléhající na jediný prior kritizovat.

Jedním z možných řešení je analýza citlivosti apriorní hustoty. Tím je myšleno to, že empirické výsledky je možno získat za použití různých priorů. Pokud jsou v tomto případě výsledky podobné (pro rozumné priory), pak čtenář, který se s těmito výsledky seznamuje, může být přesvědčen o tom, že výzkumníci s různými apriorním pohledem na věc jsou schopni po konfrontaci s daty dojít ke stejným či podobným závěrům. Pokud naopak jsou výsledky citlivé na volbu prioru, pak data jednoduše nedostačují k tomu, aby nastala shoda mezi výzkum-

níky s různými apriorními informacemi. I v tomto případě je však bayesovský přístup schopen najít z vědeckého hlediska čistý způsob řešení této situace. Existuje samostatná literatura, která hledá meze pro např. posteriorní střední hodnotu parametru. Jedná se o tzv. analýzu extrémních mezí (extreme bounds analysis). Typický závěr této literatury je přibližně v podobě, že „pro jakoukoliv volbu  $\underline{V}$  se musí  $\bar{\beta}$  nacházet mezi specifikovanou dolní a horní mezí.“

Další možností pro specifikaci prioru je případ, kdy vzhledem k neshodě o apriorní informaci je nutné sáhnout k použití tzv. **neinformativního prioru**. Opět zde existuje rozsáhlá literatura věnovaná této problematice. Dostatečné je však v tomto případě vědět, že může být v mnoha případech vhodné, aby datová informace převážila informaci apriorní. V našem konkrétním modelu lze toho dosáhnout velmi jednoduchou cestou. Stačí konkretizovat  $\underline{V}$  jakožto dostatečně malé vzhledem k  $N$  a  $\underline{V}$  dostatečně velké, což v konečném důsledku zajistí, že apriorní informace bude hrát v posteriorním vztahu velmi malou roli. Tento druh apriorní informace se nazývá **relativně neinformativní prior**.

V limitním případě můžeme vytvořit čistě neinformativní prior nastavením  $\underline{V} = 0$  a  $\underline{V}^{-1} = 0$  (tj.  $\underline{V} \rightarrow \infty$ ). Výsledkem je posteriorní hustota parametrů v podobě  $\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ , kde

$$\bar{V} = \frac{1}{\sum x_i^2}, \quad (2.19)$$

$$\bar{\beta} = \hat{\beta}, \quad (2.20)$$

$$\bar{\nu} = N, \quad (2.21)$$

$$\bar{\nu}\bar{s}^2 = \nu s^2. \quad (2.22)$$

S čistě neinformativním priorem zahrnují výše uvedené vztahy jen datovou informaci a odpovídají ve skutečnosti výsledkům aplikace metody nejmenších čtverců. Na jedné straně má tento druh prioru atraktivní vlastnosti a je jakýmsi předělem mezi klasickým a bayesiánským přístupem (díky blízkosti výsledkům odhadu metodou nejmenších čtverců). Jeho nepříjemnou vlastností je však to, že tento prior není apriorní hustotou v pravém slova smyslu, neboť jejím integrováním nezískáme jedničku. Takovýto druh prioru se nazývá nepravý prior (*improper prior*). S jeho užitím je spojena řada problémů, z nichž jeden bude ilustrován na příkladu porovnání modelů.

Nepravost tohoto neinformativního prioru lze ukázat tím, že posteriorní výsledky uvedené v rovnicích (2.19)-(2.22) lze získat kombinací věrohodnostní funkce s následující „apriorní hustotou“:

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

kde  $h$  je definováno na intervalu  $(0, \infty)$ . Integrací této hustoty na příslušném intervalu získáme výsledek rovný nekonečnu. Tento druh prioru bývá obvykle zapsán jako:

$$p(\beta, h) \propto \frac{1}{h}. \quad (2.23)$$

opět se ale nejedná z formálního hlediska o dobrý zápis vzhledem k tomu, že nepracujeme se skutečnou funkcí hustoty pravděpodobnosti. Je třeba poznamenat, že v mnohých modelech je tendence užívat neinformativní prior jakožto nepravý prior. Většinou se jedná o aplikaci uniformního rozdělení na intervalu  $(-\infty, \infty)$  (o zkoumaném parametru tak předpokládáme, že jeho realizace je stejně pravděpodobná na celé reálné ose), což samozřejmě z formálního hlediska není uniformní rozdělení, které je definováno na intervalu  $[a, b]$  pro konečné hodnoty  $a$  a  $b$ . Uniformní "neinformativní" apriorní hustota je tak hojně užívaným nepravým priorem.

## 2.5 Porovnání modelů

Předpokládejme, že máme dva regresní modely,  $M_1$  a  $M_2$ , přičemž oba se snaží popsat chování  $y$ . Tyto modely se liší ve svých vysvětlujících proměnných. Příslušným indexem rozlišujeme i parametry těchto modelů. Model  $M_j$  pro  $j = 1, 2$  je tedy založen na jednoduchém lineárním regresním modelu v podobě

$$y_i = \beta_j x_{ji} + \epsilon_{ji}, \quad (2.24)$$

pro  $i = 1, \dots, N$ . Předpoklady kladené na vysvětlující proměnné a náhodnou složku zůstávají stejné jako v předchozím oddíle, tedy  $\epsilon_{ji}$  je z i.i.d.  $N(0, h_j^{-1})$  a  $x_{ji}$  je buď nenáhodné nebo exogenní pro  $j = 1, 2$ . Pro oba modely lze zapsat přirozeně konjugované priory z normálního-gama rozdělení jako

$$\beta_j, h_j | M_j \sim NG(\underline{\beta}_j, \underline{V}_j, \underline{s}_j^{-2}, \underline{\nu}). \quad (2.25)$$

To implikuje posteriorní hustoty ve tvaru

$$\beta_j, h_j | y, M_j \sim NG(\bar{\beta}_j, \bar{V}_j, \bar{s}_j^{-2}, \bar{\nu}_j), \quad (2.26)$$

kde

$$\bar{V}_j = \frac{1}{\underline{V}_j^{-1} + \sum x_{ji}^2}, \quad (2.27)$$

$$\bar{\beta}_j = \bar{V}_j (\underline{V}_j^{-1} \underline{\beta}_j + \hat{\beta}_j \sum x_{ji}^2), \quad (2.28)$$

$$\bar{\nu}_j = \underline{\nu}_j + N \quad (2.29)$$

a  $\bar{s}_j^{-2}$  je implicitně definováno jako

$$\bar{\nu}_j \bar{s}_j^2 = \underline{\nu}_j \underline{s}_j^2 + \nu_j s_j^2 + \frac{(\hat{\beta}_j - \underline{\beta}_j)^2}{\underline{V}_j + \frac{1}{\sum x_{ji}^2}}. \quad (2.30)$$

Výrazy  $\hat{\beta}_j$ ,  $s_j^2$  a  $\nu_j$  jsou OLS odhady analogické těm, definovaným ve vztazích (2.3)–(2.5). Jinými slovy, veškeré výsledky odpovídají vztahům (2.7)–(2.12), až na dodané indexy  $j$  reflektující konkrétně specifikovaný model.

Rovnice (2.26)–(2.30) lze využít k posteriorní analýze každého z modelů. V našem případě nás však zajímá bayesovské porovnání modelů, jež se provádí pomocí posteriorních podílů šancí:

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

Apriorní pravděpodobnosti modelů  $p(M_i)$  pro  $i = 1, 2$  je třeba zvolit před tím, než se “podíváme” na data. Neinformativní volbou je přiřazení každému z modelů stejnou pravděpodobnost, tedy  $p(M_1) = p(M_2) = \frac{1}{2}$ . Marginální věrohodnost  $p(y|M_j)$  se počítá jako:

$$p(y|M_j) = \iint p(y|\beta_j, h_j)p(\beta_j, h_j)d\beta_j dh_j. \quad (2.31)$$

Oproti jiným modelům lze tento integrál vyhodnotit analyticky. Výsledkem je vztah:

$$p(y|M_j) = c_j \left( \frac{\bar{V}_j}{V_j} \right)^{\frac{1}{2}} (\bar{\nu}_j \bar{s}_j^2)^{-\frac{\bar{\nu}_j}{2}} \quad (2.32)$$

pro  $j = 1, 2$ , přičemž

$$c_j = \frac{\Gamma\left(\frac{\bar{\nu}_j}{2}\right) (\underline{\nu}_j \underline{s}_j^2)^{\frac{\bar{\nu}_j}{2}}}{\Gamma\left(\frac{\underline{\nu}_j}{2}\right) \pi^{\frac{N}{2}}} \quad (2.33)$$

a  $\Gamma(\cdot)$  je funkce gama. Posteriorní podíl šancí porovnávající  $M_1$  a  $M_2$  je:

$$PO_{12} = \frac{c_1 \left( \frac{\bar{V}_1}{V_1} \right)^{\frac{1}{2}} (\bar{\nu}_1 \bar{s}_1^2)^{-\frac{\bar{\nu}_1}{2}} p(M_1)}{c_2 \left( \frac{\bar{V}_2}{V_2} \right)^{\frac{1}{2}} (\bar{\nu}_2 \bar{s}_2^2)^{-\frac{\bar{\nu}_2}{2}} p(M_2)}. \quad (2.34)$$

Ukažme si nyní, že tomu tak skutečně je. Odvodíme si tedy výraz pro marginální hustotu. Pro přehlednost značení nebudeme uvádět dolní index  $j$ , označující příslušnost k danému modelu. Na úvod vyjdeme ze vztahů pro věrohodnostní funkci a apriorní hustotu parametrů:

$$\begin{aligned} p(y|\beta, h) &= \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{N}{2}} \exp \left[ -\frac{h}{2} (\beta - \hat{\beta})^2 \sum_{i=1}^N (x_i)^2 \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp \left[ -\frac{h\nu}{2s^{-2}} \right] \right\}, \\ p(\beta, h) &= \frac{1}{(2\pi)^{\frac{1}{2}}} h^{\frac{1}{2}} V^{-\frac{1}{2}} \exp \left[ -\frac{h}{2} V^{-1} (\beta - \underline{\beta})^2 \right] \\ &\quad \times \left[ \left( \frac{2s^{-2}}{\underline{\nu}} \right)^{\frac{\nu}{2}} \Gamma \left( \frac{\nu}{2} \right) \right]^{-1} h^{\frac{\nu-2}{2}} \exp \left( -\frac{h\underline{\nu}}{2s^{-2}} \right). \end{aligned}$$

V dalším kroku spolu pronásobíme věrohodnostní funkci a apriorní hustotu, integrační konstanty nezávislé na parametrech  $\beta$  a  $h$  dáme před odpovídající

dvojitý integrál, ve kterém oddělíme členy obsahující jen  $h$  a členy závisící na  $\beta$ :

$$\begin{aligned} p(y|M) &= \iint_{h,\beta} p(y|\beta, h)p(\beta, h)d\beta dh \\ &= \frac{1}{(2\pi)^{\frac{N+1}{2}}} V^{-\frac{1}{2}} \left[ \left( \frac{2s^{-2}}{\nu} \right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right) \right]^{-1} \\ &\quad \times \int_0^\infty h^{\frac{1+\nu+1+\nu-2}{2}} \exp\left[-\frac{h}{2}(\nu s^2 + \nu s^2)\right] \dots \\ &\quad \int_{-\infty}^\infty \exp\left\{-\frac{h}{2}\left[(\beta - \hat{\beta})^2 \sum x_i^2 + V^{-1}(\beta - \beta)^2\right]\right\} d\beta dh. \end{aligned}$$

Nejprve vyřešíme integrály podle parametru  $\beta$ , který lze při využití předchozího značení (a odvození) přepsat do podoby

$$\int_{-\infty}^\infty \exp\left\{-\frac{h}{2}\left[(\hat{\beta} - \beta)^2 \sum x_i^2 \bar{V} V^{-1} + (\beta - \bar{\beta})^2 \bar{V}^{-1}\right]\right\} d\beta.$$

Člen neobsahující parametr  $\beta$  můžeme vytknout před integrál, čímž získáme

$$\int_{-\infty}^\infty \exp\left\{-\frac{h}{2}\left[(\beta - \bar{\beta})^2 \bar{V}^{-1}\right]\right\} d\beta.$$

Výraz uvnitř integrálu odpovídá jádrové hustotě normálního rozdělení parametru  $\beta$ , kdy  $\beta \sim N(\bar{\beta}, h^{-1}\bar{V})$ . Z definice normálního rozdělení si snadno doplníme odpovídající členy integrační konstanty, které můžeme z větší části získat z výrazů původního dvojitého integrálu. Pro přehlednost si ale členy integrační konstanty doplníme samostatně (tzn. dodáme je do integrálu a jejich inverzní hodnotu pak pro korektnost necháme před daným integrálem) a až v závěru tyto výrazy spojíme s původními členy integrálu. Můžeme tedy psát

$$(2\pi)^{\frac{1}{2}} h^{-\frac{1}{2}} \bar{V}^{\frac{1}{2}} \int_{-\infty}^\infty \frac{1}{(2\pi)^{\frac{1}{2}}} h^{\frac{1}{2}} \bar{V}^{-\frac{1}{2}} \exp\left[-\frac{h}{2}(\beta - \bar{\beta})^2 \bar{V}^{-1}\right] d\beta.$$

Tento integrál je ale roven jedné (neboť se jedná o integrál funkce hustoty pravděpodobnosti), tudíž přistoupíme k integraci přes  $h$ . Všechny členy neobsahující  $h$  vytkneme před integrál, čímž získáme:

$$\begin{aligned} &\frac{1}{(2\pi)^{\frac{N+1}{2}}} V^{-\frac{1}{2}} \left[ \left( \frac{2s^{-2}}{\nu} \right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right) \right]^{-1} (2\pi)^{\frac{1}{2}} \bar{V}^{-\frac{1}{2}} \\ &\quad \times \int_0^\infty h^{\frac{\nu+\nu}{2}} h^{-\frac{1}{2}} \exp\left\{-\frac{h}{2}\left[\nu s^2 + \nu s^2 + (\hat{\beta} - \beta)^2 \sum x_i^2 \bar{V} V^{-1}\right]\right\} dh. \end{aligned}$$

Nyní vypočítáme příslušný integrál, který lze při značení zavedené v této kapitole přepsat do podoby

$$\int_0^\infty h^{\frac{\nu-2}{2}} \exp\left(-\frac{h}{2\nu s^2}\right) dh.$$

V tomto případě je výraz uvnitř integrálu jádrovou hustotou gama rozdělení pro parametr  $h$ , tedy  $h \sim G(\bar{s}^{-2}, \bar{\nu})$ . Funkce hustoty gama rozdělení je známá, a tak si můžeme doplnit odpovídající členy integrační konstanty, čímž dostaneme

$$\left[ \left( \frac{2\bar{s}^{-2}}{\bar{\nu}} \right) \Gamma \left( \frac{\bar{\nu}}{2} \right) \right] \int_0^\infty \left[ \left( \frac{2\bar{s}^{-2}}{\bar{\nu}} \right) \Gamma \left( \frac{\bar{\nu}}{2} \right) \right]^{-1} h^{\frac{\bar{\nu}-2}{2}} \exp \left( -\frac{h}{2\bar{\nu}\bar{s}^2} \right) dh.$$

Tento integrál má hodnotu jedna. Výsledkem celého našeho snažení je tak vztah pro marginální věrohodnost v podobě

$$p(y|M) = \frac{1}{(2\pi)^{\frac{N}{2}}} V^{-\frac{1}{2}} \bar{V}^{\frac{1}{2}} \left[ \left( \frac{2\bar{s}^{-2}}{\bar{\nu}} \right) \Gamma \left( \frac{\bar{\nu}}{2} \right) \right]^{-1} \left[ \left( \frac{2\bar{s}^{-2}}{\bar{\nu}} \right) \Gamma \left( \frac{\bar{\nu}}{2} \right) \right].$$

Po úpravě můžeme dojít k výrazu

$$p(y|M) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left( \frac{\bar{V}}{\bar{V}} \right)^{\frac{1}{2}} \left( \frac{\nu s^2}{2} \right)^{\frac{\bar{\nu}}{2}} \left( \frac{\nu s^2}{2} \right)^{-\frac{\bar{\nu}}{2}} \frac{\Gamma \left( \frac{\bar{\nu}}{2} \right)}{\Gamma \left( \frac{\bar{\nu}}{2} \right)},$$

který plně odpovídá vztahům (2.32) a (2.33), protože díky definici  $\bar{\nu} = N + \nu$  platí

$$\left( \frac{1}{2} \right)^{\frac{N}{2}} \left( \frac{1}{2} \right)^{\frac{\bar{\nu}}{2}} \left( \frac{1}{2} \right)^{-\frac{\bar{\nu}}{2}} = \left( \frac{1}{2} \right)^{\frac{N+\bar{\nu}-\bar{\nu}}{2}} = \left( \frac{1}{2} \right)^{\frac{0}{2}} = 1.$$

Posteriorní podíl šancí může být využit k výpočtu posteriorních pravděpodobností modelů  $p(M_j|y)$ :

$$p(M_1|y) = \frac{PO_{12}}{1 + PO_{12}},$$

$$p(M_2|y) = \frac{1}{1 + PO_{12}}.$$

Pohled na posteriorní podíl šancí jasně ukazuje, co všechno vstupuje do bayesovského porovnání modelů a co jej tedy ovlivňuje. Konkrétně tedy můžeme říct:

1. Čím větší je apriorní podíl šancí  $\frac{p(M_1)}{p(M_2)}$ , tím vyšší je podpora pro  $M_1$ .
2. Výraz  $\bar{\nu}_j \bar{s}_j^2$  obsahuje člen  $\nu_j s_j^2$ , což je součet čtverců chyb (reziduí). Součet čtverců chyb je obvyklým měřítkem souladu modelu s daty, a to z hlediska jeho schopnosti reprodukovat reálná data, kdy nižší hodnoty naznačují jeho lepší soulad. Posteriorní podíl šancí tedy oceňuje model, který lépe popisuje data.
3. Pokud všechny ostatní charakteristiky jsou stejné, posteriorní podíl šancí upřednostní model, kde je větší koherence mezi apriorní a datovou informací (tj.  $(\hat{\beta}_j - \underline{\beta}_j)^2$ , což vstupuje do  $\bar{\nu}_j \bar{s}_j^2$ ).

4. Člen  $\left(\frac{\underline{V}_i}{\underline{V}_i}\right)$  je podíl posteriorní a apriorní variance. Tento člen lze interpretovat tak, že pokud vše ostatní zůstane stejné, bude upřednostněn model s větší apriorní informací (tedy menší apriorní variancí) vztahené relativně k posteriorní informaci.

V další kapitole pak bude ukázáno, že bude upřednostněn posteriorním podílem šancí ten model, který má méně parametrů, a to za předpokladu, že ostatní hodnoty si budou rovny. Toto je důležitá vlastnost posteriorního podílu šancí, která na tomto místě (vzhledem ke stejnému počtu parametrů) nemůže být ukázána.

V případě užití neinformativní apriorní informace není marginální věrohodnost definována a stejně tak není definován i posteriorní podíl šancí. To je jeden z problémů spojených s užitím neinformativního prioru. Obvyklým řešením je nastavení  $\underline{\nu}_1 = \underline{\nu}_2$ , a to v podobě velmi malé hodnoty. Tentýž postup se uplatní i při volbě  $\underline{V}_1^{-1}$  a  $\underline{V}_2^{-1}$ . Rovněž se nastaví  $\underline{s}_1 = \underline{s}_2$ . Za těchto předpokladů je posteriorní podíl šancí definován, zjednodušuje se, a dostává se blízko k výrazu:

$$PO_{12} = \frac{\left(\frac{1}{\sum x_{1i}^2}\right)^{\frac{1}{2}} (\nu_1 s_1^2)^{-\frac{N}{2}} p(M_1)}{\left(\frac{1}{\sum x_{2i}^2}\right)^{\frac{1}{2}} (\nu_2 s_2^2)^{-\frac{N}{2}} p(M_2)}. \quad (2.35)$$

V tomto případě reflektuje posteriorní podíl šancí pouze apriorní podíl šancí, relativní kvalitu shody modelů s daty a poměr zahrnující výraz  $\frac{1}{\sum x_{ji}^2}$ , což odpovídá přesnosti posterioru pro  $M_j$ . Toto řešení problému použitím neinformativního prioru však nelze použít (jak uvidíme později) v situaci, kdy je počet parametrů v obou porovnávaných modelech odlišný.

Pokud bychom chtěli porovnávat více modelů, je možné porovnat každý pár a spočítat posteriorní pravděpodobnosti každého z modelů, pokud budeme předpokládat, že součet jejich pravděpodobností je roven jedné (a máme tak úplnou množinu relevantních modelů).

## 2.6 Predikční hustota

V této sekci se vrátíme zpět k původní specifikaci modelu s věrohodnostní funkcí a apriorní hustotou definovanou rovnicemi (2.6) až (2.7). Rovnice (2.8)–(2.12) popisují bayesiánskou metodu poznávání parametrů  $\beta$  a  $h$ , založeného na datové množině  $N$  pozorování. Nyní se zaměříme na predikci nepozorované hodnoty vysvětlované proměnné generované stejným modelem. Formálně budeme předpokládat vztah:

$$y^* = \beta x^* + \epsilon^*, \quad (2.36)$$

kde  $y^*$  není pozorován. Všechny předpoklady kladené na tento model jsou stejné jako v předchozí části, tedy  $\epsilon^*$  je nezávislé na  $\epsilon_i$  pro  $i = 1, \dots, N$  a má normální rozdělení  $N(0, h^{-1})$ . Parametr  $\beta$  je totožný s parametrem definovaným v úvodu této kapitoly. Je nutné předpokládat, že  $x^*$  je pro nás známé či pozorované, jinak by naše predikce neměla valný smysl.



Bayesiánská predikce je založena na predikční hustotě, která je dána jako:

$$p(y^*|y) = \iint p(y^*|y, \beta, h) p(\beta, h|y) d\beta dh. \quad (2.37)$$

Skutečnost, že  $\epsilon^*$  nezávisí na  $\epsilon_i$  implikuje, že  $y$  a  $y^*$  jsou na sobě nezávislé, a tedy  $p(y^*|y, \beta, h) = p(y^*|\beta, h)$ . V integrálu pro posteriorní predikční hustotu se tak vyskytuje posteriorní hustota  $p(\beta, h|y)$  a hustota  $p(y^*|\beta, h)$ . Analogickou úvahou pro odvození věrohodnostní funkce dostáváme:

$$p(y^*|\beta, h) = \frac{h^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{h}{2}(y^* - \beta x^*)^2\right]. \quad (2.38)$$

Násobením této věrohodnostní funkce predikce a posteriorní hustoty a integrací podle vztahu (2.37) získáme:

$$p(y^*|y) \propto [\bar{v} + (y^* - \bar{\beta}x^*)^2 \bar{s}^{-2} (1 + \bar{V}x^{*2})^{-1}]^{-\frac{\bar{v}+1}{2}}. \quad (2.39)$$

Opět můžeme vidět, že se jedná o jádrovou hustotu  $t$ -rozdělení se střední hodnotou  $\bar{\beta}x^*$ , rozptylem  $\frac{\bar{s}^2}{\bar{v}-2}(1 + \bar{V}x^{*2})$  a s počtem stupňů volnosti  $\bar{v}$ . Jinými slovy

$$y^*|y \sim t(\bar{\beta}x^*, \bar{s}^2\{1 + \bar{V}x^{*2}\}, \bar{v}). \quad (2.40)$$

Tyto výsledky lze využít pro bodovou predikci a míru neurčitosti asociovanou s bodovou predikcí, což je standardní chyba predikce.

Na tomto místě je dobré zavést důležitý bayesiánský koncept zvaný průměrování modelů (Bayesian model averaging). V předchozí části jsme si ukázali, jak vypočítat posteriorní pravděpodobnosti modelů  $p(M_j|y)$  pro  $j = 1, 2$ . Tuto charakteristiku lze využít při volbě či upřednostnění konkrétního modelu. Ne vždy je však žádoucí vybrat jeden model s nejvyšší posteriorní hustotou modelu a nezabývat se dále ostatními modely. Bayesiánské průměrování modelu v sobě zahrnuje výsledky všech modelů, ovšem prezentace výsledků probíhá průměrováním přes všechny modely. Z pravidel pravděpodobnosti jednoduše vyplývá:

$$p(y^*|y) = p(y^*|y, M_1)p(M_1|y) + p(y^*|y, M_2)p(M_2|y). \quad (2.41)$$

Stručně řečeno, pokud se zajímáme o predikční hustotu  $p(y^*|y)$ , neměli bychom se zaměřit na jediný model a pracovat tak např. s  $p(y^*|y, M_1)$ . Měli bychom spíše průměrovat výsledky přes oba (či obecněji všechny) relevantní modely, a to za použití vah danými posteriorními hustotami pravděpodobnosti. S využitím vlastnosti operátoru očekávání dostáváme:

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

což lze využít k výpočtu bodové predikce získané průměrováním přes oba modely. Pokud nás zajímá nějaká funkce predikovaných hodnot  $g(\cdot)$ , pak lze předchozí výsledek zobecnit do podoby:

$$E[g(y^*)|y] = E[g(y^*)|y, M_1]p(M_1|y) + E[g(y^*)|y, M_2]p(M_2|y), \quad (2.42)$$

což lze zase využít např. pro výpočet rozptylu predikce. Tyto výsledky lze zobecnit na případ více modelů nebo na případ, kde funkce, která nás zajímá, v sobě místo  $y^*$  bude obsahovat parametry. Detailněji bude bayesovské průměrování modelů diskutováno později.

## 2.7 Empirická ilustrace

Regresní model s jedinou vysvětlující proměnnou je příliš jednoduchý pro praktické použití, zvláště pak bez přítomnosti úroňové konstanty. Pro ilustraci základních principů popsaných v této kapitole tak využijeme uměle vytvořená data. Volba rozsahu datového souboru, vysvětlujících proměnných a apriorních hyperparametrů je čistě ilustrativní. Budeme postupovat v následujících krocích:

1. Zvolíme si počet uměle vytvořených dat  $N = 50$  a skutečné hodnoty parametrů  $\beta = 2$  a  $h = 1$ .
2. Vygenerujeme si jednotlivé vysvětlující proměnné,  $x_i$ , které budou pocházet z rovnoměrného (uniformního) rozdělení,  $U(0, 1)$ , pro  $i = 1, \dots, 50$ . Podobně získáme jedbotlivé hodnoty náhodných složek,  $\epsilon_i$ , které pocházejí z normálního rozdělení,  $N(0, h^{-1})$ .
3. Využijeme vytvořené vysvětlující proměnné a náhodné složky k vygenerování závisle proměnné  $y_i = \beta x_i + \epsilon_i$ .
4. Využijeme dva typy apriorních hustot - neinformativní (viz 2.23) a informativní, přirozeně konjugovanou (viz 2.7) s parametry  $\underline{\beta} = 1.5$ ,  $\underline{V} = 0.25$ ,  $\underline{\nu} = 10$  a  $\underline{s}^{-2} = 1$ .

Data pro tento příklad jsou obsahem souboru `data_NLRM1.mat` a příklad je řešen v Matlabu ve skriptu `priklad_NLRM1.mat`. Zde je možno snadno replikovat výsledky na základě „originálních“ dat odpovídajících výsledkům prezentovaným na tomto místě či si provést bayesiánskou analýzu na datech zcela nových.

**Tabulka 2.1:** Apriorní a posteriorní charakteristiky parametru  $\beta$

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

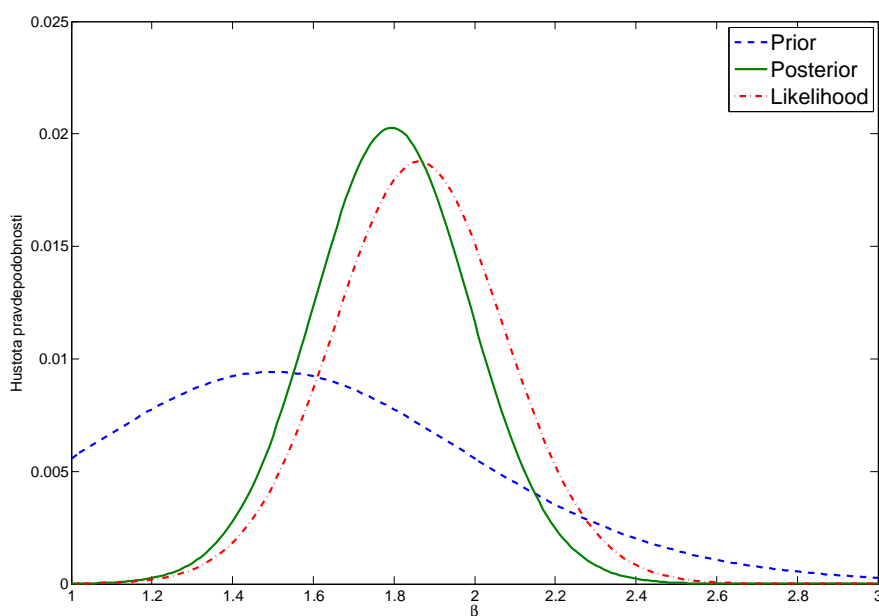
\* Jedná se o směrodatnou odchylku příslušného  $t$ -rozdělení, apriorní či posteriorní  $V$  tedy neudává přímo hodnotu rozptylu.

Tabulky 2.1–2.2 ukazují výsledky odhadů pro parametry  $\beta$  a  $h$ , s využitím vztahů (2.8)–(2.22). Obrázek 2.1 zobrazuje posteriorní hustoty parametru  $\beta$  pro

Tabulka 2.2: 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.

Obrázek 2.1: Marginální apriorní a posteriorní hustoty pro parametr  $\beta$ .

případ informativní apriorní hustoty a neinformativní apriorní hustoty. Součástí je rovněž vykreslení informativní apriorní hustoty parametru  $\beta$  (neinformativní hustota by odpovídala nekonečně ploché přímce na intervalu  $(-\infty, \infty)$ ). Z rovnice 2.13 vyplývá, že se jedná o funkce hustoty pravděpodobnosti odpovídající  $t$ -rozdělení (zmiňovaná rovnice je analogická i marginální apriorní hustotě, kdy příslušné parametry odpovídají apriorním hyperparametrům). Posteriorní charakteristiky založené na neinformativní apriorní hustotě obsahují jen informaci danou věrohodnostní funkcí a odpovídají tak klasickým závěrům při využití metody nejmenších čtverců (viz vztahy (2.19)–(2.22)). Právě proto označujeme v obrázku 2.1 výslednou marginální posteriorní hustotu pro  $\beta$  jako věrohodnostní funkci (*likelihood*).

Tabulky i obrázek ukazují zcela jasně jak bayesiánský přístup kombinuje apriorní a datovou informaci do podoby posteriorní hustoty. Na obrázku 2.1 tak vidíme, že posteriorní hustota vycházející z informativního prioru vypadá jako průměr apriorní hustoty a věrohodnostní funkce. Z tabulek 2.1 a 2.2 je patrné, že posteriorní střední hodnoty obou parametrů,  $E(\beta|y)$  a  $E(h|y)$ , které jsou založeny na totožné apriorní informaci leží mezi odpovídajícími středními hodnotami apriorní hustoty a věrohodnostní funkce (což je posteriorní hustota při neinformativním prioru). Zvolená apriorní hustota obsahuje méně informace, než té, která je obsažena v datech. To vidíme jak na obrázku, kde je apriorní hustota mnohem „roztáhlejší“ (rozptýlenější) než funkce věrohodnosti, případně stejný závěr vyčteme i z tabulek, kdy jsou apriorní směrodatné odchylky mnohem větší než ty, vycházející z „věrohodnostních“ odhadů.

Je třeba mít na paměti, že díky uměle vytvořeným datům známe skutečnou hodnotu parametrů, tedy  $\beta = 2$  a  $h = 1$ . Samozřejmě nemůžeme nikdy očekávat, že náš bodový odhad, což je např. posteriorní střední hodnota nebo příslušný OLS odhad budou přesně roven skutečné hodnotě hledaného parametru. Ovšem, posteriorní střední hodnoty jsou vcelku dost blízko svým skutečným hodnotám vzhledem ke svým posteriorním směrodatným odchylkám. Uvědomme si, že posteriorní směrodatná odchylka při použití informativní apriorní hustoty jsou trochu menší než směrodatné odchylky při využití neinformativních priorů. To odpovídá intuitivně faktu, že více informací nám obecně umožňuje přesnější odhad. Je tedy rozumné předpokládat, že posteriorní hustota kombinující jak datovou, tak i apriorní informaci, bude méně rozptýlená než posteriorní hustota založená na neinformativním prioru. Vrátime-li se zpět k vzorcům předchozích podkapitol, je tento intuitivní závěr v souladu s výrazem (2.9), který bude menší než (2.19) pokud je  $\underline{V} > 0$ . Ale pozor, tato intuice nebude platit v každém případě. Pokud budou apriorní a datová informace od sebe hodně odlišné, potom (2.12) může být mnohem větší než (2.22). Protože jak  $\bar{V}$ , tak i  $\bar{\sigma}^2$  vstupují do vztahu pro posteriorní směrodatnou odchylku parametru  $\beta$  symetricky, je možné (i když neobvyklé), že posteriorní směrodatná odchylka při informativním prioru bude větší než při použití neinformativní apriorní hustoty.

Pro ilustraci porovnání modelů předpokládejme, že nás zajímá porovnání modelu uvedeného výše s lineárním regresním modelem, který bude obsahovat pouze úrovněovou konstantu. V tomto druhém případě tedy bude vysvětlující proměnná  $x_i = 1$  pro  $i = 1, \dots, 50$ . Pro oba modely využijeme tutéž apriorní hustotu jako u modelu původního ( $NG(1.5, 0.25, 1, 10)$ ). Výsledky odhadu nabízí tabulka 2.3.

Budeme předpokládat apriorní podíl šancí roven jedné. V tomto případě využijeme vztah (2.34) k výpočtu posteriorního podílu šancí, který porovnává pravděpodobnosti těchto dvou modelů. My samozřejmě víme, že náš první model je ten pravý, a můžeme tak očekávat, že i podíl šancí tuto skutečnost dokáže vzít v úvahu. A skutečně tomu tak je, posteriorní podíl šancí je 7346. To slovně popíšeme tak, že na základě našich odhadů je první model 7346 krát pravděpodobnější než model druhý. Vyjádříme-li si tento podíl šancí pomocí pravděpodobností jednotlivých modelů, získáváme výsledek (oba modely chápeme jako jediné možné a vzájemně se vylučující), že  $p(M_1|y) = 0.9999$  a  $p(M_2|y) = 0.0001$ ,

**Tabulka 2.3:** Apriorní a posteriorní charakteristiky pro model jen s úrovníovou konstantou

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

Pokud bychom na tomto základě prováděli (např. v rámci predikcí) bayesovské průměrování modelů, potom bychom 99.99% váhy přiřadili výsledkům prvního modelu a jen 0.01% váhy výsledkům modelu druhého (viz (2.41)).

Pokud jde o analýzu zaměřenou na predikci, tu můžeme snadno provést s využitím vztahu (2.40). Zvolme si nepozorovanou hodnotu vysvětlující proměnné  $x^* = 0.5$  (nepozorovanou ve smyslu toho, že nebyla použita v rámci odhadu parametrů). Při využití informativního prioru získáváme výsledek

$$y^*|y \sim t(0.897, 0.833, 60).$$

V případě neinformativní apriorní hustoty je výsledek:

$$y^*|y \sim t(0.931, 0.7915, 50).$$

Tyto pravděpodobnosti lze využít pro prezentaci bodové predikce, směrodatné odchylky predikce či jakékoli jiné funkce predikce, kterou chceme spočítat. Z vlastností  $t$ -rozdělení tak např. vyplývá, že pro informativní apriorní hustotu jsou střední hodnota  $E(y^*|y) = 0.897$  a směrodatná odchylka predikce je 0.928 (získáme ji jako  $\sqrt{0.833 \cdot 60 / (60 - 2)}$ ). Analogicky postupujeme i pro neinformativní apriorní hustotu.

## 2.8 Shrnutí

V této kapitole jsme si prošli kompletní postup bayesiánské analýzy, tzn. věrohodnostní funkci, apriorní hustotu, posteriorní hustotu, porovnání modelů a predikci. Postup jsme si ukázali na normálním lineárním regresním modelu s jedinou vysvětlující proměnnou a tzv. přirozeně konjugovanou apriorní hustotou. Pro parametry modelu,  $\beta$  a  $h$ , má odpovídající apriorní hustota normální-gama rozdělení. Podstatou přirozeně konjugované apriorní hustoty je to, že i posteriorní hustota má normální-gama rozdělení. Pro tuto apriorní hustotu mají posteriorní hustota, predikční hustota a výrazy pro porovnání modelů analytické vyjádření. Není tedy nutné využít posteriorních simulátorů. Posteriorní hyperparametry získáváme prostým „vylepšením“ apriorních hyperparametrů o datovou informaci. Tyto postupy byly klíčové v dobách, kdy výpočetní technika nebyla na tak dobré úrovni, aby zvládala náročné posteriorní simulace. Nicméně, je zřejmé že předpoklad přirozeně konjugovaných apriorních hustot může být pro reálné aplikace hodně svazující (případně nereálný).

Doplněním této kapitoly bylo zavedení neinformativních apriorních hustot a Bayesovského průměrování modelů, což je významnou „specialitou“ bayesiánského přístupu.

## Kapitola 3

# Normální lineární regresní model s přirozeně konjugovaným priorem – více vysvětlujících proměnných

V reálných aplikacích si těžko vystačíme s regresním modelem obsahujícím jedinou vysvětlující proměnou. V této kapitole tak rozšíříme naše dosavadní znalosti o případ regresního modelu s více vysvětlujícími proměnnými. Značení a postupy v této kapitole využívají maticovou algebru, což umožňuje jednak kompaktní zápis, jednak se tak zjednodušuje prezentace samotných výsledků. Nebude asi překvapující, že výsledky a postupy prezentované v této kapitole budou ve své podstatě identické se závěry kapitoly předchozí.

### 3.1 LRM v maticovém vyjádření

Předpokládejme tedy, že máme data vysvětlované (závislé) proměnné  $y_i$  a  $k$  vysvětlujících proměnných  $x_{i1}, \dots, x_{ik}$  pro  $i = 1, \dots, N$ . Lineární regresní model bude mít podobu

$$y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i. \quad (3.1)$$

Aby se v modelu vyskytovala i úrovně konstanta, je proměnná  $x_{i1}$  implicitně nastavena na hodnotu 1 (srozumitelně pro případ, kdy úrovně konstantu mít nechceme, tak tento implicitní předpoklad nezavádíme). Mnohem kompaktněji můžeme náš model zapsat v maticové podobě, a to definováním jednotlivých

vektorů rozměru  $N \times 1$  a matice rozměru  $N \times k$ . Konkrétně

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_N \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 1 & x_{N2} & \cdots & x_{Nk} \end{bmatrix}.$$

Lineární regresní model lze maticově vyjádřit jako

$$y = X\beta + \epsilon. \quad (3.2)$$

### 3.2 Věrohodnostní funkce

Věrohodnostní funkci můžeme odvodit podobným způsobem jak je naznačeno v předchozí kapitole. Předpoklady kladené na vektor náhodných složek  $\epsilon$  a matici vysvětlujících proměnných  $X$  určují tvar věrohodnostní funkce. Maticové zobecnění těchto předpokladů je následující:

1. Vektor náhodných složek  $\epsilon$  je z vícerozměrného normálního rozdělení se střední hodnotou  $0_N$  a kovarianční maticí  $\sigma^2 I_N$ , kde  $0_N$  je  $N$ -rozměrný (sloupcový) nulový vektor a  $I_N$  je jednotková matice rozměru  $N \times N$ . Jednoduše řečeno,  $\epsilon \sim N(0_N, h^{-1} I_N)$ , kde přesnost chyby  $h = \sigma^{-2}$ .
2. Všechny prvky matice  $X$  jsou pevná čísla (tj. nenáhodné veličiny). V případě náhodných veličin předpokládáme, že tyto jsou nezávislé na všech prvcích vektoru  $\epsilon$  a jejich funkce hustoty pravděpodobnosti je  $p(X|\lambda)$ , kde  $\lambda$  je vektor parametrů, který neobsahuje  $\beta$  ani  $h$ .

Kovarianční matice příslušná vektoru parametrů či náhodných složek je matice, která obsahuje na hlavní diagonále postupně rozptyly jednotlivých prvků tohoto vektoru. Kovariance mezi jednotlivými prvky je pak obsahem prvků mimo diagonálu (kovarianční matice je symetrická matice), tedy

$$\begin{aligned} \text{var}(\epsilon) &= \begin{bmatrix} \text{var}(\epsilon_1) & \text{cov}(\epsilon_1, \epsilon_2) & \cdots & \text{cov}(\epsilon_1, \epsilon_N) \\ \text{cov}(\epsilon_1, \epsilon_2) & \text{var}(\epsilon_2) & \cdots & \cdot \\ \cdot & \text{cov}(\epsilon_2, \epsilon_3) & \cdots & \cdot \\ \cdot & \cdot & \cdots & \text{cov}(\epsilon_{N-1}, \epsilon_N) \\ \text{cov}(\epsilon_1, \epsilon_N) & \cdot & \cdots & \text{var}(\epsilon_N) \end{bmatrix} \\ &= \begin{bmatrix} h^{-1} & 0 & \cdot & \cdot & 0 \\ 0 & h^{-1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & h^{-1} \end{bmatrix}. \end{aligned}$$

Jinými slovy slovy, výraz  $\text{var}(\epsilon) = h^{-1} I_N$  je kompaktnější vyjádření předpokladu, že  $\text{var}(\epsilon_i) = h^{-1}$  (konstantní rozptyl) a  $\text{cov}(\epsilon_i, \epsilon_j) = 0$  (nezávislost jednotlivých náhodných složek) pro  $i, j = 1, \dots, N$  a  $i \neq j$ .



Z druhého předpokladu (týkajícího se matice vysvětlujících proměnných) vyplývá, že lze věrohodnostní funkci definovat jako  $p(y|X, \beta, h)$ , kdy pro zjednodušení zápisu matici  $X$  z tohoto výrazu vypouštíme.

Z definice vícerozměrného normálního rozdělení lze psát věrohodnostní funkci v podobě

$$p(y|\beta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y - X\beta)' (y - X\beta) \right] \right\}. \quad (3.3)$$

Opět je obvyklý zápis věrohodnostní funkce pomocí OLS odhadů, tedy:

$$\nu = N - k, \quad (3.4)$$

$$\hat{\beta} = (X'X)^{-1} X'y, \quad (3.5)$$

$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\nu}, \quad (3.6)$$

díky čemuž si můžeme snadno ověřit, že věrohodnostní funkci lze přepsat do podoby:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{1}{2}} \exp \left[ -\frac{h}{2} (\beta - \hat{\beta})' X'X (\beta - \hat{\beta}) \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp \left[ -\frac{h\nu}{2s^{-2}} \right] \right\}. \quad (3.7)$$

### 3.3 Apriorní hustota

S ohledem na tvar věrohodnostní funkce (3.7) je odpovídá přirozeně konjugovaná apriorní hustota normálnímu-gama rozdělení. Pokud tedy předpokládáme apriorní podmíněnou hustotu pravděpodobnosti pro  $\beta$  a apriorní hustotu pro  $h$  ve tvaru

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V}) \quad h \sim G(\underline{s}^{-2}, \underline{\nu}),$$

potom i příslušné posteriorní hustoty budou z těchto tříd rozdělení. Sdruženou apriorní hustotu pravděpodobnosti tedy zapíšeme v podobě:

$$\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}). \quad (3.8)$$

Jedná se o téměř identický zápis jako (2.7). Pouze  $\underline{\beta}$  je v tomto případě  $k$ -rozměrný vektor obsahující apriorní střední hodnoty  $k$  regresních koeficientů  $\beta_1, \dots, \beta_k$  a  $\underline{V}$  je pozitivně definitní matice rozměru  $k \times k$ , která je pevnou součástí apriorní kovarianční matice parametrů (díky normálnímu-gama rozdělení se nejedná přímo o kovarianční matici). Sdruženou apriorní hustotu pravděpodobnosti můžeme zapsat jako  $p(\beta, h) = f_{NG}(\beta, h|\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu})$ .

### 3.4 Posteriorní hustota

Posteriorní hustotu pravděpodobnosti parametrů získáme násobením apriorní hustoty s věrohodnostní funkcí. Podobně jako v předchozí kapitole vyjdeme z

definice normálního-gama rozdělení pro apriorní hustotu pravděpodobnosti a upravíme věrohodnostní funkci (3.7) roznásobením mocnin se společným základem (kdy základ opišeme a exponenty sečteme). Jediným rozdílem je zde práce s maticemi. Můžeme tedy psát:

$$\begin{aligned} p(\beta, h|y) &\propto \left\{ h^{\frac{k}{2}} \exp \left[ -\frac{h}{2} (\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) \right] \right\} \cdot \left\{ h^{\frac{\nu-2}{2}} \exp \left[ -\frac{h\nu s^2}{2} \right] \right\} \\ &\quad \times \left\{ h^{\frac{N}{2}} \exp \left[ -\frac{h}{2} \left\{ \nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right\} \right] \right\} \\ &= h^{\frac{k+\nu-2+N}{2}} \exp \left[ -\frac{h}{2} \left( \nu s^2 + \nu s^2 + (\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) \right. \right. \\ &\quad \left. \left. + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right) \right]. \end{aligned}$$

Stejně jako v předchozí kapitole označíme  $\bar{\nu} = \underline{\nu} + N$ . Parametr  $\beta$  se nachází pouze v části exponentu, konkrétně v

$$(\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}),$$

a ten snadno zapíšeme jako

$$(\hat{\beta} - \underline{\beta})' X' X \bar{V} \underline{V}^{-1} (\hat{\beta} - \underline{\beta}) + (\beta - \bar{\beta})' \bar{V}^{-1} (\beta - \bar{\beta}),$$

kde

$$\begin{aligned} \bar{V} &= (\underline{V}^{-1} + X' X)^{-1}, \\ \bar{\beta} &= \bar{V} \cdot (\underline{V}^{-1} \underline{\beta} + X' X \hat{\beta}). \end{aligned}$$

Opět nebude od věci ukázat si, proč tomu tak je. V první fázi roznásobíme obsah závorek a sdružíme členy obsahující parametr  $\beta$ , čímž dostaneme

$$\beta' \underline{V}^{-1} \beta - \beta' \underline{V}^{-1} \underline{\beta} - \underline{\beta}' \underline{V}^{-1} \beta + \underline{\beta}' \underline{V}^{-1} \underline{\beta} + \beta' X' X \beta - \beta' X' X \hat{\beta} - \hat{\beta}' X' X \beta + \hat{\beta}' X' X \hat{\beta}$$

Uspořádáním členů a zejména pak využitím vztahů pro  $\bar{V}$  a  $\bar{\beta}$  lze psát

$$\begin{aligned} &\beta' \underline{V}^{-1} \beta + \beta' X' X \beta - \beta' (\underline{V}^{-1} \underline{\beta} + X' X \hat{\beta}) - (\underline{\beta}' \underline{V}^{-1} + \hat{\beta}' X' X) \beta \\ &+ \underline{\beta}' \underline{V}^{-1} \underline{\beta} + \hat{\beta}' X' X \hat{\beta} \\ &= \beta' \bar{V}^{-1} \beta - \beta' \bar{V}^{-1} \bar{\beta} - \bar{\beta}' \bar{V}^{-1} \beta + \underline{\beta}' \underline{V}^{-1} \underline{\beta} + \hat{\beta}' X' X \hat{\beta} \end{aligned}$$

Po přidání členu  $(\bar{\beta}' \bar{V}^{-1} \bar{\beta} - \bar{\beta}' \bar{V}^{-1} \bar{\beta})$  získáme

$$(\beta - \bar{\beta})' \bar{V}^{-1} (\beta - \bar{\beta}) - \bar{\beta}' \bar{V}^{-1} \bar{\beta} + \underline{\beta}' \underline{V}^{-1} \underline{\beta} + \hat{\beta}' X' X \hat{\beta}.$$

Využíváme zde toho, že  $\underline{V}$  i  $\bar{V}$  jsou symetrické matice (stejně jako jejich inverze), tedy např.  $\underline{V} = \underline{V}'$ .<sup>4</sup> Dále se můžeme soustředit na druhou část výrazu (po

<sup>4</sup>Této vlastnosti je možno využít i v úvodu odvození, kdy platí  $(\beta' \underline{V}^{-1} \underline{\beta} = (\beta' \underline{V}^{-1} \underline{\beta})' = \underline{\beta}' \underline{V}^{-1} \beta)$ . Výsledkem součinu je totiž skalár, který i po transpozici zůstává tím samým skalárem.

znaménku mínus) a dosadit původní výraz pro  $\bar{\beta}$ , tedy

$$\begin{aligned}
& -\underline{\beta}'\underline{V}^{-1}\underline{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{\beta} + \hat{\beta}'X'X\hat{\beta} \\
& = -(\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta})'\underline{V}'\underline{V}^{-1}\underline{V}(\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}) + \underline{\beta}'\underline{V}^{-1}\underline{\beta} + \hat{\beta}'X'X\hat{\beta} \\
& = -\underline{\beta}'\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{\beta} - \underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} - \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\underline{\beta} - \hat{\beta}'X'X\underline{V}X'X\hat{\beta} \\
& \quad + \underline{\beta}'\underline{V}^{-1}\underline{\beta} + \hat{\beta}'X'X\hat{\beta} \\
& = -\underline{\beta}'\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{\beta} - \underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} - \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\underline{\beta} - \hat{\beta}'X'X\underline{V}X'X\hat{\beta} \\
& \quad + \underline{\beta}'\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{\beta} + \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\hat{\beta} \\
& = -\underline{\beta}'\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{\beta} - \underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} - \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\underline{\beta} - \hat{\beta}'X'X\underline{V}X'X\hat{\beta} \\
& \quad + \underline{\beta}'\underline{V}^{-1}\underline{V}\underline{V}^{-1}\underline{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} + \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\hat{\beta} + \hat{\beta}'X'X\underline{V}X'X\hat{\beta} \\
& = -\underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} - \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\underline{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} + \hat{\beta}'X'X\underline{V}\underline{V}^{-1}\hat{\beta} \\
& = (\hat{\beta} - \underline{\beta})'X'X\underline{V}\underline{V}^{-1}(\hat{\beta} - \underline{\beta}).
\end{aligned}$$

Platí totiž  $\underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta} = (\underline{\beta}'\underline{V}^{-1}\underline{V}X'X\hat{\beta})' = \underline{\beta}'X'X\underline{V}\underline{V}^{-1}\hat{\beta}$ . Využije se zde symetrie matic  $V$  a rovněž i skutečnost, že se jedná o transpozici skaláru.

Další postup je zase v souladu s tím, co bylo řečeno v předchozí kapitole. Parametr  $\beta$  se po naší úpravě nachází již jen ve členu  $(\beta - \bar{\beta})'\underline{V}^{-1}(\beta - \bar{\beta})$ . Díky tomu je jádrová hustota pro  $\beta|y, h$  dána jako

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

což je jádrová hustota normální hustoty pravděpodobnosti a platí tedy, že

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

Marginální posteriorní hustotu pro  $h$  lze odvodit na základě skutečnosti, že

$$p(h|y) = \int p(\beta, h|y)d\beta = \int p(h|y)p(\beta|y, h)d\beta.$$

Protože každá hustota pravděpodobnosti integrovaná přes všechny meze je rovna jedné, lze tímto vyintegrovat část sdružené posteriorní hustoty pravděpodobnosti zahrnující funkci normální hustoty pravděpodobnosti pro  $\beta$ , tedy  $p(\beta|y, h) \propto h^{\frac{k}{2}} \exp[-\frac{h}{2}(\beta - \bar{\beta})'\underline{V}^{-1}(\beta - \bar{\beta})]$ . Zůstává nám tedy

$$\begin{aligned}
p(h|y) & \propto h^{\frac{\bar{v}-2}{2}} \exp \left[ -\frac{h}{2} \left\{ \underline{\nu}s^2 + \nu s^2 + (\hat{\beta} - \underline{\beta})'X'X\underline{V}\underline{V}^{-1}(\hat{\beta} - \underline{\beta}) \right\} \right] \\
& = h^{\bar{v}-2} 2 \exp \left[ -\frac{h}{2} \underline{\nu}s^2 \right],
\end{aligned}$$

kde

$$\underline{\nu}s^2 = \underline{\nu}s^2 + \nu s^2 + (\hat{\beta} - \underline{\beta})'X'X\underline{V}\underline{V}^{-1}(\hat{\beta} - \underline{\beta}).$$

Výraz pro  $p(h|y)$  je jádrová hustota gama rozdělení, z čehož vyplývá, že

$$p(h|y) \sim G(\bar{s}^{-2}, \bar{\nu}).$$

Protože rozdělení  $\beta|y, h$  je normální a  $h|y$  je gama, je zřejmé, že sdružená hustota rozdělení pro parametry  $\beta$  a  $h$  odpovídá skutečně normálnímu-gama rozdělení, v našem značení tedy  $NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ . Apriorní i posteriorní hustota jsou z normálního-gama rozdělení, čímž je ověřena konjugovanost apriorní hustoty.

Pokud si tedy shrneme dosažené výsledky, máme posteriorní hustotu v podobě:

$$\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu}), \quad (3.9)$$

kde

$$\bar{V} = (\underline{V}^{-1} + X'X)^{-1}, \quad (3.10)$$

$$\bar{\beta} = \bar{V}(\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}), \quad (3.11)$$

$$\bar{\nu} = \underline{\nu} + N \quad (3.12)$$

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

$$\bar{\nu}\bar{s}^2 = \underline{\nu}s^2 + \nu s^2 + (\hat{\beta} - \underline{\beta})'[\underline{V} + (X'X)^{-1}]^{-1}(\hat{\beta} - \underline{\beta}). \quad (3.13)$$

Není nijak obtížné ověřit si ekvivalenci výrazů  $X'X\bar{V}V^{-1}$  a  $[\underline{V} + (X'X)^{-1}]^{-1}$ . Stačí si v prvním výrazu dosadit původní výraz za  $\bar{V}$  a uvědomit si, že  $(X'X) = \frac{1}{(X'X)^{-1}}$  a stejně tak  $\underline{V}^{-1} = \frac{1}{\underline{V}}$ . Po roznásobení členů v takto získaném jmenovateli je závěr jednoznačný.

Zajímá-li nás místo sdružené hustoty marginální hustota pro  $\beta$ , je třeba ji integrovat přes  $h$ . Výsledkem je pak vícerozměrné  $t$ -rozdělení:

$$\beta|y \sim t(\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu}), \quad (3.14)$$

z čehož vyplývá:

$$E(\beta|y) = \bar{\beta}, \quad (3.15)$$

$$var(\beta|y) = \frac{\bar{\nu}\bar{s}^2}{\bar{\nu} - 2}\bar{V}. \quad (3.16)$$

Postup důkazu je zcela identický s postupem z předchozí kapitoly a nebudu ho zde proto rozvádět. Pokud jde o přesnost chyby  $h$ , pak z normálního-gama rozdělení plyne:

$$h|y \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (3.17)$$

$$E(h|y) = \bar{s}^{-2}, \quad (3.18)$$

$$var(h|y) = \frac{2\bar{s}^{-4}}{\bar{\nu}}. \quad (3.19)$$

Výsledky jsou (jak již bylo několikrát zdůrazněno) podobné těm z předchozí kapitoly, jen skaláry nahradily vektory a matice. Např.  $\hat{\beta}$  je vektor, matice

$(X'X)^{-1}$  hraje podobnou roli jako skalár  $\frac{1}{\sum x_i^2}$ ,  $\bar{V}$  je matice rozměru  $k \times k$  apod. V kapitole 2 jsme mohli interpretovat posteriorní střední hodnotu parametru  $\beta$ ,  $\bar{\beta}$ , jako vážený průměr apriorní střední hodnoty,  $\underline{\beta}$  a OLS odhadu,  $\hat{\beta}$ , kdy váhy reflektují sílu apriorní informace obsažené v  $\underline{V}^{-1}$  a v datech ( $\sum x_i^2$ ). Stejnou úvahu lze použít i zde, s tím rozdílem, že se jedná o maticově vážený průměr apriorní a datové informace.

Opět je samozřejmě nutné rozhodnout o volbě apriorních hyperparametrů, konkrétně tedy  $\underline{\beta}$ ,  $\underline{V}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$ . V mnoha případech nám při jejich volbě pomůže ekonomická teorie, zkušenost, cit či znalost předchozích empirických studií založených na jiných datových vzorcích. V tomto ohledu nám zjednodušuje volbu apriorních hyperparametrů skutečnost, že využíváme přirozeně konjugovanou apriorní hustotu. Tento druh apriorní hustoty lze chápat a interpretovat jako výsledek fiktivní datové sady generované stejným procesem jako skutečná data. Jiný přístup k volbě apriorních hyperparametrů spočívá ve volbě řady různých hodnot v rámci citlivostní analýzy, případně můžeme pracovat s relativně neinformativním priorem. Můžeme tedy volit  $\underline{\nu}$  hodnotově výrazně nižší než je  $N$  a  $\underline{V}$  jako "dostatečně velké". Vzhledem k tomu, že pracujeme s maticemi, není výraz "dostatečně velký" zrovna snadno představitelný. Maticové zobecnění nerovnosti  $a > b$ , kde  $a$  i  $b$  jsou skaláry, je obvykle chápáno tak, že rozdíl  $A - B$  je pozitivně definitní matice (porovnáváme tedy čtvercové matice stejných rozměrů). Měřítkem velikosti matice je její determinant. Pokud tedy budeme hovořit o tom, že matice "A by měla být relativně velká vzhledem k B", máme tím na mysli to, že  $A - B$  by měla být pozitivně definitní matice s hodnotově velkým determinanem.

Čistě neinformativní prior můžeme získat volbou  $\underline{\nu} = 0$  a  $\underline{V}^{-1}$  jako dostatečně malé. To lze řešit různými způsoby. Jedním z nich je nastavení  $\underline{V}^{-1} = cI_k$ , kde  $c$  je skalár blížíci se limitně k nule. V takovémto případě bude platit  $\beta, h|y \sim NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ , kde

$$\bar{V} = (X'X)^{-1}, \quad (3.20)$$

$$\bar{\beta} = \hat{\beta}, \quad (3.21)$$

$$\bar{\nu} = N, \quad (3.22)$$

$$\bar{\nu}\bar{s}^2 = \nu s^2. \quad (3.23)$$

Všechny tyto vztahy v sobě obsahují pouze datovou informaci a odpovídají odhadům metodou nejmenších čtverců. Tento druh neinformativního prioru je nepravým priorem a lze jej tedy vyjádřit v podobě:

$$p(\beta, h) \propto \frac{1}{h}. \quad (3.24)$$

### 3.5 Porovnání modelů

Lineární regresní model s  $k$  vysvětlujícími proměnnými v sobě zahrnuje řadu možností porovnání větší škály různě specifikovaných modelů. V rámci porovnání jednotlivých specifikací modelů budeme rozlišovat dva případy: v prvním

případě se modely liší omezeními ve tvaru nerovnosti, která jsou kladena na jednotlivé parametry (či obecněji parametrický prostor, jakožto množinu parametrů), ve druhém případě jsou modely odlišeny různě specifikovanými omezeními jednotlivých parametrů ve tvaru rovnosti.

### 3.5.1 Porovnání modelů s omezením ve tvaru nerovnosti

V řadě aplikací se můžeme zaměřit na analýzu jen části parametrického prostoru. Předpokládejme např. model, ve kterém je závisle proměnná objem prodeje nějakého produktu a jednou z vysvětlujících proměnných jsou výdaje na určitou marketingovou kampaň. V takovémto případě by nás mohlo zajímat, jestli reklamní kampaň zvyšuje prodejnost výrobku, konkrétně se tedy zaměříme na to, zdali příslušný regresní koeficient (vyjadřující efekt reklamní kampaně) je kladný resp. nezáporný. V jiné modelové aplikaci nás může zajímat, zdali výnosy z rozsahu jsou rostoucí či klesající. To v kontextu regresního modelu znamená zodpovědět otázku, zdali určitá kombinace parametrů je menší či větší než jedna. Oba příklady v sobě zahrnují omezení ve tvaru nerovnosti, které klademe na určité regresní koeficienty.

Předpokládejme tedy, že naše omezení mají tvar

$$R\beta \geq r, \quad (3.25)$$

kde  $R$  je známá matice rozměru  $J \times k$  a  $r$  je známý vektor rozměru  $J \times 1$ . Takovýto předpoklad nám umožňuje formulovat  $J$  lineárních omezení ve tvaru nerovnosti, která souvisejí s jednotlivými regresními koeficienty  $\beta$ . Abychom se vyhnuli tomu, že se některá z omezení budou vzájemně duplikovat, musíme předpokládat plnou hodnotu matice  $R$ , tedy  $\text{rank}(R) = J$ . Můžeme tedy definovat dva modely  $M_1$  a  $M_2$ :

$$M_1 : R\beta \geq r$$

,

$$M_2 : R\beta \not\geq r$$

, přičemž značení v definici modelu  $M_2$  znamená, že  $J$  omezení v modelu  $M_1$  není splněno.

Pro modely definované tímto způsobem je výpočet posteriorního podílu šancí snadný a není zde ani problém s využitím neinformativního prioru:

$$PO_{12} = \frac{p(M_1|y)}{p(M_2|y)} = \frac{p(R\beta \geq r|y)}{p(R\beta \not\geq r|y)}. \quad (3.26)$$

Protože je posteriorní hustota pravděpodobnosti pro vektor parametrů  $\beta$  z více-rozměrného Studentova  $t$ -rozdělení, platí, že i posteriorní hustota lineární kombinace tohoto vektoru  $p(R\beta|y)$  je z  $t$ -rozdělení. Kvalitní software jako např. MATLAB dokáže intervalové pravděpodobnosti pro vícerozměrné  $t$ -rozdělení spočítat, tudíž i  $p(R\beta \geq r)$  lze snadno spočítat. Pokud je počet omezení  $J = 1$ , lze s úspěchem využít i statistických tabulek pro jenozměrné  $t$ -rozdělení.

### 3.5.2 Omezení ve tvaru rovnosti

Porovnání modelů, které v sobě zahrnují omezení ve tvaru rovnosti, je trošku komplikovanější a vznikají zde problémy s použitím neinformativního prioru. Lze se setkat s dvěma typy porovnání modelů:

- V prvním případě nás může zajímat porovnání modelu  $M_1$ , který v sobě obsahuje omezení  $R\beta = r$ , s modelem  $M_2$ , který toto omezení nemá. Model  $M_1$  je příkladem vnořeného (*nested*) modelu, to znamená, že model  $M_1$  získáme z modelu  $M_2$  zohledněním omezení  $R\beta = r$ .
- V druhém případě nás zajímá porovnání modelu  $M_1 : y = X_1\beta_{(1)} + \epsilon_1$  s modelem  $M_2 : y = X_2\beta_{(2)} + \epsilon_2$ , kde  $X_1$  a  $X_2$  jsou matice obsahující v případě nutnosti i zcela odlišné vysvětlující proměnné. Označení  $\beta_{(j)}$  označuje koeficienty v  $j$ -tém modelu (jelikož máme dva modely, tak  $j = 1, 2$ ). Jedná se o případ porovnání nevnořených (*non-nested*) modelů.

Obě výše uvedené typové kategorie porovnání modelů můžeme řešit v kontextu formálního zápisu

$$M_j : y_j = X_j\beta_{(j)} + \epsilon_j, \quad (3.27)$$

kde  $j = 1, 2$  označuje naše dva modely,  $y_j$  bude definováno níže,  $X_j$  je matice vysvětlujících proměnných rozměru  $N \times k_j$ ,  $\beta_{(j)}$  je  $k_j$ -rozměrný vektor regresních koeficientů a  $\epsilon_j$  je  $N$ -rozměrný vektor chyb pocházející z normálního rozdělení  $N(0_N, h_j^{-1}I_N)$ .

Případ porovnání nevnořených modelů získáme volbou  $y_1 = y_2$ . Pro případ vnořených modelů je třeba vycházet z neomezeného LRM z rovnice (3.2). Model  $M_2$  je právě tento neomezený model. Volíme tedy  $y_2 = y$ ,  $X_2 = X$  a  $\beta_{(2)} = \beta$ . Model  $M_1$ , který obsahuje omezení  $R\beta = r$ , lze definovat zahrnutím těchto omezení do vysvětlujících proměnných. Na tomto základě je pak třeba rovněž předefinovat vysvětlovanou proměnnou. Princip (bez nějakého formálního zobecnění) lze ukázat na následujících příkladech:

1. Omezení ve tvaru  $\beta_m = 0$  implikuje, že matice  $X_1$  odpovídá matici  $X$  s vypuštěnou  $m$ -tou vysvětlující proměnnou.
2. Omezení ve tvaru  $\beta_m = r$  implikuje, že matice  $X_1$  je matice  $X$  s vypuštěnou  $m$ -tou vysvětlující proměnnou a  $y_1 = y - rx_m$ , kdy vektor  $x_m$  je  $m$ -tý sloupec matice  $X$ .
3. Omezení  $\beta_2 - \beta_3 = 0$  lze získat vypuštěním příslušných vysvětlujících proměnných ( $x_2$  a  $x_3$ , což jsou druhý a třetí sloupec matice  $X$ ) a definováním nové vysvětlující proměnné ( $x_{23}$ ), která je součtem těchto vypuštěných proměnných, tedy  $x_{23} = x_2 + x_3$ .

Vícenásobná a komplikovanější omezení lze řešit zobecněním výše uvedených příkladů.

Označíme-li normální-gama apriorní rozdělení dvou modelů jako

$$\beta_{(j)}, h_j | M_j \sim NG(\underline{\beta}_j, \underline{V}_j, \underline{s}_j^{-2}, \underline{\nu}) \quad (3.28)$$

pro  $j = 1, 2$ , má posteriorní hustota podobu

$$\beta_{(j)}, h_j | y_j \sim, NG(\bar{\beta}_j, \bar{V}_j, \bar{s}_j^{-2}, \bar{\nu}_j) \quad (3.29)$$

kde

$$\bar{V}_j = (\underline{V}_j^{-1} + X_j' X_j)^{-1}, \quad (3.30)$$

$$\bar{\beta}_j = \bar{V}_j (\underline{V}_j^{-1} \underline{\beta}_j + X_j' X_j \hat{\beta}_j), \quad (3.31)$$

$$\bar{\nu}_j = \underline{\nu}_j + N \quad (3.32)$$

a  $\bar{s}^{-2}$  je implicitně definováno ve výrazu

$$\bar{\nu}_j \bar{s}_j^2 = \underline{\nu}_j \underline{s}_j^2 + \nu_j s_j^2 + (\hat{\beta}_j - \underline{\beta}_j)' [\underline{V}_j + (X_j' X_j)^{-1}] (\hat{\beta}_j - \underline{\beta}_j). \quad (3.33)$$

Výrazy  $\hat{\beta}_j$ ,  $s_j^2$  a  $\nu_j$  jsou OLS odhady analogické těm definovaným ve vztazích (3.4)–(3.6).

Odvození marginální věrohodnosti pro každý z modelů a tedy následně i posteriorního podílu šancí je obdobné jako v předchozí kapitole. Marginální věrohodnost tak lze spočítat jako

$$p(y_j | M_j) = c_j \left( \frac{|\bar{V}_j|}{|\underline{V}_j|} \right)^{\frac{1}{2}} (\bar{\nu}_j \bar{s}_j^2)^{-\frac{\bar{\nu}_j}{2}} \quad (3.34)$$

pro  $j = 1, 2$ , přičemž

$$c_j = \frac{\Gamma\left(\frac{\bar{\nu}_j}{2}\right) (\underline{\nu}_j \underline{s}_j^2)^{\frac{\underline{\nu}_j}{2}}}{\Gamma\left(\frac{\underline{\nu}_j}{2}\right) \pi^{\frac{N}{2}}}. \quad (3.35)$$

Posteriorní podíl šancí porovnávající modely  $M_1$  a  $M_2$  odpovídá výrazu

$$PO_{12} = \frac{c_1 \left( \frac{|\bar{V}_1|}{|\underline{V}_1|} \right)^{\frac{1}{2}} (\bar{\nu}_1 \bar{s}_1^2)^{-\frac{\bar{\nu}_1}{2}} p(M_1)}{c_2 \left( \frac{|\bar{V}_2|}{|\underline{V}_2|} \right)^{\frac{1}{2}} (\bar{\nu}_2 \bar{s}_2^2)^{-\frac{\bar{\nu}_2}{2}} p(M_2)}. \quad (3.36)$$

Faktory ovlivňující posteriorní podíl šancí byly diskutovány v předchozí kapitole 2. Posteriorní podíl šancí tedy závisí na apriorním podílu šancí, zvýhodňuje soulad modelu s daty, koherenci mezi apriorní a datovou informací a šetrnost pokud jde o počet vysvětlujících proměnných.

Problém nám však opět může způsobit použití neinformativní apriorní hustoty. Tento druh prioru jsme zavedli způsobem, kdy  $\underline{\nu} = 0$  a  $\underline{V}^{-1} = cI_k$ , přičemž  $c$  se limitně blížílo nule. Volně řečeno,  $\underline{\nu} = 0$  implikuje, že zde není apriorní informace o přesnosti chyby  $h$  a nastavení  $c$  limitně se blížíci nule znamená neexistenci apriorní informace o regresních koeficientech  $\beta$  (jejich rozptyl je nekonečně velký). V následující části budeme tyto dvě různé specifikace analyzovat odděleně. Důležitým závěrem bude to, že je rozumné využít neinformativního prioru pro  $h_j$ , kde  $j = 1, 2$ , ale není rozumné užít neinformativních priorů pro



$\beta_{(j)}$ . Důvod je takový, že přesnost chyby je parametr společný pro oba modely a má v obou modelech stejnou interpretaci. Naopak  $\beta_{(1)}$  a  $\beta_{(2)}$  nejsou stejné (respektive nemusí být) a v případech, kdy počty parametrů se budou v obou modelech lišit, tedy  $k_1 \neq k_2$ , bude použití neinformativních priorů působit vážné problémy v chování posteriorního podílu šancí.

Obecně lze formulovat následující pravidlo (platné pro jakýkoliv model, tedy nejen regresní): pokud porovnáme modely využitím posteriorního podílu šancí, pak je žádoucí použití neinformativních priorů pro parametry, které jsou společné oběma modelům. Čistě informativní priory je nutné užít pro všechny ostatní parametry.

Pokud nastavíme  $\nu_1 = \nu_2 = 0$  (přesněji oba členy se budou limitně k nule blížit stejným tempem), potom se nám vztah v rovnici (3.36) zjednoduší, neboť  $c_1 = c_2$ . Posteriorní podíl šancí má přesto stále rozumnou interpretaci, neboť v sobě obsahuje měřítko souladu modelu s daty, tj.  $s_j^2$ , jakožto koherenci mezi apriorní a datovou informací, apod.

Použijeme-li neinformativní priory pro  $\beta_{(j)}$ , nastanou problémy zejména v případě, kdy  $k_1 \neq k_2$ . V případě porovnání nevnořených modelů může být interpretace parametrů  $\beta_{(1)}$  a  $\beta_{(2)}$  zcela odlišná. V případě vnořených modelů nám restrikce kladené na model  $M_1$  zajistí, že  $\beta_{(1)}$  má menší dimenzi než  $\beta_{(2)}$  a tedy  $k_1 < k_2$ . Rozdílnost počtu vysvětlujících proměnných,  $k_1 \neq k_2$ , je tedy obvyklá. Problém s interpretací posteriorních podílů šancí nastává díky členu  $|\underline{V}_j|$ . Pokud  $\underline{V}_j^{-1} = cI_k$ , potom  $|\underline{V}_j| = \frac{1}{c^{k_j}}$ . Blíží-li se  $c$  k nule, potom se členy zahrnující  $c$  v (3.36) vzájemně nevyruší. Pokud je apriorní podíl šancí kladný a konečný, nastane v případě nerovnosti  $k_1 < k_2$  situace, kdy  $PO_{12}$  je nekonečno. V případě, kdy  $k_1 > k_2$ , bude podíl šancí  $PO_{12}$  roven nule. Jinými slovy, posteriorní podíl šancí vždy upřednostní model s menším počtem parametrů bez ohledu na data, v limitním případě pak tak bude zcela upřednostněn šetrnější model (z hlediska počtu použitých parametrů resp. vysvětlujících proměnných). Tento výsledek nás nutí k upřednostnění použití informativního prioru přinejmenším u koeficientů, které jsou odlišné v obou porovnávaných modelech.

V případě rovnosti počtu vysvětlujících proměnných ( $k_1 = k_2$ ) má posteriorní podíl šancí podobu

$$PO_{12} = \frac{(|X_1'X_1|)^{\frac{1}{2}} (\nu_1 s_1^2)^{-\frac{N}{2}} p(M_1)}{(|X_2'X_2|)^{\frac{1}{2}} (\nu_2 s_2^2)^{-\frac{N}{2}} p(M_2)}. \quad (3.37)$$

Hodnota tohoto výrazu bude záviset na měrných jednotkách jednotlivých veličin. Pokud by vysvětlující proměnné v modelu  $M_1$  byly měřeny v jednotkách dolarů a toto měřítko bychom následně změnili na tisíce dolarů, přičemž v matici vysvětlujících proměnných druhého modelu ( $X_2$ ) bychom tuto změnu nezohlednili, změnil by se i náš posteriorní podíl šancí. To je nepříjemná vlastnost, která opět vede k upřednostnění informativního prioru i v případě, kdy  $k_1 = k_2$ . V případě informativního prioru totiž tento problém nenastává. Jako příklad uvažujme regresní model závislosti ceny domu na různých charakteristikách, kdy  $x_2$  může být rozloha obytné plochy v metrech čtverečních. Parametr  $\beta_2$  můžeme interpretovat jako měřítko toho, o kolik dolarů vzroste cena domu, jestliže zvýšíme

jeho obytnou plochu o jeden metr čtvereční a ostatní charakteristiky zůstanou nezměněny. Z ohledem na to pak volíme příslušný hyperparametr apriorní hustoty (resp. apriorní hustotu jako takovou) pro  $\beta_2$  (jeho apriorní střední hodnotu a rozptyl). Pokud by se měřítko  $x_2$  změnilo na stovky metrů čtverečních, pak se i interpretace parametru  $\beta_2$  změnila na měřítko toho, o kolik dolarů vzroste cena domu, zvýší-li se jeho obytná plocha o sto metrů čtverečních a ostatní charakteristiky zůstanou nezměněny. Tuto změnu reflektuje i volba apriorní hustoty. Pokud tedy volíme použití informativní apriorní hustoty, implicitně bereme do úvahy i příslušné měrné jednotky.<sup>5</sup> To však neplatí při použití neinformativního prioru.

Shrneme-li obsah této části kapitoly, pak je nutné zdůraznit, že užití neinformativního prioru je akceptovatelné, pokud se zaměříme na odhad parametrů či predikci a chceme zůstat "objektivní", tedy přesněji nechceme (nebo neumíme) do našich výpočtů vnášet nějakou apriorní informaci. V případě výpočtu posteriorního podílu šancí však volba neinformativního prioru akceptovatelná není.

### 3.5.3 Intervaly nejvyšší posteriorní hustoty (HPDI)

Tradiční techniky bayesovského porovnání modelů jsou založeny na intuitivní myšlence, že  $p(M_j|y)$  je souhrnem veškerých našich znalostí a nejistoty týkajících se modelu  $M_j$  po konfrontaci s daty. Výpočet rozumných posteriorních pravděpodobností modelů v sobě obvykle zahrnuje práci s informativními prioru. Pokud bychom však přesto chtěli testovat či porovnávat modely za použití neinformativních priorů, můžeme využít jiných technik. Tyto techniky však již nejsou do té míry intuitivní jako bayesovské pravděpodobnosti modelu a jejich oprávněnost je dána *ad hoc*. V této části zavedeme pojem „interval nejvyšší posteriorní hustoty“ (Highest Posterior Density Interval - HPDI, či jenom HDI) a ukážeme si, jak jej lze využít pro *ad hoc* porovnání vnořených modelů.

Následující definice jsou definovány v kontextu vektoru parametrů  $\beta$  normálního lineárního regresního modelu, nicméně jsou dostatečně obecné a lze je tak využít pro parametry jakéhokoliv modelu. Předpokládejme tedy, že prvky vektoru regresních koeficientů  $\beta$  se mohou nacházet v intervalu  $(-\infty, \infty)$ , což označíme jako  $\beta \in R^k$  (zde  $R$  označuje množinu reálných čísel). Nechť  $\omega = g(\beta)$  je  $m$ -rozměrný vektor funkcí  $\beta$  definovaný na množině  $\Omega$ , kdy  $m \leq k$ . Nechť  $C$  je oblast v rámci  $\Omega$ , tedy  $C \subseteq \Omega$ .

**Definice 3.1** (Přijatelné množiny (Credible Sets)). Množina  $C \subseteq \Omega$  je  $100(1 - \alpha)\%$  přijatelná množina vzhledem k  $p(\omega|y)$  jestliže

$$p(\omega \in C|y) = \int_C p(\omega|y)d\omega = 1 - \alpha$$

<sup>5</sup>Připomněme si v této souvislosti tvrzení o tom, že volbu apriorní hustoty nám ulehčuje skutečnost, že apriorní hustota odpovídá (pro „přirozeně konjugovaný“ případ) procesu generujícího fiktivní datové sady. Právě tento generující proces se odvíjí od použitého měřítka vysvětlujících proměnných, což musí být samozřejmě zohledněno v apriorní hustotě.

Jako příklad předpokládejme  $\omega = g(\beta) = \beta_j$ , tedy jediný regresní koeficient. Potom 95% přijatelný interval pro  $\beta_j$  je interval  $[a, b]$  takový, že

$$p(a \leq \beta_j \leq b|y) = \int_a^b p(\beta_j|y)d\beta_j = 0.95$$

Existuje obvykle nekonečně mnoho možných přijatelných intervalů. Předpokládejme příklad, kdy  $\beta_j|y$  pochází z  $N(0, 1)$ . Potom ze statistických tabulek pro standardizované normální vyplývá, že 95% přijatelný interval je  $[-1.96, 1.96]$  stejně jako  $[-1.75, 2.33]$ ,  $[-1.64, \infty]$  atd. Abychom si z této nekonečné množiny intervalů vybrali jediného reprezentant, provádíme volbu obvykle tak, že zvolíme ten interval, který má nejmenší rozsah. Pro standardní normální rozdělení je  $[-1.96, 1.96]$  nejkratší přijatelný interval. Pro tento druh výběru užíváme označení interval s nejvyšší posteriorní hustotou (*Highest Posterior Density Interval*).

**Definice 3.2** (Interval nejvyšší posteriorní hustoty (HPDI)).  $(1 - \alpha) \cdot 100\%$  interval nejvyšší posteriorní hustoty pro množinu  $\omega$  je  $(1 - \alpha) \cdot 100\%$  přijatelný interval množiny  $\omega$ , pro který platí, že má nejmenší rozsah oproti jiným  $(1 - \alpha) \cdot 100\%$  přijatelným intervalům množiny  $\omega$ .

Kromě bodových odhadů parametrů (v rámci bayesovského odhadu) je obvyklé uvádět i tyto intervaly nejvyšší posteriorní hustoty. Můžeme tak například uvádět posteriorní střední hodnotu a 95% HPDI pro  $\beta_j$ . Jsme si tak z 95 % jistí, že  $\beta_j$  leží uvnitř HPDI. Lze jej rovněž využít pro *ad hoc* porovnání modelů. Předpokládejme v této souvislosti dva NLRM a můžeme se zajímat o to, zdali  $j$ -tá vysvětlující proměnná má být zahrnuta do modelu. Dva modely které budeme uvažovat budou

$$M_1 : \beta_j = 0,$$

$$M_2 : \beta_j \neq 0.$$

Posteriorní analýzu lze provést (analyticky) standardním způsobem a HPDI pro  $\beta_j$  lze spočítat využitím vlastností Studentova  $t$ -rozdělení. Pokud tento HPDI neobsahuje nulu, potom je to důkaz či skutečnos hovořící v neprospěch modelu  $M_1$ . Pokud by závěr zněl, že HPDI obsahuje nulu, potom bychom měli důkaz ve prospěch modelu  $M_1$ . Tento postup lze zobecnit standardním způsobem pro ověření platnosti soustavy lineárních omezení ( $R\beta = r$ ).

Tento přístup je podobný přístupu klasické ekonometrie v rámci standardního testování hypotéz. Obvykle se tak testuje omezení  $\beta_j = 0$  (nevýznamnost parametru) pomocí konfidenčních intervalů (intervalů spolehlivosti) pro  $\beta_j$ . Pokud tento interval obsahuje nulu, pak je nulová hypotéza přijata, v opačném případě je zamítnuta. Tato podobnost je jen velmi hrubou, rámcovou intuicí. Konfidenční intervaly mají oproti HPDI velmi odlišnou interpretaci.

HPDI je obecným nástrojem, který existuje vždy, když existuje posteriorní hustota pravděpodobnosti. Lze jej tak využít i při použití neinformativní apriorní hustoty. Ačkoliv je však oprávněnost jejich využití při porovnání modelů smysluplná, má jen neformální charakter, který nemá pevnou oporu v teorii

pravděpodobnosti (oproti posteriornímu podílu šancí, či konfidenčním intervallům v klasické ekonometrii a statistice).

)ú

### 3.6 Predikční hustota

V této části si jen rozšíříme poznatky o predikční hustotě z minulé kapitoly pro případ více vysvětlujících proměnných. Předpokládejme, že nás zajímá predikční analýza  $T$  pozorovaných hodnot závisle proměnné  $y^* = (y_1^*, \dots, y_T^*)'$ , generovaných jako

$$y^* = X^* \beta + \epsilon^*, \quad (3.38)$$

kde  $\epsilon^*$  je nezávislé na  $\epsilon$  a má normální rozdělení  $N(0, h^{-1}I_T)$ . Matice  $X^*$  je matice rozměru  $T \times k$  analogická matici  $X$  a obsahující  $k$  vysvětlujících proměnných pro každé z  $T$  pozorování mimo dostupný datový vzorek.

Kroky vedoucí k odvození predikční hustoty pro  $y^*$  jsou zobecněním vztahů (2.37)–(2.40). Bayesiánská předpověď je založena na

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

Skutečnost, že  $\epsilon^*$  je nezávislé na  $\epsilon$ , současně implikuje nezávislost  $y$  na  $y^*$  a tedy  $p(y^*|y, \beta, h) = p(y^*|\beta, h)$ . Dále platí

$$p(y^*|\beta, h) = \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \exp \left[ -\frac{h}{2} (y^* - X^* \beta)' (y^* - X^* \beta) \right]. \quad (3.39)$$

Násobením vztahu (3.39) a posteriorní hustoty v (3.9) a následnou integrací získáme vícerozměrnou  $t$ -predikční hustotu pravděpodobnosti ve tvaru

$$y^*|y \sim t(X^* \bar{\beta}, \bar{s}^2 \{I_T + X^* \bar{V} X^{*'}\}, \bar{\nu}). \quad (3.40)$$

Na tomto základě lze provést veškerou predikční analýzu normálního lineárního modelu s přirozeně konjugovaným priorem.

### 3.7 Monte Carlo integrace

Porovnání modelů, predikci a posteriorní analýzu týkající se vektoru  $\beta$  lze provést analyticky za využití výsledků předchozích částí kapitoly. Protože marginální posteriorní hustota pro vektor parametrů  $\beta$  odpovídá vícerozměrnému Studentovu  $t$ -rozdělení, jsou i lineární kombinace vektoru  $\beta$  z vícerozměrného  $t$ -rozdělení. Pokud tak definujeme matici  $R$  podobně jako ve vztahu (3.25), lze provést posteriorní analýzu i pro lineární kombinaci našich parametrů, tedy  $R\beta$ . Marginální posteriorní hustota pro přesnost chyby  $h$  odpovídá gama rozdělení, tudíž lze využít vlastnosti tohoto rozdělení k analýze přesnosti chyby.

V některých případech nás však nemusí zajímat vektor  $\beta$  či  $R\beta$ , ale pozornost může být zaměřena na nelineární funkci parametrů  $\beta$ ,  $f(\beta)$ . Budeme předpokládat, že  $f(\cdot)$  je skalární funkce, ovšem techniku zmíněnou v tomto oddíle je možno rozšířit i pro případ vektorový.

Obecně nemusí platit, že posteriorní hustota pro  $f(\beta)$  bude spadat do třídy hustot se známými analytickými vlastnostmi. V tom případě je nutné využít simulační metody. Nejjednodušším simulačním algoritmem je Monte Carlo integrace zmiňovaná již v kapitole 1. V kontextu normálního lineárního regresního modelu lze zapsat základní teorém Monte Carlo integrace následovně:

**Teorém 3.1** (Monte Carlo integrace). Nechť  $\beta^{(s)}$  pro  $s = 1, \dots, S$  je náhodný výběr (vzorek) z  $p(\beta|y)$  a  $g(\cdot)$  je funkce a definujeme

$$\hat{g}_S = \frac{1}{S} \sum_{s=1}^S g(\beta^{(s)}), \quad (3.41)$$

potom  $\hat{g}_S$  konverguje k  $E[g(\beta)|y]$  pro  $S$  jdoucí k nekonečnu.

ú)

Není třeba nechat se zmást zavedením dvou funkcí  $f(\cdot)$  a  $g(\cdot)$ . Pokud budeme definovat  $g(\cdot) = f(\cdot)$  můžeme získat odhady  $E[f(\beta)|y]$  pro jakoukoliv funkci parametrů  $f(\cdot)$ . Pokud chceme spočítat nějaké další posteriorní charakteristiky funkce parametrů  $f(\beta)$ , je třeba funkci  $g(\cdot)$  zavést. Pro výpočet rozptylu  $var[f(\beta)|y]$  je třeba definovat  $g(\cdot) = f(\cdot)^2$  a spočítat pomocí vztahu (3.41)  $E[f(\beta)^2|y]$ .

Rovnice (3.41) nám říká, že pro daný náhodný výběr z posteriorního rozdělení příslušnému parametrům  $\beta$  je možno analyzovat jakoukoliv funkci těchto parametrů. Prakticky lze algoritmus výpočtu (např. pro MATLAB) shrnout do následujících kroků:

1. Vygenerujeme náhodný výběr  $\beta^{(s)}$  z posteriorní hustoty pro  $\beta$  dané vztahem (3.14), a to za použití generátoru náhodných čísel pro vícerozměrné  $t$ -rozdělení.
2. Spočítáme  $g(\beta^{(s)})$  a uchováme si tento výsledek.
3. Opakujeme kroky 1 a 2 celkem  $S$ -krát.
4. Spočítáme celkový průměr  $S$  náhodných výběrů  $g(\beta^{(1)}), \dots, g(\beta^{(S)})$ .

Těmito kroky získáme odhad  $E[g(\beta)|y]$  pro jakoukoliv funkci parametrů, která nás zajímá. Volbou  $S$  můžeme řídit velikost chyby aproximace, které se dopouštíme, neboť pouze jen volba  $S = \infty$  by nám dala skutečnou střední hodnotu funkce parametrů, která nás zajímá. Jak již bylo popsáno v kapitole 1, můžeme snadno získat i číselné vyjádření chyby aproximace, a to využitím centrální limitní věty. Přesněji řečeno, vyjdeme z definice chyby odhadu, a konkrétně využijeme tu vlastnost, že

$$\sqrt{S} \{\hat{g}_S - E[g(\beta)|y]\} \rightarrow N(0, \sigma_g^2) \quad (3.42)$$

pro  $S$  jdoucí k nekonečnu, přičemž  $\sigma_g^2 = \text{var}[g(\beta)|y]$ . Intuitivně nám tento vztah říká, že střední hodnota chyby odhadu je nulová, a variabilita chyby odhadu (tedy míra nejistoty spojená s tím, že máme konečný počet vzorků) se snižuje s růstem počtu generovaných vzorků  $S$ . Rozptyl chyby odhadu  $\sigma_g^2$  lze odhadnout Monte Carlo integrací a tento odhad označíme  $\hat{\sigma}_g^2$ . Použitím tohoto odhadu, vztahu (3.42) a vlastností normálního rozdělení můžeme psát:

$$\Pr \left\{ E[g(\beta)|y] - 1.96 \frac{\hat{\sigma}_g}{\sqrt{S}} \leq \hat{g}_S \leq E[g(\beta)|y] + 1.96 \frac{\hat{\sigma}_g}{\sqrt{S}} \right\} \approx 0.95. \quad (3.43)$$

Další úpravou můžeme zjistit přibližný 95% konfidenční interval pro střední hodnotu funkce parametrů, která nás zajímá,  $E[g(\beta)|y]$ , ve tvaru

$$\left[ \hat{g}_S - 1.96 \frac{\hat{\sigma}_g}{\sqrt{S}}, \hat{g}_S + 1.96 \frac{\hat{\sigma}_g}{\sqrt{S}} \right].$$

Tento výraz lze využít jako měřítko přesnosti našeho odhadu  $E[g(\beta)|y]$  nebo jako vodítko pro výběr velikosti počtu generovaných vzorků  $S$ . Alternativně lze využít i numerickou standardní chybu  $\frac{\hat{\sigma}_g}{\sqrt{S}}$ , která obsahuje stejnou informaci v poněkud kompaktnější podobě.

### 3.8 Empirická ilustrace

K ilustraci bayesiánské analýzy v rámci modelu vícenásobné regrese využijeme data obsahující údaje o prodejních cenách  $N = 546$  domů prodaných ve Windsoru (Kanada) v roce 1987. Data jsou obsahem souboru `hprice.txt`. Naším zájmem je nalézt faktory, které ovlivňují prodejní cenu domu. Prodejní cena je tedy vysvětlovanou (závisle) proměnnou. Využijeme čtyři vysvětlující proměnné (datový soubor jich obsahuje více, a tak je samozřejmě možné model dále rozšiřovat), konkrétně celkovou rozlohu domu (včetně pozemku), počet ložnic, počet koupelen a počet pater. Definujeme tedy:

- $y_i$  = prodejní cena  $i$ -tého domu (v Kanadských dolarech),
- $x_{i2}$  = rozloha  $i$ -tého domu (ve čtverečních stopách),
- $x_{i3}$  = počet ložnic v  $i$ -tém domě,
- $x_{i4}$  = počet koupelen v  $i$ -tém domu,
- $x_{i5}$  = počet pater v  $i$ -tém domu.

Pokud bychom skutečně prováděli výzkum faktorů ovlivňující cenu domů ve Windsoru, asi bychom měli nějaké povědomí o tamním trhu s nemovitostmi, a na tomto základě bychom mohli stanovit parametry odpovídající apriorní hustoty. Další možností je zeptat se místního realitního agenta, který by nám mohl tuto apriorní informaci poskytnout. Mohli bychom se tak tohoto agenta zeptat na několik otázek v podobě: „Jakou cenou byste ohodnotil jednopatrový dům s

rozlohou 4000 čtverečních stop, se dvěma ložnicemi a jednou koupelnou?"; „Jakou cenou byste ohodnotil dvoupatrový dům s rozlohou 6000 čtverečních stop, se třemi ložnicemi a dvěma koupelnami?“, apod. Protože máme pět neznámých regresních koeficientů, odpovědi na pět otázek v této podobě nám dá pět rovnic o pět neznámých. Jejich řešením bychom získali agentův odhad regresních koeficientů. Ty pak mohou být využity pro nastavení apriorní střední hodnoty parametrů  $\beta$ .

Pro ilustrační účely použijeme pouze hrubý odhad apriorní hustoty. Ceny domů ve Windsoru v roce 1987 mají velkou variabilitu, nicméně většina domů je prodávána za cenu 50000 až 150000 dolarů. Regresní model, který dobře vyrovná data, by mohl mít chyby v řádu několika tisíc dolarů, řekněme, maximálně 10000 dolarů. To nám napovídá, že směrodatná odchylna,  $\sigma$ , by mohla být asi 5000. Protože jsou náhodné chyby normálně rozděleny se střední hodnotou nulovou, potom, pokud předpokládáme, že  $\sigma = 5000$ , bude 95 % chyb v absolutní hodnotě menší než  $1.96 \times 5000 = 9800$  dolarů. Protože je  $h = \frac{1}{\sigma^2}$ , bude rozumný apriorní odhad pro  $h$  roven  $\frac{1}{5000^2} = 4.0 \times 10^{-8}$ . Nastavíme tedy  $\underline{s}^{-2} = 4.0 \times 10^{-8}$ . To je však jen velmi hrubý odhad, a tak mu chceme přiřadit jen malou váhu, a to nastavením  $\underline{\nu}$  na hodnotu mnohem menší než velikost vzorku,  $N$ . Protože je  $N = 546$ , bude volba  $\underline{\nu} = 5$  relativně neinformativní. Volně řečeno, říkáme, že apriorní informace o  $h$  by měla odpovídat 1 % váhy datové informace (tj.  $\frac{\underline{\nu}}{N} \approx 0.01$ ).

Pro regresní koeficienty volíme:

$$\underline{\beta} = \begin{bmatrix} 0.0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{bmatrix}.$$

Připomeňme si, že regresní koeficienty lze interpretovat tak, že nám říkají, „jestliže vysvětlující proměnná  $j$  vzroste o jednotku a ostatní vysvětlující proměnné zůstanou konstantní, cena tohoto domu bude mít tendenci zvýšit se o  $\beta_j$  dolarů.“ Naše apriorní střední hodnota by tak měla implikovat tvrzení v podobě: „jestliže porovnáme dva domy, které jsou identické, až na to, že první dům má o jednu ložnici více než druhý, potom očekáváme, že první dům bude o 5000 dolarů dražší než druhý,“ nebo, „jestliže počet koupelen vzroste o jednu a ostatní charakteristiky se nezmění, budeme očekávat, že cena domu vzroste o 10000 dolarů,“ apod.

Všechny tyto odhady regresních koeficientů jsou velmi hrubé, a tak nám dává smysl přiřadit jim relativně vysokou variabilitu. Předpokládejme například, že naše apriorní informace o úrovně konstantě je nejistá. V tomto případě bychom mohli chtít, aby  $\text{var}(\beta_1) = 10000^2$  (tzn. apriorní směrodatná odchylna je 10000), a tudíž přiřazujeme zhruba 95% apriorní pravděpodobnost oblasti  $[-20000, 20000]$ , což je vcelku široký interval.<sup>6</sup> Pokud si budeme hodně jistí, že

<sup>6</sup>Používáme zde aproximativní pravidlo, které nám říká, že 95 % pravděpodobnosti funkce hustoty je obsaženo v oblasti dvojnásobné směrodatné odchylny od své střední hodnoty. Tato

vliv rozlohy bude mezi 0 a 20 dolary za stopu čtvereční, zvolíme  $var(\beta_2) = 25$  (tzn. apriorní směrodatná odchylka pro  $\beta_2$  je 5). Pro ostatní koeficienty volíme  $var(\beta_3) = 2500^2$  a  $var(\beta_4) = var(\beta_5) = 5000^2$ . Tyto hodnoty hyperparametrů nám říkají, že například nejlepší apriorní odhad pro  $\beta_4$  je 10000 a domníváme se tedy, že se tato hodnota bude nacházet s vysokou pravděpodobností v intervalu  $[0, 20000]$ .

Pro tuto volbu rozptylů si můžeme představit apriorní kovarianční matici. Vlastnosti normálního-gama rozdělení nám implikují, že kovarianční matice pro vektor parametrů  $\beta$  má podobu:

$$var(\beta) = \frac{\nu s^2}{\nu - 2} \underline{V},$$

Protože  $\frac{\nu s^2}{\nu - 2} = 41666666 \frac{2}{3}$ , implikují volby  $var(\beta_j)$  pro  $j = 1, \dots, 5$ :

$$\underline{V} = \begin{bmatrix} 2.40 & 0 & 0 & 0 & 0 \\ 0 & 6.0 \times 10^{-7} & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 & 0 \\ 0 & 0 & 0 & 0.60 & 0 \\ 0 & 0 & 0 & 0 & 0.60 \end{bmatrix}.$$

Poznamenejme, že jsme všechny apriorní kovariance nastavili na hodnotu nula. To je obvyklá volba, protože je apriori velmi obtížné odhadnout jaká by mohla být vzájemná závislost mezi těmito parametry. Naše apriorní informace o možných hodnotách  $\beta_j$  by tak mohla být ta, že jsou nekorelovány s  $\beta_j$  pro  $i \neq j$ . V řadě případů se jedná o rozumný předpoklad. Tím jsme dokončili specifikaci informativní apriorní přirozeně konjugované hustoty parametrů našeho modelu.

Předchozí odstavce nám ilustrovaly, jak lze v praxi postupovat při nastavení priorů. Stanovení apriorních představ může být obtížný a obsahuje velkou část práce. Nicméně tato práce se vyplatí, neboť nás nutí přemýšlet o našem modelu a o interpretaci příslušných parametrů. Pokud se nacházíme v situaci, že nás nic nenapadá, nemáme tedy apriorní informaci, nebo ji nechceme použít, můžeme bayesiánskou analýzu provést s využitím neinformativní apriorní hustoty (3.24).

Tabulky 3.1 a 3.2 ukazují apriorní a posteriorní výsledky a informace za použití jak informativní, tak i neinformativní apriorní hustoty. Veškeré výpočty jsou náplní souborů `chapter03.m` resp. `chapter03_neinf.m`. Tyto soubory vyžadují další podpůrné funkce, zejména pak `nlrn_ncp.m`, která pro výpočet posteriorních hustot využívá vztahy (3.9)–(3.19). Na základě vztahů (3.20)–(3.23) lze provést posteriorní analýzu pro neinformativní hustotu. V našem případě jsme však využili funkce pro informativní apriorní hustoty, s tím, že jsme použili pro  $\beta$  vysoce neinformativní apriorní hyperparametry. Tabulka 3.1 potvrzuje, že naše apriorní hustota je relativně neinformativní, protože výsledky založené na informativní apriorní hustotě jsou podobné těm založeným na hustotě neinformativní. V předchozí kapitole jsme viděli, že posteriorní střední hodnota jednoho regresního koeficientu při využití informativního prioru leží mezi apriorní

aproximace je téměř dokonalá pro normální rozdělení či rozdělení s podobným průběhem jako je ono normální (např.  $t$ -rozdělení).



střední hodnotou a OLS odhadem. V tabulce 3.1 je vidět tendence, že posteriorní střední hodnota vycházející z informativní apriorní hustoty leží mezi apriorní střední hodnotou a OLS odhadem. Vzpomeňme si, že OLS odhad je identický posteriorní střední hodnotě založené na neinformativní apriorní hustotě (viz (3.21)). Ovšem ne každá posteriorní střední hodnota založená na informativním prioru leží mezi apriorní střední hodnotou a OLS odhadem (viz výsledek pro  $\beta_1$ ). To je z toho důvodu, že posteriorní střední hodnota je maticově váženým průměrem apriorní střední hodnoty a OLS odhadu (viz (3.11)). Maticové vážení nám neimplikuje, že každý jednotlivý koeficient leží mezi svou apriorní střední hodnotou a OLS odhadem.

**Tabulka 3.1:** Apriorní a posteriorní střední hodnoty pro  $\beta$  (směrodatné odchylky v závorkách)

	Prior	Posterior	
	Informativní	(Inf. prior)	(Neinf. prior)
$\beta_1$	0	-4035.05	-4009.55
	(10000)	(3530.16)	(3590.11)
$\beta_2$	10	5.43	5.43
	(5)	(0.37)	(0.37)
$\beta_3$	5000	2886.81	2824.61
	(2500)	(1184.93)	(1210.43)
$\beta_4$	10000	16965.24	17105.18
	(5000)	(1708.02)	(1728.18)
$\beta_5$	10000	7641.23	7634.90
	(5000)	(997.02)	(1004.34)

**Tabulka 3.2:** Apriorní a posteriorní střední hodnoty pro  $h$  (směrodatné odchylky v závorkách)

	Prior	Posterior	
	Informativní	(Inf. prior)	(Neinf. prior)
Stř. hodnota	$4 \times 10^{-8}$	$3.05 \times 10^{-9}$	$3.03 \times 10^{-9}$
Sm. odchylka	$2.53 \times 10^{-8}$	$1.84 \times 10^{-10}$	$1.83 \times 10^{-10}$

Tabulka 3.2 ukazuje apriorní a posteriorní výsledky pro  $h$ . Pro tento parametr opět můžeme vidět, že datová informace převyšuje apriorní informaci. Posteriorní výsledky jsou tedy podobné těm získaným na základě neinformativního prioru.

Obsah písemné zprávy shrnující výsledky tabulek 3.2 a 3.2 by odpovídal standardní interpretaci regresních koeficientů. Např. by zde mohla být obsažena

tvrzení: „Bez ohledu na to, jestli použijeme informativní nebo neinformativní apriorní hustotu, zjišťujeme, že posteriorní střední hodnota pro parametr  $\beta_4$  je přibližně 17000. Náš bodový odhad tak vypovídá o tom, že pokud srovnáme dva totožné domy, až na to, že jeden má o jednu ložnici více než druhý, můžeme očekávat, že první dům bude zhruba o 17000 dolarů hodnotnější než druhý.“ Toto tvrzení lze vyjádřit i stroze v podobě, že „bodový odhad mezního vlivu počtu ložnic na cenu domu je zhruba 17000 dolarů.“

Tabulka 3.3 obsahuje výsledky vztahující se k různým metodám porovnání modelů, diskutovaných v této kapitole. Všechny výsledky lze použít pro objasnění otázky, jestli jednotlivé koeficienty jsou nulové. Sloupec označený jako  $p(\beta_j > 0|y)$  využívá (3.14) a vlastnosti  $t$ -rozdělení k výpočtu pravděpodobnosti, že jednotlivé koeficienty jsou kladné. Užitečnost těchto pravděpodobností je obsahem části 3.5.1. Sloupec označený jako „Posteriorní podíl šancí pro  $\beta_j = 0$ “ obsahuje logicky posteriorní podíl šancí porovnávající model s omezením, že daný parametr je roven nule oproti neomezené alternativě. K výpočtu posteriorního podílu šancí pro modely  $M_1 : \beta_j = 0$  a  $M_2 : \beta_j \neq 0$  se používají metody z části 3.5.2 (i to je obsahem funkce `nlrn_ncp.m`).

**Tabulka 3.3:** Porovnání modelů zahrnující parametr  $\beta$

	$p(\beta_j > 0 y)$	95% HPDI	99% HPDI	P. podíl šancí pro $\beta_j = 0$
<i>Informativní apriorní hustota</i>				
$\beta_1$	0.13	[-10969.27,2899.17]	[-13159.75,5089.64]	4.14
$\beta_2$	1.00	[4.71,6.15]	[4.49,6.38]	0.00
$\beta_3$	0.99	[559.29,5214.34]	[-175.96,5949.58]	0.39
$\beta_4$	1.00	[13610.20,20320.27]	[12550.37,21380.10]	0.00
$\beta_5$	1.00	[5682.82,9599.65]	[5064.16,10218.30]	0.00
<i>Neinformativní apriorní hustota*</i>				
$\beta_1$	0.13	[-11061.65,3042.54]	[-13289.45,5270.34]	—
$\beta_2$	1.00	[4.71,6.15]	[4.48,6.38]	—
$\beta_3$	0.99	[446.96,5202.27]	[-304.15,5953.38]	—
$\beta_4$	1.00	[13710.50,20499.85]	[12638.10,21572.25]	—
$\beta_5$	1.00	[5662.07,9607.73]	[5038.84,10230.96]	—

\* Jako neinformativní apriorní hustota byla použita apriorní hustota blížící se čistě neinformativní hustotě. Podíl šancí tak bylo možno spočítat, ale není zde uváděn.

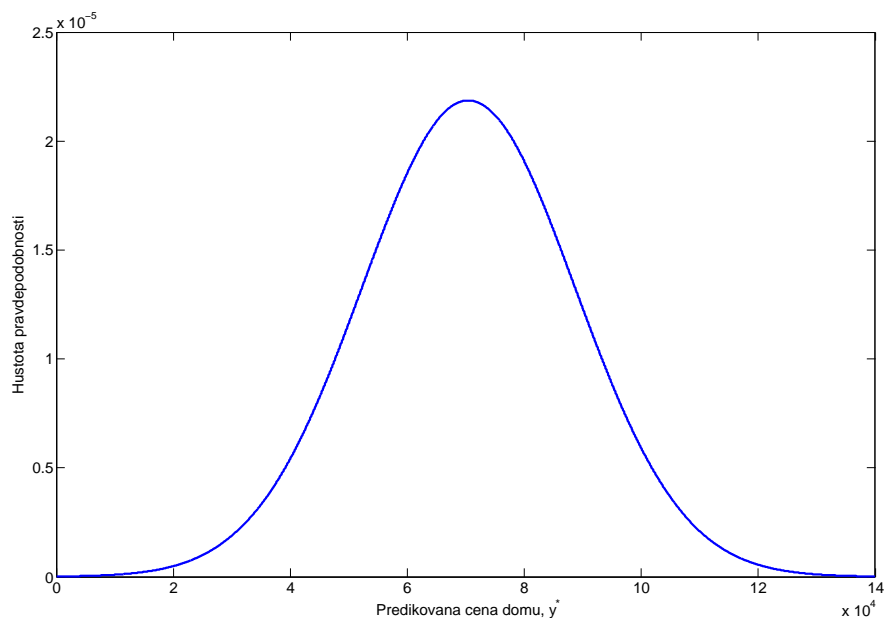
Omezený model využívá informativní apriorní hustotu, která je stejná jako neomezená apriorní hustota, s tou výjimkou, že  $\underline{\beta}$  a  $\underline{V}$  jsou postupně matice rozměru  $4 \times 1$  respektive  $4 \times 4$ , po vypuštění apriorní informace vztahující se k vypuštěnému  $\beta_j$ . Apriorní podíl šancí je nastaven na hodnotu jedna (výsledkem je tedy Bayesův faktor). Zbylé dva sloupce tabulky 3.3 obsahují 95% a 99% intervaly nejvyšší posteriorní hustoty (HPDI) pro každý z parametrů  $\beta_j$  (využita je zde jednoduchá funkce `HPDI_nlrn_ncp`, využívající vlastnosti posteriorní hustoty odpovídající  $t$ -rozdělení). Jak bylo popsáno v části 3.5.3, HPDI lze použít testování omezení ve tvaru rovnosti. Tyto intervaly mají své oprávnění (byť *ad hoc*) i v případě použití neinformativní apriorní hustoty. Nezapomeňme, že posteriorní podíl šancí vyžaduje použití informativních priorů (přinejmenším v případě společných parametrů porovnávaných modelů). Přestože jsme prak-

ticky v algoritmu využili „téměř“ neinformativní hustoty, posteriorní podíl šancí v tomto případě neprezentujeme.

Výsledky tabulky 3.3 jsou konzistentní s těmi z tabulky 3.1. V této poslední jmenované tabulce jsme viděli, že posteriorní střední hodnoty pro  $\beta_2$ ,  $\beta_4$  a  $\beta_5$  jsou kladné a velmi vysoké relativně ke své směrodatné odchylce, což je silný důkaz pro to, že všechny tyto koeficienty jsou nenulové a kladné. Bez ohledu na to, jestli použijeme informativní nebo neinformativní prior, tabulka 3.3 říká, že  $p(\beta_j > 0|y)$  je v podstatě rovna jedné pro  $j = 2, 4, 5$ , a žádný z intervalů nejvyšší posteriorní hustoty neobsahuje nulu. Pro informativní prior jsou posteriorní podíly šancí, porovnávací modely  $M_1 : \beta_j = 0$  vzhledem k  $M_2 : \beta_j \neq 0$  pro  $j = 2, 4, 5$ , velmi malé, což nám indikuje, že neomezený model má mnohem větší pravděpodobnost oproti modelům omezeným. Výsledky pro  $\beta_1$  a  $\beta_3$  jsou smíšené. Většina důkazů hovoří pro to, že  $\beta_3 \neq 0$ . Nicméně, 99% HPDI pro tento parametr obsahuje nulu. Pokud bychom tedy použili postup výběru modelů nastíněný v části 3.5.3, budou výsledky záviset na volbě intervalu nejvyšší posteriorní hustoty. 95% interval by implikoval  $\beta_3 \neq 0$ , ale 99% HPDI říká, že  $\beta_3 = 0$ . Tato nejistota je zobrazena i v posteriorním podílu šancí, který říká, že omezený model je 0.39 krát pravděpodobnější než neomezený model. Pokud použijeme tento posteriorní podíl šancí k výpočtu posteriorní pravděpodobnosti modelu, zjistíme, že  $P(M_1 : \beta_3 = 0|y) = 0.28$  (předpokládáme zde pro jednoduchost, že omezený a neomezený model tvoří úplnou množinu, i když s ohledem na nulové podíly šancí u ostatních modelů, nám stačí „ignorovat“ model s možností nulové úrovně konstanty). Jinými slovy, existuje 28% šance, že  $\beta_3 = 0$  a 72% šance, že tomu tak není. Pokud zde máme tuto nejistotu, je rozumné využít bayesovské průměrování modelů. Alternativně můžeme zvolit buď neomezený nebo omezený model. V každém případě je zde však významná pravděpodobnost, že zvolíme špatný model.

Pro ilustraci, jak provést predikci v normální lineárním regresním modelu, budeme předpokládat případ, kdy nás zajímá predikce prodejní ceny domu s rozlohou 5000 čtverečních stop, dvěma ložnicemi, dvěma koupelnami a jedním patrem. S využitím (3.40) můžeme v případě informativního prioru pracovat s rozdělení  $t(70468, 3.33 \times 10^8, 551)$ . Pro neinformativní prior je rozdělení predikční hustoty  $t(70631, 3.35 \times 10^8, 546)$ . Některou z těchto hustot (jsou podobné) můžeme využít k prezentaci informace pro klienta, který chce prodat dům s výše uvedenými charakteristikami. Můžeme říct, že náš nejlepší odhad prodejní ceny je 70000 dolarů, nicméně je zde velká nejistota spojená s tímto odhadem, kdy predikční směrodatná odchylka je zhruba 18000 dolarů. Predikční hustota je vykreslena na obrázku 3.1.

Část 3.7 byla věnována Monte Carlo integraci. Monte Carlo integrace není v případě normálního lineárního regresního modelu a přirozeně konjugovanou apriorní hustotou nezbytná, pokud nás nezajímají nelineární funkce regresních koeficientů. Známe posteriorní vlastnosti  $\beta$  (viz tabulka 3.1), a Monte Carlo integraci na tomto místě není třeba provádět. Pro ilustraci si však s jejím využitím spočítáme posteriorní střední hodnotu a směrodatnou odchylku parametru  $\beta_2$ . Z tabulky 3.1 víme, že střední hodnota je 5.43 a směrodatná odchylka 0.37. Tyto analytické hodnoty lze využít pro hodnocení kvality Monte Carlo integrace. Z



**Obrázek 3.1:** Predikční hustota ceny domu (přirozeně konjugovaná apriorní hustota).

důvodu úspory místa si výsledky ukážeme jen pro informativní apriorní hustotu.

Monte Carlo integraci lze provést na základě náhodných výběrů z posteriorního rozdělení  $\beta$ , a to následným zprůměrováním nějaké funkce (která nás zajímá) těchto výběrů (viz (3.41)). Z (3.14) víme, že  $p(\beta|y)$  odpovídá  $t$ -rozdělení. Můžeme tedy vytvořit program, který bude opakovat výběry z této hustoty a následně je zprůměruje (případně zprůměruje nějakou funkci těchto výběrů).

Tabulka 3.4 ukazuje posteriorní střední hodnotu a směrodatnou odchylku pro  $\beta_2$  spočítanou různými způsoby (využita je funkce `MCI_nlrn_ncp.m`). Řádek označený jako „Analyticky“ je přesný výsledek získaný z (3.14)–(3.16). Ostatní řádky ukazují výsledky na základě Monte Carlo integrace s různým počtem replikací. Je zde uvedena rovněž numerická standardní chyba (NSE), která nám poskytuje pohled na přesnost Monte Carlo aproximace střední hodnoty parametru,  $E(\beta_2|y)$  resp. jeho nějaké funkce (viz diskuze v části 3.7).

Jak bychom asi očekávali, přesnost aproximace jak posteriorní střední hodnoty, tak i směrodatné odchylky se zvyšuje s rostoucím počtem replikací.<sup>7</sup> Z empirického hlediska závisí volba  $S$  na námi požadované přesnosti. Při předběžném průzkumu dat někdy stačí hrubý odhad, a tedy volba  $S = 10$  nebo  $S = 100$ . Pro hodně přesné výsledky je však třeba nastavení  $S = 10000$  nebo

<sup>7</sup>V rámci našeho programu není možno zreplikovat přesně dané výsledky, protože je zde zcela náhodně nastavován generátor náhodných čísel, a to dle času spuštění programu.

**Tabulka 3.4:** Posteriorní výsledky pro parametr  $\beta_2$  spočítané alternativním způsobem

	Stř. hodnota	Směrodatná odchylka	NSE
Analyticky	5.43	0.37	—
Počet replikací			
$S = 10$	5.44	0.27	0.085
$S = 100$	5.43	0.38	0.038
$S = 1000$	5.44	0.36	0.011
$S = 10000$	5.44	0.36	0.004
$S = 100000$	5.43	0.37	0.001

dokonce  $S = 100000$ . Numerická standardní chyba je dobrým měřítkem přesnosti každé aproximace. Aporixmativní posteriorní střední hodnoty jsou zřídka odlišné o více než jednu numerickou standardní chybu od skutečné posteriorní střední hodnoty dané v řádku „Analyticky“.

Je potřeba zdůraznit, že rostoucí  $S$  sice zvýší přesnot Monte Carlo aproximace střední hodnoty  $E(\beta_2|y)$ , nicméně toto zvýšení není lineární v  $S$ . Tabulka 3.4 ukazuje výsledky pro  $S = 100000$ , které nejsou desetkrát přesnější než ty pro  $S = 10000$ . Analyticky, numerická standardní chyba,  $\frac{\hat{\sigma}_g}{\sqrt{S}}$ , se snižuje tempem  $\frac{1}{\sqrt{S}}$ . Výsledky pro  $S = 100000$  by tedy měly být jen zhruba  $\sqrt{10} = 3.16$  krát přesnější než ty pro  $S = 10000$ .

### 3.9 Shrnutí

V této kapitole jsme si prošli postup bayesiánské analýzy, zahrnující věrohodnostní funkci, apriorní hustotu, posteriorní hustotu, porovnání modelů a predikci, a to pro případ normálního lineárního regresního modelu s přirozeně konjugovanou apriorní hustotou a  $k$  vysvětlujícími proměnnými. Tato kapitola v podstatě odpovídá kapitole předešlé, až na to, že jsme zde využili značení s využitím matic, abychom se tak vypořádali s „komplikacemi“ vznikajícími v důsledku více vysvětlujících proměnných ( $k > 1$ ). Zavedli jsme si rovněž koncept intervalů nejvyšší posteriorní hustoty a ukázali jsme si, jak využít Monte Carlo integraci pro posteriorní analýzu nelineárních funkcí regresních parametrů.



## Kapitola 4

# Normální lineární regresní model s jinými priory

Pro reálné aplikace je přirozeně konjugovaná apriorní hustota poněkud omezující a není tak pravděpodobné, že bychom si s ní vystačili. V této kapitole tak budeme pracovat s normálním lineárním regresním model s nezávislým normálním-gama apriorním rozdělením. Nezávislost rozdělení parametru  $\beta$  a  $h$  má totiž zásadní důsledky, neboť posteriorní hustota, predikční hustota a posteriorní podíl šancí již nebudou mít známou analytickou formu. Bude tedy nutné využít posteriorních simulátorů. V této kapitole tak bude zaveden mocný nástroj posteriorních simulací – *Gibbsův vzorkovač* (*Gibbs sampler*). V této souvislosti si vysvětlíme problematiku tzv. konvergenčních diagnostik, pomocí kterých budeme ověřovat, zda-li je počet simulovaných a využitelných vzorků dostatečný natolik, abychom mohli hovořit o reprezentativním výběru z posteriorní hustoty pravděpodobnosti. Uvidíme rovněž jak vypočítat posteriorní podíl šancí pro vnořené modely, a to za užití tzv. *Savage-Dickeyho poměr hustot* (*Savage-Dickey density ratio*). Jako druhou variantu apriorní hustoty si ukážeme takový typ hustoty, který v sobě bude obsahovat určité restriktce kladené na vektor parametrů  $\beta$ . V této souvislosti bude zavedena technika zvaná *importance sampling*. Český ekvivalent by mohlo být „vzorkování dle důležitosti“ popř. „vážené vzorkování“, nicméně v textu bude uváděn spíše anglický výraz. V obou analyzovaných situacích budou věrohodnostní funkce totožné a budou odpovídat funkci z kapitoly 3, neboť se stále pohybujeme v intencích normálního lineárního regresního modelu.

## 4.1 NLRM s nezávislou normální-gama apriorní hustotou

### 4.1.1 Apriorní hustota

Předpokládejme tedy, že rozdělení vektoru parametrů  $\beta$  není podmíněno přesností chyby  $h$  a že jsou tedy tyto parametry vzájemně nezávislé. Platí tedy, že sdružená apriorní hustota pravděpodobnosti je rovna  $p(\beta, h) = p(\beta)p(h)$ , kde  $p(\beta)$  odpovídá normálnímu rozdělení a  $p(h)$  rozdělení gama:

$$p(\beta) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\underline{V}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) \right], \quad (4.1)$$

$$p(h) = c_G^{-1} h^{\frac{\nu-2}{2}} \exp \left( -\frac{h\underline{\nu}}{2\underline{s}^{-2}} \right), \quad (4.2)$$

kde  $c_G$  je integrační konstanta pro funkci hustoty pravděpodobnosti odpovídající rozdělení gama. Značení je už tradiční,  $\underline{\beta}$  je apriorní střední hodnota parametru  $\beta$ , apriorní střední hodnota a stupně volnosti pro přesnost chyby  $h$  jsou  $\underline{s}^{-2}$  a  $\underline{\nu}$ . Matice  $\underline{V}$  je přímo apriorní kovarianční matice vektoru parametrů  $\beta$  (v případě přirozeně konjugované apriorní hustoty byl rozptyl parametrů podmíněn přesností chyby, tedy  $\text{var}(\beta|h) = h^{-1}\underline{V}$ ).

### 4.1.2 Posteriorní hustota

Posteriorní hustota je proporcionální součinu věrohodnostní funkce a apriorní hustoty. Pokud budeme abstrahovat od členů nezávislých na  $\beta$  nebo  $h$  získáváme jádrovou hustotu v podobě:

$$p(\beta, h|y) \propto \left\{ \exp \left[ -\frac{1}{2} \{ h(y - X\beta)'(y - X\beta) + (\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) \} \right] \right\} \\ \times h^{\frac{N+\nu-2}{2}} \exp \left[ -\frac{h\underline{\nu}}{2\underline{s}^{-2}} \right]. \quad (4.3)$$

Tato sdružená hustota pravděpodobnosti pro  $\beta$  a  $h$  nemá tvar známé hustoty pravděpodobnosti a nelze jí tedy snadným způsobem využít pro posteriorní analýzu. Pro nalezení např. posteriorní střední hodnoty  $E(\beta)$  je nutné využít nějakého posteriorního simulátoru. Ačkoliv tedy sdružená hustota pravděpodobnosti nemá obvyklou formu (ve smyslu známého rozdělení), podmíněné posteriorní hustoty již tuto formu mít budou. Podmíněnou posteriorní hustotu  $p(\beta|y, h)$  snadno získáme, budeme-li zacházet se vztahem (4.3) jako s funkcí parametrů  $\beta$  pro pevně dané  $h$ . Formálně řečeno, pravidla pravděpodobnosti implikují  $p(\beta|y, h) = \frac{p(\beta, h|y)}{p(h|y)}$ . Jelikož však  $p(h|y)$  nezávisí na vektoru  $\beta$ , dává nám samotná  $p(\beta, h|y)$  jádrovou hustotu  $p(\beta|y, h)$ . Protože je však hustota definovaná svou jádrovou hustotou, jsme schopni z tvaru sdružené posteriorní hustoty  $p(\beta, h|y)$  vyvodit informaci o podmíněné hustotě  $p(\beta|y, h)$  za předpokladu, že



budeme fixovat  $h$  (bereme tento výraz jako proměnnou, jejíž hodnota je daná a známá).

Po maticových úpravách analogických těm z předchozích kapitol můžeme definovat:

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

kde

$$\bar{V} = (\underline{V}^{-1} + hX'X)^{-1}, \quad (4.4)$$

$$\bar{\beta} = \bar{V}(\underline{V}^{-1}\underline{\beta} + hX'y), \quad (4.5)$$

$$Q = hy'y + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - \bar{\beta}'\bar{V}^{-1}\bar{\beta}.$$

Další kombinací s rovnicí (4.3) a vypuštěním členů, které nezahrnují vektor  $\beta$  (včetně výrazu  $Q$ ), lze psát:

$$p(\beta|y, h) \propto \exp\left[-\frac{1}{2}(\beta - \bar{\beta})'\bar{V}^{-1}(\beta - \bar{\beta})\right], \quad (4.6)$$

což je samozřejmě jádrová hustota vícerozměrného normálního rozdělení. Jinými slovy

$$\beta|y, h \sim N(\bar{\beta}, \bar{V}). \quad (4.7)$$

Podmíněnou posteriorní hustotu přesnosti chyby  $p(h|y, \beta)$  získáme z (4.3) jako funkci přesnosti chyby  $h$ . Lze tedy odvodit:

$$p(h|y, \beta) \propto h^{\frac{N+\nu-2}{2}} \exp\left[-\frac{h}{2}\{(y - X\beta)'(y - X\beta)\}\right]. \quad (4.8)$$

Můžeme si snadno ověřit, že se jedná o jádrovou hustotu gama rozdělení a tedy

$$h|y, \beta \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (4.9)$$

kde

$$\bar{\nu} = N + \nu, \quad (4.10)$$

$$\bar{s}^2 = \frac{(y - X\beta)'(y - X\beta) + \nu s^2}{\bar{\nu}}. \quad (4.11)$$

Tyto vztahy vypadají podobně jako v případě normálního lineárního regresního modelu s přirozeně konjugovanou apriorní hustotou. Neformálně zde samozřejmě platí podobná úvaha o tom, jak je kombinována datová a apriorní informace. Je však třeba zdůraznit, že vztahy (4.4)-(4.11) se nevztahují přímo k posteriorní hustotě, která nás zajímá, tedy  $p(\beta, h|y)$ , ale vztahují se k podmíněným hustotám  $p(\beta|y, h)$  a  $p(h|y, \beta)$ . Jelikož však  $p(\beta, h|y) \neq p(\beta|y, h)p(h|y, \beta)$ , neříkají nám podmíněné posteriorní hustoty (4.8) a (4.9) vše o sdružené posteriorní hustotě  $p(\beta, h|y)$ . Existuje však posteriorní simulátor zvaný Gibbsův vzorkovač, který využívá náhodných výběrů z podmíněných hustot k vytvoření množiny náhodných vzorků z odpovídající sdružené hustoty pravděpodobnosti. Ty pak lze využít k odhadům posteriorních středních hodnot apod.

### 4.1.3 Gibbsův vzorkovač

Princip tohoto nástroje budeme prezentovat ve zcela obecné podobě. Mějme tedy dán  $p$ -rozměrný vektor parametrů  $\theta$ , věrohodnostní funkci  $p(y|\theta)$ , apriorní hustotu  $p(\theta)$  a posteriorní hustotu  $p(\theta|y)$ . V případě lineárního regresního modelu je  $p = k + 1$  a  $\theta = (\beta', h)'$ . Rozdělme si navíc vektor parametrů  $\theta$  do několika ( $B$ ) bloků, tedy  $\theta = (\theta'_{(1)}, \theta'_{(2)}, \dots, \theta'_{(B)})'$ , kde  $\theta_{(j)}$  je skalár nebo vektor pro  $j = 1, 2, \dots, B$ . V lineárním regresním modelu je obvyklé nastavit počet bloků  $B = 2$ , tedy první blok parametrů  $\theta_{(1)} = \beta$  a druhý blok  $\theta_{(2)} = h$ .

Monte Carlo integrace poskytuje odhad  $E[g(\theta)|y]$  na základě náhodných výběrů z posteriorní hustoty pravděpodobnosti  $p(\theta|y)$ , a to pro jakoukoliv funkci parametrů  $g(\theta)$ . Ne vždy je však možno provést náhodný výběr přímo z  $p(\theta|y)$ . Často je však možné provést náhodné výběry z podmíněných hustot pravděpodobnosti

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

Množinu těchto rozdělení nazýváme *plně podmíněnými posteriorními rozděleními*, neboť definují posteriorní hustotu pro každý blok jako hustotu podmíněnou ostatními bloky.

V NLRM s nezávislou normální-gama apriorní hustotou je podmíněná hustota  $p(\beta|y, h)$  normální a hustota  $p(h|y, \beta)$  má gama rozdělení, z nichž je velmi snadné provést náhodný výběr. Náhodné výběry z takto definovaných plně podmíněných rozdělení nám dají sekvenci vzorků  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(S)}$ , ze kterých jsme schopni získat Monte Carlo integraci  $E[g(\theta)|y]$ .

Předpokládejme tedy případ, kdy  $B = 2$  a mějme již dán jeden počáteční náhodný výběr z  $p(\theta_{(2)}|y)$ , který označíme  $\theta_{(2)}^{(0)}$ . Dolní index indikuje blok, horní index pak číslo vzorku. Protože  $p(\theta|y) = p(\theta_{(1)}|y, \theta_{(2)})p(\theta_{(2)}|y)$ , je tak zřejmé, že výběr z  $p(\theta_{(1)}|y, \theta_{(2)}^{(0)})$  je řádným výběrem  $\theta_{(1)}$  z  $p(\theta|y)$ . Označme tento výběr  $\theta_{(1)}^{(1)}$ . Neboť  $p(\theta|y) = p(\theta_{(2)}|y, \theta_{(1)})p(\theta_{(1)}|y)$ , platí, že náhodný výběr z  $p(\theta_{(2)}|y, \theta_{(1)}^{(1)})$  je platným výběrem  $\theta_{(2)}$  z  $p(\theta|y)$ . Tedy vektor parametrů  $\theta^{(1)} = (\theta_{(1)}^{(1)'}, \theta_{(2)}^{(1)'})'$  je řádným výběrem z  $p(\theta|y)$ . Takto lze postupovat do nekonečna. Jsme-li tedy schopni úspěšně najít  $\theta_{(2)}^{(0)}$ , potom sekvenční výběr z posteriorního rozdělení  $\theta_{(1)}$  podmíněného předchozím výběrem  $\theta_{(2)}$  a výběr  $\theta_{(2)}$  podmíněný takto získaným  $\theta_{(1)}$  nám dává řadu náhodných výběrů (vzorků) z posteriorního rozdělení. Strategie sekvenčních výběrů z plně podmíněných posteriorních rozdělení je nazývána Gibbsův vzorkovač.

Problémem je samozřejmě získání počátečního výběru  $\theta_{(2)}^{(0)}$ . Pokud bychom uměli získávat náhodné výběry z  $p(\theta_{(2)}|y)$ , mohli bychom jej přímo využít spolu s podmíněnou hustotou pro první blok  $p(\theta_{(1)}|\theta_{(2)}, y)$  v rámci Monte Carlo integrace a Gibbsův vzorkovač by nebylo potřeba využít.<sup>8</sup> Lze však ukázat, že za

<sup>8</sup>Připomeňme si, že pravidla pravděpodobnosti nám říkají, že sdruženou hustotu pravděpodobnosti můžeme přepsat do podoby  $p(\theta_{(1)}, \theta_{(2)}|y) = p(\theta_{(1)}|\theta_{(2)}, y)p(\theta_{(2)}|y)$ .

podmínky splnění tzv. slabých podmínek nehraje počáteční výběr  $\theta_{(2)}^{(0)}$  žádnou roli v tom smyslu, že Gibbsův vzorkovač bude vždy konvergovat k sekvenci výběrů z  $p(\theta|y)$ . Obvykle tak zvolíme počáteční hodnotu  $\theta_{(2)}^{(0)}$  necháme proběhnout Gibbsův vzorkovač v rámci  $S$  replikací. Poté prvních  $S_0$  vzorků odstraníme (tzv. *burn-in replications*) a zbývajících  $S_1$  vzorků použijeme k odhadu střední hodnoty funkce parametrů, která nás zajímá,  $E[g(\theta)|y]$ , přičemž  $S_0 + S_1 = S$ .

Pokud jde o ony slabé podmínky zmiňované výše, stačí nám pro naše potřeby vědět, že nejtypičtějším příkladem, kdy nejsou splněny je případ, kdy posteriorní hustota je definována ve dvou různých oblastech, které nejsou vzájemně propojeny. V tom případě Gibbsův vzorkovač poskytne výběry jen z jedné z těchto oblastí (do druhé oblasti se nebude schopen dostat). To samozřejmě není případ normálního-gama rozdělení.

Gibbsův vzorkovač lze tedy algoritmicky definovat (pro obecný počet  $B$  bloků) následovně:

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

Získáme tímto  $S$  výběrů,  $\theta^{(s)}$  pro  $s = 1, \dots, S$ . Prvních  $S_0$  výběrů vyhodíme, abychom eliminovali efekt počáteční volby  $\theta^{(0)}$  a zbývajících  $S_1$  výběrů můžeme zprůměrovat, abychom obdrželi požadované posteriorní charakteristiky. Pokud nás zajímá funkce parametrů  $g(\cdot)$  a

$$\widehat{g}_{S_1} = \frac{1}{S_1} \sum_{s=S_0+1}^S g(\theta^{(s)}), \quad (4.12)$$

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

Tato strategie bude fungovat pro jakoukoliv volbu bloků, nicméně ve většině případů se tato volba nabízí sama. Centrální limitní věta nám rovněž umožňuje přibližné určení chyby aproximace. Nicméně oproti samotné Monte Carlo integraci zde vznikají dva problémy:

1. V rámci Gibbsova vzorkovače se musíme ujistit, že volba  $\theta^{(0)}$  nemá vliv na získané výsledky.

2. Sekvence výběrů není i.i.d. (stejněměrně a nezávisle rozdělena). Konkrétně, vektory  $\theta^{(s)}$  a  $\theta^{(s-1)}$  nejsou vzájemně nezávislé, neboť  $\theta_{(j)}^{(s)}$  závisí na  $\theta_{(l)}^{(s-1)}$  pro  $j = 1, \dots, B-1$  a  $l > j$ .

Prakticky tedy je nutné pro dosažení požadované úrovně přesnosti vygenerovat mnohem více výběrů při využívání Gibbsova vzorkovače než by tomu bylo v případě Monte Carlo integrace.

#### 4.1.4 Markov Chain Monte Carlo diagnostiky

Skutečnost, že stav Gibbsova vzorkovače pro výběr  $s$  (tj.  $\theta^{(s)}$ ) závisí na jeho stavu při výběru  $s-1$  (tj.  $\theta^{(s-1)}$ ) znamená, že sekvence tvoří tzv. Markovský řetězec (*Markov chain*). Existuje mnoho dalších posteriorních simulátorů s touto vlastností. Tento druh posteriorních simulátorů má obecné označení jako *Markov Chain Monte Carlo* (MCMC) algoritmy. Existuje zde celá řada možností jak měřit chybu aproximace v rámci MCMC algoritmu a mnoho dalších diagnostik, které analyzují, zda-li odhadnuté výsledky jsou spolehlivé. Souhrně je lze označit jako *MCMC diagnostiky*.

První MCMC diagnostikou je numerická standardní chyba. V předchozím textu byla odvozena na základě centrální limitní věty. V kontextu MCMC metod lze rovněž odvodit numerickou standardní chybu, nicméně vzhledem k tomu, že náhodné výběry nejsou v rámci těchto algoritmů nezávislé, je třeba použít jiný typ centrální limitní věty. Jedná se o to, že za slabých podmínek nutných pro konvergenci Gibbsova vzorkovače získáváme známou podobu centrální limitní věty:

$$\sqrt{S_1} \{ \hat{g}_{S_1} - E[g(\theta)|y] \} \rightarrow N(0, \sigma_g^2) \quad (4.13)$$

pro  $S_1$  jdoucí k nekonečnu. V tomto případě má  $\sigma_g^2$  mnohem složitější formu než v případě (3.42) a v literatuře zatím nebyl publikován dostatečně ověřený způsob jejího odhadu. Intuitivně by  $\sigma_g^2$  měla zohledňovat skutečnost, že  $\theta^{(s)}$  pro  $s = 1, \dots, S$  je vzájemně korelovaná řada. Geweke [14] na tomto základě navrhuje myšlenku převzatou z literatury pojednávající o časových řadách a nabízí odhad  $\sigma_g^2$  v podobě:

$$\hat{\sigma}_g^2 = \frac{S(0)}{S_1} \quad (4.14)$$

Oprávnění tohoto odhadu je spíše neformální nicméně v praxi se osvědčuje.  $S(0)$  je spektrální hustota řady  $\theta^{(s)}$  pro  $s = S_0 + 1, \dots, S$  vyhodnocená v 0. Podstatné pro nás nemusí být co přesně spektrální hustota vyjadřuje, stačí nám, že odhad  $\sigma_g^2$  existuje a počítačové programy jej dokáží spočítat. Lze tak obdržet numerickou standardní chybu  $\frac{\hat{\sigma}_g}{\sqrt{S_1}}$ . Interpretace je stejná jako v předchozí kapitole.

Geweke [14] nabízí rovněž diagnostiku založenou na intuici, že pokud byl učiněn dostatečně velký výběr, odhad  $g(\theta)$  založený na první polovině výběru by měl být zhruba stejný jako odhad založený na polovině druhé. Pokud jsou tyto odhady rozdílné, nasvědčuje to tomu, že máme málo výběrů (a odhad je tak nepřesný) nebo vliv počátečního výběru  $\theta^{(0)}$  ještě neodezněl a ovlivňuje

nám první polovinu výběru. Obecněji tedy rozdělme našich  $S$  výběrů z Gibbsova vzorkovače na prvních  $S_0$  vzorků, které zahazujeme, a  $S_1$  vzorků, které ponecháváme. Tento vzorek pak dále rozdělíme na první sadu  $S_A$  výběrů, prostřední  $S_B$  a poslední  $S_C$ . V praxi se osvědčila volba  $S_A = 0.1S_1$ ,  $S_B = 0.5S_1$  a  $S_C = 0.4S_1$ . Pro potřeby diagnostiky nebudeme používat prostřední množinu vzorků  $S_B$ . Tím si zajistíme větší pravděpodobnost, že první a poslední sada výběrů bude na sobě nezávislá. Nechť  $\hat{g}_{S_A}$  a  $\hat{g}_{S_C}$  jsou odhady  $E[g(\theta)|y]$  za použití prvních  $S_A$  vzorků (po vyhození  $S_0$  výběrů) a  $S_C$  vzorků. Definujeme  $\frac{\hat{\sigma}_A}{\sqrt{S_A}}$  a  $\frac{\hat{\sigma}_C}{\sqrt{S_C}}$  jakožto numerické standardní chyby těchto odhadů. Potom lze využít centrální limitní větu analogickou (4.13):

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

kde  $CD$  je konvergenční diagnostika daná výrazem:

$$CD = \frac{\hat{g}_{S_A} - \hat{g}_{S_C}}{\frac{\hat{\sigma}_A}{\sqrt{S_A}} + \frac{\hat{\sigma}_C}{\sqrt{S_C}}} \quad (4.15)$$

V empirických aplikacích využívajících Gibbsův vzorkovač lze tuto konvergenční diagnostiku spočítat a porovnat s kritickými hodnotami standardizovaného normálního rozdělení. Vysoké hodnoty  $CD$  ukazují na to, že  $\hat{g}_{S_A}$  a  $\hat{g}_{S_C}$  jsou od sebe vzdálené a je tedy potřeba více replikací. Pokud naopak tato diagnostika nasvědčuje tomu, že jsme provedli dostatečný počet replikací, lze konečné statistiky posteriorní hustoty vypočítat na základě celého vzorku  $S_1$  výběrů.

Předchozí MCMC diagnostiky jsou vcelku informativní, pokud jde o otázku, zda-li Gibbsův vzorkovač pracuje dobře a zda-li jsme zadali dostatečný počet replikací pro dosažení žádoucí úrovně přesnosti. Není zde ovšem vždy zajištěno, že nám budou dávat spolehlivé výsledky. Příkladem může být případ bimodální posteriorní hustotou pravděpodobnosti. Předpokládejme, že posteriorní hustota je kombinací dvou normálních rozdělení nacházejících se v různých částech parametrického prostoru. Je možné, že Gibbsův vzorkovač, který začíná blízko jedné ze středních hodnot těchto normálních rozdělení zde setrvá a všechny vzorky budou jen z této oblasti. Numerické standardní odchylky budou vypadat rozumně, výše zmiňované konvergenční diagnostiky budou indikovat konvergenci, ovšem ve skutečnosti všechny naše výsledky budou postrádat jedno z normálních rozdělení, které zahrnuje posteriorní hustota. Toto není případ normálního lineárního regresního modelu s nezávislým normálním-gama apriorním rozdělením. Některé kombinace normálních modelů však v sobě mohou zahrnovat multimodální posteriorní hustotu.

Druhým příkladem, kdy by výsledky Gibbsova vzorkovače byly zavádějící a zmiňované MCMC diagnostiky nás na to neupozorní, je případ, kdy náš počáteční výběr  $\theta^{(0)}$  je extrémě vzdálen od oblasti parametrického prostoru s největší posteriorní pravděpodobností. Pokud je míra korelace ve vzorcích vysoká, bylo by třeba enormně velkého počtu výběrů Gibbsovým vzorkovačem, abychom se dostali do oblasti vyšší posteriorní pravděpodobnosti. Ve většině případů tento problém bude indikovat  $CD$  diagnostika, neboť  $\hat{g}_{S_A}$  a  $\hat{g}_{S_C}$  se budou od sebe lišit

(jak se bude Gibbsův vzorkovač postupně vzdalovat od  $\theta^{(0)}$ ), ovšem v neobvyklých případech tomu tak být nemusí.

Gibbsův vzorkovač je často označován jakožto procházející posteriorní rozdělení, přičemž nejvíce navštěvuje oblasti s vysokou posteriorní pravděpodobností a méně pak oblasti s pravděpodobností menší. Ve výše uvedených případech však tomu tak nebylo, což způsobilo nevěrohodnost MCMC diagnostik. Ostatně, Gibbsův vzorkovač nám těžko může něco říct o oblastech parametrického prostoru, které nenavštívil.

Tyto dva případy nastaly z důvodu, že neodezněl efekt počátečního výběru  $\theta^{(0)}$ . Obvyklou praxí je nechat vzorkovač běžet několikrát po sobě s užitím různých  $\theta^{(0)}$ . Pokud Gibbsův vzorkovač dosáhne ve všech případech podobné výsledky, můžeme si být jistí, že jsme provedli dostatečný počet replikací a zároveň jsme i vyhodili dostatečný počet prvních vzorků, což v souhrnu anulovalo vliv počátečního výběru. Tato myšlenka byla formalizována v rámci další MCMC diagnostiky.

Nechť  $\theta^{(0,i)}$  pro  $i = 1, \dots, m$  označuje  $m$  počátečních hodnot z různých oblastí parametrického prostoru. Jedná se o tzv. *overdispersed starting values*, tedy silně rozptýlené počáteční hodnoty. Nechť  $\theta^{(s,i)}$  pro  $s = 1, \dots, S$  označuje  $S$  výběrů navzorkovaných Gibbsovým vzorkovačem z  $i$ -té počáteční hodnoty a  $\hat{g}_{S_i}^{(i)}$  označuje odpovídající odhad  $E[g(\theta)|y]$ . Intuitivně, pokud by efekt počátečních hodnot odezněl, každá z těchto  $m$  sekvencí by měla být stejná resp. podobná. Tedy rozptyl spočtený mezi sekvencemi by neměl být velký vzhledem k rozptylu v rámci sekvence. Obvyklým odhadem rozptylu sekvence je:

$$s_i^2 = \frac{1}{S_1 - 1} \sum_{s=S_0+1}^S \left[ g(\theta^{(s,i)}) - \hat{g}_{S_i}^{(i)} \right]^2 \quad (4.16)$$

což se označuje jako vnitřní rozptyl sekvence. Lze pak definovat průměrný rozptyl vnitřních rozptylů sekvencí jako

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2 \quad (4.17)$$

Obdobně lze ukázat, že mezisekvenční rozptyl je možno odhadnout jako:

$$B = \frac{S_1}{m-1} \sum_{i=1}^m (\hat{g}_{S_i}^{(i)} - \hat{g})^2 \quad (4.18)$$

kde

$$\hat{g} = \frac{1}{m} \sum_{i=1}^m \hat{g}_{S_i}^{(i)} \quad (4.19)$$

Poznamenejme, že  $W$  je odhadem  $\text{var}[g(\theta)|y]$ . Lze ukázat, že dalším odhadem rozptylu,  $\text{var}[g(\theta)|y]$ , je

$$\widehat{\text{var}}[g(\theta)|y] = \frac{S_1 - 1}{S_1} W + \frac{1}{S_1} B. \quad (4.20)$$

Pokud Gibbsův vzorkovač nezkonvergoval, potom  $W$  podhodnotí  $\text{var}[g(\theta)|y]$ . Intuitivně řečeno, pokud Gibbsův vzorkovač prošel jenom část posteriorní hustoty, měl by podhodnotit její varianci. Naopak  $B$  je založeno na sekvencích s rozptýlenými počátečními hodnotami. Tato rozptýlenost by mělo implikovat, že pokud Gibbsův vzorkovač zkonvergoval,  $\widehat{\text{var}}[g(\theta)|y]$  nadhodnocuje  $\text{var}[g(\theta)|y]$ . Obvyklá MCMC konvergenční diagnostika má podobu:

$$\widehat{R} = \frac{\widehat{\text{var}}[g(\theta)|y]}{W} \quad (4.21)$$

Hodnoty by měly být větší než jedna, přičemž hodnoty blízké jedné indikují úspěšnou konvergenci.  $\sqrt{\widehat{R}}$  je označována jako *estimated potential scale reduction*. Lze ji interpretovat jako mez toho, jak vzdálené mohou být odhady směrodatné odchylky  $g(\theta)$  díky nedostatečné konvergenci. V literatuře jsou hodnoty  $\widehat{R}$  větší než 1.2 indikátorem nedostatečné konvergence.

#### 4.1.5 Porovnání modelů: Savage-Dickey density ratio

Pro NLRM s normální-gama apriorní hustotou neexistuje analytická podoba marginální věrohodnosti. Marginální věrohodnost je tak dána standardním výrazem:

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

kde  $p(\beta, h)$  má podobu danou výrazy (4.1) a (4.2) a  $p(y|\beta, h)$  je věrohodnostní funkce normálního lineárního regresního modelu definována v předchozí kapitole. Obecná simulační metoda pro výpočet marginální věrohodnosti je tzv. metoda Gelfanda a Deye. Ta může být někdy dosti komplikovaná. Na tomto místě bude představena jednodušší metoda, která není sice obecně uplatnitelná, ale umožňuje zápis (a následně i výpočet) Bayesova faktoru pro vnořené modely. Tento zápis je nazýván jako *Savage-Dickey density ratio* (*Savageho-Dickeyho poměr hustot*). Je možno tohoto vztahu využít pouze pro porovnání vnořených modelů a to pouze v případě určitých typů apriorních hustot. Pokud jej však můžeme využít, nabízí se velmi jednoduchý způsob jak vypočítat Bayesův faktor a tedy i posteriorní podíl šancí.

Myšlenka stojící v pozadí je následující: Předpokládejme neomezenou verzi modelu  $M_2$  s vektorem parametrů  $\theta = (\omega', \psi)'$ . Věrohodnostní funkce a apriorní hustota je dána hustotami  $p(y|\omega, \psi, M_2)$  a  $p(\omega, \psi, M_2)$ . Omezená verze modelu  $M_1$  má  $\omega = \omega_0$ , kde  $\omega_0$  je vektor konstant. Parametry vektoru  $\psi$  jsou neomezené pro oba modely. Věrohodnostní funkce a apriorní hustota pravděpodobnosti jsou  $p(y|\psi, M_1)$  a  $p(\psi|M_1)$ . Protože  $\omega$  je rovna  $\omega_0$  v rámci modelu  $M_1$  není třeba specifikovat apriorní hustotu pro tento vektor parametrů.

**Teorém 4.1** (Savage-Dickey density ratio). Předpokládejme, že apriorní hustoty obou modelů splňují vztah:

$$p(\psi|\omega = \omega_0, M_2) = p(\psi, M_1) \quad (4.22)$$

potom  $BF_{12}$ , Bayesův faktor porovnávající  $M_1$  a  $M_2$  má podobu:

$$BF_{12} = \frac{p(\omega = \omega_0|y, M_2)}{p(\omega = \omega_0|M_2)} \quad (4.23)$$

kde  $p(\omega = \omega_0|y, M_2)$  a  $p(\omega = \omega_0|M_2)$  jsou postupně neomezená posteriorní a apriorní hustota pravděpodobnosti vyhodnocené v bodě  $\omega_0$ .

Vztah (4.23) je nazýván jako Savage-Dickey density ratio (Savage-Dickeyeho poměr hustot). Důkaz tohoto teorému a poněkud komplikovanější výraz pro Bayesův faktor pro případ, kdy apriorní hustoty nespĺňují vztah (4.22) lze nalézt v článku Verdinelliho a Wassermana [28].

V mnoha případech je rozumné využít stejnou apriorní hustotu pro parametry, které jsou stejné v obou modelech (tj.  $p(\psi|M_2) = p(\psi|M_1)$ ). Ve skutečnosti je podmínka (4.22) mnohem slabší, neboť vyžaduje shodu apriorních hustot pro  $\psi$  pouze v bodě  $\omega = \omega_0$ .

Savage-Dickey density ratio v sobě obsahuje informaci týkající se modelu  $M_2$ , tudíž se nemusíme zatěžovat posteriorní analýzou pro model  $M_1$ . Navíc vztah (4.23) zahrnuje pouze apriorní a posteriorní hustoty, se kterými se velmi snadno pracuje. Přímý výpočet marginální věrohodnosti není potřebný.

Pro ilustraci využití Savage-Dickeyeho poměru hustot v rámci NLRM v této kapitole předpokládáme omezený model  $M_1$  zahrnující omezení  $\beta = \beta_0$ . Případ omezení  $R\beta = r$  je jen jednoduchým rozšířením této ilustrace. Neomezený model  $M_2$  je model diskutovaný v této kapitole. Bayesův faktor porovnávající tyto modely je:

$$BF_{12} = \frac{p(\beta = \beta_0|y, M_2)}{p(\beta = \beta_0|M_2)} \quad (4.24)$$

Jmenovatel tohoto výrazu lze snadno spočítat, neboť marginální apriorní hustota pro  $\beta$  je normální a lze ho spočítat přímo, tedy:

$$p(\beta = \beta_0|M_2) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\underline{V}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\beta_0 - \underline{\beta})' \underline{V}^{-1} (\beta_0 - \underline{\beta}) \right]. \quad (4.25)$$

Spočítat číselník výrazu (4.24) je o něco obtížnější. Ačkoliv víme, že hustota  $p(\beta|y, h, M_2)$  odpovídá normálnímu rozdělení, neznáme podobu  $p(\beta|y, M_2)$ . Ovšem využitím zákonů pravděpodobnosti a výsledků Gibbsova vzorkovače lze  $p(\beta = \beta_0|y, M_2)$  snadno odhadnout. Gibbsovým vzorkovačem získáme náhodné vzorky  $\beta^{(s)}$  a  $h^{(s)}$  pro  $s = S_0 + 1, \dots, S$ . Pokud pak zprůměrujeme  $p(\beta = \beta_0|y, h^{(s)}, M_2)$  přes všechny výběry  $h^{(s)}$ , získáme požadovaný odhad  $p(\beta = \beta_0|y, M_2)$ . Přesněji řečeno,

$$\frac{1}{S_1} \sum_{s=S_0+1}^S p(\beta = \beta_0|y, h^{(s)}, M_2) \rightarrow p(\beta = \beta_0|y, M_2) \quad (4.26)$$

pro  $S_1$  jdoucí k nekonečnu. Protože platí

$$p(\beta = \beta_0|y, h^{(s)}, M_2) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\bar{V}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\beta_0 - \bar{\beta})' \bar{V}^{-1} (\beta_0 - \bar{\beta}) \right] \quad (4.27)$$



průměr pravé strany (4.26) tak získáme velmi snadno. Abychom pochopili, proč platí vztah (4.26), poznamenejme, že zákony pravděpodobnosti implikují

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

Jelikož  $p(\beta = \beta_0 | y, h, M_2)$  v sobě neobsahuje nadále  $\beta$  (je zde nyní vektor konstant  $\beta_0$ ), jedinou náhodnou veličinou v rámci integrálu je  $h$ . Můžeme tak 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 však jsou použitelné právě pro výpočet charakteristik jako je  $E[g(h) | y]$ . Vztah (4.26) nám dává odhad  $p(\beta = \beta_0 | y, M_2)$  podobně jako (4.12) poskytuje odhad  $E[g(\theta) | y]$  pro jakýkoliv vektor parametrů  $\theta$  a jeho funkci  $g(\theta)$ .

Na závěr je dobré poznamenat, že je celá řada modelů, pro které je Gibbsův vzorkovač vhodným posteriorním simulátorem. Pro tyto modely je většinou velmi snadné spočítat i Savage-Dickey density ratio. Jedná se tedy o silný a rozšířený nástroj pro výpočet Bayesova faktoru.

#### 4.1.6 Předpověď

Stejně jako v předchozí kapitole nás bude zajímat predikce pro  $T$  nepozorovaných hodnot závisle proměnné  $y^* = (y_1^*, \dots, y_T^*)'$  danými jako:

$$y^* = X^* \beta + \epsilon^* \quad (4.28)$$

kde  $\epsilon^*$  je nezávislé na  $\epsilon$  a je z  $N(0, h^{-1} I_T)$  a  $X^*$  je matice rozměru  $T \times k$  analogická matici  $X$ , obsahující  $k$  vysvětlujících proměnných pro každé z  $T$  pozorování pro predikci. Predikční hustota je dána jako

$$p(y^* | y) = \iint p(y^* | y, \beta, h) p(\beta, h | y) d\beta dh \quad (4.29)$$

Skutečnost, že  $\epsilon^*$  je nezávislé na  $\epsilon$ , implikuje, že  $y$  a  $y^*$  jsou vzájemně nezávislé, a tudíž  $p(y^* | y, \beta, h) = p(y^* | \beta, h)$ , což lze zapsat jako

$$p(y^* | \beta, h) = \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \exp \left[ -\frac{h}{2} (y^* - X^* \beta)' (y^* - X^* \beta) \right] \quad (4.30)$$

Integrál ve vztahu (4.29) nelze pro NLRM s nezávislou normální-gama apriorní hustotou řešit analyticky. S úspěchem zde využijeme simulačních metod. V podstatě jakákoliv prediktivní charakteristika může být zapsána v podobě  $E[g(y^*) | y]$  pro příslušnou funkci  $g(\cdot)$ . Například výpočet střední hodnoty predikce  $y_i^*$  implikuje  $g(y^*) = y_i^*$ , výpočet rozptylu predikce vyžaduje znalost prediktivní střední hodnoty a  $E[y_i^* | y]$ , tedy bude nás zajímat  $g(y^*) = y_i^{*2}$ . Zajímá nás tedy výpočet

$$E[g(y^*) | y] = \int g(y^*) p(y^* | y) dy^* \quad (4.31)$$

Tento vztah je samozřejmě obdoba vztahu pro analýzu parametru  $\theta$

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)dy^* \quad (4.32)$$

pro jakoukoliv funkci  $g(\theta)$ . Stačí zaměnit  $\theta$  a  $y^*$ . Monte Carlo integrace a Gibbsův vzorkovač využívají skutečnosti, že pokud  $\theta^{(s)}$  pro  $s = 1, \dots, S$  jsou náhodné výběry z posteriorní hustoty, potom

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

konverguje k  $E[g(\theta)|y]$  pro  $S$  jdoucí k nekonečnu.<sup>9</sup> Logicky, pokud jsme schopni nalézt  $y^{*(s)}$  pro  $s = 1, \dots, S$  jakožto výběry z  $p(y^*|y)$ , potom

$$\hat{g}_Y = \frac{1}{S} \sum_{s=1}^S g(y^{*(s)}) \quad (4.33)$$

bude konvergovat k  $E[g(y^*)|y]$ .

Následující postup nám poskytne potřebné vzorky  $y^*$ . Pro každé  $\beta^{(s)}$  a  $h^{(s)}$ , které nám dává Gibbsův vzorkovač, vezmeme výběr  $y^{*(s)}$  z predikční hustoty  $p(y^*|y, \beta^{(s)}, h^{(s)})$ . Protože se jedná o hustotu odpovídající normálnímu rozdělení (viz (4.30)), je tato strategie jednoduše realizovatelná. Máme tedy  $\beta^{(s)}$ ,  $h^{(s)}$  a  $y^{*(s)}$  pro  $s = 1, \dots, S$ . Zákony pravděpodobnosti nám říkají, že  $p(\beta, h, y^*|y) = p(y^*|y, \beta, h)p(\beta, h|y)$  a tedy strategie výběru nejdříve z posteriorní hustoty a potom z  $p(y^*|y, \beta, h)$  nám dává výběr z  $p(\beta, h, y^*|y)$ . Náš soubor výběrů  $\beta^{(s)}$ ,  $h^{(s)}$  a  $y^{*(s)}$  lze tedy využít k vyhodnocení jakékoliv posteriorní charakteristiky využitím vztahu (4.12) a prediktivní charakteristiky za využití (4.33).<sup>10</sup>

Postup uvedený v této části lze využít pro jakýkoliv model, kde je využít posteriorní simulátor pro získání vzorků z  $p(\theta|y)$  a  $p(y^*|y, \theta)$  má podobu, se kterou lze snadno pracovat.

### 4.1.7 Empirická ilustrace

K ilustraci použití Gibbsova vzorkovače pro normální lineární regresní model s nezávislou normální-gama apriorní hustotou použijeme data obsahující údaje o prodejních cenách  $N = 546$  domů prodaných ve Windsoru (Kanada) v roce 1987. Data jsou obsahem souboru `hprice.txt` (stejná data jako v kapitole 3). V této kapitole jsme se rovněž zabývali postupem volby apriorních hyperparametru. Je

<sup>9</sup>V případě Gibbsova vzorkovače je žádoucí prvních  $S_0$  vzorků vyhodit, tudíž sumace půjde od  $S_0 + 1$  do  $S$ .

<sup>10</sup>Tento závěr využívá obecné pravidlo, že pokud máme výběry ze sdružené hustoty pravděpodobnosti  $p(\theta, y^*|y)$ , potom samostatné výběry  $\theta$  jsou výběry z marginálního rozdělení  $p(\theta|y)$  a samotné výběry  $y^*$  jsou výběrem z  $p(y^*|y)$ .

logické, že i zde použijeme obdobný postup, tzn.  $\underline{\nu} = 5$ ,  $\underline{s}^{-2} = 4.0 \times 10^{-8}$  a

$$\underline{\beta} = \begin{bmatrix} 0.0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{bmatrix}.$$

Tyto hodnoty jsou stejné jako hodnoty z předchozí kapitoly a mají podobnou interpretaci, jen s tím rozdílem, že  $\underline{V}$  bude odpovídat přímo kovarianční matici parametrů  $\beta$ , tedy

$$\text{var}(\beta) = \underline{V}.$$

Jen pro připomenutí, v případě přirozeně konjugované apriorní hustoty jsme měli

$$\text{var}(\beta) = \frac{\underline{\nu} s^2}{\underline{\nu} - 2} \underline{V}.$$

Abychom měli srovnatelné apriorní hustoty s těmi z předchozí kapitoly, zvolíme

$$\underline{V} = \begin{bmatrix} 10000^2 & 0 & 0 & 0 & 0 \\ 0 & 5^2 & 0 & 0 & 0 \\ 0 & 0 & 2500^2 & 0 & 0 \\ 0 & 0 & 0 & 5000^2 & 0 \\ 0 & 0 & 0 & 0 & 5000^2 \end{bmatrix}.$$

Všimněme si, že v případě nezávislé normální-gama apriorní hustoty je obvykle snadné stanovení  $\underline{V}$ , neboť odpovídá apriorní kovarianční matici vektoru parametrů  $\beta$ . V případě přirozeně konjugované apriorní hustoty však máme apriorní závislost mezi  $\beta$  a  $h$ , což znamená, že rozptyl  $\beta$  závisí na volbě priorů pro  $h$  a  $\underline{V}$ .

Bayesiánskou analýzu v tomto modelu provedem snadno s využitím Gibbsova vzorkovače. Příslušný skript je podobný tomu, který byl využit v předchozí kapitole pro Monte Carlo integraci, kdy pouze musíme sekvenčně generovat výběry z podmíněných hustot  $p(\beta|h)$  a  $p(h|\beta)$  oproti jednoduchému výběru  $p(\beta)$  z MC integrace. Příslušný soubor se skriptem je označen jako `priklad_NLRMjiny_Gibbs.m`. Mohou být vyžadovány další funkce LeSageho toolboxu [22].

Tabulka 4.1 obsahuje empirické výsledky našeho modelu vztahované k parametru  $\beta$ , včetně MCMC konvergenčních diagnostik. Nastavili jsme počáteční výběr přesnosti chyby jako inverzní hodnotu k OLS odhadu rozptylu  $\sigma^2$  (tzn.  $h^{(0)} = \frac{1}{s^2}$ ). Vyhodili jsme prvních  $S_0 = 1000$  replikací a ponechali  $S_1 = 10000$  replikací. Pro přehlednost neukazujeme výsledky pro  $h$ .

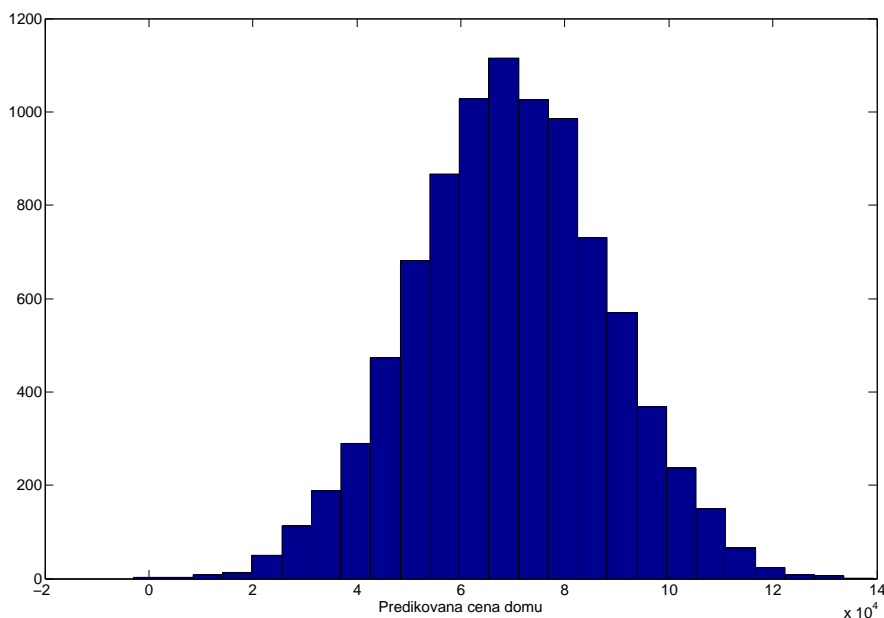
Posterioní střední hodnoty a směrodatné odchylky jsou velmi podobné těm z tabulky 3.1, což odpovídá skutečnosti, že jsme v obou kapitolách použili podobné informativní apriorní hustoty. Sloupec označený jako „NSE“ obsahuje numerickou standardní chybu aproximace  $E(\beta_j|y)$  pro  $j = 1, \dots, 5$ . Pro odhad spektrální hustoty  $S(0)$  byla využita funkce `momentg.m`, nicméně v rámci skriptu

**Tabulka 4.1:** Apriorní a posteriorní střední hodnoty pro parametr  $\beta$  (směrodatné odchylky v závorkách)

	Prior	Posterior	NSE	Gewekeho CD	P. podíl šancí pro $\beta_j = 0$
$\beta_1$	0 (10000)	-4119.01 (3251.44)	34.273	-0.065	1.39
$\beta_2$	10 (5)	5.45 (0.36)	0.004	-1.474	0.00
$\beta_3$	5000 (2500)	3228.83 (1080.46)	11.389	1.073	0.18
$\beta_4$	10000 (5000)	16136.64 (1605.11)	16.919	0.144	0.00
$\beta_5$	10 (5)	7685.55 (987.20)	10.406	-0.597	0.00

byla použita i funkce Matlabu pro odhad spektrální hustoty. Interpretace NSE zůstává stejná, měří tedy přesnost našeho odhadu. Pokud požadujeme vyšší úroveň přesnosti, je třeba zvýšit počet replikací,  $S_1$ . SLOupec označený jako  $CD$  je hodnota Gewekovy konvergenční diagnostiky. V našem případě porovnává odhad  $E(\beta_j|y)$  založený na prvních 1000 ponechaných vzrocích s posledními 4000. Pokud nám efekt počátečních podmínek odezněl a generovali jsme dostatečný počet výběrů, měly by být tyto dva odhady podobné. Vzhledem k tomu, že  $CD$  má asymptoticky standardizované normální rozdělení, platí, že konvergence MCMC algoritmu nastala v případě, pokud je  $CD$  menší než 1.96 v absolutní hodnotě pro všechny parametry (1.96 odpovídá 97.5% kvantilu standardizovaného normálního rozdělení). Z tabulky vidíme, že zkonvergovaly odhady pro všechny parametry.

Tabulka 4.1 obsahuje rovněž i posteriorní podíly šancí porovnávající dva regresní modely:  $M_1 : \beta_j = 0$  vzhledem k  $M_2 : \beta_j \neq 0$ . Stejně jako v kapitole 3, omezený model používá informativní apriorní hustotu, která je stejná jako apriorní hustota neomezeného modelu s tou výjimkou, že  $\beta$  je vektor rozměru  $4 \times 1$  a  $V$  je matice rozměru  $4 \times 4$ , kdy informaci o  $\beta_j$  vynecháváme. Apriorní podíl šancí je jednička. Informace z porovnání modelů v tabulce 4.1 je kvalitativně podobná jako informace z tabulky 3.3. Existuje zde tedy jasný důkaz toho, že  $\beta_2$ ,  $\beta_4$  a  $\beta_5$  jsou nenulové, nicméně je zde určitá nejistota pokud jde o to, jestli je nulový parametr  $\beta_1$  nebo  $\beta_3$ . Pokud porovnáme empirické výsledky z tabulky 4.1 s výsledky z kapitoly 3, jsou výsledky porovnání modelů hodnotově více odlišné než výsledky např. odhadu středních hodnot posteriorních hustot. To je vcelku obvyklý jev. Informativní apriorní hustota má tendenci mít větší vliv na posteriorní podíl šancí než na posteriorní střední hodnoty a směrodatné odchylky. Odlišnosti apriorních hustot kapitol 3 a 4 se tak projevují více v posteriorním



**Obrázek 4.1:** Predikční hustota ceny domu (Gibbsův vzorkovač).

podílu šancí než v posteriorních středních hodnotách.

Snadno si rovněž sestrojíme predikční hustotu pro dům s danými charakteristikami. Budeme opět předpokládat případ, kdy nás zajímá predikce prodejní ceny domu o rozloze 5000 čtverečních stop, dvěma ložnicemi, dvěma koupelnami a jedním patrem. Bohužel v tomto případě nemáme analytické výsledky. Nicméně vlastnosti predikční hustoty lze získat snadnou modifikací skriptu pro simulaci Gibbsovým vzorkovačem. Pokud do něj přidáme řádek kódu pro generování  $y^{*(s)}$  podmíněného  $\beta^{(s)}$  a  $h^{(s)}$  s využitím vztahu (4.30) a uložíme-li si výsledky výběrů  $y^{*(s)}$  pro  $s = S_0 + 1, \dots, S$ , můžeme spočítat jakoukoli charakteristiku predikční hustoty s využitím (4.33). Na tomto základě zjistíme, že posteriorní střední hodnota domu s těmito specifickými charakteristikami je 69499 dolarů a predikční směrodatná odchylka činí 18111. Tyto hodnoty jsou podobné těm z minulé kapitoly.

Pro jednorozměrný (maximálně dvojrozměrný) případ mohou být efektivním způsobem prezentace empirických výsledků grafické metody. Obrázek 4.1 vykresluje naši predikční hustotu. Obrázek je prostým histogramem všech náhodných výběrů,  $y^{*(s)}$  pro  $s = S_0 + 1, \dots, S$ . Aproximace je to z toho důvodu, protože histogram můžeme brát jako diskrétní aproximaci spojitě funkce predikční hustoty. Tneto obrázek nám tak nejen umožňuje hrubý odhad predikční střední hodnoty, ale rovněž nám dává i náhled na rozptýlenost predikovaných hodnot, což je vyjádřeno tloušťkou konců „rozdělení“. Tento obrázek nám napovídá, že daná datová množina nám neumožňuje dostatečnou (přesnou) predikční

analýzu. Ačkoliv je naše nejlepší predikce ceny domu o rozloze 5000 čtverečních stop, dvou ložnicích, dvou koupelnách a jednom patru asi 70000 dolarů, predikční hustota dává i nezanedbatelnou pravděpodobnost cenám do 30000 dolarů nebo nad 110000 dolarů.

## 4.2 NLRM s omezeními ve tvaru nerovnosti

V této části bude diskutován lineární regresní model, ve kterém budeme požadovat omezení koeficientů ve tvaru nerovnosti. Tento požadavek lze zapsat v podobě  $\beta \in A$ , kde  $A$  je příslušná relevantní oblast parametrického prostoru. Bayesiánské řešení tohoto problému je jednoduché, neboť omezení kladená na parametry lze zakomponovat do apriorní hustoty pravděpodobnosti. Pro posteriorní analýzu se využívá technika zvaná *importance sampling*, tedy něco jako vzorkování dle důležitosti či vážené vzorkování. Pro některé typy nerovnostních omezení (např. lineární omezení  $\beta_j > 0$ ) je možno využít ještě jednodušších metod. Nicméně, importance sampling je i tak dosti jednoduchá a průhledná metoda posteriorní analýzy a lze ji využít pro jakýkoliv typ omezení ve tvaru nerovnosti a v podstatě i pro jakýkoliv typ modelů.

### 4.2.1 Apriorní hustota

Pro zavedení nerovnostních omezení je obvyklé využít apriorní hustoty. Tedy tvrzení  $\beta \in A$  je ekvivalentní tvrzení, že oblast parametrického prostoru mimo  $A$  je apriori vyloučená, a tedy měly by jí být přiřazeny apriorní váhy hodnoty 0. Tuto apriorní informaci lze kombinovat s dalšími apriorními informacemi, například je možná kombinace s nezávislým apriorním normálním-gama rozdělením, nebo přirozeně konjugovaným priorem. V rámci přiblížení konceptu importance sampling ji budeme kombinovat s přirozeně konjugovanou apriorní hustotou danou ve vztahu (3.8). Speciálním případem je pak neinformativní prior daným vztahem (3.24). Tento neinformativní prior je užitečné využít v případě, kdy chceme zavést omezení ve tvaru nerovnosti na parametry  $\beta$ , ale nemáme žádnou další apriorní informaci.

Apriorní hustota je tedy dána jako

$$p(\beta, h) \propto f_{NG}(\beta, h | \underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}) 1(\beta \in A) \quad (4.34)$$

kde  $\underline{\beta}$ ,  $\underline{V}$ ,  $\underline{s}^{-2}$ ,  $\underline{\nu}$  jsou apriorní hyperparametry a  $1(\beta \in A)$  je indikační funkce, která nabývá hodnot 1, pokud  $\beta \in A$  a 0 jinak.

V případě přirozeně konjugovaného prioru odpovídá marginální apriorní hustota parametru  $\beta$   $t$ -rozdělení (obdobně jako marginální posteriorní hustota):

$$p(\beta) \propto f_t(\beta | \underline{\beta}, \underline{s}^2 \underline{V}, \underline{\nu}) 1(\beta \in A) \quad (4.35)$$

Neinformativní variantu přirozeně konjugovaného prioru získáváme nastavením  $\underline{\nu} = 0$  a  $\underline{V}^{-1} = cI_k$ , kde  $c$  se blíží nule:

$$p(\beta, h) \propto \frac{1}{h} 1(\beta \in A) \quad (4.36)$$

### 4.2.2 Posteriorní hustota

Odvození posteriorní hustoty je analogické odvození v kapitole 3 s jedinou výjimkou, kdy musíme brát v úvahu omezení ve tvaru nerovnosti. Sdružená posteriorní hustota pro  $\beta$  a  $h$  tedy bude z normálního-gama rozdělení a marginální posteriorní hustota pro  $\beta$  bude odpovídat vícerozměrnému  $t$ -rozdělení (neinformativní prior pak byl speciálním případem přirozeně konjugovaného prioru). Získáváme tedy totožné výsledky, kdy výsledné hustoty budou ohraničené. Přesněji,  $p(\beta, h|y)$  bude ohraničená v rámci oblasti  $\beta \in A, (\beta|y)$  je vícerozměrné  $t$ -rozdělení omezené na stejnou oblast  $\beta \in A$ :

$$p(\beta|y) \propto f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})1(\beta \in A) \quad (4.37)$$

kde  $\bar{\beta}$ ,  $\bar{V}$ ,  $\bar{s}^{-2}$  a  $\bar{\nu}$  jsou definovány rovnicemi (3.10)-(3.13). Posteriorní hustota pro neinformativní apriorní hustotu má tutéž podobu, kde  $\bar{\beta}$ ,  $\bar{V}$ ,  $\bar{s}^{-2}$  a  $\bar{\nu}$  jsou definovány rovnicemi (3.20)-(3.23)

Pokud bychom kombinovali naše omezení s nezávislou normální-gama apriorní hustotou, pak by vztah pro  $p(\beta|y, h)$  v (4.6) měl být násoben  $1(\beta \in A)$ .

### 4.2.3 Importance sampling

Pro některou volbu  $A$  jsou k dispozici analytické výsledky. Pro jiné lze využít Gibbsův vzorkovač. Pro obecnou volbu  $A$  však nelze využít žádný z výše uvedených přístupů. Ukážeme si tedy přístup posteriorní simulace zvaný importance sampling. Jelikož se jedná o obecně uplatnitelnou metodu, využijeme obecného značení. Budeme tedy využívat vektor parametrů  $\theta$ , věrohodnostní funkci  $p(y|\theta)$ , apriorní hustotu  $p(\theta)$  a posteriorní hustotu  $p(\theta|y)$ .

Monte Carlo integrace v sobě zahrnuje náhodné výběry z  $p(\theta|y)$ , ovšem v celé řadě modelů je není snadné získat. Předpokládejme na místo toho, že náhodné výběry  $\theta^{(s)}$  pro  $s = 1, \dots, S$  pocházejí z hustoty pravděpodobnosti  $q(\theta)$ , ze které je snadné tyto výběry získat. Tato hustota se nazývá *importance function*. Je zřejmé, že průměrování těchto výběrů nezískáme požadované výsledky, neboť

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

nebude konvergovat k  $E[g(\theta)|y]$  pro  $S \rightarrow \infty$ . Pro lepší představu předpokládejme situaci, kdy  $q(\theta)$  a  $p(\theta|y)$  mají podobnou střední hodnotu, nicméně  $q(\theta)$  má vyšší variabilitu než  $p(\theta|y)$ . Náhodné výběry z  $q(\theta)$  budou obsahovat příliš mnoho výběrů z okrajů hustoty  $p(\theta|y)$  a příliš málo výběrů z blízkosti střední hodnoty hustoty pravděpodobnosti  $p(\theta|y)$ . Importance sampling koriguje tento nesoulad tím, že přiřazuje menší váhu odlehlým výběrům a větší váhu výběrům blízko střední hodnoty. Jednoduše řečeno, namísto jednoduchého průměru bereme vážený průměr.

**Teorém 4.2** (Importance sampling). Necht  $\theta^{(s)}$  pro  $s = 1, \dots, S$  je náhodný výběr z  $q(\theta)$  a definujme

$$\hat{g}_S = \frac{\sum_{s=1}^S w(\theta^{(s)})g(\theta^{(s)})}{\sum_{s=1}^S w(\theta^{(s)})} \quad (4.38)$$

kde

$$w(\theta^{(s)}) = \frac{p(\theta = \theta^{(s)}|y)}{q(\theta = \theta^{(s)})} \quad (4.39)$$

potom  $\hat{g}_S$  koverguje k  $E[g(\theta)|y]$  pro  $S$  jdoucí k nekonečnu (při splnění slabých podmínek)<sup>11</sup>.

Jelikož váhy se objevují v čitateli i jmenovateli (4.38), je třeba 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 (4.39) nahradit výrazem

$$w(\theta^{(s)}) = \frac{p^*(\theta = \theta^{(s)}|y)}{q^*(\theta = \theta^{(s)})} \quad (4.40)$$

a výše zmíněný teorém bude stále platit.

Jak se na první pohled může zdát, importance sampling je krásným řešením jakéhokoliv problému posteriorní simulace. Zdá se, že můžeme brát náhodné vzorky z obvyklé a známé hustoty  $q(\theta)$  a využitím (4.38) pro odhad  $E[g(\theta)|y]$ . Bohužel v praxi to není tak jednoduché. Pokud  $q(\theta)$  nebude dobře aproximovat  $p(\theta|y)$ , mohou nastat případy, kdy  $w(\theta^{(s)})$  je nula pro téměř každý výběr. To znamená, že vážený průměr bude v sobě zahrnovat velmi málo výběrů. To by znamenalo, že  $S$  by muselo být enormně vysoké, abychom získali dostatečně přesné odhady  $E[g(\theta)|y]$ . Importance sampling tak může být neproveditelné, pokud  $q(\theta)$  není vhodně zvolena. Jelikož vhodný výběr  $q(\theta)$  s sebou může přinášet hodně práce a rozdílné  $q(\theta)$  jsou použitelné pro různé třídy modelů, většinou se v rámci možností využívá jiných strategií, jako je Gibbsův vzorkovač. Pokud zvolíme počet bloků, Gibbsův vzorkovač dovoluje vytvářet náhodné výběry z podmíněných hustot pravděpodobnosti (a monitorovat konvergenci). Importance sampling v sobě obsahuje hledání a ověření obvyklé třídy importance funkcí (např. z třídy normálních rozdělání), a po té je třeba v rámci této třídy příslušnou funkci vhodně zadefinovat (např. vybrat střední hodnotu a rozptyl normálního rozdělání) abychom dobře aproximovali  $p(\theta|y)$ . Zejména pokud je  $\theta$  vícedimenzionální, může být extrémně obtížné najít vhodnou importance funkci. Obecnou strategii pro výběr rozumné importance funkce nabízí Geweke [13].

Pro normální lineární regresní model omezený požadavky ve formě nerovností se nabízí importance function v podstatě sama a importance sampling tak lze provést vcelku jednoduše. Předpokládejme co se stane, když zvolíme

$$q(\beta) = f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{v}) \quad (4.41)$$

<sup>11</sup>Tyto podmínky obnášejí zejména to, že  $q(\theta)$  podporuje  $p(\theta|y)$  a  $E[g(\theta|y)]$  existuje.



Z této importance function je snadné provádět náhodné výběry, neboť se jedná o vícerozměrné  $t$ -rozdělení. Využitím (4.37) a (4.40) můžeme váhy spočítat jako

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

a (4.38) lze využít pro odhad  $E[g(\beta)|y]$  pro jakoukoliv funkci parametrů  $g(\cdot)$ , která nás zajímá. Všechny tyto váhy jsou buď rovny jedné, pokud  $\beta^{(s)} \in A$ , a nebo nula, pokud  $\beta^{(s)} \notin A$ . Tento postup tedy zahrnuje náhodný výběr z neomezené posteriorní hustoty a vyhození výběrů, které nesplňují omezení (mají nulovou váhu). Tento postup je tedy snadné využít, pokud  $A$  není příliš malá oblast, kdy by veškeré vzorky byly vyhozeny.

Numerická standardní chyba se snadno vypočte využitím centrální limitní věty.

**Teorém 4.3** (Numerická standardní chyba). Využitím definice a značení v rámci teorému 4.2 platí

$$\sqrt{S}\{\hat{g}_S - E[g(\theta)|y]\} \rightarrow N(0, \sigma_g^2) \quad (4.42)$$

pro  $S$  jdoucí k nekonečnu, přičemž  $\sigma_g^2$  je možno konzistentně odhadnout jako

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

Numerickou standardní chybu,  $\frac{\hat{\sigma}_g}{\sqrt{S}}$ , lze tedy vypočítat, což nám napomůže ve volbě  $S$ .

#### 4.2.4 Porovnání modelů

Omezení ve tvaru nerovnosti obvykle neumožňuje přímý výpočet marginální věrohodnosti takového modelu. Předpokládejme případ, ve kterém  $M_1$  je normální lineární regresní model s přirozeně konjugovaným priorem s omezeními ve tvaru nerovnosti (tj.  $\beta \in A$ ). Nechť  $M_2$  je tentýž model, přičemž předchozí restriktace neplatí (tj.  $\beta \notin A$ ). Jelikož omezení ve tvaru nerovnosti jsou obvyklou implikací ekonomické teorie, porovnání modelů v této podobě je pro nás více než zajímavé. Ekonomická teorie může implikovat  $\beta \in A$  a  $p(M_1|y)$  tak označuje pravděpodobnost, že ekonomická teorie je v souladu s daty, že tedy platí.

Příklad modelů tohoto typu (a jejich porovnání) byl diskutován v kapitole 3. Prakticky lze tedy využít neomezený NLRM s přirozeně konjugovaným priorem k výpočtu  $p(M_1|y) = p(\beta \in A|y)$  a  $p(M_2|y) = 1 - p(M_1|y)$ . Pokud mají omezení ve tvaru nerovnosti lineární podobu,  $p(\beta \in A)$  lze vypočítat analyticky. Alternativně lze využít i importance sampling, v rámci neomezeného modelu,  $p(\beta \in A|y) = E[g(\theta)|y]$ , kde  $g(\theta) = 1(\beta \in A)$ . Právě posteriorní simulace je ideální pro vyhodnocení takovýchto charakteristik. Můžeme tedy vzít náhodné výběry u neomezené posteriorní hustoty ( $f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{v})$ ) a jednoduše spočítat podíl těch, které splňují  $\beta \in A$ . Tento podíl je odhadem  $p(\beta \in A|y)$ . Ovšem

výběr z  $f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})$  je právě to, co bylo diskutováno v předchozí části při ilustraci techniky importance sampling. Jejím provedením a zachováním informace o tom, kolik vzorků jsme nechali a kolik vyhodili (přičítali jim nulovou váhu), získáme základ pro výpočet  $p(M_1|y)$  a  $p(M_2|y)$ .

Savage-Dickey density ratio lze využít k porovnání vnořených modelů, na které jsou kladeny tytéž nerovnostní omezení. Nechť  $M_2$  je NLRM s přirozeně konjugovaným priorem a nerovnostními omezeními, jehož posteriorní hustota je dána rovnicí (4.37) a nechť  $M_1$  je stejný jako  $M_2$  s tou výjimkou, že  $\beta = \beta_0$ . Pokud předpokládáme v obou modelech tutéž apriorní hustotu pro  $h$ , Savage-Dickey density ratio nám říká, že Bayesův faktor lze spočítat jako

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

Naneštěstí vyhodnocení Bayesova faktoru není tak jednoduché, jak by se mohlo znát, neboť výsledky v (4.35) a (4.37) jsou apriorními a posteriorními jádrovými hustotami (není zde znaménko rovnosti). Formálně mají apriorní a posteriorní hustoty podobu:

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

kde  $\underline{c}$  a  $\bar{c}$  jsou apriorní a posteriorní konstanty zajišťující integrovatelnost těchto hustot na hodnotu jedna. Savage-Dickey ratio má tedy podobu

$$BF_{12} = \frac{\bar{c}f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})}{\underline{c}f_t(\beta|\underline{\beta}, \underline{s}^2\underline{V}, \underline{\nu})} \quad (4.43)$$

Tento vztah vyžaduje vyhodnocení dvou hustot pravděpodobnosti z vícerozměrného  $t$ -rozdělení v bodě  $\beta = \beta_0$  a výpočet konstant  $\underline{c}$  a  $\bar{c}$ . Pro některé typy hypotéz je výpočet těchto konstant snadný. Předpokládejme případ jednorozměrného omezení mající podobu  $\beta_j > 0$ . V tomto případě je možno využít statistických tabulek  $t$ -rozdělení (nebo jejich počítačové podoby) pro získání těchto integračních konstant. Pro obecnější nerovnosti je třeba využít metody předchozího odstavce, která počítala pravděpodobnost  $p(M_1|y)$ , jež odpovídala pravděpodobnosti, že omezení  $\beta \in A$  platí. Ovšem  $\bar{c} = \frac{1}{p(M_1|y)}$ , protože

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

a  $p(M_1|y) = \int f_t(\beta|\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})1(\beta \in A)d\beta$ . Výpočet  $\underline{c}$  je podobný, jen je třeba využít importance sampling v rámci apriorní hustoty pravděpodobnosti.

#### 4.2.5 Předpověď

Postup vyjádřený ve vztazích (4.28) až (4.33) lze využít i v tomto případě. Výběry z importance function je třeba vážit způsobem popsáním v (4.38) a (4.39).

V rámci obecného značení označme  $\theta^{(s)}$  jakožto náhodný výběr z importance function a  $y^{*(s)}$  jako náhodný výběr z  $p(y^*|y, \theta^{(s)})$  pro  $s = 1, \dots, S$ . Potom

$$\hat{g}_Y = \frac{\sum_{s=1}^S w(\theta^{(s)})g(y^{*(s)})}{\sum_{s=1}^S w(\theta^{(s)})} \quad (4.44)$$

konverguje k  $E[g(y^*)|y]$  pro  $S$  jdoucí k nekonečnu, kdy  $w(\theta^{(s)})$  je dána výrazem (4.39) nebo (4.40). Tento postup pro výpočet predikčních charakteristik je možno využít všude tam, kde je prováděno importance sampling, tedy i v případě normálního lineárního regresního modelu s přirozeně konjugovaným priorrem s omezením ve tvaru nerovnosti.

### 4.2.6 Empirická ilustrace

K ilustraci techniky importance sampling opět použijeme dana o prodejních cenách  $N = 546$  domů prodaných ve Windsoru (Kanada) v roce 1987. Data jsou tradičně obsahem souboru `hprice.txt`. Připomeňme si, že závisle proměnná je prodejní cena domu a vysvětlující proměnné jsou rozloha domu, počet ložnic, počet koupelen a počet pater. Budeme předpokládat, že všechny tyto veličiny mají kladný vlv na cenu domu. Navíc předpokládejme, že máme informaci, že  $\beta_2 > 5$ ,  $\beta_3 > 2500$ ,  $\beta_4 > 5000$  a  $\beta_5 > 5000$  a tuto informaci chceme samozřejmě včlenit do naší apriorní hustoty. To nám bude definovat oblast omezení parametrů,  $A$ . Apriorní hustota je součinem  $1(\beta \in A)$  normální-gama apriorní hustoty. Předpokládáme, že se nyní pohybujeme v rámci modelu s přirozeně konjugovanou normální-gama apriorní hustotou. Musíme si tedy zvolit hyperparametry  $\underline{\beta}$ ,  $\underline{V}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$ . Zvolíme si stejné hodnoty jako v kapitole 3. To znamená, že volíme  $\underline{s}^{-2} = 4.0 \times 10^{-8}$ ,  $\underline{\nu} = 5$ ,

$$\underline{\beta} = \begin{bmatrix} 0.0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{bmatrix}$$

a

$$\underline{V} = \begin{bmatrix} 2.40 & 0 & 0 & 0 & 0 \\ 0 & 6.0 \times 10^{-7} & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 & 0 \\ 0 & 0 & 0 & 0.60 & 0 \\ 0 & 0 & 0 & 0 & 0.60 \end{bmatrix}.$$

Přestože bychom byli schopni využít přímo Monte Carlo integraci založenou na výběrech z omezené ho normálního rozdělení (díky jednoduchosti omezení), zaměříme se na využití techniky importance sampling. Příslušný programový kód je jednoduchý a je obsažen v souboru `priklad_NLRMjiny_IS.m`. Jako importance function můžeme využít (4.41) což je ale stejná funkce jako je posteriorní hustota pravděpodobnosti z kapitoly 2. Váhy pro importance sampling získáme jednoduše na základě vztahu (4.37). Při dané volbě importance function

je tak jasné, že váhy budou nabývat hodnoty 1 (pokud výběr splňuje omezení) nebo 0 (pokud výběr omezení nespĺňuje). Na základě váženého průměru jsme schopni získat posteriorní vlastnosti vektoru parametrů  $\beta$  (resp. každého jeho prvku). Jsme schopni spočítat i numerickou standardní chybu. Tabulka 4.2 obsahuje posteriorní střední hodnoty, standardní odchylky a numerické standardní chyby parametrů  $\beta$  spolu s hodnotami posteriorního podílu šancí porovnávající model s  $\beta_j = \underline{\beta}_j$  s modelem zahrnujícím pouze zavedená nerovnostní omezení. Tato volba modelů, které porovnáváme je pouze ilustrativní, kdy k výpočtu posteriorního podílu šancí využíváme vztah (4.43). Protože  $\beta_j = \underline{\beta}_j$  je jedno-rozměrné omezení, můžeme  $\underline{c}$  a  $\bar{c}$  spočítat na základě vlastností jednorozměrného  $t$ -rozdělení. Tabulka 4.2 je založena na 10000 replikacích (tj.  $S = 10000$ ).

**Tabulka 4.2:** Posteriorní výsledky pro parametr  $\beta$

	Stř. hodnota	Směrodatná odchylka	NSE	P. podíl šancí pro $\beta_j = \underline{\beta}_j$
$\beta_1$	-5658.15	3011.44	41.245	1.20
$\beta_2$	5.50	0.30	0.004	0.00
$\beta_3$	3571.50	777.15	10.644	0.49
$\beta_4$	16638.59	1671.19	22.889	0.00
$\beta_5$	7454.92	925.41	12.675	0.22

Výsledky jsou velmi podobné těm z tabulek 3.1 nebo 4.1. Všimněme si, že pro parametry  $\beta_4$  a  $\beta_5$  měly zavedené restriktce nepatrný dopad. Posteriorní střední hodnota (resp. směrodatná odchylka) pro  $\beta_4$  a  $\beta_5$  jsou 16965.24 (1708.02) respektive 7634.90 (1004.34). Veškerá posteriorní hustota se tak nachází v oblasti  $\beta_4 > 5000$  a  $\beta_5 > 5000$  a dodání takovéhoho dplňku k apriorní hustotě tak nepřineslo žádnou novou informaci.

Omezení ve tvaru nerovnosti však ovlivnila  $\beta_2$  a  $\beta_3$ , kdy došlo ke zvýšení jejich posteriorní střední hodnoty. Pokud a priori odřízneme oblast  $\beta_2 < 5$  a  $\beta_3 < 2500$ , není překvapující, že se střední hodnota zvýší. Došlo rovněž ke snížení posteriorních směrodatných odchylek, což nám naznačuje, že dodání další apriorní informace snížilo naši posteriorní nejistotu o hodnotách daných parametřů.

Numerické standardní chyby nám říkají, že jsme dosáhli rozumné přesnosti odhadů, ale pokud bychom chtěli přesnost vyšší, tak stačí (stejně jako u jakýchkoli jiných posteriorních simulátorů) zvýšit počet replikací,  $S$ . Porovnáním s tabulkou 3.4 však ukazují, že numerické standardní chyby (a tedy aproxima-tivní chyby odhadu) jsou v rámci importance sampling vyšší než u Monte Carlo integrace. Např. se obdobným počtem replikací 10000 byla NSE odhadu  $E(\beta_2|y)$  rovna v případě Monte Carlo integrace hodnotě 0.004, po zaokrouhlení směrem nahoru, kdežto u importance sampling 0.004, po zaokrouhlení směrem dolů. I když tento rozdíl nemusí být až tak patrný, je vcelku logický, protože Monte Carlo integrace vychází z výběřů přímo z posteriorní hustoty, kdežto impor-

tance sampling generuje výběry z importance function, což je jen aproximace posteriorní hustoty. Importance sampling tak je numericky méně efektivní.

Posteriorní podíly šancí jsou v souladu s tím co nám poskytují výsledky odhadu posteriorní střední hodnoty a směrodatné odchylky. S výjimkou úrovně konstanty, nemáme důkaz pro tvrzení, že  $\beta_j = \underline{\beta}$ . Pro parametry  $\beta_3$  a  $\beta_5$  dává posteriorní podíl šancí vyšší pravděpodobnost platnosti restrikcí. Pro tyto koeficienty totiž není posteriorní střední hodnota příliš vzdálená od  $\underline{\beta}_j$  (vzhledem k příslušné směrodatné odchylce), posteriorní podíl šancí tak je svou hodnotou vcelku odpovídající.

Predikční hodnotu ceny domu s danými charakteristikami lze spočítat tak jak je zmiňováno v části 4.2.5. Každý výběr v rámci importance sampling tak lze využít k výběru  $y^{*(s)}$  pro  $s = 1, \dots, S$ . Tyto výběry se pak odpovídajícím způsobem zprůměrují (viz (4.44)) pro získání požadovaných predikčních charakteristik. V předchozí empirické ilustraci v části 4.1.7 jsme generovali výběry z  $p(y^*|\beta^{(s)}, h^{(s)})$ . To bylo vcelku snadné, neboť se jednalo o hustotu odpovídající normálnímu rozdělení. Je snadné převzít podobnou strategii i pro náš příklad, přestože v tomto případě musíme rozšířit naši importance function pro generování výběrů z  $h^{(s)}$ . Logickou volbou pro tuto funkci bude normální-gama posteriorní hustota z (3.9). Alternativně lze využít analogické postupy přechodu z (3.39) k (3.40), což nám implikuje, že

$$p(y^*|y, \beta) = p(y^*|\beta) = f_t(y^*|X^*\beta, \bar{s}^2 I_T, \bar{v}).$$

Výběry z  $p(y^*|\beta^{(s)})$  lze tak generovat z  $t$ -rozdělení. Pokud nás překvapuje to, kam se poděly naše omezení kladená na  $\beta$ m stačí si uvědomit, že predikční výběry z  $p(y^*|\beta^{(s)})$  jsou podmíněny výběry  $\beta$  na základě importance sampling. Tyto výběry již v sobě daná omezení zahrnují. Pokud tedy použijeme tento postup pro získání charakteristik predikční hustoty prodejní ceny domu s rozlohou 5000 čtverečních stop, se dvěma ložnicemi, dvěma koupelnami a jedním patrem, zjistíme, že predikční střední hodnota a směrodatná odchylka jsou postupně 69639 a 18311. Tyto výsledky jsou podobné těm z předchozích empirických ilustrací využívajících tuto datovou sadu.

## 4.3 Shrnutí

V této kapitole jsme si popsali metody bayesovské posteriorní analýzy, predikční analýzy a porovnání modelů pro normální lineární regresní model s dvěma typy apriorních hustot. První z nich byla nezávislá normální-gama apriorní hustota a druhá přirozeně konjugovaná s omezením ve tvaru nerovnosti. Tyto apriorní hustoty se nám mohou hodit v řadě empirických aplikací. Důvod však byl i ten, že nám dovolily zavést dvě důležité metody bayesovského výpočtu, a to v nám známém kontextu. První z těchto metod byl Gibbsův vzorkovač. Oproti Monte Carlo integraci, zahrnující náhodné výběry ze sdružené hustoty pravděpodobnosti, využívá Gibbsův vzorkovač sekvenční výběry z plně podmíněných apriorních hustot. Tyto výběry lze brát jako výběry ze sdružené hustoty

pravděpodobnosti, nicméně je třeba dávat si pozor v důsledku toho, že se nejedná o nezávislé výběry a mohou být ovlivněny volbou počátečních podmínek (počátečního výběru). Odstranění těchto problémů lze ověřit s pomocí MCMC konvergenčních diagnostik.

Druhou výpočetní metodou bylo importance sampling. Tento algoritmus umožňuje generovat výběry z „importance function“, tyto výběry následně odpovídajícím způsobem převáží pro zohlednění toho, že „importance function“ a posteriorní hustota nejsou identické. V rámci této kapitoly byl zaveden i Savageho-Dickeyeho poměr hustot, což je obvyklý způsob zápisu Bayesova faktoru pro vnořené modely.

Máme tak k dispozici již tři algoritmy posteriorní simulace: Monte Carlo integraci, Gibbsův vzorkovač a importance sampling. Otázka, který z nich využít, závisí na podstatě řešeného problému. Pokud dokážeme snadno generovat náhodné výběry z posteriorní hustoty, Monte Carlo integrace je tou správnou volbou. Pokud je generování vzorků s posteriorní hustoty obtížné, nicméně je snadné generování z podmíněných hustot, nabízí se Gibbsův vzorkovač. Pokud nelze použít ani Monte Carlo integraci, ani Gibbsův vzorkovač, nicméně můžeme posteriorní hustotu snadno aproximovat, je rozumnou volbou importance sampling.

# Kapitola 5

## Nelineární regresní model

### 5.1 Úvod

V předchozích kapitolách jsme pracovali s lineárním regresním modelem v podobě:

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

pro jednotlivá pozorování  $i = 1, \dots, N$ . Tento model je užitečný nejen v případech, kdy vztah mezi vysvětlovanou a vysvětlujícími proměnnými je lineární, ale i v případech, kdy tento vztah lze převést do lineární podoby. Příkladem může být Cobb-Douglasova produkční funkce vyjadřující závislost výstupu  $y$  na vstupech  $x_2, \dots, x_k$

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

Logaritmováním obou stran této rovnice a přidáním náhodné složky získáváme regresní model:

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

kde  $\beta_1 = \ln(\alpha_1)$ . Tento model je již lineární v logaritmech jak vysvětlované tak i vysvětlujících proměnných a lze tak uplatnit postupy z předchozích kapitol.

Některé funkční formy však do lineární podoby snadno transformovat nelze. Příkladem skutečně nelineárního funkčního tvaru je produkční funkce s konstantní elasticitou substituce (CES funkce), která má podobu

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

V této kapitole se zaměříme na bayesiánskou analýzu regresních modelů, v kterých vystupují vysvětlující proměnné nelineární formou. Empirická ilustrace bude vedena v intencích CES produkční funkce, tedy nelineární regresní model

bude mít podobu:

$$y_i = \left( \sum_{j=1}^k \gamma_j x_{ij}^{\gamma_{k+1}} \right)^{\frac{1}{\gamma_{k+1}}} + \epsilon_i. \quad (5.1)$$

Notace bude standardní, tedy  $\epsilon$  a  $y$  jsou  $N$ -rozměrné vektory obsahující náhodnou složku a pozorování závisle proměnné,  $X$  je matice rozměru  $N \times k$  obsahující pozorování  $k$  vysvětlujících proměnných. Tradiční předpoklady jsou:

1.  $\epsilon$  je z normálního rozdělení  $N(0_N, h^{-1}I_N)$ .
2. Všechny prvky matice  $X$  jsou pevná čísla (tj. nenáhodné proměnné) nebo pokud jsou to náhodné veličiny, jsou nezávislé se všemi prvky vektoru  $\epsilon$ , přičemž jejich funkce hustoty pravděpodobnosti je  $p(X|\lambda)$ , kde  $\lambda$  je vektor parametrů, který neobsahuje žádný s ostatních parametrů modelu.

Hlavní myšlenky této kapitoly budou platit pro obecný nelineární regresní model mající podobu

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

kde  $X_i$  je  $i$ -tý řádek matice  $X$ ,  $f(\cdot)$  je funkce závislejší na  $X_i$  a vektoru parametrů  $\gamma$ . Tento model můžeme v maticové podobě zapsat jako:

$$y = f(X, \gamma) + \epsilon, \quad (5.2)$$

kde  $f(X, \gamma)$  je  $N$ -rozměrný vektor funkcí s  $i$ -tým prvkem daným obecně nelineární funkcí  $f(X_i, \gamma)$ . Přesná implementace algoritmu posteriorní simulace diskutované v této kapitole bude záviset na tvaru  $f(\cdot)$ . Základní koncept tak bude nejprve rozebírán za použití vztahu (5.2).

Nelineární model je velmi užitečný v řadě modelových aplikací. V rámci něho si ukážeme řadu technik, které budou aplikovatelné v podstatě v jakémkoliv modelu. Lineární regresní model byl velmi speciálním případem, který umožňoval prezentovat výsledky posteriorní analýzy v analytickém vyjádření. Dokonce i při použití apriorních hustot, které analytickou posteriorní analýzu neumožňují, jsou k dispozici speciální techniky posteriorní analýzy (např. Gibbsův vzorkovač a Savage-Dickey density ratio). Některé modely však použití těchto speciálních technik neumožňují a je tedy důležité vyvinout obecné metody využitelné v jakémkoliv modelu. Nelineární regresní model dovoluje zavedení těchto obecných metod v kontextu velmi podobnému lineárnímu regresnímu modelu. V rámci posteriorní simulace zavedeme významnou třídu posteriorních simulátorů zvaných *Metropolis-Hastings algoritmy*. Zavedeme rovněž obecnou metodu výpočtu marginální věrohodnosti vyvinutou Gelfandem a Deyem [12]. Měřítkem kvality modelu pak bude charakteristika zvaná posteriorní predikční  $p$ -hodnota (*posterior predictive p-value*).



## 5.2 Věrohodnostní funkce

Z definice vícerozměrného normálního rozdělení lze zapsat věrohodnostní funkci nelineárního regresního modelu jako

$$p(y|\gamma, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp \left[ -\frac{h}{2} \{y - f(X, \gamma)\}' \{y - f(X, \gamma)\} \right] \right\}. \quad (5.3)$$

V rámci lineárního regresního modelu bylo možno zapsat tento výraz pomocí kvantit OLS odhadů, což nahrávalo použití přirozeně kojugované apriorní hustoty. Takovému zjednodušení zde není možné, pokud  $f(\cdot)$  nemá vyložené specifický tvar.

## 5.3 Apriorní hustota

Volba apriorní hustoty bude záviset na podobě  $f(\cdot)$  a významu parametrů  $\gamma$ . Například pro případ CES produkční funkce odpovídá  $\gamma_{k+1}$  odpovídá pružnosti substituce mezi jednotlivými vstupy. V tomto případě by výzkumník pravděpodobně měl apriorní informaci o tom, jakých hodnot by tento parametr měl nabývat. Volba apriorní hustoty tak závisí na kontextu, ve kterém probíhá empirické analýza. V této kapitole budeme používat buď obecnou apriorní hustotu  $p(\gamma, h)$  nebo neinformativní prior využívaný v rámci LRM:

$$p(\gamma, h) = \frac{1}{h}. \quad (5.4)$$

Tato apriorní hustota je uniformní pro  $\gamma$  a  $\ln(h)$ .

## 5.4 Posteriorní hustota

Posteriorní hustota je proporcionální součinu věrohodnostní funkce a apriorní hustoty a lze ji zapsat jako

$$p(\gamma, h|y) \propto p(\gamma, h) \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp \left[ -\frac{h}{2} \{y - f(X, \gamma)\}' \{y - f(X, \gamma)\} \right] \right\}. \quad (5.5)$$

Obecně není možno tento výraz dále zjednodušit. Vše závisí na specifikaci  $p(\gamma, h)$  a  $f(\cdot)$ . Posteriorní hustota nebude nabývat tvaru hustoty obvyklého rozdělení. Pokud využijeme apriorní hustotu z (5.4), lze  $h$  analyticky "vyintegrovat" a marginální posteriorní hustota pro  $\gamma$  je

$$p(\gamma|y) \propto [\{y - f(X, \gamma)\}' \{y - f(X, \gamma)\}]^{-\frac{N}{2}}. \quad (5.6)$$

V případě, kdy  $f(\cdot)$  byla lineární, bylo možno tento výraz přepsat do podoby jádrové hustoty  $t$ -rozdělení.

## 5.5 Metropolis-Hastings algoritmus

Neexistence analytických posteriorních výsledků vyžaduje využití posteriorního simulátoru. Pro některé tvary  $f(\cdot)$  je možno odvodit Gibbsův vzorkovač. V určitých případech, kdy se sama nabízí vhodná aproximace  $p(\gamma|y)$  je možno využít importance sampling. Zde si zavedeme třetí možnost – Matropolis-Hastings algoritmus. V rámci něho je možně představit si celou třídu algoritmů, jež lze využít k vytvoření posteriorních simulátorů pro širokou škálu modelů. Opět využijeme obecné značení  $\theta$  jakožto vektor parametrů,  $p(y|\theta)$  pak bude věrohodnostní funkce a samozřejmě  $p(\theta)$  a  $p(\theta|y)$  budou postupně apriorní a posteriorní hustota.

Metropolis-Hastings algoritmus má mnoho společného s importance sampling. Obě techniky se využívají v situacích, kdy je obtížné z posteriorní hustoty přímo generovat náhodné vzorky. V rámci importance sampling se nabízí možnost generovat náhodné výběry z importance function, v rámci Metropolis-Hastings algoritmu je analogická funkce nazývána *candidate generating density*, tedy kandidátská hustota. Necht  $\theta^*$  označuje výběr právě z této hustoty pravděpodobnosti, který označíme  $q(\theta^{(s-1)}; \theta)$ . Toto označení je interpretováno tak, že kandidátský výběr  $\theta^*$  je realizací náhodné veličiny  $\theta$  jejíž hustota závisí na  $\theta^{(s-1)}$ . Jinými slovy, stejně jako u Gibbsova vzorkovače současný výběr závisí na předchozím výběru. M-H algoritmus je tedy také Markov Chain Monte Carlo (MCMC) algoritmem a řada výběrů  $\theta^{(s)}$  pro  $s = 1 \dots, S$  je označována jako řetězec (chain).

V rámci importance sampling jsme rozdílnoť importance function a posteriorní hustoty korigovala vážením jednotlivých vzorků. V případě M-H algoritmu mají všechny výběry stejnou váhu, nicméně ne všechny kandidátské výběry (kandidáti) jsou akceptovány. Jinými slovy, pokud  $g(\cdot)$  je funkce, která nás zajímá, je možno získat odhad  $E[g(\theta)|y]$ , označovaný  $\hat{g}_S$ , jednoduchým průměrováním našich výběrů:

$$\hat{g}_S = \frac{1}{S} \sum_{s=1}^S g(\theta^{(s)}) \quad (5.7)$$

Metropolis-Hastings algoritmus tak má vždy následující podobu:

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

Tímto postupem získáme odhad  $E[g(\theta)|y]$  pro jakoukoliv funkci, která nás zajímá.

Stejně jako u Gibbsova vzorkovače, M-H algoritmus obvykle vyžaduje znalost počáteční hodnoty  $\theta^{(0)}$ . Pro omezení efektu této počáteční hodnoty je rozumné vyhodit prvních  $S_0$  vzorků. Je zde rovněž vhodné využít MCMC diagnostik (z nichž některé byly prezentovány v kapitole 4) pro ověření toho, zda-li bylo generováno dostatečné množství výběrů a zda-li byl vyhozen dostatečný počet prvních vzorků.

Před tím, než si uvedeme přesný vztah pro akceptační pravděpodobnost  $\alpha(\theta^{(s-1)}, \theta^*)$ , bude užitečné prodiskutovat vlastnosti, které by dobrá akceptační funkce měla mít.

V předchozí kapitole byla zmínka o tom, že MCMC algoritmus lze intuitivně chápat jako proces procházející posteriorní hustotou, kdy vybírá vzorky nejčastěji z oblastí vysoké posteriorní pravděpodobnosti a proporcionálně méně vzorků z oblastí nízké posteriorní hustoty pravděpodobnosti. Kandidátská hustota pravděpodobnosti není identická s posteriorní hustotou a z ní brané výběry by tedy nepokryly adekvátně oblast parametrického prostoru. Matropolis-Hastings algoritmus řeší tento nesoulad tím, že neakceptuje každý kandidátský výběr. Odvozuje se tak akceptační pravděpodobnost, která je nejvyšší v oblastech, kde je i posteriorní hustota vysoká a nejnižší v oblastech nízké posteriorní pravděpodobnosti. Intuitivně tedy v případě, kdy  $\theta^{(s-1)}$  je z oblasti nízké posteriorní pravděpodobnosti, bude algoritmus směřovat pryč od  $\theta^{(s-1)}$ . To znamená, že současná pozice řetězce je v oblasti nízké pravděpodobnosti, a je tedy pravděpodobné, že kandidátský výběr, který nás posouvá ze současné pozice, bude s velkou pravděpodobností akceptován. Pokud naopak  $\theta^{(s-1)}$  je v oblasti vysoké posteriorní pravděpodobnosti, bude algoritmus směřovat k tomu, že v této pozici zůstane (v rámci Kroku 3 je možno nastavit  $\theta^{(s)} = \theta^{(s-1)}$ ). Pokud algoritmus zůstává na místě, dává se tomuto bodu vysoké posteriorní hustoty implicitně vyšší váha podobně jako tomu je v případě vah u importance sampling. Podobné závěry lze učinit i u kandidátského výběru  $\theta^*$ . Pro daný stav  $\theta^{(s-1)}$  chceme kandidáta  $\theta^*$ , který bude akceptován s vysokou pravděpodobností, pokud je v oblasti vyšší posteriorní hustoty než  $\theta^{(s-1)}$ . Kandidáty  $\theta^*$  v oblastech nižší pravděpodobnosti pak budeme chtít s vysokou pravděpodobností zamítnout.

Předchozí odstavec poskytl intuici pro akceptační pravděpodobnost, která závisí na  $\theta^*$  a  $\theta^{(s-1)}$  způsobem, který má tendenci posunout řetězec z oblasti nízké posteriorní pravděpodobnosti do oblasti s pravděpodobností vyšší. Samozřejmě je zde nutné zdůraznit spojení "má tendenci". Není žádoucí, aby řetězec neustále zůstával v oblastech vysoké posteriorní pravděpodobnosti. Naopak chceme, aby byly prozkoumány i oblasti nízké pravděpodobnosti (i když samozřejmě ne tak často). Způsob konstrukce akceptační pravděpodobnosti by tedy měl zajistit, aby se řetězec obvykle (ne vždy) posouval z oblasti nízké pravděpodobnosti do oblasti pravděpodobnosti vyšší.

Výborným úvodem do M-H algoritmu je Chib a Greenberg [6], který zahrnuje i odvození akceptační pravděpodobnosti, která zajišťuje konvergenci Metropolis-Hastings algoritmu k posteriorní hustotě. Akceptační pravděpodobnost má po-

dobu:

$$\alpha(\theta^{(s-1)}, \theta^*) = \min \left[ \frac{p(\theta = \theta^* | y) q(\theta^*; \theta = \theta^{(s-1)})}{p(\theta = \theta^{(s-1)} | y) q(\theta^{(s-1)}; \theta = \theta^*)}, 1 \right] \quad (5.8)$$

Poznamenejme, že  $p(\theta = \theta^* | y)$  je označení posteriorní hustoty vyhodnocené v bodě  $\theta = \theta^*$ , přičemž  $q(\theta^*; \theta)$  je hustota pro náhodnou veličinu  $\theta$ , a tudíž  $q(\theta^*; \theta = \theta^{(s-1)})$  je tato hustota vyhodnocená v bodě  $\theta = \theta^{(s-1)}$ . Lze ověřit, že akceptační pravděpodobnost má žádoucí vlastnosti diskutované výše. Operátor minima je uplatněn proto, aby akceptační pravděpodobnost nemohla být větší než jedna.

Stejně jako importance sampling i M-H algoritmus je na první pohled skvělé řešení pro jakýkoliv problém posteriorní simulace. Zdá se, že je možné generovat náhodné vzorky v podstatě z jakéhokoliv obvyklého rozdělení,  $q(\theta^{(s-1)}; \theta)$  a přijmout nebo zamítnout kandidátské výběry za použití (5.8) k získání sekvence výběrů  $\theta^{(s)}$  pro  $s = 1, \dots, S$ , který lze použít pro odhad  $E[g(\theta) | y]$ . V praxi to však takto jednoduché není. Pokud kandidátská hustota není dobře vybrána, může dojít k tomu, že fakticky všichni kandidáti budou zamítáni a řetězec bude setrvávat po dlouhou dobu na jednom místě. Výběru kandidátské hustoty je tedy třeba věnovat pečlivou pozornost a vždy je třeba používat MCMC diagnostiky k verifikaci konvergence algoritmu. Existuje celá řada možných strategií výběru kandidátských hustot. V následujících částech si popíšeme dvě nejobvyklejší.

### 5.5.1 Independence Chain M-H algoritmus

Jak již název napovídá *Independence Chain Metropolis-Hastings algoritmus* využívá kandidátských hustot, které jsou nezávislé na výběrech. Tedy funkce  $q(\theta^{(s-1)}; \theta) = q^*(\theta)$  a kandidátská hustota tak nezávisí na  $\theta^{(s-1)}$ . Tento přístup je užitečný v případech, kdy existuje vhodná aproximace posteriorní hustoty. Tuto vhodnou aproximaci můžeme využít jako kandidátskou hustotu. Pokud tak učiníme, zjednodušuje se akceptační pravděpodobnost na

$$\alpha(\theta^{(s-1)}, \theta^*) = \min \left[ \frac{p(\theta = \theta^* | y) q^*(\theta = \theta^{(s-1)})}{p(\theta = \theta^{(s-1)} | y) q^*(\theta = \theta^*)}, 1 \right] \quad (5.9)$$

Independence Chain M-H algoritmus je úzce svázán s importance sampling. To lze ukázat, pokud budeme definovat váhy importance sampling:

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

akceptační pravděpodobnost v (5.9) lze tak zapsat jako

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

Akceptační pravděpodobnost je tak jednoduše poměr vah importance sampling vyhodnocených v posledním a kandidátském výběru.

Pokud jde o nelineární regresní model, použitelnost tohoto algoritmu závisí na tom, zda-li  $f(\cdot)$  má tvar takový, že lze najít vhodnou kandidátskou hustotu. Neexistuje zcela obecný postup na výběr aproximativní hustoty. Pokud však nějakou zvolíme, je třeba užít MCMC diagnostiky pro verifikaci konvergence výsledného algoritmu.

Jedním z postupů je využití výsledků klasické metody maximální věrohodnosti pro nalezení vhodné  $q^*(\theta)$ . Čistě bayesovsky zaměřený čtenář může tento odstavec přeskóčit a věnovat se až praktickým doporučením. Nicméně, klasická ekonometrie vychází z toho, že estimátor metody maximální věrohodnosti  $\hat{\theta}_{ML}$  je za určitých podmínek asymptoticky normální s asymptotickou kovarianční maticí danou jako

$$\text{var}(\hat{\theta}_{ML}) = I(\theta)^{-1}$$

kde  $I(\theta)$  je informační matice definovaná jako záporná očekávaná (střední) hodnota druhé derivace logaritmované věrohodnostní funkce (přičemž operátor očekávání resp. střední hodnoty je brán vzhledem k  $y$ ):

$$I(\theta) = -E \left[ \frac{\partial^2 \ln(p(y|\theta))}{\partial \theta \partial \theta'} \right]$$

Slovně řečeno, pokud je velikost vzorku dostatečně velká, inverze informační matice nám dává dobrou představu o podobě  $p(y|\theta)$ . Dokonce i kdyby nebylo možno informační matici přímým způsobem spočítat,  $\frac{\partial^2 \ln(p(y|\theta))}{\partial \theta \partial \theta'}$  půjde vyčísřit (ručně anebo využitím numerických procedur pro derivaci, které jsou dostupné např. v Matlabu) a lze ji tak využít pro získání aproximativního vyjádření  $\text{var}(\hat{\theta}_{ML})$ , což označíme jako  $\widehat{\text{var}}(\hat{\theta}_{ML})$ . Výraz  $\frac{\partial^2 \ln(p(y|\theta))}{\partial \theta \partial \theta'}$  je označován jako *Hessián*, a lze tak mluvit o odhadu  $\widehat{\text{var}}(\hat{\theta}_{ML})$  jako o "záporném inverzním Hessiánu".

Výsledky předchozího odstavce implikují pro naše potřeby, že pokud je velikost vzorku dostatečně velká a apriorní informace relativně neinformativní, potom posteriorní hustota může být aproximativně z normálního rozdělení se střední hodnotou  $\hat{\theta}_{ML}$  a kovariační maticí aproximativně rovnou  $\widehat{\text{var}}(\hat{\theta}_{ML})$ . Pro některé druhy modelů existují počítačové programy, které počítají přímo takové veličiny maximální věrohodnosti. Alternativně lze využít např. prostředí Matlabu, který dokáže nalézt optima uživatelem specifikované funkce. To je možné využít pro počítání charakteristik odhadů metodou maximální věrohodnosti. Pokud jsme schopni naprogramovat proceduru, maximalizující věrohodnostní funkci a hledající  $\widehat{\text{var}}(\hat{\theta}_{ML})$ , je mnohdy žádoucí na místo toho maximalizovat posteriorní hustotu, tedy nalézt  $\hat{\theta}_{\max}$  a vzít druhé derivace posteriorní hustoty k nalezení aproximace k  $\widehat{\text{var}}(\hat{\theta}_{\max})$ . Tento postup je vhodný při použití informativní apriorní hustoty, neboť je zde posteriorní hustota aproximována mnohem lépe než jen za využití věrohodnostních odhadů. Asymptotické výsledky nasvědčují tomu, že posteriorní hustota bude aproximativně z  $f_N(\theta|\hat{\theta}_{\max}, \widehat{\text{var}}(\hat{\theta}_{\max}))$ . V následující části budeme pracovat s výsledky odhadů

metodou maximální věrohodnosti, které však lze nahradit odhady z maximalizace posteriorní hustoty,  $\widehat{\theta}_{\max}$  a  $\widehat{\text{var}}(\widehat{\theta}_{\max})$ , pokud je máme k dispozici.

Volba  $q^*(\theta) = f_N(\theta|\widehat{\theta}_{ML}, \widehat{\text{var}}(\widehat{\theta}_{ML}))$  může být v řadě případů velmi úspěšná. Obvyklejší je však volba  $t$ -rozdělení jako kandidátské hustoty, tedy  $q^*(\theta) = f_t(\theta|\widehat{\theta}_{ML}, \widehat{\text{var}}(\widehat{\theta}_{ML}), \nu)$ . Důvod je ten, že v praxi je důležité, aby kandidátská hustota měla své konce přinejmenším stejně tak tlusté jako posteriorní hustota. Geweke [13] zdůvodňoval tuto volbu v kontextu importance sampling, nicméně stejné zdůvodnění platí i pro Independence Chain M-H algoritmus.

Normální hustota má velmi úzké konce. Oproti tomu  $t$ -hustota má konce mnohem tlustší, zejména pro malé hodnoty  $\nu$ . Užitečnými vlastnostmi  $t$ -rozdělení je to, že pro  $\nu \rightarrow \infty$  se toto rozdělení blíží normálnímu a pro malá  $\nu$  jsou jeho konce velmi tlusté. Ve skutečnosti,  $t$ -rozdělení pro  $\nu = 1$  je Cauchyho rozdělení, které má natolik tlusté konce, že jeho střední hodnota je nekonečno (i když jeho medián a modus jsou konečné hodnoty). V některých případech je možné vyšetřením posteriorní hustoty najít takové hodnoty  $\nu$ , které zajistí, že konce kandidátské hustoty převýší konce hustoty posteriorní. Z opatrnosti se však volí malé hodnoty  $\nu$  a MCMC diagnostiky se využijí k verifikaci konvergenčních vlastností algoritmu.

Je důležité zdůraznit, že existují případy, kdy je použití  $t$ -rozdělení neadekvátní pro generování kandidátů. Například, pokud je posteriorní hustota multimodální, potom unimodální  $t$ -hustota nebude obvykle dobře pracovat. Rovněž pokud je posteriorní hustota definována v omezené oblasti (např. Gamma rozdělení je definováno pouze v oblasti kladných reálných čísel), potom  $t$ -rozdělení (definované na celé množině reálných čísel) nemusí pracovat nejlépe, pokud posteriorní hustota není ostře vymezena uvnitř této oblasti.

V případě nelineárního regresního modelu vyžaduje maximalizace věrohodnostní funkce (posteriorní hustoty) napsání programu pro vyhodnocení (5.3) nebo (5.5). Obychom získali odhad  $\widehat{\text{var}}(\widehat{\theta}_{ML})$ , je třeba provést druhou derivaci (5.3) nebo využít podprogramu pro numerickou derivaci, který je dostupný v rámci každého relevantního počítačového programového prostředí. Tento postup bude záviset na přesné specifikaci  $f(\cdot)$ .

Z předchozí diskuze je patrné, že nalezení aproximativní hustoty pro Independent Chain M-H algoritmus popřípadě importance sampling může být umění. Pro většinu modelů však existují asymptotické výsledky, které nám říkají, že tím, jak se velikost datového vzorku blíží nekonečnu, blíží se posteriorní hustota normálnímu rozdělení. Pro modely této třídy (máme-li dostatečný počet pozorování) je vhodnou aproximací funkce hustoty pravděpodobnosti  $f_t(\theta|\widehat{\theta}_{ML}, \widehat{\text{var}}(\widehat{\theta}_{ML}), \nu)$ .

### 5.5.2 Random Walk Chain M-H algoritmus

*Random Walk Chain Metropolis-Hastings algoritmus* je užitečný v případě, kdy nejsme schopni najít dobrou aproximaci pro posteriorní hustotu. V rámci Independence Chain M-H algoritmu (podobně jako u importance sampling), bereme

výběry z hustoty, která je podobná posteriorní hustotě a akceptační pravděpodobnost (nebo vážení v případě importance sampling) se využije pro korekci rozdílu mezi posteriorní a aproximační hustotou. V rámci Random Walk Chain M-H algoritmu nezkoušíme aproximovat posteriorní hustotu, místo toho vybíráme kandidátskou hustotu, která má široký záběr a bere proporcionálně náhodné výběry v různých oblastech posteriorní hustoty.

Formálně řečeno, tento M-H algoritmus generuje kandidáty podle následujícího schématu:

$$\theta^* = \theta^{(s-1)} + z \quad (5.10)$$

kde  $z$  se nazývá *increment random variable*, tedy přírůstková náhodná veličina. Předpoklad v (5.10) implikuje, že kandidáti jsou generováni jako náhodná procházka, tj. kandidáti jsou vybíráni v náhodném směru ze současného bodu. Akceptační pravděpodobnost zajistí, že řetězec se posune žádoucím směrem. Poznamenejme, že  $\theta^*$  a  $\theta^{(s-1)}$  vstupují do vztahu (5.10) symetricky a vždy bude platit  $q(\theta^*; \theta = \theta^{(s-1)}) = q(\theta^{(s-1)}; \theta = \theta^*)$ . To znamená, že akceptační pravděpodobnost lze zapsat jako

$$\alpha(\theta^{(s-1)}, \theta^*) = \min \left[ \frac{p(\theta = \theta^* | y)}{p(\theta = \theta^{(s-1)} | y)}, 1 \right] \quad (5.11)$$

a je zřejmé, že random walk chain má tendenci posunovat se do oblastí vyšší posteriorní hustoty.

Výběr hustoty pro  $z$  determinuje přesnou formu kandidátské hustoty. Obvyklým a vhodným výběrem je vícerozměrné normální rozdělení. V tomto případě (5.10) určuje střední hodnotu tohoto normálního rozdělení, což je  $\theta^{(s-1)}$  a nám nezbyvá nic jiného než zvolit kovarianční matici, kterou označíme jako  $\Sigma$ :

$$q(\theta^{(s-1)}; \theta) = f_N(\theta | \theta^{(s-1)}, \Sigma) \quad (5.12)$$

V rámci tohoto přístupu je nutné zvolit  $\Sigma$ , a to tak, aby akceptační pravděpodobnost nebyla ani moc vysoká, ani moc nízká. Pokud je akceptační pravděpodobnost obvykle velmi nízká, potom kandidátské výběry jsou téměř vždy zamítány a řetězec se jen zřídka pohne. To není moc dobrá situace, neboť implikuje, že  $S$  musí být extrémně vysoké, aby řetězec prošel celou posteriorní hustotou. Malá akceptační pravděpodobnost indikuje, že  $\Sigma$  je příliš "velká" a většina kandidátů pochází z oblastí konců posteriorního rozdělení. Druhým extrémem je situace, kdy akceptační pravděpodobnost je blízko jedné ( $\Sigma$  je příliš "malá"). V tomto případě leží  $\theta^*$  a  $\theta^{(s-1)}$  velmi blízko u sebe a akceptační pravděpodobnost bude blízko jedné. I v tomto případě by  $S$  muselo být velmi obrovské číslo, aby řetězec prozkoumal celé posteriorní rozdělení.

Neexistuje obecné pravidlo, které nám říká, jaká by měla být optimální míra akceptačního poměru (tedy podíl počtu akceptovaných ze všech výběrů). Ve speciálním případě, kdy posteriorní i kandidátská hustota jsou obě normální byl vypočítán optimální akceptační podíl 0.45 pro případ jednodimenzionálního problému a o něco menší hodnota pro vícedimenzionální problém. Pokud se počet dimenzí blíží nekonečnu, optimální akceptační podíl se přibližuje hodnotě

0.23. Jiným hrubým a často zmiňovaným pravidlem je akceptační poměr na hodnotě 0.5. Obecně, pokud si zvolíme  $\Sigma$ , které zajistí, že akceptační podíl bude v tomto rozmezí, je nepravděpodobné, že bychom se touto volbou dopouštěli závažné chyby. Vždy je však nutné verifikovat konvergenci algoritmu.

Tvrzení, že by  $\Sigma$  mělo být vybráno tak, aby akceptační podíl spadl do intervalu od 0.2 do 0.5 je dostatečná informace pro její výběr v případě, kdy  $\theta$  a tedy i  $\Sigma$  je skalár. V tomto případě lze experimentovat s různými hodnotami  $\Sigma$  a opakovat Random Walk Chain M-H algoritmus dokud nebudeme dostávat rozumné akceptační podíly. V případě, kdy  $\theta$  je  $p$ -dimenzionální je tento přístup poněkud obtížnější, neboť  $\Sigma$  bude mít  $\frac{p(p+1)}{2}$  prvků. V tomto případě je vhodné nastavit  $\Sigma = c\Omega$ , kde  $c$  je skalára  $\Omega$  je odhad posteriorní kovarianční matice vektoru  $\theta$ . Pak lze experimentovat s různými hodnotami  $c$ , abychom získali rozumné akceptační podíly. Tento přístup vyžaduje najít  $\Omega$  a odhadnout  $\text{var}(\theta|y)$ . To lze učinit dvěma způsoby. Z pohledu výzkumníka je nejjednodušší začít se  $\Sigma = cI_p$  a zkoušet najít hodnoty  $c$ , které nebudou implikovat nepužitelné hodnoty akceptačních pravděpodobností (to je případ, kdy kandidáti jsou akceptováni s pravděpodobností 0.000001 nebo 0.99999, pokud bychom samozřejmě neměli k dispozici výkonné počítače, které by byly schopny generovat biliony vzorků). Tuto hodnotu  $c$  je pak možné použít pro velmi hrubý odhad  $\Omega$ . Pak je možné nastavit  $\Sigma = c\Omega$  a zkoušet hledat nové hodnoty  $c$ , které přinášejí ještě rozumnější akceptační pravděpodobnosti. Výsledky po tomto případě je pak zase možné využít pro získání lepší matice  $\Omega$ , kterou dále využijeme pronalezení lepších hodnot pro  $\Sigma$  atd. Tento postup lze opakovat až do nalezení vhodné matice  $\Sigma$ . Jedná se o jednoduchý postup z toho hlediska, že pokud si vytvoříme základní program pro Random Walk Metropolis-Hastings algoritmus, není už nutné žádné další programování. Ovšem jsou zde určité nároky na výpočetní dobu.

Alternativně lze  $\Omega$  nastavit na hodnotu  $\text{var}(\widehat{\theta}_{ML})$ , teda jako odhad variance odhadu metodou maximální věrohodnosti. Tento odhad je však nutno zjistit, což vyžaduje další programování.

### 5.5.3 Metropolis-within-Gibbs

Metropolis-Hastings algoritmus poskytuje posteriorní simulátor pro  $p(\theta|y)$ . V předchozí kapitole jsme zavedli Gibbsův vzorkovač, který pro případ dvou bloků, kde  $\theta = (\theta'_{(1)}, \theta'_{(2)})'$ , vyžadoval sekvenční výběry z  $p(\theta_{(1)}|y, \theta_{(2)})$  a  $p(\theta_{(2)}|y, \theta_{(1)})$ . Pro NLRM s nezávislou normální-gama apriorní hustotou bylo snadné Gibbsův vzorkovač implementovat, protože  $p(\beta|y, h)$  odpovídalo normálnímu rozdělení a  $p(h|y, \beta)$  odpovídalo rozdělení gama. V nelineárním regresním modelu, neinformativní apriorní hustota nebo nezávislá gama apriorní hustota pro  $h$  bude implikovat, že  $p(h|y, \gamma)$  odpovídá gama rozdělení. Naopak  $p(\gamma|y, h)$  bude proporcionální k (5.5) a tedy nebude mít podobu obvyklé hustoty, ze které by bylo možno generovat náhodné výběry. Zdá se tedy, že zde není možno využít Gibbsův vzorkovač zahrnující  $p(h|y, \gamma)$  a  $p(\gamma|y, h)$ . Lze však ukázat, že pokud využijeme Metropolis-Hastings algoritmus pro  $p(\gamma|y, h)$ , výsledné nasimulované vzorky  $\gamma^{(s)}$  a  $h^{(s)}$  pro  $s = 1, \dots, S$  jsou platnými posteriorními výběry. Formálně



řečeno, využití Metropolis-Hastings algoritmu pro jednu či obě podmíněné posteriorní hustoty využitě v Gibbsově vzorkovači,  $p(\theta_{(1)}|y, \theta_{(2)})$  a  $p(\theta_{(2)}|y, \theta_{(1)})$ , je zcela akceptovatelné a užitečné. Toto je možné využít i když Gibbsův vzorkovač zahrnuje více než dva bloky. Tento *Metropolis-within-Gibbs* algoritmus je obvyklou technikou, neboť řada modelů má jednak posteriorní hustoty, ze kterých je snadné generovat náhodné vzorky, ovšem jedna či dvě podmíněné hustoty standardní formu nemají, a právě pro ně je možné využít M-H algoritmus.

## 5.6 Měřítko kvality modelu: Posteriorní predikční p-hodnota

Typickou bayesiánskou metodou pro porovnání modelu je posteriorní podíl šancí, což je relativní pravděpodobnost dvou plně specifikovaných modelů. Mohou však nastat situace, kdy je žádoucí prověřit kvalitu modelu v jakémsi absolutním slova smyslu a nikoliv tedy relativně vzhledem k alternativnímu modelu. Rovněž v řadě případů může být žádoucí využít nepravý, neinformativní prior, kdy ovšem posteriorní podíl šancí nemusí být spočítatelný respektive může být nesmyslný, pokud je tato apriorní hustota použita pro parametry, které nejsou společné oběma modelům. V těchto situacích je rozumnou alternativou posteriornímu podílu šancí posteriorní predikční p-hodnota.

V dalším výkladu budeme rozlišovat mezi aktuálně pozorovanými daty  $y$  a pozorovatelnými daty  $y^\circ$ , které by mohly být generovány modelem (tj.  $y^\circ$  je náhodný vektor rozměru  $N \times 1$  s funkcí hustoty pravděpodobnosti  $p(y^\circ|\theta)$ , což je věrohodnostní funkce bez zahrnutí  $y$ ). Nechť  $g(\cdot)$  je funkce, která nás zajímá. Potom  $p(g(y^\circ)|y)$  je souhrnem veškeré informace, kterou nám dává model o  $g(y^\circ)$  po shlédnutí dat. Jinými slovy nám ukazuje, jaký druh dat může náš model generovat. Pro pozorovaná data můžeme přímo spočítat  $g(y)$ . 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. Formálně jsme schopni získat pravděpodobnosti okrajových oblastí způsobem podobným p-hodnotě z klasické statistiky. Přesněji řečeno, posteriorní predikční p-hodnota je pravděpodobnost modelu generovat datovou sadu, která bude mít extrémnější vlastnosti než ta, kterou skutečně pozorujeme (analogicky s p-hodnotou klasické statistiky či ekonometrie je možné prezentovat buď jednostrannou nebo oboustrannou p-hodnotu).

Hustotu pravděpodobnosti  $p(g(y^\circ)|y)$  je možno spočítat použitím simulačních metod obdobných pro predikční analýzu. Lze tedy psát:

$$p(g(y^\circ)|y) = \int p(g(y^\circ)|\theta, y)p(\theta|y)d\theta = \int p(g(y^\circ)|\theta)p(\theta|y)d\theta \quad (5.13)$$

poslední rovnice vyplývá ze skutečnosti, že díky podmíněnosti vektorem parametrů  $\theta$ , aktuální data nepřinášejí žádnou dodatečnou informaci o  $y^\circ$ . 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  $\theta$ .

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. Pro porovnání modelů se však (je-li to možné) využívá posteriorní podíl šancí.

Přístupy využívající posteriorní predikční p-hodnoty vyžadují výběr funkce, která nás zajímá,  $g(\cdot)$ . Přesný výběr  $g(\cdot)$  bude záviset na empirické aplikaci. Pokud se vrátíme k nelineárnímu regresnímu modelu z této kapitoly, potom máme

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

pro  $i = 1, \dots, N$ . Vzhledem k předpokladům kladeným na náhodnou složku platí:

$$p(y^\circ | \gamma, h) = f_N(y^\circ | f(X, \gamma), h^{-1} I_N) \quad (5.14)$$

kde  $f(X, \gamma)$  je  $N$ -rozměrný vektor definovaný v (5.2). Poznamenejme, že pro daný vektor parametrů je simulace hodnot  $y^\circ$  vcelku snadné, neboť postačuje generovat náhodné výběry z vícerozměrného normálního rozdělení. Tato jednoduchost je obvyklá pro řadu modelů, což usnadňuje výpočet posteriorní predikční p-hodnoty.

Pro nelineární regresní model s neinformativní apriorní hustotou uvedenou v (5.4) je možno předchozí výraz dále zjednodušit, neboť  $h$  lze vyintegrovat. Přesněji řečeno, odvozením obdobným přechodu z (5.5) do (5.6) lze ukázat, že

$$p(y^\circ | \gamma) = f_t(y^\circ | f(X, \gamma), \bar{s}^2 I_N, N) \quad (5.15)$$

kde

$$\bar{s}^2 = \frac{[y - f(X, \gamma)]' [y - f(X, \gamma)]}{N} \quad (5.16)$$

Výběry  $y^\circ$  lze tak získat využitím vícerozměrného  $t$ -rozdělení (podmíněného vektorem  $\gamma$ ). Tyto výběry lze interpretovat jako reprezentativní typů dat, které je schopen generovat náš model. Přístup využívající posteriorní predikční p-hodnotu využívá myšlenky, že pokud je model rozumně formulován, měla by skutečně pozorovaná data spadat do kategorie dat, která dokáže náš model běžně generovat. Nalezení příslušného percentilu, který vytváří  $g(y)$  v rámci hustoty  $p(g(y^\circ) | y)$  je záležitostí využití formální metriky.

Pro konkrétnější představu si nastíníme možnosti pro volbu  $g(\cdot)$ . Kvalita modelu se často posuzuje skrze tzv. analýzu reziduí. V rámci klasické ekonometrie se počítají odhady  $\epsilon_i$ , které se nazývají rezidua. Poté lze zkoumat vlastnosti těchto reziduí pro ověření toho, jestli předpoklady kladené na model a techniky pro odhad jeho parametrů jsou splněny. V bayesovském kontextu jsou chyby  $\epsilon_i$  pro  $i = 1, \dots, N$  dány jako

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

O těchto chybových členech jsme vyslovili řadu předpokladů, Předpokládali jsme, že mají normální rozdělení, a to i.i.d.  $N(0, h^{-1})$ . Mnohdy je možné, že tyto předpoklady nemusí být v souladu s pozorováním, a tudíž je žádoucí otestovat je. Tento předpoklad v sobě zahrnuje skutečnost, že chybové členy jsou vzájemně nezávislé, mají stejný rozptyl apod. My se zatím zaměříme pouze na předpoklad jejich normality. Typickou vlastností normálního rozdělení je, že má nulovou šikmost a specifickou špičatost. Šikmost a špičatost je měřena pomocí třetích a čtvrtých momentů. Pro standardizované normální rozdělení je třetí moment nulový a čtvrtý moment má hodnotu tři. Předpoklad normality tedy implikuje, že následující obvykle užívané míry šikmosti a převisu špičatosti by měly být obě nulové:

$$Skew = \frac{\sqrt{N} \sum_{i=1}^N \epsilon_i^3}{\left[ \sum_{i=1}^N \epsilon_i^2 \right]^{\frac{3}{2}}} \quad (5.17)$$

$$Kurt = \frac{N \sum_{i=1}^N \epsilon_i^4}{\left[ \sum_{i=1}^N \epsilon_i^2 \right]^2} - 3 \quad (5.18)$$

Tyto míry šikmosti a špičatosti nelze vypočítat přímým způsobem, neboť  $\epsilon_i$  je nepozorovatelné. Klasická ekonometrie nahradí  $\epsilon_i$  příslušnými OLS rezidui a provede test šikmosti a špičatosti.

Bayesiánská analogie klasického přístupu vyžaduje výpočet očekávané hodnoty (5.17) nebo (5.18) a ověření, zda-li je tato hodnota přijatelná. Formálně tedy,

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

je něco, co dokážeme spočítat, pokud máme posteriorní simulátor. *Skew* je tedy funkce parametrů modelu a dat. Její posteriorní střední hodnota tak může být spočítána jako jakákoliv posteriorní střední hodnota funkce, která by nás zajímala. Rovněž  $E[Kurt|y]$  můžeme spočítat analogickým způsobem. Pokud platí předpoklad normality,  $E[Skew|y]$  a  $E[Kurt|y]$  by měly být shruba nulové.

Nyní se vraťme zpět k tématu posteriorní predikční p-hodnoty.  $E[Skew|y]$  a  $E[Kurt|y]$  jsou funkce pozorovaných dat a lze je spočítat využitím posteriorního simulátoru. Pro jakákoliv pozorovatelná data  $y^\circ$ ,  $E[Skew|y]$  a  $E[Kurt|y]$  lze spočítat stejným způsobem. Pokud tyto dvě posledně zmiňované funkce spočítáme pro celou řadu pozorovatelných datových vzorků, můžeme získat rozdělení hodnot šikmosti a převisu špičatosti, které je schopen model generovat. Pokud  $E[Skew|y]$  nebo  $E[Kurt|y]$  leží hluboko na koncích rozdělení  $E[Skew|y^\circ]$  respektive  $E[Kurt|y^\circ]$  jedná se o důkaz svědčící proti předpokladu normality. Je třeba zdůraznit, že  $E[Skew|y]$  a  $E[Kurt|y]$  jsou prostá čísla, oproti tomu  $E[Skew|y^\circ]$  respektive  $E[Kurt|y^\circ]$  jsou obě náhodné veličiny s rozdělením pravděpodobnosti počítaným dle (5.13). V kontextu našeho předchozího značení nastavíme  $g(y) = E[Skew|y]$  nebo  $E[Kurt|y]$  a  $g(y^\circ) = E[Skew|y^\circ]$  nebo  $E[Kurt|y^\circ]$ .

V praxi vyžaduje programový kód pro výpočet posteriorní predikční p-hodnoty pro nelineární regresní model využívající neinformativní apriorní hustotu níže uvedené kroky. Příklad výpočtu převisu špičatosti nebo jiné funkce, která by nás mohla zajímat, je naprosto podobný. Tyto kroky předpokládají, že máme odvozený posteriorní simulátor (např. M-H algoritmus).

- *Krok 0:* Vezmeme výběr  $\gamma^{(s)}$  využitím posteriorního simulátoru.
- *Krok 1:* Vygenerujeme reprezentativní datovou sadu  $y^\circ$  z  $p(y^\circ|\gamma^{(s)})$  užitím (5.15).
- *Krok 2:* Definujeme  $\epsilon_i^{(s)} = y_i - f(X_i, \gamma^{(s)})$  pro  $i = 1, \dots, N$  a vyhodnotíme (5.17) v tomto bodě, abychom získali  $Skew^{(s)}$ .
- *Krok 3:* Definujeme  $\epsilon_i^{\circ(s)} = y_i^{\circ(s)} - f(X_i, \gamma^{(s)})$  pro  $i = 1, \dots, N$  a vyhodnotíme (5.17) v tomto bodě, abychom získali  $Skew^{\circ(s)}$ .
- *Krok 4:* Opakujeme Krok 1, 2, 3 a 4 celkem  $S$  krát.
- *Krok 5:* Spočítáme průměr  $S$  výběrů  $Skew^{(1)}, \dots, Skew^{(S)}$  pro odhad  $E[Skew|y]$ .
- *Krok 6:* Spočítáme podíl  $S$  vzorků  $Skew^{\circ(1)}, \dots, Skew^{\circ(S)}$ , které jsou menší než odhad  $E[Skew|y]$ . Pokud je toto číslo menší než 0.5, potom je to odhad posteriorní predikční p-hodnoty. V opačném případě je tímto odhadem jedna mínus toto číslo.

Neexistuje nějaké pevné pravidlo, které nám řekne kolik přesně by tato hodnota měla být, aby svědčila v neprospěch našeho modelu. Užitečným hrubým pravidlem je vzít posteriorní predikční p-hodnotu menší než 0.05 (nebo 0.01) jako důkaz svědčící proti modelu. Pokud tedy např. posteriorní predikční p-hodnota pro šikmost je rovna 0.05, potom můžeme říct, že model generuje měřítko šikmosti vyšší než tu, kterou skutečně pozorujeme, pouze v pěti procentech případů. Je tedy nepravděpodobné, že by model generoval pozorovaná data.

## 5.7 Metoda Gelfanda-Deye

V rámci porovnání modelů je v případě nelineární specifikace často žádoucí porovnávat různé funkční tvary  $f(\cdot)$ . Ve většině případů se bude jednat o porovnání vnořených modelů. Alternativně je možno porovnávat lineární a nelineární specifikaci. Například pro CES produkční funkci zavedenou v úvodu této kapitoly se model stává lineárním pro  $\gamma_{k+1} = 1$  a lineární model je tak vnořeným modelem modelu nelineárního, tedy  $M_1 : \gamma_{k+1} = 1$  a  $M_2 : \gamma$  neomezeno. Při porovnávání vnořených odelů tohoto typu je obvyklým nástrojem Savage-Dickey density ratio. Pro nelineární regresní model, jestliže apriorní hustota pro  $\gamma$  má jednoduchou podobu, lze Savage-Dickeyeho poměr snadno spočítat (tzn. pokud posteriorní hustota pro  $\gamma$ , která musí být vyhodnocena v bodě  $\gamma_{k+1}$ , může mít

podobu, se kterou lze jednoduše pracovat<sup>12</sup>). Například, pokud apriorní hustota má podobu

$$p(\gamma, h) \propto \frac{p(\gamma)}{h}$$

potom posteriorní hustota pro  $\gamma$  bude mít podobu danou v (5.6) násobenou  $p(\gamma)$ . Pokud je však apriorní hustota komplikovanější, může být nemožné vyhodnotit Savage-Dickey denstiy ratio.

Pro porovnání nevnořených modelů, či vnořených modelů, kde nelze snadno využít Savage-Dickey density ratio, je třeba obecnější metody vyhodnocení posteriorního podílu šancí. Nejen v těchto případech lze využít metodu Gelfanda a Deye [12].

Metoda Gelfanda a Deye je založena na skutečnosti, že inverzi marginální věrohodnosti pro model  $M_i$ , který závisí na vektoru parametrů  $\theta$ , lze zapsat jako  $E[g(\theta)|y, M_i]$  pro specifickou volbu  $g(\cdot)$ . Právě posteriorní simulátory jako např. M-H algoritmus jsou vytvořeny právě pro odhad takovýchto charakteristik. Následující teorém ukazuje nutnou volbu pro  $g(\cdot)$ .

**Teorém 5.1** (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  $\Theta$ . Pokud  $f(\theta)$  je funkce hustoty pravděpodobnosti definovaná na oblasti obsahující  $\Theta$ , potom*

$$E \left[ \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)} \middle| y, M_i \right] = \frac{1}{p(y|M_i)} \quad (5.19)$$

Důkaz

$$\begin{aligned} E \left[ \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)} \middle| y, M_i \right] &= \int \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)} p(\theta|y, M_i) d\theta \\ &= \int \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)} \frac{p(\theta|M_i)p(y|\theta, M_i)}{p(y|M_i)} d\theta \\ &= \frac{1}{p(y|M_i)} \int f(\theta) d\theta \\ &= \frac{1}{p(y|M_i)} \end{aligned}$$

Tento teorém se zdá být na první pohled velmi silný v tom smyslu, že pro jakoukoliv p.d.f  $f(\theta)$  můžeme nastavit:

$$g(\theta) = \frac{f(\theta)}{p(\theta|M_i)p(y|\theta, M_i)} \quad (5.20)$$

a využít výsledky posteriorního simulátoru k odhadu  $E[g(\theta)|y, M_i]$ . Pro úspěšné použití této metody je potřeba obezřetné volby  $f(\theta)$ . Například Geweke [16]

<sup>12</sup>Je třeba poznamenat, že existují algoritmy, které při daných výstupech posteriorního simulátoru  $\gamma^{(s)}$  pro  $s = 1, \dots, S$  dokáží aproximovat  $p(\gamma|y)$ . Tuto aproximaci lze využít při výpočtu čitatele Savage-Dickeyeho poměru. Tyto algoritmy spadají do oblasti neparametrických odhadů hustoty a vyžadují specifické znalosti přesahující zaměření tohoto textu.

poukazuje na to, že  $\frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)}$  musí být shora omezena, tedy musí nabývat konečných hodnot pro jakoukoliv volbu  $\theta$ .

Geweke [16] doporučuje následující strategii pro volbu  $f(\theta)$ , která se v praxi ukázala jako velmi účinná. Tato strategie v sobě zahrnuje definování  $f(\theta)$  jakožto funkce normální hustoty pravděpodobnosti s ohraničenými okraji. Důvod je ten, že je obtížné ověřit, zda-li  $\frac{f(\theta)}{p(\theta|M_i)p(y|\theta,M_i)}$  je konečné na okrajích hustoty normálního rozdělení. Odříznutím těchto konců je  $f(\theta)$  pro tyto potenciálně problematické oblasti nulová. Formálně řečeno, necht'  $\hat{\theta}$  a  $\hat{\Sigma}$  jsou odhady  $E(\theta|y, M_i)$  a  $\text{var}(\theta|y, M_i)$  získané z posteriorní simulace. Dále necht' pro určitou oblast pravděpodobnosti  $p \in (0, 1)$  označuje  $\hat{\Theta}$  oblast definičního oboru funkce  $f(\theta)$ , která je definována jako:

$$\hat{\Theta} = \{\theta : (\hat{\theta} - \theta)' \hat{\Sigma}^{-1} (\hat{\theta} - \theta) \leq \chi_{1-p}^2(k)\} \quad (5.21)$$

kde  $\chi_{1-p}^2(k)$  je  $(1-p)$  procentní kvantil rozdělení chí-kvadrát s  $k$  stupni volnosti, přičemž  $k$  je počet prvků vektoru  $\theta$ . Geweke potom doporučuje volbu  $f(\theta)$  jakožto funkci vícerozměrné normální hustoty pravděpodobnosti omezené do oblasti  $\hat{\Theta}$ :

$$f(\theta) = \frac{1}{p(2\pi)^{\frac{k}{2}}} |\hat{\Sigma}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\hat{\theta} - \theta)' \hat{\Sigma}^{-1} (\hat{\theta} - \theta) \right] 1(\theta \in \hat{\Theta}) \quad (5.22)$$

kde  $1(\cdot)$  je indikační funkce. Očekává se, že nejlepších výsledků dosahujeme při volbě nízkých hodnot  $p$  (např.  $p = 0.01$ ), neboť potom při odhadu marginální věrohodnosti zahrnujeme mnohem více vzorků. Jak ovšem dále zdůrazňuje Geweke, dodatečný náklad zkoušení různých hodnot  $p$  je velmi malý. Standardním způsobem jsme schopni spočítat i numerickou standardní chybu, kterou lze využít při vyhodnocení přesnosti odhadu marginální věrohodnosti metodou Gelfanda-Deye. Implementace této metody je obsažena např. v balíku BACC.

Je třeba zdůraznit, že obecná metoda Gelfanda-Deye pro výpočet marginální věrohodnosti funguje pro jakýkoliv model. V praxi jsou reálnými požadavky pouze to, aby byl k dispozici posteriorní simulátor a aby byly známy  $p(\theta|M_i)$  a  $p(y|\theta, M_i)$ . Právě druhý z požadavků není až tak triviální, neboť v některých případech známe pouze jádrové hustoty apriorní hustoty a (nebo) věrohodnostní funkce, a tedy ne plnou p.d.f. V těchto případech metodu Gelfanda-Deye nelze použít. Gewekova implementace této metody funguje pouze v případě, kdy definiční obor posteriorní hustoty obsahuje oblast definovanou v (5.21), tj.  $\hat{\Theta} \in \Theta$ . Pokud tomu tak není, nabízí Geweke [16] návrhy postupu, jak tuto implementaci mírně pozměnit. Gewekův přístup je tedy o něco méně obecný, nicméně je přesto uplatnitelný pro širokou škálu modelů.

## 5.8 Predikce

Predikční analýza probíhá analogickým způsobem nastíněným v předchozí kapitole. Výběry z  $p(\gamma|y)$ , které nám poskytuje M-H algoritmus, využijeme pro

tvorbu vzorků z podmíněné predikční hustoty  $p(y^*|y, \gamma^{(s)})$ . Jejich průměrováním pak získáme odhad jakékoliv charakteristiky predikční hustoty, která nás může zajímat. Pokud například využijeme neinformativní apriorní hustotu danou v (5.4) a s využitím techniky analogické pro odvození (3.40) lze ověřit, že

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

kde  $\bar{s}^2$  je definovaná v (5.16). Výběry z  $y^*$  podmíněné vektorem  $\gamma$  lze tedy získat velmi snadno.

## 5.9 Empirická ilustrace

K ilustraci bayesiánské analýzy nelineárního regresního modelu využijeme příklad z mikroekonomie. Využijeme data z  $N = 123$  společností týkající se jejich výstupu,  $y$ , a vstupů v podobě práce,  $x_1$ , a kapitálu,  $x_2$ . Data jsou obsahem textového souboru `ch5data.out`. Abychom nemuseli řešit problém jednotek jednotlivých veličin, jsou všechny proměnné standardizované, tedy jejich směrodatná odchylka je jedna. Pozorování každé proměnné tedy byla vydělena svou směrodatnou odchylkou. Standardizace se někdy dělá proto, že výsledná interpretace koeficientů probíhá v intencích právě směrodatných odchylek proměnných. Např. v lineárním regresním modelu lze parametr  $\beta_j$  interpretovat tak, že nám říká: „Pokud se vysvětlující  $j$ -tá proměnná zvýší o jednu směrodatnou odchylku, vysvětlovaná proměnná má tendenci zvýšit se o  $\beta_j$  směrodatných odchylek.“

Budeme pracovat s CES funkcí v podobě (viz (5.1)):

$$y_i = \gamma_1 + (\gamma_2 x_{i1}^{\gamma_4} + \gamma_3 x_{i2}^{\gamma_4})^{\frac{1}{\gamma_4}} + \epsilon_i.$$

Volba aditivně přidané úrovně konstanty má jediný účel, a to ten, že omezení  $\gamma_4 = 1$  povede k lineárnímu regresnímu modelu.

Začneme porovnáním výsledků posteriorní analýzy tohoto modelu, a to na základě použití Independence Chain a Random Walk Chain Metropolis-Hastings algoritmů. Kompletní analýza je obsahem soubor `priklad_CES.m`. V tomto případě využijeme neinformativní apriorní hustotu danou výrazem (5.4). Změříme se zatím jen na marginální posteriorní hustotu pro vektor parametrů  $\gamma = (\gamma_1, \dots, \gamma_4)'$ , danou výrazem (5.6). Ke konstrukci kandidáty generujících hustot pro tyto dva algoritmy, využijeme optimalizační techniku k nalezení  $\widehat{\gamma}_{\max}$ , tedy modu marginální posteriorní hustoty,  $p(\gamma|y)$ . K dispozici je možnost využití optimalizačního algoritmu z ekonometrického toolbox LeSageho [22] nebo funkci pro neomezenou optimalizaci (minimalizaci) `fminunc.m`, která je součástí *Optimisation toolboxu* Matlabu. Obě funkce poskytují Hesián, který využijeme ke konstrukci kovarianční matice odhadu parametrů,  $\widehat{\text{var}}(\widehat{\gamma}_{\max})$  (viz část 5.5.1). Optimalizaci můžeme obejít i způsobem postupného vylepšování střední hodnoty a kovarianční matice, kdy vyjdeme např. z počátečního vektoru jedniček a jednotkové kovarianční matice a provedeme M-H algoritmus pro zkrácený počet iterací s neextrémními průměrnými akceptačními pravděpodobnostmi (viz část 5.5.2).

Funkce `CES_post` je námi vytvořená funkce odpovídající jádrové marginální posteriorní hustotě. Jedná se o záporný logaritmus jádra marginální posteriorní hustoty (algoritmus tak hledá minimum logaritmu této funkce, což odpovídá i nalezení maxima původní nelogaritmované funkce).

Tabulka 5.1 ukazuje výsledky na základě implementace obou algoritmů. Příslušný skript matlabovského souboru odpovídá krokům popsaným v části 5.5. Algoritmy se liší jen ve svých kandidátských hustotách, a tedy ve výrazech pro akceptační pravděpodobnost (viz (5.9) a (5.11)). Akceptační pravděpodobnosti jsou zapsány s ohledem na to, že pracujeme se záporným logaritmem příslušné posteriorní hustoty a logaritmem kandidátské hustoty. Independence Chain algoritmus generuje kandidáty z vícerozměrného  $t$ -rozdělení,  $f_t(\gamma|\widehat{\gamma}_{\max}, \widehat{var}(\widehat{\gamma}_{\max}), 5)$  a Random Walk algoritmu bere kandidáty z vícerozměrného normálního rozdělení,  $f_N(\gamma|\gamma^{s-1}, \widehat{var}(\widehat{\gamma}_{\max}))$ . S kandidátskými hustotami lze experimentovat použitím různých kovariančních maticí v obou kandidátských hustotách. Jak již bylo naznačeno, matici  $\widehat{var}(\widehat{\gamma}_{\max})$  jsme mohli aproximovat kovarianční maticí  $var(\gamma|y)$  získanou na základě výchozího běhu každého z algoritmů. Dále si můžeme vyhrát s použitím  $c \cdot \widehat{var}(\widehat{\gamma}_{\max})$  a experimentováním s různými volbami skaláru  $c$  pro vylepšení výkonnosti algoritmů. Pro Independence Chain Metropolis-Hastings algoritmus si můžeme zkusit měnit i parametr vyjadřující počet stupňů volnosti v kandidáty generující hustotě. V případě obou algoritmů jsme zvolili počet generovaných vzorků  $S = 25000$  a počet vyhozených vzorků  $S_0 = 5000$ .

**Tabulka 5.1:** Posteriorní výsledky na základě dvou M-H algoritmů

	Independence Chain		Random Walk Chain	
	Stř. hodnota	Rozptyl	Stř. hodnota	Rozptyl
$\gamma_1$	1.022	0.003	1.019	0.003
$\gamma_2$	0.706	0.011	0.710	0.012
$\gamma_3$	1.004	0.024	0.983	0.023
$\gamma_4$	1.393	0.081	1.336	0.058

Akceptační poměr u Independence Chain algoritmu byl 45 % u Random Walk algoritmu pak 21 %. V obou případech jsou výsledné odhady velmi podobné. S ohledem na posteriorní směrodatnou odchylku (odmocnina z prezentovaného rozptylu) jsou odhadnuté posteriorní střední hodnoty skoro stejné. Odhadnuté posteriorní rozptyly jsou již trochu odlišné, nicméně ne moc dramaticky. Pokud bychom chtěli ještě přesnější výsledky, mohli bychom ještě vyladit kandidátské hustoty nebo zvýšit počet replikací. Tyto algoritmy jsou výpočetně náročné, i když v našem případě čas výpočtu na čtyřjádrovém procesoru Intel Core 2 Quad (Q9550), 8Gb RAM a 64bitovým operačním systémem zabrala doba výpočtu asi 40 sekund (celkově pro oba algoritmy). S rozvojem počítačů je samozřejmě mnohem snadnější generovat více a více vzorků, pro velký počet

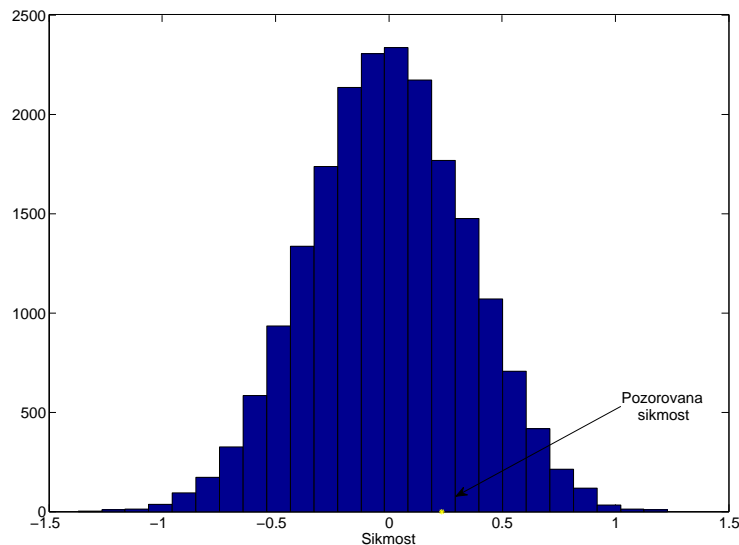


replikací stačí nechat běžet počítač přes noc či přes víkend. To má svůj význam pro empirickou praxi. Pro dosažení požadované úrovně přesnosti máme možnost buď zlepšovat algoritmus posteriorní simulace, nebo pracovat s méně efektivním algoritmem na více replikacích. Práce nad vytvořením lepšího algoritmu zabírá hlavně čas náš, práce s více replikacemi zabere čas počítače. Většina z nás si asi více váží času svého než počítače a tak raději prezentuje výsledky na základě „dostatečně dobrého“ algoritmu, než na základě algoritmu „nejlepšího“ pro danou aplikaci.

Výsledky v tabulce 5.1 jsou získány na základě nepravé, neinformativní apriorní hustoty. Z tohoto důvodu nelze využít posteriorní podíl šancí pro porovnání tohoto modelu s modely jinými. Místo toho použijeme posteriorní predikční  $p$ -hodnotu, pro pohled na to, jak dobře je model v souladu s daty. Část 5.6 popisuje, jak tyto  $p$ -hodnoty získat. V každém kroku tak generujeme umělá data z (5.15) a spočítáme měřítko šikmosti a převisu špičatosti, tedy (5.17) a (5.18). Na tomto základě jsme schopni získat posteriorní predikční hustotu pro tato dvě měřítká. Formálně počítáme  $E[Skew|y^\circ]$  a  $E[Kurt|y^\circ]$ , což jsou náhodné veličiny, protože  $y^\circ$  je náhodná veličina. Můžeme rovněž spočítat měřítko šikmosti a špičatosti pro námi pozorovaná data,  $E[Skew|y]$  a  $E[Kurt|y]$ , což již nejsou náhodné veličiny, protože  $y$  není náhodné. Umístění  $E[Skew|y]$  a  $E[Kurt|y]$  v rámci posteriorních predikčních hustot  $E[Skew|y^\circ]$  a  $E[Kurt|y^\circ]$  nám dává dostatečný náhled na to, jako dobře model popisuje data.

Obrázky 5.1 a 5.2 nám vykreslují aproximace výše zmiňovaných predikčních hustot. Jedná se o histogramy výběrů, které jsme v části 5.6 označovali jako  $Skew^{(1)}, \dots, Skew^{(S)}$  a  $Kurt^{(1)}, \dots, Kurt^{(S)}$ . V těchto obrázcích jsou rovněž označeny hodnoty  $E[Skew|y^\circ]$  a  $E[Kurt|y^\circ]$  jako „Pozorovaná šikmost“ a „Pozorovaná špičatost“. Jinými slovy, posteriorní predikční hustota pro šikmost (nebo špičatost) nám říká, jaké hodnoty šikmosti (nebo špičatosti) má náš model tendenci generovat. Obrázek 5.1 nám naznačuje, že nelineární regresní model z tohoto příkladu v podstatě ve všech případech generuje data s hodnotou šikmosti v absolutní hodnotě menší než jedna. Kdyby nám skutečná data ukázala hodnotu šikmosti v absolutní hodnotě větší než jedna, byl by to důkaz toho, že model není vhodný pro aplikaci na tato data. Ve skutečnosti je  $E[Skew|y] = 0.233$ , což je v blízkosti středu vykreslené hustoty z obrázku 5.1. Odpovídající  $p$ -hodnota je 0.51. V rámci našeho skritu se jedná o oboustrannou  $p$ -hodnotu). Toto číslo nám říká, že 51 % uměle vygenerovaných dat bude vykazovat v absolutní hodnotě vyšší hodnotu šikmosti než aktuální data. Šikmost pozorovaných dat je tak v souladu s tím, co by nám měl generovat náš nelineární regresní model. Podobný závěr lze učinit pro posteriorní predikční hodnotu převisu špičatosti. Pozorovaná data vracejí hodnotu  $E[Kurt|y] = -0.326$ , což je konzistentní se špičatostí data generovatelných modelem. Posteriorní predikční  $p$ -hodnota 0.43, což nám říká, že 43 % uměle vytvořených datových vzorků bude mít extrémnější hodnotu špičatosti (v absolutní hodnotě). Můžeme tedy učinit závěr, že nelineární regresní model použitý v tomto příkladě dobře vystihuje data, alespoň pokud jde o jejich šikmost a špičatost. Samozřejmě lze prezentovat i jednostranné varianty  $p$ -hodnoty.

Doposud jsme se zaměřili na empirickou analýzu vektoru parametrů  $\gamma$  a



**Obrázek 5.1:** Posteriorní predikční hustota šikmosti.

nic jsme si neřekli o parametru  $h$ . Pokud by nás zajímala posteriorní analýza parametru  $h$ , potřebovali bychom posteriorní simulátor, který by nám poskytl výběry  $h^{(s)}$ . Kromě toho, metoda Gelfanda-Deye pro výpočet marginální věrohodnosti vyžaduje výběry z úlného vektoru parametrů,  $\theta = (\gamma', h)'$ . Z těchto důvodů (a pro ilustraci dalšího algoritmu posteriorní simulace) si v krátkosti odvodíme Metropolis-within-Gibbs algoritmus (viz část 5.5.3), který bude generovat výběry z  $p(h|y, \gamma)$  a  $p(\gamma|y, h)$ .

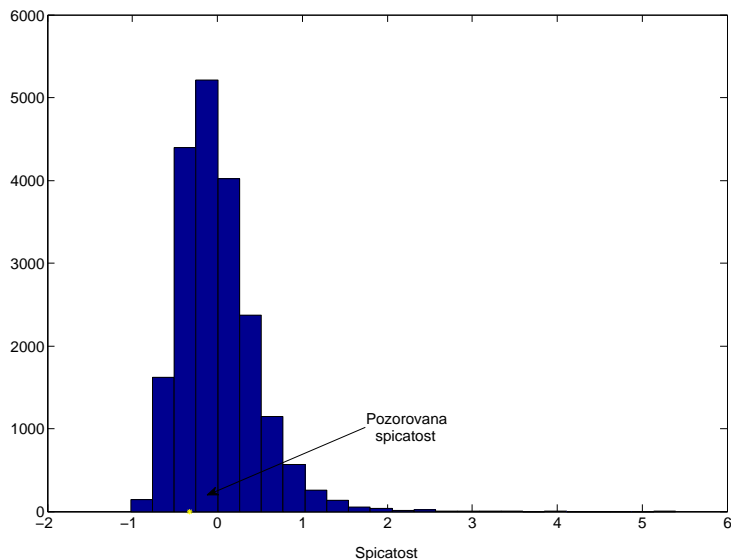
Porovnání modelů na základě posteriorního podílu šancí si budeme ilustrovat na příkladu analýzy toho, jestli je pro náš příklad adekvátní lineární regresní model. Náš omezený model bude  $M_1 : \gamma_4 = 1$ . Neomezený model  $M_2 : \gamma_4 \neq 1$ . K výpočtu posteriorního podílu šancí porovnávajícího tyto modely potřebujeme informativní apriorní hustotu. Použijeme normální-gama apriorní hustotu pro oba modely (viz kapitola 4, část 4.1). Pro model  $M_2$  je apriorní hustota pro  $\gamma$  nezávislá na apriorní hustotě pro  $h$  a má podobu

$$\gamma \sim N(\underline{\gamma}, \underline{V}).$$

Apriorní hustota pro  $h$  odpovídá hustotě

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

Zvolíme  $\underline{\gamma} = (1, 1, 1, 1)'$ ,  $\underline{V} = 0.25I_4$ ,  $\underline{\nu} = 12$  a  $\underline{s}^{-2} = 10.0$ . Vzhledem k pravděpodobným rozměrům mezního produktu práce a kapitálu a díky způsobu standardizace dat jsou tyto volby vcelku rozumné, i když relativně neinformativní.



**Obrázek 5.2:** Posteriorní predikční hustota špičatosti.

Pro model  $M_1$  použijeme tutéž apriorní hustotu s tím, že  $\gamma$  bude mít jen tři prvky, tudíž  $\underline{\gamma}$  bude bez svého posledního řádku a podobně  $\underline{V}$  bude bez posledního řádku a sloupce.

Pro lineární regresní model využijeme veškeré postupy z kapitoly 4, části 4.1, včetně Gibbsova vzorkovače pro posteriorní simulaci. Pro nelineární regresní model s touto apriorní hustotou si odvodíme Metropolis-within-Gibbs algoritmus. Ke konstrukci tohoto algoritmu si musíme odvodit podmíněné hustoty  $p(h|y, \gamma)$  a  $p(\gamma|y, h)$ . Využitím stejného postupu jako pro odvození (4.8) až (4.11) zjistíme, že

$$h|y, \gamma \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (5.23)$$

kde

$$\begin{aligned} \bar{\nu} &= N + \nu, \\ \bar{s}^2 &= \frac{[y - f(X, \gamma)]'[y - f(X, \gamma)] + \nu s^2}{\bar{\nu}}. \end{aligned}$$

S využitím (5.5) a s vědomím toho, že  $p(\gamma|y, h) \propto p(\gamma, h|y)$ , můžeme vidět,

že

$$p(\gamma|y, h) \propto \exp \left[ -\frac{h}{2} \{y - f(X, \gamma)\}' \{y - f(X, \gamma)\} \right] \exp \left[ -\frac{1}{2} (\gamma - \underline{\gamma})' \underline{V}^{-1} (\gamma - \underline{\gamma}) \right]. \quad (5.24)$$

Tato podmíněná posteriorní hustota nemá podobu známého rozdělení, ze kterého bychom mohli snadno generovat náhodné výběry. Nicméně, můžeme použít Metropolis-Hastings algoritmus pro  $p(\gamma|y, h)$ , který v kombinaci s výběry z gama rozdělení z (5.23) dává Metropolis-within-Gibbs algoritmus. V našem příkladě použijeme pro výběry z (5.24) Random Walk Chain Metropolis-Hastings algoritmus. Tento algoritmus je stejný jako v úvodní části tohoto příkladu, jen s tím rozdílem, že akceptační pravděpodobnost je počítána s využitím (5.24).

Výpočet marginální věrohodnosti metodou Gelfanda a Deye je vcelku jasný, pokud máme výstup z takových posteriorních simulátorů jako jsme využili pro modely  $M_1$  a  $M_2$ . V části 5.7 lze nalézt potřebné detaily. Samotný výpočet je obsahem souboru `priklad_GelfandDey.m`. Po tom, co získáme výběry posteriorních simulátorů, musíme vyhodnotit pro tyto výběry věrohodnostní funkci (funkce `CES_like.m` a `CES_like_lin.m`) a apriorní hustotu (`CES_like_lin.m`), viz (5.20). Protože je pro oba modely apriorní hustota v podobě nezávislé normální-gama hustoty, vyhodnocení apriorní hustoty je jasné. Pro model  $M_1$  je věrohodnostní funkce dána výrazem (3.3) a pro  $M_2$  je dán v (5.3). Z Matlabovských funkcí lze vidět, že napsání příslušných kódů není nijak obtížné. Stejně tak lze snadno naprogramovat funkci označenou jako  $f(\theta)$  (viz (5.20)), zahrnující jen (5.21) a (5.22). Jakmile vyhodnotíme apriorní hustotu, věrohodnostní funkci a  $f(\cdot)$  v každém posteriorním výběru, můžeme spočítat  $g(\cdot)$  (5.20) a zprůměrovat získané výsledky, čímž získáme odhad marginální věrohodnosti (resp. její inverzi).

Oba posteriorní simulátory (jak Gibbsův vzorkovač pro lineární model, tak i Metropolis-within-Gibbs algoritmus pro nelineární model) generovaly  $S = 25000$  vzorků, z čehož jsme  $S_0 = 5000$  vzorků vyhodili. Marginální věrohodnost byla spočítána pro  $p = 0.01$  (viz (5.21)). Pro jiné hodnoty ale výsledek nebude příliš odlišný. Výsledný odhad Bayesova faktoru je 1.09. Lineární model tak je tak stejně pravděpodobný jako model nelineární. S ohledem na výsledky odhadu  $\gamma_4$  a příslušný rozptyl to není překvapující výsledek. Pro věrohodnost všech předchozích výsledků je třeba ověřit konvergenci algoritmů pomocí MCMC diagnostik.

## 5.10 Shrnutí

V této kapitole jsme se zabývali bayesiánskou analýzou nelineárního regresního modelu. Aspekt nelinearity znamená, že nemáme k dispozici analytické výsledky, a to i pro případ jednoduchých neinformativních apriorních hustot. Obecně není možné vyvinout Gibbsův vzorkovač pro posteriorní simulaci. Tyto vlastnosti nás motivovaly pro zavedení nové třídy velmi obecných algoritmů posteriorní

simulace, třídu Metropolis-Hastings algoritmů. Dvě obvyklé varianty jsou Independence Chain a Random Walk Chain Metropolis-Hastings algoritmy a obě jsme si detailně popsali. První z nich pracuje velmi dobře v případě, kdy existuje dobrá aproximace posteriorní hustoty. Druhý algoritmus lze použít v případě, kdy tato aproximace neexistuje resp. ji nelze nalézt. Nelineární regresní model jsme využili k zavedení nové metody hodnocení modelu nebo modelů. Jedná se o prediktivní  $p$ -hodnotu, což je obecný nástroj pro vyjádření kvality modelu z hlediska jeho souladu s daty. Obecný nástroj pro výpočet marginální věrohodnosti je metoda Gelfanda a Deye, který byl rovněž detailně vysvětlena.

Na tomto místě jsme si fakticky představili veškeré základní nástroje a koncepty využívané v následujících kapitolách. Pokud jde o posteriorní simulátory, tak jsme se seznámili s Monte Carlo integrací, importance sampling, Gibbsovým vzorkovačem a Metropolis-Hastings algoritmem. V rámci porovnání modelů jsme pokryli metodu Gelfanda a Deye pro výpočet marginální věrohodnosti, stejně jako snadnější, ale méně obecnou metodu Savage-Dickeyeho poměru hustot. Rovněž byl popsán přístup využívající predikční  $p$ -hodnotu. Tento přístup lze nejlépe popsat jako způsob souladu modelu s daty, nicméně jej lze využít i k porovnání modelů.

Zbytek kapitol se zabývá problematikou implementace doposud probraných obecných nástrojů a konceptů na konkrétní typy modelů. Je dobré poznamenat, že k dispozici máme širokou paletu nástrojů, kdy po některé modely je možno použít vícero z nich. V případě nelineárního modelu tak můžeme použít jak importance sampling, tak i Metropolis-Hastings algoritmus. Pro porovnání modelů můžeme použít posteriorní podíl šancí nebo posteriorní predikční  $p$ -hodnotu. Použití konkrétního nástroje tak může být spíše otázkou preferencí nebo zvyku. Konkrétně zvolený přístup či nástroj v dalších kapitolách tak nelze automaticky brát jako jediný možný nebo dokonce ten nejlepší.



## Kapitola 6

# Lineární regresní model s obecnou kovarianční maticí chyb

### 6.1 Úvod

V této kapitole se vrátíme k normálnímu lineárnímu regresnímu modelu

$$y = X\beta + \epsilon \quad (6.1)$$

V předchozích kapitolách jsme předpokládali, že  $\epsilon$  je z  $N(0_N, h^{-1}I_N)$ . Předpoklad, že náhodná složka má nulovou střední hodnotu, pokud by tomu tak snad nebylo, je možno tuto nenulovou střední hodnotu zahrnout do úrovně konstanty (tedy pokud máme model bez úrovně konstanty a předpokládáme-li nenulovou střední hodnotu chybového členu, je tento model možno zapasat jako model s úrovní konstantou a náhodnou složkou, která již nulovou střední hodnotu mít bude). Předpoklad kovarianční matice  $h^{-1}I_N$  však již tak samozřejmý není, a v řadě případů jej bude žádoucí uvolnit, rovněž i předpoklad u normalitě náhodné složky nemusí být vždy adekvátní modelové situaci.

V rámci této kapitoly budou všechny modely založeny na (6.1) a na následujících předpokladech:

1.  $\epsilon$  je z vícerozměrného normálního rozdělení s střední hodnotou  $0_N$  a kovarianční maticí  $h^{-1}\Omega$ , kde  $\Omega$  je pozitivně definitní matice rozměru  $N \times N$ .
2. Všechny prvky  $X$  jsou buď pevně daná čísla (tj. nenáhodné veličiny) nebo pokud se jedná o náhodné veličiny, pak musí být nezávislé vzhledem k  $\epsilon$  s funkcí hustoty pravděpodobnosti  $p(X|\lambda)$ , kde  $\lambda$  je vektorem parametrů, které nezahrnují  $\beta$  nebo  $h$ .

Různé modely diskutované v této kapitole se budou odlišovat v přesně podobě  $\Omega$ . Budeme se zabývat některými specifickými tvary, se kterými se můžeme

setkat v řadě aplikací. Začneme předpokladem *heteroskedasticity*, což je případ, kdy se rozptyly náhodných složek budou v rámci jednotlivých pozorování lišit. Budeme předpokládat dva typy heteroskedasticity: v jednom případě bude její podoba známa, ve druhém pak bude neznámá. Poslední případ nám umožní uvolnit předpoklad normality a ukážeme si, jak určitý model s heteroskedasticitou v neznámé podobě je ekvivalentní LRM s náhodnou složkou ze Studentova  $t$ -rozdělení. Tento model nám umožní zavést koncept *hierarchického prioru*. Dále bude předpokládat případ, kdy náhodné chyby jsou vzájemně korelovány. Přesněji, budeme analyzovat model s autoregresními (AR) chybovými členy. Autoregresní modely jsou významnými modely v rámci analýzy časových řad. Posledním typem modelů v této kapitole budou tzv. SUR modely, neboli *seemingly unrelated regression models*, tedy modely zdánlivě nesouvisejících regresí.

## 6.2 Model s obecnou maticí $\Omega$

Na úvod se zmíníme o obecném řešení našeho problému. Jelikož předpokládáme, že  $\Omega$  je pozitivně definitní matice, platí, že bude existovat matice  $P$  rozměru  $N \times N$ , s vlastností, že  $P\Omega P' = I_N$ . Vynásobíme-li obě strany (6.1) maticí  $P$ , dostaneme transformovaný model

$$y^* = X^* \beta + \epsilon^* \quad (6.2)$$

kde  $y^* = Py$ ,  $X^* = PX$  a  $\epsilon^* = P\epsilon$ . Lze ověřit, že po této transformaci je  $\epsilon^*$  z normálního rozdělení  $N(0_N, h^{-1}I_N)$ . Tento transformovaný model je tak identický s modely diskutovanými v předchozích kapitolách. To má dvě důležité implikace:

1. Pokud je  $\Omega$  známa, jsme schopni přetransformovat data a provést bayesiánskou analýzu metodami probíranými v předchozích kapitolách.
2. Pokud je  $\Omega$  neznámá, vztah (6.2) nám napovídá jak provést bayesovskou analýzu. Podmíněno maticí  $\Omega$ , podmíněné posteriorní hustoty pro  $\beta$  a  $h$  budou mít obdobnou podobu jako v předchozích kapitolách. Pokud je apriorní hustota pro  $\beta$  a  $h$   $NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu})$ , potom veškeré výsledky kapitol 2 a 3 jsou aplikovatelné, podmíníme-li je  $\Omega$ , a jsme tedy schopni generovat náhodné výběry. V této kapitole se zaměříme na nezávislou normální-gama apriorní hustotu. Můžeme tak vytvořit Gibbsův vzorkovač pro sekvenční výběry z  $p(\beta|y, h, \Omega)$ ,  $p(h|y, \beta, \Omega)$  a  $p(\Omega|y, \beta, h)$ . Obě dvě posteriorní podmíněné hustoty budou normální a gama (jako v kapitole 4), přičemž  $p(\Omega|y, \beta, h)$  závisí na přesné formě  $\Omega$ . Veškerá nová odvození se tak budou týkat pouze poslední z hustot.



### 6.2.1 Věrohodnostní funkce

Využitím vlastností vícerozměrného normálního rozdělení můžeme věrohodnostní funkci zapsat jako:

$$p(y|\beta, h, \Omega) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} |\Omega|^{-\frac{1}{2}} \left\{ \exp \left[ -\frac{h}{2} (y - X\beta)' \Omega^{-1} (y - X\beta) \right] \right\} \quad (6.3)$$

případně za použití transformovaných dat

$$p(y^*|\beta, h, \Omega) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y^* - X^*\beta)' (y^* - X^*\beta) \right] \right\} \quad (6.4)$$

V kapitole 3 bylo ilustrováno, jak zapsat věrohodnostní funkci pomocí OLS kvantit (3.4-3.7). Obdobným odvozením zde jsme schopni odvodit věrohodnostní funkci v kontextu GLS (General Least Squares – zobecněná metoda nejmenších čtverců) kvantit:

$$\nu = N - k \quad (6.5)$$

$$\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} y^* = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad (6.6)$$

$$\begin{aligned} s^2(\Omega) &= \frac{(y^* - X^* \hat{\beta}(\Omega))' (y^* - X^* \hat{\beta}(\Omega))}{\nu} \\ &= \frac{(y - X \hat{\beta}(\Omega))' \Omega^{-1} (y - X \hat{\beta}(\Omega))}{\nu} \end{aligned} \quad (6.7)$$

$$\begin{aligned} p(y|\beta, h, \Omega) &= \frac{1}{(2\pi)^{\frac{N}{2}}} \\ &\times \left\{ h^{\frac{1}{2}} \exp \left[ -\frac{h}{2} (\beta - \hat{\beta}(\Omega))' X' \Omega^{-1} X (\beta - \hat{\beta}(\Omega)) \right] \right\} \\ &\times \left\{ h^{\frac{\nu}{2}} \exp \left[ -\frac{h\nu}{2s(\Omega)^{-2}} \right] \right\} \end{aligned} \quad (6.8)$$

### 6.2.2 Apriorní hustota

Použijeme normální-gama apriorní hustotu pro  $\beta$  a  $h$  a obecné značení  $p(\Omega)$  pro označení apriorní hustoty pro  $\Omega$ :

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

kde

$$p(\beta) = f_N(\beta|\underline{\beta}, \underline{V}), \quad (6.9)$$

$$p(h) = f_G(h|\underline{\nu}, \underline{s}^{-2}). \quad (6.10)$$

### 6.2.3 Posteriorní hustota

Posteriorní hustota je tradičně poroporcionální součinu apriorní hustoty a věrohodnostní funkce v podobě

$$\begin{aligned}
 p(\beta, h, \Omega|y) &\propto p(\Omega) \\
 &\times \left\{ \exp \left[ -\frac{1}{2} \left\{ h(y^* - X^*\beta)'(y^* - X^*\beta) \right. \right. \right. \\
 &\quad \left. \left. \left. + (\beta - \underline{\beta})' \underline{V}^{-1} (\beta - \underline{\beta}) \right\} \right] \right\} \\
 &\times h^{\frac{N+\nu-2}{2}} \exp \left[ -\frac{h\nu}{2\underline{s}^{-2}} \right]
 \end{aligned} \tag{6.11}$$

Sdružená hustota pravděpodobnosti nemá tvar žádného obvyklého rozdělení. Alternativní vyjádření této posteriorní hustoty jsou tedy nadbytečná. Snadno však lze odvodit některé podmíněné posteriorní hustoty, a to obdobným způsobem jako tomu bylo v kapitole 4. Je tedy možné odvodit podmíněnou posteriorní hustotu pravděpodobnosti pro  $\beta$ , což je vícerozměrné normální rozdělení:

$$\beta|y, h, \Omega \sim N(\bar{\beta}, \bar{V}) \tag{6.12}$$

kde

$$\bar{V} = (\underline{V}^{-1} + hX'\Omega^{-1}X)^{-1}, \tag{6.13}$$

$$\bar{\beta} = \bar{V}(\underline{V}^{-1}\underline{\beta} + hX'\Omega^{-1}X\hat{\beta}(\Omega)). \tag{6.14}$$

Podmíněná posteriorní hustota pro  $h$  odpovídá gama rozdělení:

$$h|y, \beta \sim G(\bar{s}^{-2}, \bar{\nu}) \tag{6.15}$$

kde

$$\bar{\nu} = N + \nu, \tag{6.16}$$

$$\bar{s}^2 = \frac{(y - X\beta)'\Omega^{-1}(y - X\beta) + \nu s^2}{\bar{\nu}}. \tag{6.17}$$

Posteriorní podmíněná hustota pro  $\Omega$  má jádrovou hustotu v podobě

$$p(\Omega|y, \beta, h) \propto p(\Omega)|\Omega|^{-\frac{1}{2}} \left\{ \exp \left[ -\frac{h}{2} (y - X\beta)'\Omega^{-1}(y - X\beta) \right] \right\} \tag{6.18}$$

Obecně tato posteriorní podmíněná hustota nenabývá podoby známého rozdělení. Dále budeme předpokládat již konkrétní podobu pro  $\Omega$  a odvodíme tomu odpovídající posteriorní simulátory. Pokud budeme schopni generovat výběry z  $p(\Omega|y, \beta, h)$ , jsme schopni využít Gibbsův vzorkovač, neboť  $p(\beta|y, h, \Omega)$  odpovídá normálnímu rozdělení a  $p(h|y, \beta, \Omega)$  rozdělení gama.

## 6.3 Heteroskedasticita ve známé podobě

Případ heteroskedasticity nastává v případech, kdy rozptyl náhodných složek se liší mezi jednotlivými pozorováními. Opakem heteroskedasticity je případ homoskedasticity, která byla předpokladem modelů probíraných v předchozích kapitolách. Jako motivace, proč může být žádoucí předpokládat heteroskedasticitu chybového členu, si uvedme několik příkladů. Předpokládejme mikroekonomický případ, kde závislou proměnnou jsou tržby firem. Zde můžeme reálně předpokládat, že chybový člen bude proporcionální velikosti firmy, tedy pro malé firmy bude menší než pro velké. Jiným příkladem může být model, který analyzuje data různých států, kdy z důvodu, že rozvinuté země budou mít lepší institucionální zabezpečení pro sběr statistických dat než země rozvojové, můžeme se zde opět setkat s případem, kdy chybové složky budou pro rozvinuté země výrazně nižší.

V notaci našeho regresního modelu, heteroskedasticita nastává v případě, kdy

$$\Omega = \begin{bmatrix} \omega_1 & 0 & \cdot & \cdot & 0 \\ 0 & \omega_2 & 0 & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \omega_N \end{bmatrix}. \quad (6.19)$$

Jinými slovy tedy máme stejný model jako v kapitolách 2 až 4, s tím rozdílem, že nyní předpokládáme  $\text{var}(\epsilon_i) = h^{-1}\omega_i$  pro  $i = 1, \dots, N$ .

Výše uvedené motivační příklady naznačují, že často známe nebo alespoň předpokládáme, jakou formu bude heteroskedasticita nabývat. Například,  $\omega_i$  by mohla záviset na tom, zda-li firma  $i$  je malá nebo velká, nebo zda-li země  $i$  je rozvinutá nebo rozvojová. Budeme obecně předpokládat, že

$$\omega_i = h(z_i, \alpha) \quad (6.20)$$

kde  $h()$  je kladná funkce závisající na parametrech  $\alpha$  a  $p$ -rozměrném vektoru dat,  $z_i$ . Vektor  $z_i$  může obsahovat některé či všechny vysvětlující proměnné  $x_i$ . Obvyklá volba pro  $h()$ , která nám zajistí, že všechny rozptyly náhodných složek budou kladné je:

$$h(z_i, \alpha) = (1 + \alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_p z_{ip})^2 \quad (6.21)$$

Samozřejmě závěry této části kapitoly budou platit i pro jiné volby funkce  $h()$ .

Apriorní hustota, věrohodnostní funkce a posteriorní hustota odpovídají funkcím hustoty pravděpodobnosti prezentovaným v úvodu této kapitoly, s tím, že jsme schopni vyjádřit i výraz pro  $\Omega$  daný vztahem (6.19). Je však třeba upozornit, že  $\Omega$  nyní závisí na parametrech  $\alpha$ .

Pro bayesiánskou analýzu tohoto „heteroskedastického“ modelu potřebujeme mít posteriorní simulátor. Vhodným simulátorem je zde algoritmus Metropolis-within-Gibbs. Jak již bylo poznamenáno,  $p(\beta|y, h, \alpha)$  má normální rozdělení a  $p(h|y, \beta, \alpha)$  odpovídá gama rozdělení. Pro kompletní posteriorní simulátoru

tedy potřebujeme metodu pro generování vzorků z  $p(\alpha|y, \beta, h)$ . Pokud dosadíme (6.19) a (6.20) do (6.18), nebude mít výsledný výraz pro  $p(\alpha|y, \beta, h)$  podobu známé hustoty rozdělení. Ovšem jsme schopni vyvinout Metropolis-Hastings algoritmus. V následující ilustraci je použit Random Walk Chain M-H algoritmus, ačkoliv jsou možné i jiné algoritmy. Lze spočítat Bayesův faktor pro jakoukoliv hypotézu (např.  $\alpha_1 = \dots = \alpha_p = 0$ , tedy zda-li je v modelu přítomna heteroskedasticita), a to s využitím přístupu Gelfanda-Deye. Rovněž lze testovat kvalitu modelu za pomoci  $p$ -hodnot nebo intervalů nejvyšší posteriorní hustoty. Predikční analýzu je rovněž možno provést standardním způsobem.

### 6.3.1 Empirická ilustrace

K ilustraci problému heteroskedasticity ve známé podobě využijeme data o cenách domů z kapitoly 3. Náš model budeme chápat jako normální lineární regresní model s heteroskedasticitou ve známé podobě a k jeho identifikaci tak využijeme Gibbsův vzorkovač. Využívaná data jsou podrobně popsána v části 3.8. Předpokládáme, že heteroskedasticita bude mít podobu odpovídající vztahu (6.21), kdy  $z_i = (x_{i2}, \dots, x_{ik})'$ . Apriorní hustoty pro parametry  $\beta$  a  $h$  jsou dány vztahy (6.9) a (6.10). Jako apriorní hyperparametry použijeme stejné parametry jako v kapitole 4, části 4.1.7. Pro vektor parametrů  $\alpha$  použijeme neinformativní apriorní hustotu v podobě

$$p(\alpha) \propto 1.$$

Všimněme si, že se jedná o nepravou apriorní hustotu, což má za následek to, že nejsme schopni pro ověření hypotéz zahrnující prvky vektoru  $\alpha$  smysluplné Bayesovy faktory. Z tohoto důvodu si ukážeme, kromě posteriorních středních hodnot a směrodatných odchylek, 95% intervaly nejvyšší posteriorní hustoty (HPDI). To vše je obsahem tabulky 6.1.

**Tabulka 6.1:** Posteriorní výsledky pro  $\beta$ ,  $h$  a  $\alpha$

	Stř. hodnota	Sm. odchylka	95% HPDI
$\beta_1$	-5406.98	2981.53	[-11267.91, 376.88]
$\beta_2$	6.1438	0.4223	[5.3270, 6.9771]
$\beta_3$	3107.33	1018.74	[1117.58, 5107.99]
$\beta_4$	14417.35	1669.22	[11156.38, 17714.82]
$\beta_5$	7866.99	929.76	[6034.01, 9691.73]
$h$	0.0000	0.0000	[0.0000, 0.0000]
$\alpha_1$	0.0006	0.0001	[0.0003, 0.0009]
$\alpha_2$	0.6401	0.2907	[0.1030, 1.2395]
$\alpha_3$	0.7057	0.3786	[0.0641, 1.5376]
$\alpha_4$	-0.3678	0.3087	[-0.9528, 0.2785]

Posteriorním simulátorem je Metropolis-within-Gibbs algoritmus, který ge-

neruje vzorky  $\beta$  a  $h$  z rozdělení daného vztahy (6.12) resp. (6.15). Výběry z  $p(\alpha|y, \beta, h)$  jsou dány v (6.18), kdy známe přesnou podobu matice  $\Omega$  (která je funkcí parametrů  $\alpha$ ). Rovnice (6.18), vyhodnocená v kandidátském a posledním výběru, je použita pro výpočet akceptační pravděpodobnosti (viz kapitola 5, vztah (5.11)). Rozptyl kandidátské hustoty, označený jako  $\Sigma$  ve vztahu (5.12), je zvolen na základě volby  $\Sigma = cI$  a na základě následného experimentování s různými hodnotami skalární proměnné  $c$ , a to až do bodu nalezení přijatelné akceptační pravděpodobnosti. Posteriorní simulátor je pak spuštěn s využitím této hodnoty pro odhad posteriorního rozptylu  $\alpha$ ,  $\widehat{var}(\alpha|y)$ . Potom je zvolena matice  $\Sigma = c\widehat{var}(\alpha|y)$  a opět se provede několik experimentálních odhadů pro různé hodnoty  $c$ , až nalezneme hodnotu, která povede k průměrné akceptační pravděpodobnosti mezi 0.2 až 0.3 (případně až do 0.5). To vše je provedeno na krátkých vzorcích, neboť simulace zabírá nějaký ten čas. Finální běh algoritmu je proveden na základě 50000 replikací, kdy je prvních 20000 zahozeno. Samotný matlabovský skript je obsahem souboru `priklad_hetero_znama.m`, čas odhadu na čtyřjádrovém procesoru Intel Core 2 Quad (s frekvencí 2.83 GHz) zabral 65 minut. Podobné odhady by bylo možno získat i při nižším počtu replikací. Pravdou rovněž je, že i samotný algoritmus by bylo možné dále výpočetně zefektivnit, snahou však bylo využít značení a obecné postupy co nejvíce odpovídající textu předchozí části kapitoly. Jako užitečné se jeví upozornit na úskalí numerických chyb např. při vyhodnocování jádrové hustoty (ta je obsahem funkce `a_post.m`). V tomto případě bylo přímo nutné využít nejen logaritmus jádrové hustoty (což je velmi běžný postup), ale i faktu, že matice  $\Omega$  je diagonální matice, z čehož vyplývá, že logaritmus determinantu této matice je roven součtu logaritmů prvků na hlavní diagonále. V opačném případě MATLAB nevracel správné výsledky hodnoty logaritmu determinantu takto rozsáhlé matice (součin 526 malých čísel na diagonále může vést při numerickém násobení k hodnotě nekonečno a logaritmus nekonečna nekonečnem zůstane). MCMC diagnostiky indikují konvergenci Metropolis-within-Gibbs algoritmu a numerické standardní chyby pak naznačují chybu aproximace, které jsou velmi malé relativně k posteriorním směrodatným odchylkám jednotlivých parametrů.

Tabulka 6.1 (hodnoty pro parametr  $h$  vypadají nulověnám říká, že heteroskedasticita bude pro danou datovou sadu relevantním faktorem. To vyplývá ze skutečnosti, že 95% intervaly nejvyšší posteriorní hustoty neobsahují nulu pro parametry  $\alpha_1$ ,  $\alpha_2$  a  $\alpha_3$ . To tedy znamená, že rozloha domu, počet ložnic a počet koupelen disponující významnou vysvětlující silou v rovnici heteroskedasticity. Skutečnost, že jsou všechny tyto koeficienty kladné, nám dále říká, že rozptyl náhodných složek pro velké domy má tendenci být vyšší než pro domy menší. V předchozích kapitolách byl problém heteroskedasticity ignorován (v rámci odhadu modelu cen domů). Pokud porovnáme výsledky z tabulky 6.1 a z tabulky 4.1, vidíme, jaký vliv toto opomenutí mělo. Posledně zmiňovaná tabulka obsahuje výsledky pro homoskedastickou verzi modelu, nicméně využívá tatáž data a tytéž apriorní hustoty pro parametry  $\beta$  a  $h$ . Lze vidět, že zahrnutí heteroskedasticity má určitý vliv na posteriorní hustotu vektoru parametrů  $\beta$ . Můžeme například vidět, že střední hodnota parametru  $\beta_4$  byla 16136 pro

homoskedastický model a 14417 pro model v heteroskedastickém formátu. Pro mnohá využití však i takovýto rozdíl může být zanedbatelná a lze tak dospět k závěru, že zahrnutí heteroskedasticity nijak výrazně výsledky odhadu neovlivní. Ale záleží skutečně na vlastní interpretaci výsledků.

## 6.4 Heteroskedasticita v neznámé podobě

V této části budeme hledat odpověď na otázku, jak postupovat, pokud víme, že v datech je přítomna heteroskedasticita náhodných složek, ale neznáme její přesnou podobu. Jinými slovy, předpokládáme, že platí (6.19), nicméně již nechceme (resp. neumíme) specifikovat přesnou funkční podobu pro  $\omega_i$  jako ve vztahu (6.20). Na první pohled se zdá, že odhadnout na základě  $N$  pozorování  $N + k + 1$  parametrů (tj.  $\beta$ ,  $h$  a  $\omega = (\omega_1, \dots, \omega_N)'$ ) je v podstatě nemožné. Nicméně uvidíme, že rozšířením techniky z předchozí části této kapitoly, jsme schopni řešit i tento problém. Metoda, kterou si zde v této souvislosti ukážeme, bude užitečná ze dvou hledisek:

1. Tato metoda zahrnuje využití tzv. *hierarchických priorů*. Tyto priority hrají významnou roli v současném rozvoji bayesiánské statistické teorie a své místo nalézají čím dál více i v ekonometrii. Využívají se jako způsob tvorby více flexibilních na parametry bohatých modelů pro statistickou analýzu.
2. Tento model nám dovolí zavést koncept vztahující se k flexibilnímu ekonometrickému modelování a konkrétně pak nám umožní uvolnit předpoklad normality náhodné složky.

Začneme tedy formulací apriorní hustoty  $p(\omega)$  pro  $N$ -rozměrný vektor parametrů  $\omega$ . Obvykle se pracuje s přesností chyby než jejím rozptylem, tudíž budeme definovat  $\lambda \equiv (\lambda_1, \lambda_2, \dots, \lambda_N)' \equiv (\omega_1^{-1}, \omega_2^{-1}, \dots, \omega_N^{-1})'$ . Předpokládáme následující apriorní hustotu pro  $\lambda$ :

$$p(\lambda) = \prod_{t=1}^N f_G(\lambda_t | 1, \nu_\lambda) \quad (6.22)$$

Poznamenejme, že apriorní hustota pro  $\lambda$  závisí na hyperparametru  $\nu_\lambda$ , který si volíme, a předpokládá se, že každé  $\lambda_i$  pochází z téhož rozdělení. Jinými slovy, parametry  $\lambda_i$  jsou i.i.d. výběry z gama rozdělení. Tento předpoklad je nutný pro vypořádání se s problémy vysoké dimenzionality vektoru  $\lambda$ . Intuitivně řečeno, pokud bychom chtěli pracovat s  $\lambda_1, \dots, \lambda_N$  jako s  $N$  zcela nezávislými a neomezenými parametry, nebudeme mít dostatek pozorování k jejich odhadům. Náš předpoklad umožňuje to, že rozptyly náhodných složek se budou vzájemně lišit, ovšem budou pocházet ze stejného rozdělení. Máme tak velmi flexibilní model s dostatečně pevnou strukturou pro možnou statistickou analýzu.

Můžeme se podívat nad tím, proč by  $\lambda_i$  měly být i.i.d. výběry z gama rozdělení se střední hodnotou 1.0. Důvod je ten, že náš model s věrohodností

funkcí (6.3) a apriorními hustotami danými vztahy (6.9) (6.10) a (6.22) je naprosto stejný jako lineární regresní model s náhodnými složkami z i.i.d. Studentova  $t$ -rozdělení s  $\nu_\lambda$  stupni volnosti. Jinými slovy, předpokládáme-li

$$p(\epsilon_i) = f_t(\epsilon_i|0, h^{-1}, \nu_\lambda) \quad (6.23)$$

pro  $i = 1, \dots, N$ , odvozením tomu odpovídající věrohodnostní funkce a kombinací s apriorními hustotami pro  $\beta$  a  $h$  získáme identickou posteriorní hustotu. Důkaz a další vysvětlení nabízí Geweke [15]. Studentovo  $t$ -rozdělení je podobné normálnímu rozdělení, má ovšem tlustší konce a je více flexibilnější. Ve skutečnosti je normální rozdělení speciálním případem Studentova  $t$ -rozdělení pro  $\nu_\lambda \rightarrow \infty$ . Máme tedy model, který nám dává flexibilnější rozdělení náhodných složek, a to bez toho abychom opustili rámec normálního lineárního regresního modelu. Navíc můžeme těžit z výpočetních metod vyvinutých výše, abychom vytvořili posteriorní simulátor pro lineární regresní model s nezávislými  $t$ -rozdělenými chybami.

Je třeba poznamenat, že zde diskutovaný model je mixem specifických normálních rozdělení. Intuitivně, pokud je normální rozdělení příliš restriktivní, jsme schopni vytvořit mnohem flexibilnější rozdělení váženým průměrem více normálních rozdělení. Čím více normálním rozdělení "namixujeme", tím flexibilnější bude výsledné rozdělení, čímž bude schopno aproximovat takřka jakékoli rozdělení s vysokým stupněm přesnosti. Takováto směs normálních modelů je tedy mocným nástrojem v situacích, kdy ekonomická teorie nenabízí specifikaci věrohodnostní funkce a my chceme být dostatečně flexibilní. Naše chápání heteroskedasticity v neznámé podobě je ekvivalentní poměrnému mixu normálních rozdělení (*scale mixture of Normals*). Tím je myšleno, že předpoklad, že  $\epsilon_i$  jsou  $N(0, h^{-1}\lambda_i^{-1})$  s apriorní hustotou pro  $\lambda_i$  danou v (6.22) je ekvivalentní předpokladu, že rozdělení chyb je váženým průměrem (či mixem) různých normálních rozdělení, které mají různé rozptyly (tj. různá měřítka – *scales*) a stejné střední hodnoty (tj. všechny chyby mají nulovou střední hodnotu). Pokud je tato směs vytvořena za použití gama hustot  $f_G(\lambda_i|1, \nu_\lambda)$ , je výsledek ekvivalentní  $t$ -rozdělení. Využitím jiných hustot než  $f_G(\lambda_i|1, \nu_\lambda)$  jsme schopni získat jiná rozdělení, ještě mnohem flexibilnější.

Předchozí analýza předpokládala, že  $\nu_\lambda$  bylo známé. Prakticky to nemusí být rozumný předpoklad, a je tedy žádoucí brát ho jako neznámý parametr. V bayesovském konceptu vyžaduje každý parametr znalost apriorní hustoty, a tudíž zde zatím použijeme značení  $p(\nu_\lambda)$ . Pokud to učiníme, je apriorní hustota pro  $\lambda$  specifikována ve dvou krocích, jedním z nich je (6.22), druhým pak  $p(\nu_\lambda)$ . Alternativně lze apriorní hustotu pro  $\lambda$  zapsat jako  $p(\lambda|\nu_\lambda)p(\nu_\lambda)$ . Apriorní hustoty zapsána ve dvou či více krocích, jako je uvedeno zde, se nazývají hierarchické apriorní hustoty. Zápis apriorní hustoty jako hierarchický prior je obvyklým způsobem zápisu apriorní informace. Není to však nutné, neboť zákony pravděpodobnosti implikují, že hierarchický prior lze zapsat v nehierarchickém pojetí. V našem případě můžeme použít  $p(\lambda) = \int p(\lambda|\nu_\lambda)p(\nu_\lambda)d\nu_\lambda$  pro odvození nehierarchické verze apriorní hustoty pro  $\lambda$ .

Ve všech předchozích empirických ilustracích byly prezentovány posteriorní střední hodnoty jako bodové odhady parametrů a posteriorní směrodatné od-

chyby jako měřítka neurčitosti spojená s bodovými odhady. Střední hodnoty a směrodatné odchylky však nemusí existovat pro každou hustotu pravděpodobnosti. Zde uváděný model je právě takovým modelem, kde střední hodnoty a směrodatné odchylky nemusí nutně existovat. Přesněji, Geweke [15] ukázal, že pokud použijeme neinformativní apriorní hustotu pro  $\beta$  (tj.  $p(\beta) \propto 1$  na intervalu  $(-\infty, \infty)$ ), potom posteriorní střední hodnota nebude existovat, pokud  $p(\nu_\lambda)$  není nulová na intervalu  $(0, 2)$ . Posteriorní směrodatná odchylka nebude existovat v případě, pokud  $p(\nu_\lambda)$  nebude nulová na intervalu  $(0, 4)$ . Pokud tedy chceme využít neinformativní prior pro  $\beta$ , měli bychom užít apriorní hustotu, která bude vylučovat malé hodnoty  $\nu_\lambda$ , jinak bude nutné prezentovat posteriorní mediány a mezikvartilová rozpětí (které budou existovat pro jakoukoliv p.d.f). Pokud využijeme normální informativní apriorní hustotu pro  $\beta$  jako (6.9), posteriorní střední hodnoty i směrodatné odchylky budou existovat.

Rovněž se nemusí vyplatit použít neinformativní apriorní hustotu pro  $\nu_\lambda$ . Naivní výzkumník, který by chtěl být dostatečně neinformativní, by mohl použít nepravou uniformní apriorní hustotu

$$p(\nu_\lambda) \propto 1 \quad \nu_\lambda(0, \infty)$$

s úmyslem dát stejnou apriorní váhu každému intervalu o stejné délce. Ovšem Studentovo  $t$ -rozdělení s  $\nu_\lambda$  stupni volnosti se blíží normálnímu rozdělení pro  $\nu_\lambda \rightarrow \infty$ . Prakticky je Studentovo  $t$ -rozdělení identické normálnímu pro  $\nu_\lambda > 100$ . Předchozí naivní "neinformativní" apriorní hustota všechnu váhu právě do tohoto intervalu, tzn.  $\frac{p(\nu_\lambda \leq 0)}{p(\nu_\lambda > 100)} = 0$ . Tento neinformativní prior je tedy více než informativní, říká, že chybové složky jsou normálně rozděleny! Toto je ilustrace jednoho z mnoha problémů, které může přinést použití neinformativních apriorních hustot.

### 6.4.1 Bayesovský výpočet

V této části si odvodíme Gibbsův vzorkovač pro posteriorní analýzu  $\beta$ ,  $h$ ,  $\lambda$  a  $\nu_\lambda$ . Gibbsův vzorkovač vyžaduje odvození plně podmíněných posteriorních rozdělení těchto parametrů. Některé z těchto hustot již byly odvozeny, konkrétně  $p(\beta|y, h, \lambda)$  a  $p(h|y, \beta, \lambda)$  jsou dány v (6.12) a (6.15).<sup>13</sup> V této části se tedy zaměříme na  $p(\lambda|y, \beta, h, \nu_\lambda)$  a  $p(\nu_\lambda|y, \beta, h, \lambda)$ . Odvození podmíněné hustoty pro  $\lambda$  je zřejmé, stačí dosadit apriorní hustotu danou (6.22) do obecného tvaru podmíněné posteriorní hustoty daného vztahem (6.18). Prozkoumáním výsledné hustoty zjistíme, že veličiny  $\lambda_i$  jsou navzájem nezávislé (podmíněné ostatními parametry modelu) a každá z podmíněných posteriorních hustot pro  $\lambda_i$  má tvar gama rozdělení. Formálně tak máme

$$p(\lambda|y, \beta, h, \nu_\lambda) = \prod_{i=1}^N p(\lambda_i|y, \beta, h, \nu_\lambda) \quad (6.24)$$

<sup>13</sup>Formálně by plně podmíněné hustoty pravděpodobnosti měly být  $p(\beta|y, h, \lambda, \nu_\lambda)$  a  $p(h|y, \beta, \lambda, \nu_\lambda)$ . Nicméně za podmínky  $\lambda$  nám  $\nu_\lambda$  nepřináší žádnou novou informaci a tudíž  $p(\beta|y, h, \lambda, \nu_\lambda) = p(\beta|y, h, \lambda)$  a  $p(h|y, \beta, \lambda, \nu_\lambda) = p(h|y, \beta, \lambda)$



a

$$p(\lambda_i|y, \beta, h, \nu_\lambda) = f_G\left(\lambda_i \mid \frac{\nu_\lambda + 1}{h\epsilon_i^2 + \nu_\lambda}, \nu_\lambda + 1\right) \quad (6.25)$$

Poznamenejme, že za podmínky, kdy známé  $\beta$  (což předpokládá podmíněná pravděpodobnost), jsme schopni náhodnou složku  $\epsilon_i$  spočítat, tudíž parametry gamma rozdělení v předchozím výrazu lze spočítat v rámci Gibbsova vzorkovače.

Doposud jsme se nezmínili o apriorní hustotě pro  $\nu_\lambda$ , jejíž přesná podoba není relevantní pro podmíněné posteriorní hustoty ostatních parametrů. Podoba  $p(\nu_\lambda)$  samozřejmě ovlivní  $p(\nu_\lambda|y, \beta, h, \lambda)$  a je tak nutné ji specifikovat. Jelikož musí platit, že  $\nu_\lambda > 0$ , využijeme jako apriorní hustotu exponenciální rozdělení. Exponenciální rozdělení není nic jiného než gama rozdělení s dvěma stupni volnosti. Můžeme tedy psát

$$p(\nu_\lambda) = f_G(\nu_\lambda|\underline{\nu}_\lambda, 2) \quad (6.26)$$

Můžeme pracovat i s jinými apriorními hustotami, kdy dojde jen k drobným úpravám následujícího algoritmu posteriorní simulace.

Podmíněnou hustotu  $p(\nu_\lambda|y, \beta, h, \lambda)$  je snadné odvodit, jelikož  $\nu_\lambda$  nevstupuje do věrohodnostní funkce, a lze si tak ověřit, že  $p(\nu_\lambda|y, \beta, h, \lambda) = p(\nu_\lambda|\lambda)$ . Z Bayesova teorému vyplývá, že

$$p(\nu_\lambda|\lambda) \propto p(\lambda|\nu_\lambda)p(\nu_\lambda)$$

a jádrovou podmíněnou posteriorní hustotu snadno získáme vynásobením (6.22) a (6.26). Získáme tedy

$$p(\nu_\lambda|y, \beta, h, \lambda) \propto \left(\frac{\nu_\lambda}{2}\right)^{\frac{N\nu_\lambda}{2}} \Gamma\left(\frac{\nu_\lambda}{2}\right)^{-N} \exp(-\eta\nu_\lambda) \quad (6.27)$$

kde

$$\eta = \frac{1}{\underline{\nu}_\lambda} + \frac{1}{2} \sum_{i=1}^N [\ln(\lambda_i^{-1}) + \lambda_i]$$

Nejedná se v tomto případě o žádnou standardní hustotu rozdělení. Pro náhodné výběry z tohoto rozdělení je tak nutné využít Metropolis-Hastings algoritmus. Je však třeba zdůraznit, že Geweke [15] doporučuje jiný užitečný algoritmus zvaný *acceptance sampling*. Tento algoritmus je velmi užitečný v případech, kdy chceme získat náhodné vzorky z nestandardního jednorozměrného rozdělení, které je navíc omezeno resp. ohraničeno. Obecnou diskuzi k této technice nabízí Devroye [8].

Pro řadu hypotéz (např.  $\beta_j = 0$ ) můžeme využít v rámci porovnání modelů Savage-Dickey density ratio. Ne vždy je to však snadné či možné. V mnoha případech nás může zajímat, zda-li zde existuje náznak odchýlení rozdělení náhodné složky od normality. V tomto případě bychom rádi porovnali  $M_1 : \nu_\lambda \rightarrow \infty$  a  $M_2 : \nu_\lambda$  jako konečné číslo. Tyto modely není vůbec snadné dát do kontextu vnořených modelů, pro které je jako dělaný Savage-Dickeyho poměr hustot. Bayesův faktor (porovnávající tyto dva modely) lze však vypočítat za použití

metody Gelfanda-Deye. To vyžaduje posteriorní simulátor pro každý z těchto modelů, tj. posteriorní simulátor z kapitoly 4 části 4.1 pro model  $M_1$  a posteriorní simulátor popisovaný na tomto místě pro model  $M_2$ . Alternativně lze vypočítat prediktivní  $p$ -hodnoty nebo HPDI pro analýzu kvality a vhodnosti modelu. Predikční analýza probíhá standardním způsobem popsáním v kapitole 4.

## 6.4.2 Empirická ilustrace

BUDE DOPLNĚNO!!!

## 6.5 Autokorelace náhodných složek

Řada proměnných (časových řad) je korelována v čase díky faktorům jakými jsou setrvačnost v preferencích či čas potřebný k přizpůsobení se novým podmínkám. Tato korelace mezi hodnotami jedné proměnné v různých časech (proto hovoříme o autokorelaci) se může projevovat skrze chybový člen. Je tedy žádoucí předpokládat podobu kovarianční matice, která bude tuto skutečnost reflektovat. V předchozích kapitolách jsme předpokládali, že  $\epsilon$  je  $N(0_N, h^{-1}I_N)$ . Tento předpoklad jsme v předchozích oddílech uvolnili předpokladem, že kovarianční matice chyb je diagonální. I v tomto případě jsme předpokládali, že chybové členy jsou vzájemně nekorelované, tedy platí  $E(\epsilon_i, \epsilon_j) = 0$  pro  $i \neq j$ . Nyní tento předpoklad opustíme.

Indexem  $t$  budeme označovat čas, tedy  $y_t$  pro  $t = 1 \dots, T$  bude označovat pozorování závisle proměnné pro období 1 až  $T$  (např. roční pozorování hrubého domácího produktu). Jednoduchým způsobem, jak dovolit autokorelaci chyb, je předpoklad, že jsou generovány *autoregresním procesem řádu 1* resp. *AR(1) procesem*:

$$\epsilon_t = \rho\epsilon_{t-1} + u_t m \quad (6.28)$$

kde  $u_t$  je i.i.d.  $N(0, h^{-1})$ . Tato specifikace umožňuje aby chybový člen v jednom období závisel na členu v předchozím období.

Literatura zabývající se problematikou časových řad uvádí řadu pojmů, nástrojů a technik, které napomáhají formálnějšímu pochopení vlastností různých časových řad. Zde si uvedeme některé z nich, přičemž budeme uvádět obecné označení časové řady jako  $z_t$ . Bayesovský přístup k časovým řadám analyzuje Bauwens, Lubrano a Richard [1], nebayesovský pak Enders [10]. V této části budeme předpokládat, že  $z_t = \epsilon_t$ . Standardně se předpokládá, že proces generující časovou řadu probíhá od období  $-\infty$  do  $\infty$ . My pak pozorujeme tento proces pro období  $t = 1 \dots, T$ . O řadě  $z_t$  řekneme, že je *kovariančně stacionární* (či *slabě stacionární*), pokud pro každé  $t$  a  $s$  platí:

$$\begin{aligned} E(z_t) &= E(z_{t-s}) = \mu, \\ \text{var}(z_t) &= \text{var}(z_{t-s}) = \gamma_0, \\ \text{cov}(z_t, z_{t-s}) &= \gamma_s, \end{aligned}$$

kde  $\mu$ ,  $\gamma_0$  a  $\gamma_s$  jsou všechno konečné hodnoty. O časové řadě řekneme, že je kovariančně či slabě stacionární, pokud má konstantní střední hodnotu a rozptyl a kovariance mezi jakýmkoli dvěma pozorováními závisí pouze na počtu období, který je mezi nimi. Většina časových řad tuto vlastnost má a pokud ne, lze je obvykle *diferencovat* na stacionární řady. *První diference*  $z_t$  se označuje jako  $\Delta z_t$  a je definována jako

$$\Delta z_t = z_t - z_{t-1}$$

Podobně můžeme definovat diferenci  $m$ -tého řádu pro  $m > 1$  jako

$$\Delta^m z_t = \Delta^{m-1} z_t - \Delta^{m-1} z_{t-1}$$

Abychom pochopili ekonomický význam diferencování, předpokládejme, že  $z_t$  je logaritmus cenové hladiny. Potom  $\Delta z_t$  je (aproximativně) procentní změna cen, tedy inflace. Druhá diference  $\Delta^2 z_t$  bude procentní změna v míře inflace. Práce z logaritmy veličin a jejich diferencemi je v makroekonomických modelech běžná.

Obvyklým nástrojem pro analýzu vlastností stacionárních časových řad je  $\gamma_s$ , nebo-li tzv. *autokovariační funkce*. S ní je úzce svázána *autokorelační funkce*, která počítá korelace mezi pozorováními vzdálenými od sebe  $s$  období (tj. je definována jako  $\frac{\gamma_s}{\gamma_0}$  pro  $s = 0, \dots, \infty$ ). Jedná se o funkce proměnné  $s$  a je obvykle vhodné vykreslit si některou z nich pro analýzu jejich vývoje při zvyšujícím se  $s$ . Například pro makroekonomické proměnné je autokorelační funkce klesající v  $s$ , neboť události poslední doby mají obvykle větší dopad na současnou makroekonomickou situaci než události starších období.

Nyní se vrátíme k našemu AR(1) procesu. Pro analýzu jeho vlastností je obvyklé zapsat  $\epsilon_t$  jako funkci  $u_{t-s}$  pro  $s = 0, \dots, \infty$ . To můžeme udělat pomocí zápisu  $\epsilon_{t-1} = \rho\epsilon_{t-2} + u_{t-1}$  a substitucí tohoto výrazu do (6.28), tím dostáváme

$$\epsilon_t = \rho^2\epsilon_{t-2} + \rho u_{t-1} + u_t$$

Pokud dále provedem substituci pro  $\epsilon_{t-2}$  dostaneme rovnici zahrnující  $\epsilon_{t-3}$ . Dalšími substitucemi získáme výraz

$$\epsilon_t = \sum_{s=0}^{\infty} \rho^s u_{t-s} \quad (6.29)$$

Při tomto zápisu si můžeme všimnout, že může nastat problém při výpočtu střední hodnoty, rozptylu a kovariance  $\epsilon_t$ , neboť  $\rho^s$  bude nekonečno pro  $|\rho| > 1$ . Rovněž pro  $\rho = 1$  se bude jednat o nekonečnou sumu konečných členů. Ve skutečnosti je podmínka  $|\rho| < 1$  nutnou podmínkou pro stacionaritu časové řady.

Při podmínce  $|\rho| < 1$  lze ověřit, že  $E(\epsilon_t) = 0$ ,

$$\gamma_0 = \text{var}(\epsilon_t) = h^{-1} \sum_{s=0}^{\infty} \rho^{2s} = \frac{1}{h(1-\rho^2)},$$

$$\gamma_s = \text{cov}(\epsilon_t, \epsilon_{t-s}) = \frac{\rho^s}{h(1-\rho^2)}.$$

Poznamenejme, že pro  $|\rho| < 1$  autokovarianční funkce  $\gamma_s$  klesá s rostoucím  $s$ . Intuitivně se tak v rámci AR(1) postupně vytrácí vliv minulých hodnot na hodnotu současnou.

Tyto poznatky lze využít pro zápis kovarianční matice pro  $\epsilon$  jako  $h^{-1}\Omega$ , kde

$$\Omega = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdot & \rho^{T-1} \\ \rho & 1 & \rho & \cdot & \cdot \\ \rho^2 & \rho & \cdot & \cdot & \rho^2 \\ \cdot & \cdot & \cdot & \cdot & \rho \\ \rho^{T-1} & \cdot & \rho^2 & \rho & 1 \end{bmatrix} \quad (6.30)$$

AR(1) model lze rozšířit i o vliv starších období resp. zpoždění (*lags*) vyššího řádu. Můžeme tedy definovat autoregresní proces řádu  $p$  nebo-li *AR(p) proces* jako

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \dots + \rho_p \epsilon_{t-p} + u_t \quad (6.31)$$

Výše uvedené metody lze pak využít pro výpočet střední hodnoty, rozptylu a autokovarianční funkce. Jak dále uvidíme, pro bayesiánskou analýzu procesu *AR(p)* nebude nutné znát přesnou podobu autokovarianční funkce. Nebudeme ji zde tedy odvozovat. Stačí jen poznamenat, že proces *AR(p)* má podobné vlastnosti jako AR(1), ale je více flexibilnější.

Zavedeme si ještě tzv. *operátor zpoždění*, označovaný jako  $L$ . Jeho vlastností je to, že  $L\epsilon_t = \epsilon_{t-1}$  nebo obecněji  $L^m \epsilon_t = \epsilon_{t-m}$ . AR(p) proces je možno zapsat jako

$$(1 - \rho_1 L - \dots - \rho_p L^p) \epsilon_t = u_t$$

nebo

$$\rho(L) \epsilon_t = u_t$$

kde  $\rho(L) = (1 - \rho_1 L - \dots - \rho_p L^p)$  je polynom řádu  $p$  pro operátor zpoždění. Lze ověřit, že proces *AR(p)* je stacionární, pokud kořeny rovnice  $\rho(z) = 0$  jsou všechny v absolutní hodnotě větší než jedna. Pro další potřeby definujeme  $\rho = (\rho_1, \dots, \rho_p)'$  a nechť  $\Phi$  označuje stacionární oblast tohoto modelu.

### 6.5.1 Bayesovský výpočet

Posteriorní simulátor, který nám umožní bayesovskou analýzu NLRM s chybami generovanými *AR(p)* procesem, dokážeme odvodit ze vztahů platných pro obecnou nespécifikovanou matici  $\Omega$ , které jsou dány v (6.12), (6.15) a v (6.18). Pokud provedeme jednu specifickou aproximaci, budou mít tyto podmíněné posteriorní hustoty jednoduchou formu. Tato aproximace v sobě zahrnuje úpravu resp. ošetření počátečních podmínek. Než se však dostaneme k tomu, co se za touto aproximací a úpravou skrývá, je dobré se zamyslet nad tím, jak by bylo možné transformovat model podobně jako v případě (6.1). To můžeme provést tak, že specifikujeme přesný tvar  $\Omega$  za předpokladu přítomnosti chyb v podobě *AR(p)* procesu, a odvodíme matici  $P$  takovou, pro kterou platí  $P\Omega P' = I$ . Pro alternativní postup si zapíšeme regresní model jako

$$y_t = x_t' \beta + \epsilon_t \quad (6.32)$$

kde  $x_t = (1, x_{t2}, \dots, x_{tk})'$ . Přenásobením obou stran rovnice výrazem  $\rho(L)$  a zavedením  $y_t^* = \rho(L)y_t$  a  $x_t^* = \rho(L)x_t$  získáváme

$$y_t^* = x_t^{*'}\beta + u_t. \quad (6.33)$$

Předpokládali jsme, že  $u_t$  je i.i.d.  $N(0, h^{-1})$  a tedy transformovaný model je normální lineární regresní model s i.i.d. chybami. Poznamenejme ovšem, co se stane s těmito transformovanými hodnotami pro  $t \leq p$ . Například  $y_1^*$  závisí na  $y_0, \dots, y_{1-p}$ . Jelikož naše data máme naměřená pro  $t = 1, \dots, T$ , tyto tzv. počáteční podmínky  $y_0, \dots, y_{1-p}$  nejsou pozorovány. Ošetření počátečních podmínek je choulostivá otázka, zejména v případě, kdy AR proces není stacionární (či se nestacionárnímu procesu blíží). Bližší podrobnosti nabízí Bauwens, Lubrano a Richard [1]. V našem případě budeme předpokládat stacionaritu chyb, tudíž ošetření počátečních podmínek nebude mít takovou důležitost. Obvyklým způsobem řešení tohoto problému je práce s věrohodnostní funkcí založenou spíše na datech od  $t = p + 1, \dots, T$  než  $t = 1, \dots, T$ . Pokud je  $p$  relativně malé vzhledem k  $T$ , bude výsledná aproximativní věrohodnostní funkce velmi blízká skutečné věrohodnostní funkci. Protože  $y_t^*$  a  $x_t^*$  pro  $t = p + 1, \dots, T$  nezávisí na nepozorovaných zpožděných hodnotách, lze transformaci danou v (6.33) velmi snadno provést.

Pro jednoduchost značení nebudeme zavádět speciální označení pro věrohodnostní funkci, posteriorní hustotu apod. vycházející z dat od  $t = p + 1, \dots, T$ . Pro zbytek tohoto oddílu budeme interpretovat  $y, y^*, \epsilon$  a  $\epsilon^*$  jako vektory rozměru  $T - p$ , tedy odstranili jsme prvních  $p$  pozorování. Matice  $X$  a  $X^*$  budou rozměrů  $(T - p) \times k$ . Za těchto předpokladů lze odvodit Gibbsův vzorkovač s využitím výsledků předchozích oddílů. Intuitivně,  $p(\beta|y, h, \rho)$  a  $p(h|y, \beta, \rho)$  jsou dány v (6.12) a (6.15). Podmíněnou hustotu  $p(\rho|y, \beta, h)$  lze odvodit, pokud si uvědomíme, že za podmínky  $\beta$  a  $h$  je  $\epsilon_t$  pro  $t = p + 1, \dots, T$  známé a (6.31) je normální lineární regresní model (se známým rozptylem chyb) a s koeficienty danými vektorem  $\rho$ . Pomocí výsledků bayesiánské analýzy z předchozích kapitol jsme tak schopni odvodit  $p(\rho|y, \beta, h)$ .

Využijeme-li tedy nezávislou normální-gama apriorní hustotu pro  $\beta$  a  $h$  danou v (6.9) a (6.10), lze výsledky z úvodního oddílu 6.2 modifikovat pro naše potřeby následujícím způsobem:

$$\beta|y, h, \rho \sim N(\bar{\beta}, \bar{V}) \quad (6.34)$$

kde

$$\bar{V} = (\underline{V}^{-1} + hX^{*'}X^*)^{-1} \quad (6.35)$$

a

$$\bar{\beta} = \bar{V}(\underline{V}^{-1}\underline{\beta} + hX^{*'}y^*) \quad (6.36)$$

Posteriorní hustota pro  $h$  podmíněná ostatními parametry modelu odpovídá gama rozdělení:

$$h|y, \beta, \rho \sim G(\bar{s}^{-2}, \bar{\nu}) \quad (6.37)$$

kde

$$\bar{\nu} = T - p + \nu \quad (6.38)$$

a

$$\bar{s}^2 = \frac{(y^* - X^*\beta)'(y^* - X^*\beta) + \nu s^2}{\bar{\nu}} \quad (6.39)$$

Posteriorní hustota pro  $\rho$  závisí na své apriorní hustotě, která může reflektovat jakýmkoli způsobem naši nedatovou informaci. Zde budeme předpokládat, že se jedná o vícerozměrné normální rozdělení omezené na stacionární oblast. Tedy,

$$p(\rho) \propto f_N(\rho|\underline{\rho}, \underline{V}_\rho)1(\rho \in \Phi) \quad (6.40)$$

kde  $1(\rho \in \Phi)$  je indikační funkce, která je rovna jedné pro stacionární oblast a nula jinak. S touto apriorní hustotou je snadné odvodit podmíněnou posteriorní hustotu

$$p(\rho|y, \beta, h) \propto f_N(\rho|\bar{\rho}, \bar{V}_\rho)1(\rho \in \Phi) \quad (6.41)$$

kde

$$\bar{V}_\rho = (\underline{V}_\rho^{-1} + hE'E)^{-1}, \quad (6.42)$$

$$\bar{\rho} = \bar{V}_\rho(\underline{V}_\rho^{-1}\underline{\rho} + hE'\epsilon) \quad (6.43)$$

a  $E$  je matice rozměru  $(T - p) \times p$  s  $t$ -tým řádkem daným jako  $\epsilon_{t-1}, \dots, \epsilon_{t-p}$ .

Gibbsův vzorkovač v sobě zahrnuje sekvenční výběry z (6.34), (6.37) a (6.41). Skutečnost, že podmíněná posteriorní hustota pro  $\rho$  je omezené vícerozměrné normální rozdělení, přináší drobnou komplikaci. Ovšem výběry z omezeného rozdělení lze provést pomocí výběrů z neomezeného rozdělení a jednoduchým vypuštěním vzorků, které spadají mimo stacionární oblast. Pokud  $\bar{\rho}$  leží uvnitř této oblasti (nebo alespoň nepříliš daleko od ní), bude tento postup více než úspěšný. Alternativně lze odvodit Metropolis-Hastings algoritmus.

Pro naše potřeby jsme schopni provést i predikční analýzu standardním způsobem. Pro ověření kvality modelu můžeme spočítat predikční  $p$ -hodnoty nebo HPDI. Bayesův faktor pro ověření jakýchkoli hypotéz, které nás zajímají, můžeme vypočítat použitím Savage-Dickeyho poměru hustot nebo za pomoci metody Gelfanda-Deye. Je však třeba připomenout, že Savage-Dickey density ratio vyžaduje znalost celých funkcí hustoty pravděpodobnosti (ne jen jádrových hustot)  $p(\rho|y, \beta, h)$  a  $p(\rho|y)$ . Pro  $p = 1$  lze integrační konstantu snadno vypočítat, neboť  $p(\rho|y, \beta, h)$  odpovídá jednorozměrnému omezenému normálnímu rozdělení a vlastnosti tohoto rozdělení jsou dobře známy (viz Poirier [26], strana 115). Pro  $p > 1$  je však stacionární oblast nelineární a analytické vyjádření  $p(\rho|y, \beta, h)$  je velmi obtížné. Ovšem snadno lze k aproximativnímu výpočtu integrační konstanty využít posteriorní simulaci. Hustota odpovídající (6.41) je

$$p(\rho|y, \beta, h) = \frac{f_N(\rho|\bar{\rho}, \bar{V}_\rho)1(\rho \in \Phi)}{\int_{\Phi} f_N(\rho|\bar{\rho}, \bar{V}_\rho)d\rho}$$

Běžný posteriorní simulátor umožňuje získávat vzorky z  $f_N(\rho|\bar{\rho}, \bar{V}_\rho)$  a vyhodit ty výběry, které jsou mimo stacionární oblast. Ovšem  $\int_{\Phi} f_N(\rho|\bar{\rho}, \bar{V}_\rho)d\rho$  není nic jiného než podíl vzorků (z celkového počtu), který nám zůstane. V rámci Gibbsova vzorkovače tak v každém běhu můžeme evidovat počet pokusů potřebných

k získání akceptovatelného výběru. Aproximací výrazu  $1 - \int_{\Phi} f_N(\rho|\bar{\rho}, \bar{V}_\rho) d\rho$  je tedy podíl celkového množství zamítnutých vzorků na celkovém počtu vzorků. Obecně lze integrační konstantu jakékoliv omezené hustoty nalézt za pomoci výběrů z neomezené verze této hustoty a spočítáním podílu vzorků nacházejících se v oblasti dané příslušným omezením. V závislosti na zvolené apriorní hustotě je použití této strategie nutné i pro výpočet integrační konstanty apriorní hustoty.

### 6.5.2 Empirická ilustrace

BUDE DOPLNĚNO!

## 6.6 Modely zdánlivě nesouvisejících regresí

Posledním modelem v této kapitole je tzv. „Seemingly Unrelated Regressions“ (SUR) model. Tedy model zdánlivě nesouvisejících regresí. Jedná se vícerovnícový model zajímavý sám o sobě. V ekonomii jsou vícerovnícové modely běžné v řadě kontextů. Například při analýze spotřeby nás může zajímat odhad rovnic pro každou z kategorií spotřeby (tj. potraviny, výrobky dlouhodobé spotřeby apod.). V mikroekonomických aplikacích se můžeme pokoušet o odhad rovnic poptávky po výrobních faktorech, a to pro každý výrobní faktor v produkční funkci.<sup>14</sup> V mnoha případech se nedopustíme příliš velké chyby, budeme-li pracovat s jednotlivými rovnicemi samostatně, čímž můžeme využít techniky odvozené v předchozích kapitolách. Ovšem práce se všemi modely najednou může vylepšit výsledky odhadů.

SUR model můžeme zapsat jako

$$y_{mi} = x_{mi}'\beta_m + \epsilon_{mi} \quad (6.44)$$

kde  $i = 1, \dots, N$  označuje  $N$  pozorování pro  $m = 1, \dots, M$  rovnic,  $y_{mi}$  tedy označuje  $i$ -té pozorování závisle proměnné v rovnici  $m$ ,  $x_{mi}$  je  $k_m$ -rozměrný vektor obsahující  $i$ -té pozorování vektoru vysvětlujících proměnných v  $m$ -té rovnici a nakonec  $\beta_m$  je  $k_m$ -rozměrný vektor regresních koeficientů pro  $m$ -tou rovnici.<sup>15</sup> Je třeba zdůraznit, že takovéto označení umožňuje, aby se počet vysvětlujících proměnných v rámci jednotlivých rovnic lišil, nicméně některé či všechny vysvětlující proměnné mohou být v některých rovnicích totožné.

Tento SUR model můžeme přepsat do obvyklejšího maticového vyjádření.

<sup>14</sup>V rámci klasické ekonometrie je druhem SUR modelu redukováný tvar modelu simultánních rovnic. Podobně i v analýze časových řad patří tzv. vektorový autoregresní model (VAR model) do skupiny SUR modelů.

<sup>15</sup>Je třeba upozornit, že se trochu změnilo značení. V tomto oddíle je  $x_{mi}$  vektor, kdy první index označuje číslo rovnice. V předchozích kapitolách označovalo  $x_{ij}$   $i$ -té pozorování pro  $j$ -tou vysvětlující proměnnou.

Stačí zapsat všechny rovnice podle pozorování do vektorů a matic jako

$$y_i = \begin{pmatrix} y_{1i} \\ \vdots \\ y_{Mi} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix} \quad X_i = \begin{pmatrix} x'_{1i} & 0 & \cdot & \cdot & 0 \\ 0 & x'_{2i} & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & x'_{Mi} \end{pmatrix}$$

a definovat  $k = \sum_{m=1}^M k_m$ . Díky tomuto značení lze ověřit, že (6.44) lze přepsat do podoby

$$y_i = X_i \beta + \epsilon_i. \quad (6.45)$$

Nyní stačí dát všechna pozorování dohromady

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$$

a psát

$$y = X\beta + \epsilon$$

SUR model tak lze zapsat jako nám dobře známý lineární regresní model.

Pokud bychom předpokládali, že  $\epsilon_{mi}$  je i.i.d.  $N(0, h^{-1})$  pro všechna  $i$  a  $m$ , měli bychom skutečně normální lineární regresní model, analyzovaný v předchozích kapitolách. V mnoha případech je však obvyklé předpokládat, že náhodné složky jsou korelovány v rámci pozorování, tedy předpokládáme, že  $\epsilon_i$  je i.i.d.  $N(0, H^{-1})$  pro  $i = 1, \dots, N$ , kde  $H$  je matice přesností chyb rozměru  $M \times M$ . S tímto předpokladem lze ukázat, že  $\epsilon$  je z  $N(0, \Omega)$ , kde  $\Omega$  je blokově-diagonální matice rozměru  $NM \times NM$ :

$$\Omega = \begin{pmatrix} H^{-1} & 0 & \cdot & \cdot & 0 \\ 0 & H^{-1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & H^{-1} \end{pmatrix}. \quad (6.46)$$

SUR model tak náleží do třídy modelů analyzovaných v této kapitole. Jediným rozdílem je to, že zde nevystupuje  $h$ . To nepřináší žádné podstatné rozdíly, neboť  $h$  byl skalár, který byl vytknut v předchozích oddílech jen kvůli zvyklosti. Pro tento typ modelů toto vytknutí obvyklé není (nicméně pokud bychom chtěli, mohli bychom to udělat).

### 6.6.1 Apriorní hustota

V této části si rozšíříme nezávislé normální-gama rozdělení do podoby nezávislého normálního-Wishartova rozdělení:

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



kde

$$p(\beta) = f_N(\beta | \underline{\beta}, \underline{V}) \quad (6.47)$$

a

$$p(H) = f_W(H | \underline{\nu}, \underline{H}) \quad (6.48)$$

Wishartovo rozdělení je maticovým zobecněním gama rozdělení. Pro analýzu apriorní hustoty je důležité poznamenat, že  $E(H) = \underline{\nu} \underline{H}$  a neinformativnosti lze docílit volbou  $\underline{\nu} = 0$  a  $\underline{H}^{-1} = 0_{M \times M}$ .

Pro tento model existuje i řada jiných apriorních hustot. Zjeměna je to přirozeně konjugovaná apriorní normální-Wishartova hustota. Pro tuto apriorní hustotu je možno získat analytické výsledky a není zde nutné využití posteriorního simulátoru. Ovšem tento druh hustoty byl shledán jako příliš restriktivní. Jeho implikací je například to, že apriorní kovariance mezi koeficienty v každé dvojici rovnic (tj.  $\beta_m$  a  $\beta_j$  pro  $j \neq m$ ) jsou všechny proporcionální téže matici. Z toho odůvodu se v empirických pracích objevuje pouze neinformativní varianta přirozeně konjugované apriorní hustoty. Rovněž byla učiněna řada pokusů pro odvození méně restriktivních přirozeně konjugovaných hustot.

### 6.6.2 Bayesovský výpočet

Bayesovský výpočet pro tento model lze provést Gibbsovým vzorkovačem při využití vztahů pro podmíněnou posteriorní hustotu (6.12) a (6.18) založených na apriorních hustotách (6.47) a (6.48). Obě z těchto podmíněných posteriorních hustot zahrnují inverzi  $NM \times NM$ -rozměrné matice  $\Omega$ , což je výpočetně náročné. Ovšem bloková struktura matice  $\Omega$  umožňuje provést tuto inverzi částečně analyticky. Pokud to učiníme,  $p(\beta | y, H)$  a  $p(H | y, \beta)$  budou mít obvyklou podobu, tedy:

$$\beta | y, H \sim N(\bar{\beta}, \bar{V}) \quad (6.49)$$

kde

$$\bar{V} = \left( \underline{V}^{-1} + \sum_{i=1}^N X_i' H X_i \right)^{-1} \quad (6.50)$$

a

$$\bar{\beta} = \bar{V} \left( \underline{V}^{-1} \underline{\beta} + \sum_{i=1}^N X_i' H y_i \right) \quad (6.51)$$

Podmíněná posteriorní hustota pro  $H$  odpovídá Wishartovu rozdělení:

$$H | y, \beta \sim W(\bar{\nu}, \bar{H}) \quad (6.52)$$

kde

$$\bar{\nu} = N + \underline{\nu} \quad (6.53)$$

a

$$\bar{H} = \left[ \underline{H}^{-1} + \sum_{i=1}^N (y_i - X_i \beta)(y_i - X_i \beta)' \right]^{-1} \quad (6.54)$$

Protože generátory náhodných čísel z Wishartova rozdělení jsou k dispozici (např. varianta pro Matlab je součástí LeSageho Ekonometrického toolboxu), lze snadno implementovat Gibbsův vzorkovač, který bude generovat vzorky z  $p(\beta|y, H)$  a  $p(H|y, \beta)$ .

Predikční analýza bude probíhat standardním způsobem. Rovněž tak lze ověřit kvalitu modelu pomocí posteriorní predikční  $p$ -hodnoty a intervalů nejvyšší posteriorní hustoty. Pro výpočet posteriorního podílu šancí můžeme využít Savage-Dickey density ratio.

### **6.6.3 Empirická ilustrace**

BUDE DOPLNĚNO!

## **6.7 Shrnutí**

BUDE DOPLNĚNO!

## Kapitola 7

# Lineární regresní model s panelovými daty

### 7.1 Úvod

V předchozích kapitolách jsme předpokládali případy, kdy jedna datová položka byla dostupná pro každou pozorovatelnou jednotku. Například,  $y_i$  byl skalár, který obsahoval jediné pozorování závisle proměnné. V ekonomii (a nejen v ní) však existuje řada případů, kdy pro jednu proměnnou máme několik pozorování. Mikroekonomickým příkladem by tak mohl být případ produkce firem, kdy můžeme mít údaje o vstupech a výstupech za několik let, a to pro více firem. V literatuře zabývající se problémem ekonomického růstu se často pracuje s modely, které mají časovou i prostorovou dimenzi (časovou řadu údaje pro velkou množinu zemí). V rámci finančního modelování lze například pracovat s cenami akcií více firem obchodovaných v několika dnech čiměsících. V oblasti marketingu můžeme disponovat daty o nákupech jednotlivých zákazníků realizovaných při jejich návštěvách v jednom obchodě. Všechny tyto příklady jsou charakteristické dostupností  $T$  pozorování pro každého z  $N$  jednotlivců či firem (nebo jiných jednotek). V ekonometrii je tento typ dat nazýván *panelovými daty*, ačkoli ve statistické literatuře nesou název *longitudinální data*.

V této kapitole budou diskutovány modely a metody bayesovské analýzy právě pro tento typ dat. Tato kapitola nepřináší žádné nové výpočetní techniky, naopak jsou zde kombinovány aspekty různých modelů a metod posteriorní simulace z předchozích kapitol. Více se zde zaměříme na koncept hierarchických apriorních hustot. Tato kapitola je strukturována dle předpokladů kladených na regresní koeficienty. Začneme předpokladem, že regresní koeficienty jsou stejné pro všechny jednotlivce (souhrnný model – *pooled model*). Potom tento předpoklad uvolníme a dovolíme, aby se úrovněová konstanta mohla mezi jednotlivci lišit (modely individuálních vlivů – *individual effects models*). Uvolnění všech koeficientů pak je představováno modelem náhodných koeficientů (*random coefficients model*). Podrobněji je analyzován speciální případ modelu individuálních

vlivů, tzv. *stochastic frontier model*. Zavedeme si rovněž novou metodu výpočtu marginální věrohodnosti, nazývanou dle svého autora Chibova metoda [5]. Tato metoda je aplikovatelná pokud je posteriorní simulace prováděna Gibbsovým vzorkovačem a je užitečná v případech vysoké dimenzionality parametrického prostoru.

Na úvod si ujasníme značení v této kapitole. Nechť  $y_{it}$  a  $\epsilon_{it}$  označuje  $t$ -té pozorování (pro  $t = 1, \dots, T$ ) závisle proměnné a náhodné složky pro  $i$ -tého jednotlivce, kde  $i = 1, \dots, N$ . Vektory  $y_i$  a  $\epsilon_i$  tedy budou obsahovat  $T$  pozorování závisle proměnné a chyb pro  $i$ -tého jednotlivce. V některých regresních modelech budeme rozlišovat mezi úroňovou konstantou a ostatními koeficienty. Budeme tedy definovat  $X_i$  jakožto matici rozměru  $T \times k$  obsahující  $T$  pozorování každé z  $k$  vysvětlujících proměnných (včetně úroňové konstanty) pro  $i$ -tého jednotlivce. Matice  $\tilde{X}_i$  bude matice rozměru  $T \times (k - 1)$  odpovídající matici  $X_i$  bez sloupce odpovídajícímu úroňové konstantě. Tedy,  $X_i = [\iota_T \tilde{X}_i]$ . Pokud dáme pozorování pro  $N$  jednotlivců dohromady, získáváme  $TN$ -rozměrné vektory:

$$y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \epsilon_N \end{bmatrix}$$

Podobně i agregace všech pozorování vysvětlujících proměnných vede k matici rozměru  $TN \times K$ :

$$X = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ X_N \end{bmatrix}$$

## 7.2 Souhrnný model (pooled model)

V rámci tohoto modelu předpokládáme že stejný regresní vztah platí pro všechny jednotlivce, tedy

$$y_i = X_i \beta + \epsilon_i \quad (7.1)$$

pro  $i = 1, \dots, N$ , kde  $\beta$  je  $k$ -rozměrný vektor regresních koeficientů včetně úroňové konstanty. Tvar věrohodnostní funkce závisí na předpokladech kladených na náhodnou složku. V této kapitole budeme předpokládat pro  $i, j = 1, \dots, N$ :

1.  $\epsilon_i$  má vícerozměrné normální rozdělení se střední hodnotou  $0_T$  a kovarianční maticí  $h^{-1}I_T$ .
2.  $\epsilon_i$  a  $\epsilon_j$  jsou nezávislé pro  $i \neq j$ .
3. Všechny prvky  $X_i$  jsou pevná čísla (tj. nenáhodné veličiny) nebo v případě, že jsou náhodnými veličinami, jsou nezávislé na všech prvcích  $\epsilon_j$  a mají hustotu pravděpodobnosti  $p(X_i|\lambda)$  kde  $\lambda$  je vektor parametrů, který neobsahuje  $\beta$  ani  $h$ .

Tyto předpoklady jsou podobné předpokladům definovaným v předchozích kapitolách. Poznamenejme jen, že předpokládáme, že  $\epsilon_{it}$  a  $\epsilon_{is}$  jsou vzájemně nezávislé pro  $t \neq s$ . Předpoklad nezávislosti chyb pozorování pro jednoho jednotlivce v čase nemusí být vždy rozumný. Můžeme tedy obecněji předpokládat, že  $\epsilon_i$  má kovarianční matici  $\Omega$ . Tento případ lze pak řešit podobným způsobem jako SUR model v předchozí kapitole.

Předpoklad nezávislosti chyb v čase i mezi jednotlivci navzájem redukuje tento model na lineární regresní model diskutovaný v kapitolách 2, 3 a 4. Data za všechny jednotlivce jsou tak shrnuty do jedné velké regrese.

Předchozí předpoklady tedy implikují věrohodnostní funkci v podobě:

$$p(y|\beta, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i\beta)' (y_i - X_i\beta) \right] \right\}$$

Což lze přepsat do podoby

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

což je věrohodnostní funkce z kapitol 3 a 4 pro  $TN$  pozorování. V rámci empirické ilustrace je využita nezávislá apriorní normální-gama hustota, kde  $\beta \sim N(\beta, \underline{V})$  a  $h \sim G(\underline{s}^{-2}, \underline{\nu})$ . Lze tedy využít metody Gibbsova vzorkovače z kapitoly 4.

### 7.3 Modely individuálních vlivů (individual effects models)

Souhrnný model nemusí být vždy vhodný a předpoklad stejných koeficientů pro všechny jednotlivce nemusí být rozumný. Předpokládejme například příklad z oblasti marketingu kde  $y_{it}$  je prodej nápoje značky  $i$  v čase  $t$ . Prodeje mohou záviset na vysvětlujících proměnných, které je snadné pozorovat, jako např na ceně. Existují však i nezachytitelné kvantit jako je věrnost značce. Tomu odpovídající model tak může mít podobu:

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it}$$

V našem příkladu by  $x_{it}$  byla cena  $i$ -tého nápoje v čase  $t$ . Skutečnost, že úrovně konstanta se mezi jednotlivými značkami liší ( $\alpha$  má index  $i$ ), může znamenat, že v ní je obsažen právě onen zmiňovaný efekt věrnosti značce. Je tím umožněno, že dvě značky nápojů se stejnou cenou mohou mít rozdílnou očekávanou prodejnost. Tento typ modelů označujeme jako modely individuálních vlivů, kdy  $\alpha_i$  je označováno jako individuální vliv (*individual effect*). Podobná terminologie je i v klasické ekonometrii, kde dva typy modelů individuálních vlivů jsou označovány jako *random effect models* a *fixed effects models*.

### 7.3.1 Věrohodnostní funkce

Věrohodnostní funkce tohoto modelu je založena na regresní rovnici:

$$y_i = \alpha_i \iota_T + \tilde{X}_i \tilde{\beta} + \epsilon_i \quad (7.2)$$

kde je zřejmé, že  $\alpha_i$  označuje úrovnovou konstantu regresní rovnice  $i$ -tého jednotlivce a  $\tilde{\beta}$  je vektorem zbylých koeficientů, o kterých předpokládáme, že jsou stejné pro všechny jednotlivce. Rovnice (7.2) spolu s předpoklady na náhodnou složku formulovanými za rovnici (7.1) implikuje věrohodnostní funkci v podobě:

$$p(y|\alpha, \tilde{\beta}, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) \right] \right\} \quad (7.3)$$

přičemž  $\alpha = (\alpha_1, \dots, \alpha_N)'$ .

### 7.3.2 Apriorní hustota

V rámci bayesovské analýzy lze použít jakýkoliv druh apriorní hustoty, včetně neinformativních. Zde budeme předpokládat dva typy apriorních hustot.

#### Nehierarchická apriorní hustota

Regresní model v (7.2) lze zapsat jako:

$$y = X^* \beta^* + \epsilon \quad (7.4)$$

kde  $X^*$  je matice rozměru  $TN \times (N + k - 1)$  dána jako

$$X^* = \begin{bmatrix} \iota_T & 0_T & \cdot & \cdot & 0_T & \tilde{X}_1 \\ 0_T & \iota_T & \cdot & \cdot & \cdot & \tilde{X}_2 \\ \cdot & 0_T & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0_T & \cdot \\ 0_T & \cdot & \cdot & \cdot & \iota_T & \tilde{X}_N \end{bmatrix}$$

a

$$\beta^* = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_N \\ \tilde{\beta} \end{bmatrix}$$

Tento způsob zápisu jasně ukazuje, že individual effects model lze psát jako regresní model, který byl řešen v kapitole 3 a 4 a jakákoukoli apriorní hustotu zavedenou v těchto kapitolách lze použít i pro  $(\beta^*, h)$ . Pro srovnání s klasickou ekonometrií, tento nehierarchický prior vede k modelu analogickému fixed effects modelu. Matice  $X^*$  totiž obsahuje vysvětlující proměnné připojené k matici obsahující umělou (dummy) proměnnou pro každého jednotlivce.

V této kapitole využijeme nezávislou normální-gama apriorní hustotu, tedy

$$\beta^* \sim N(\underline{\beta}^*, \underline{V}) \quad (7.5)$$

a

$$h \sim G(\underline{s}^{-2}, \underline{\nu}) \quad (7.6)$$

### Hierarchická apriorní hustota

V moderní statistice roste zájem o modely, kde vektor parametrů má velkou dimenzi. Jak bylo řečeno v kapitole 6, oddílu 6.4, hierarchické apriorní hustoty lze využít k řešení problémů způsobených vysokou dimenzí parametrického prostoru. Model individuálních vlivů je modelem s parametrickým prostorem obsahujícím  $N + k$  parametrů (tj.  $N$  úrovnových konstant v  $\alpha$ ,  $k - 1$  koeficientů v  $\tilde{\beta}$  a přesnost chyb  $h$ ). Pokud je  $T$  relativně malé vzhledem k  $N$ , je počet parametrů relativně vysoký vzhledem k velikosti celkového vzorku.<sup>16</sup> To nasvědčuje tomu, že hierarchická apriorní hustota je zcela přiměřená řešenému problému.

Obvykle se předpokládá, že pro  $i = 1, \dots, N$  je

$$\alpha_i \sim N(\mu_\alpha, V_\alpha) \quad (7.7)$$

kde  $\alpha_i$  a  $\alpha_j$  jsou vzájemně nezávislé pro  $i \neq j$ . Hierarchická struktura apriorní hustoty nastává v případě, kdy předpokládáme, že  $\mu_\alpha$  a  $V_\alpha$  jsou neznámé parametry, které vyžadují svou vlastní apriorní hustotu. Předpokládáme, že  $\mu_\alpha$  a  $V_\alpha$  jsou navzájem nezávislé a platí

$$\mu_\alpha \sim N(\underline{\mu}_\alpha, \underline{\sigma}_\alpha^2) \quad (7.8)$$

a

$$V_\alpha^{-1} \sim G(\underline{V}_\alpha^{-1}, \underline{\nu}_\alpha) \quad (7.9)$$

Jak hierarchická, tak i nehierarchická apriorní hustota dovoluje různé úrovnové konstanty pro každého z jednotlivců. Hierarchická struktura díky dodatečnému předpokladu o rozdělení parametrů úrovnové konstanty (je-li v souladu s daty) umožňuje přesnější odhad.

Pro zbylé parametry předpokládáme nehierarchickou apriorní hustotu nezávislého normálního-gama rozdělení:

$$\tilde{\beta} \sim N(\underline{\beta}, \underline{V}_\beta) \quad (7.10)$$

a

$$h \sim G(\underline{s}^{-2}, \underline{\nu}) \quad (7.11)$$

Opět pro srovnání s klasickou ekonometrií, model s takto definovanou hierarchickou apriorní hustotou vede k tzv. random effects modelu.

<sup>16</sup>Modely panelových dat jsou obvykle používány pro data získaná v rámci výběrových šetření. To většinou obsahuje dotazování velkého počtu osob (např.  $N = 10000$ ) co rok či dva (tedy  $T = 5$ ). Otázky se týkají jejich výdajů a spotřebních zvyklostí. V těchto případech je  $N$  nesrovnatelně velké vzhledem k  $T$ .

### 7.3.3 Bayesovský výpočet

#### Posteriorní analýza při nehierarchické apriorní hustotě

Za předpokladu nehierarchické apriorní hustoty dané v (7.5) a (7.6) máme lineární regresní model s nezávislou normální-gama apriorní hustotou. Posteriorní analýza je tak analogické té z kapitoly 4. Můžeme tedy využít Gibbsův vzorkovač na základě znalosti podmíněných hustot:

$$\beta^* | y, h \sim N(\bar{\beta}^*, \bar{V}) \quad (7.12)$$

a

$$h | y, \beta^* \sim G(\bar{s}^{-2}, \bar{v}), \quad (7.13)$$

kde

$$\begin{aligned} \bar{V} &= (\underline{V}^{-1} + hX^{*'}X^*)^{-1}, \\ \bar{\beta}^* &= \bar{V}(\underline{V}^{-1}\underline{\beta}^* + hX^{*'}y), \\ \bar{v} &= TN + \underline{v}, \\ \bar{s}^2 &= \frac{\sum_{i=1}^N (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})'(y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) + \underline{v}S^2}{\bar{v}}. \end{aligned}$$

Analýza konvergence, predikční analýza a porovnání modelů lze provést metodami předchozích kapitol. Je třeba poznamenat, že může nastat numerický problém, pokud je  $N$  příliš velké, neboť  $\bar{V}$  je matice rozměru  $(N + k - 1) \times (N + k - 1)$  a je ji třeba invertovat. Počítačový algoritmus pro maticovou inverzi tak může být nespolehlivý. V takových případech je vhodné využít teorém o inverzi dělené matice, který snižuje dimenzi invertovaných matic (viz příloha A).

#### Posteriorní analýza při hierarchické apriorní hustotě

Odvození posteriorní hustoty pro hierarchickou apriorní hustotu danou v (7.7) až (7.11) není nijak obtížné. Toto odvození v sobě zahrnuje násobení věrohodnostní funkce a apriorních hustot a analýzu výsledného výrazu pro  $\tilde{\beta}$ ,  $h$ ,  $\alpha$ ,  $\mu_\alpha$  a  $V_\alpha$ , abychom tak našli jádrové hustoty pro každou z podmíněných posteriorních hustot. Pak již není nic jednoduššího než využít Gibbsův vzorkovač.

Relevantní posteriorní rozdělení pro  $\tilde{\beta}$  a  $h$  podmíněné veličinou  $\alpha$  jsou odvozeny analogickým způsobem jako v případě lineárního regresního modelu s nezávislou normální-gama apriorní hustotou, tedy<sup>17</sup>

$$\tilde{\beta} | y, h, \alpha, \mu_\alpha, V_\alpha \sim N(\bar{\beta}, \bar{V}_\beta) \quad (7.14)$$

a

$$h | y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha \sim G(\bar{s}^{-2}, \bar{v}), \quad (7.15)$$

<sup>17</sup>Poznamenejme, že rovnice pro  $p(\tilde{\beta} | y, h, \alpha, \mu_\alpha, V_\alpha)$  a  $p(h | y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha)$  nezávisí na  $\mu_\alpha$  a  $V_\alpha$ , a jsou tak ekvivalentní  $p(\tilde{\beta} | y, h, \alpha)$  a  $p(h | y, \tilde{\beta}, \alpha)$ . Úplné značení používáme jen z toho důvodu, abychom zdůraznili, že Gibbsův vzorkovač v sobě zahrnuje výběry z plně podmíněných posteriorních rozdělení.



kde

$$\begin{aligned}\bar{V}_\beta &= \left( \underline{V}_\beta^{-1} + h \sum_{i=1}^N \tilde{X}_i' \tilde{X}_i \right)^{-1}, \\ \bar{\beta} &= \bar{V}_\beta \left( \underline{V}_\beta^{-1} \underline{\beta} + h \sum_{i=1}^N \tilde{X}_i' [y_i - \alpha_i \iota_T] \right), \\ \bar{\nu} &= TN + \underline{\nu}, \\ \bar{s}^2 &= \frac{\sum_{i=1}^N (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) + \underline{\nu} \underline{s}^2}{\bar{\nu}}.\end{aligned}$$

Podmíněná posteriorní hustota pro každé  $\alpha_i$  je nezávislá na  $\alpha_j$  pro  $i \neq j$  je dána jako

$$\alpha_i | y, \tilde{\beta}, h, \mu_\alpha, V_\alpha \sim N(\bar{\alpha}_i, \bar{V}_i), \quad (7.16)$$

kde

$$\begin{aligned}\bar{V}_i &= \frac{V_\alpha h^{-1}}{TV_\alpha + h^{-1}}, \\ \bar{\alpha}_i &= \frac{V_\alpha (y_i - \tilde{X}_i \tilde{\beta})' \iota_T + h^{-1} \mu_\alpha}{TV_\alpha + h^{-1}}.\end{aligned}$$

Nakonec pak podmíněné posteriorní hustoty pro hierarchické parametry  $\mu_\alpha$  a  $V_\alpha$  jsou

$$\mu_\alpha | y, \tilde{\beta}, h, \alpha, V_\alpha \sim N(\bar{\mu}_\alpha, \bar{\sigma}_\alpha^2) \quad (7.17)$$

a

$$V_\alpha^{-1} | y, \tilde{\beta}, h, \alpha, \mu_\alpha, V_\alpha \sim G(\bar{V}_\alpha^{-1}, \bar{\nu}_\alpha), \quad (7.18)$$

kde

$$\begin{aligned}\bar{\sigma}_\alpha^2 &= \frac{V_\alpha \underline{\sigma}_\alpha^2}{V_\alpha + N \underline{\sigma}_\alpha^2} \\ \bar{\mu}_\alpha &= \frac{V_\alpha \underline{\mu}_\alpha + \underline{\sigma}_\alpha^2 \sum_{i=1}^N \alpha_i}{V_\alpha + N \underline{\sigma}_\alpha^2}, \\ \bar{\nu}_\alpha &= \underline{\nu}_\alpha + N, \\ \bar{V}_\alpha &= \frac{\sum_{i=1}^N (\alpha_i - \mu_\alpha)^2 + V_\alpha \underline{\nu}_\alpha}{\bar{\nu}_\alpha}.\end{aligned}$$

Je dobré si všimnout, že Gibbsův vzorkovač vyžaduje v našem případě výběry jen z normálního a gama rozdělení. Ačkoliv tedy vzorce vypadají dosti složitě, není obtížné napsat počítačový skript či program, který bude provádět posteriorní simulaci tohoto modelu. Predikční analýza a konvergenční testy lze opět provést standardním způsobem.

## 7.4 Model náhodných koeficientů

Souhrnný (pooled) model předpokládá, že všem sledovaným jednotkám příslušela tatáž regresní přímka respektive rovina. Model individuálního vlivu pak umožnil to, že sklon této roviny zůstává pro všechny jednotlivce stejný, ale úrovně členy se mohou lišit. Zde prezentovaný model uvolňuje předpoklad o společném sklonu regresní roviny. Je tedy možno psát

$$y_i = X_i\beta_i + \epsilon_i, \quad (7.19)$$

kde  $\beta_i$  je  $k$ -rozměrný vektor regresních koeficientů (včetně úrovněového členu). Tato rovnice platí pro  $i = 1, \dots, N$ , tedy celý model zahrnuje  $Nk + 1$  parametrů (tj.  $k$  regresních koeficientů pro každého z  $N$  jednotlivců a přesnost chyby  $h$ ). Pokud  $T$  není relativně vysoké vzhledem k  $N$ , je velmi obtížné odhadnout všechny parametry modelu s nějakým stupněm přesnosti. Proto je obvyklé využít hierarchických apriorních hustot pro regresní koeficienty. Takovýto model je nazáván modelem náhodných koeficientů (*random coefficients model*).

Jako motivaci jeho využití se můžeme vrátit k příkladu z marketingu pracujícímu s prodeji různých značek nápojů (vysvětlovaná proměnná) v závislosti na faktorech obsažených v  $X_i$ , kterými byla úrovněová konstanta a cena dané značky. Pokud se  $\beta_i$  liší mezi značkami, model umožňuje, aby dvě značky s identickou cenou měly rozdílné očekávané prodeje (díky rozdílným úrovněovým konstantám), a navíc umožňuje i odlišný marginální efekt změny ceny mezi značkami (díky rozdílným koeficientům u ceny). Pokud zde existuje věrnost značce, pak zde tyto odlišnosti mezi jednotlivými značkami mohou nastat. Využití hierarchických apriorních hustot dává těmto odlišnostem jakousi podobu či strukturu, neboť je modeluje jakožto parametry pocházející ze společných rozdělení. Model náhodných koeficientů dovoluje každé jednotlivé značce, aby se nějakým způsobem odlišovala, nicméně hierarchická apriorní hustota říká, že tato odlišnost nebude přílišná, že se tedy bude jednat o jakousi homogenní skupinu. Takovýto model je užitečný i v řadě jiných praktických aplikacích.

### 7.4.1 Věrohodnostní funkce

Předpoklady kladené na chybový člen (diskutované v úvodu kapitoly) a tvar regresního modelu napovídá, jakou podobu bude mít věrohodnostní funkce:

$$p(y|\beta, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i\beta_i)' (y_i - X_i\beta_i) \right] \right\}, \quad (7.20)$$

kde  $\beta = (\beta_1', \dots, \beta_N)'$  označuje všechny regresní koeficienty pro všechny jednotlivce.

### 7.4.2 Hierarchická apriorní hustota

Obvyklou hierarchickou apriorní hustotou je ta, která předpokládá, že  $\beta_i$  pro  $i = 1, \dots, N$  jsou nezávislé výběry z normálního rozdělení. Předpokládáme tedy

$$\beta_i \sim N(\mu_\beta, V_\beta). \quad (7.21)$$

Druhá fáze hierarchické apriorní hustoty je dána jako

$$\mu_\beta \sim N(\underline{\mu}_\beta, \underline{\Sigma}_\beta) \quad (7.22)$$

a

$$V_\beta^{-1} \sim W(\underline{\nu}_\beta, V_\beta^{-1}). \quad (7.23)$$

Připomeňme si, že Wishartovo rozdělení (zmiňované v kapitole 6 při analýze SUR modelů) je maticovým zobecněním gama rozdělení. Rozdělení v (7.23) je parametrizováno tak, že  $E(V_\beta^{-1}) = \underline{\nu}_\beta V_\beta^{-1}$ . Neinformativní variantu lze získat nastavením  $\underline{\nu}_\beta = 0$ .

Pro přesnost chyby využíváme apriorní hustotu v podobě gama rozdělení:

$$h \sim G(\underline{s}^{-2}, \underline{\nu}). \quad (7.24)$$

### 7.4.3 Bayesovský výpočet

Posteriorní analýzu lze opět provést za využití Gibbsova vzorkovače (podobně jako v případě předchozího modelu). Odvození požadovaných podmíněných posteriorních hustot není obtížné a lze je získat násobením věrohodnostní funkce (7.20) s apriorními hustotami v (7.21)-(7.24). Analýza výsledných výrazů nám odkrývá jádrové hustoty relevantních podmíněných posteriorních rozdělení.

Podmíněné posteriorní hustoty parametrů  $\beta_i$ , pro  $i = 1, \dots, N$ , jsou vzájemně nezávislé a platí

$$\beta_i | y, h, \mu_\beta, V_\beta \sim N(\bar{\beta}_i, \bar{V}_i), \quad (7.25)$$

kde

$$\begin{aligned} \bar{V}_i &= (hX_i'X_i + V_\beta^{-1})^{-1}, \\ \bar{\beta}_i &= \bar{V}_i(hX_i'y_i + V_\beta^{-1}\mu_\beta). \end{aligned}$$

Pro hierarchické parametry  $\mu_\beta$  a  $V_\beta$  jsou relevantní podmíněné posteriorní hustoty:

$$\mu_\beta | y, \beta, h, V_\beta \sim N(\bar{\mu}_\beta, \bar{\Sigma}_\beta) \quad (7.26)$$

a

$$V_\beta^{-1} | y, \beta, h, \mu_\beta \sim W(\bar{\nu}_\beta, [\bar{\nu}_\beta \bar{V}_\beta]^{-1}), \quad (7.27)$$

kde

$$\begin{aligned}\bar{\Sigma}_\beta &= \left( NV_\beta^{-1} + \underline{\Sigma}_\beta^{-1} \right)^{-1}, \\ \bar{\mu}_\beta &= \bar{\Sigma}_\beta \left( V_\beta^{-1} \sum_{i=1}^N \beta_i + \underline{\Sigma}_\beta^{-1} \underline{\mu}_\beta \right), \\ \bar{\nu}_\beta &= N + \underline{\nu}_\beta, \\ \bar{V}_\beta &= \sum_{i=1}^N (\beta_i - \mu_\beta)(\beta_i - \mu_\beta)' + \underline{V}_\beta.\end{aligned}$$

Výraz  $\sum_{i=1}^N \beta_i$  je třeba chápat jako  $k$ -rozměrný vektor obsahující součty prvků  $\beta_i$ .

Podmíněná posteriorní hustota pro přesnost chyby má obvyklou podobu:

$$h|y, \beta, \mu_\beta, V_\beta \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (7.28)$$

kde

$$\begin{aligned}\bar{\nu} &= TN + \underline{\nu}, \\ \bar{s}^{-2} &= \frac{\sum_{i=1}^N (y_i - X_i \beta_i)' (y_i - X_i \beta_i) + \underline{\nu} s^2}{\bar{\nu}}.\end{aligned}$$

Gibbsův vzorkovač v tomto případě vyžaduje výběry z normálního, gama a Wishartova rozdělení. Není tedy obtížné tento problém počítačově zpracovat. Predikční analýza a konvergenční testy lze opět provést standardním způsobem.

## 7.5 Porovnání modelů: Chibova metoda výpočtu marginální věrohodnosti

Doposud nebyla pozornost věnována porovnání modelů třídy tohoto typu. Pro řadu typů modelů lze využít metody z předchozích kapitol. Ppro analýzu pevně daných reskricí, zahrnujících vektor parametrů  $\tilde{\beta}$  v modelech individuálních vlivů, lze využít Savage-Dickeyho poměru hustot podobně, jako je tomu v kapitole 4, části 4.1.5. Ti, kdo nechtějí počítat posteriorní podíl šancí (např. z důvodu použití neinformativní apriorní hustoty), si mohou spočítat intervaly nejvyšší posteriorní hustoty a predikční  $p$ -hodnotu. Některé druhy porovnání modelů, které jsou obsahem této kapitoly, je však obtížné provést. Uvažujme například situaci, kdybychom chtěli porovnat model individuálních vlivů s hierarchickou apriorní hustotou a souhrnný model. Lze ukázat, že posledně jmenovaný model odpovídá modelu individuálních vlivů v případě, kdy  $V_\alpha = 0$ . Skutečnost, že jeden model je vnořeným modelem druhého modelu nám napovídá, že by bylo možné k výpočtu Bayesova faktoru v tomto případě využít Savage-Dickeyho poměr hustot. Pokud se podíváme na vztah (7.18), vidíme, že je tento přístup spojen velkými problémy. K výpočtu Savage-Dickeyho poměru hustot by bylo

potřeba zvolí  $V_\alpha^{-1} = \infty$ . Asi by bylo možno vyřešit tento problém volbou  $V_\alpha^{-1}$  jakožto velkého, konečného čísla, nicméně, v tomto případě by byla výsledkem jen hrubá aproximace skutečné hodnoty bayesova faktoru. Savage-Dickeyho poměr hustot tak je v tomto případě nedostačující metodou. Bylo by možné, na druhé straně, pokusit se o použití metody Gelfanda a Deye, což už by byla akceptovatelná metoda pro výpočet marginální věrohodnosti. V případě problému zahrnujících vektory parametrů vysokých dimenzí může být metoda Gelfanda a Deye dosti nepřesná, a to zejména v kontextu odpovídající volby funkce  $f(\theta)$  (viz kapitola 5, vztahy (5.21) a (5.22)). Obecně je výpočet marginální věrohodnosti v případě vysoké dimenze parametrického prostoru vcelku obtížný a tomuto problému je v literatuře věnována velká pozornost. V této části si ukážeme jeden přístup, který lze v mnoha případech využít k efektivnímu výpočtu marginální věrohodnosti. Jedná se zejména o případy vysoké dimenze parametrického prostoru.

Stejně jako v předchozích částech kapitol, tak i zde se vrátíme k obecnému značení, kde  $\theta$  je vektor parametrů a  $p(y|\theta)$ ,  $p(\theta)$  a  $p(\theta|y)$  jsou postupně věrohodnostní funkce, apriorní hustota a posteriorní hustota. Chibova metoda k výpočtu marginální věrohodnosti, popsaná v práci Chib [5], vychází z velmi jednoduchého zjištění. Bayesovo pravidlo říká

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

Nicméně,  $p(y)$  nezávisí na parametrech  $\theta$ , tedy pravá strana rovnice může být vyhodnocena v jakémkoliv bodě  $\theta^*$ , a výsledkem bude marginální věrohodnost. Pro jakýkoliv bod  $\theta^*$  tak získáváme

$$p(y) = \frac{p(y|\theta^*)p(\theta^*)}{p(\theta^*|y)}, \quad (7.29)$$

kterou Chib nazývá jako *základní identita marginální věrohodnosti* (*basic marginal likelihood identity*). Všimněme si, že všechny hustoty na pravé straně rovnice (7.29) jsou vyhodnoceny v jendo bodě. Například  $p(\theta^*)$  je zkrácený zápis pro  $p(\theta = \theta^*)$ . Pokud bychom znali přesnou podobu věrohodnostní funkce, apriorní hustoty a posteriorní hustoty (tedy ne jen jejich jádra, ale přesnou podobu funkcí hustoty pravděpodobnosti), potom můžeme marginální věrohodnost spočítat na základě vyhodnocení těchto funkcí v jakémkoliv bodě a s využitím vztahu (7.29). V mnoha případech známe přesnou podobu věrohodnostní funkce a apriorní hustoty. Obvykle však neznáme přesnou podobu posteriorní hustoty. K implementaci Chibovy metody si tak musíme ukázat, jak vyhodnotit posteriorní hustotu v podě (tj. spočítat  $p(\theta^*|y)$ ). Chibova práce popisuje tento postup pro různé případy. My se zaměříme na případ, který je relevantní pro modely z této kapitoly.

## 7.6 Empirická ilustrace

BUDE DOPLNĚNO!

## 7.7 Analýza efektivity a model stochastických hranic

V této části využijeme ekonomickou teorii k prezentaci *modelu stochastických mezí* (*stochastic frontier model*, který spadá do kategorie modelů individuálních vlivů, ovšem s odlišnou hierarchickou apriorní hustotou než tou prezentovanou v části 7.4.2. Tento model je důležitý sám o sobě, protože se využívá v případě studií zaměřených na efektivitu produkce firem nebo jiného typu ekonomických agentů. Odvození modelu stochastických mezí je rovněž dobrou ukázkou toho, jak aplikovaná ekonomie může využít ekonomickou teorii ke konstrukci ekonometrického modelu.

### 7.7.1 Úvod do modelu stochastických hranic

Principy, na kterých je tato třída modelů postavena, lze ukázat na příkladu ekonomického modelu produkce, kdy výstup firmy  $i$  v čase  $t$ ,  $Y_{it}$ , je vyráběn s využitím vektoru vstupů,  $X_{it}^*$ , kde  $i = 1, \dots, N$  a  $t = 1, \dots, T$ . Firmy využívají běžnou, nejlepší možnou dostupnou technologii, pomocí které přetavují vstupy do podoby výstupu. Tato technologie je závislá na neznámých parametrech,  $\beta$ , a je dána jako:

$$Y_{it} = f(X_{it}^*; \beta). \quad (7.30)$$

To je tzv. *hranice výrobních možností* (*production frontier*) a říká nám, jaký maximální objem výstupu může být dosažen z daného objemu vstupů. Ve skutečnosti může skutečný produkt firmy spadnout pod takovouto maximální dosažitelnou úroveň. Odchylka skutečného výstupu od maximálně dosažitelného bývá chápána jako měřítko neefektivity a spadá do oblasti zájmu v řadě praktických aplikací. Formálně si můžeme vztah (7.30) rozšířit do podoby

$$Y_{it} = f(X_{it}^*; \beta) \tau_i, \quad (7.31)$$

kde  $0 < \tau_i \leq 1$  je míra efektivity, která je specifická pro jednotlivé firmy, a  $\tau_i = 1$  znamená, že firma  $i$  je plně efektivní. Např. hodnota  $\tau_i = 0.75$  znamená, že firma  $i$  produkuje jen 75% výstupu, který by mohla produkovat, kdyby pracovala dle nejlépe dostupné technologie. V rámci této specifikace předpokládáme, že každá firma má svou konkrétní úroveň efektivity, která je neměnná v čase. Tento předpoklad však můžeme uvolnit, kdy odkaz na příslušnou detailnější literaturu nabízí Koop [19].

V ekonometrické praxi se v modelu dále uvažuje náhodný chybový člen,  $\zeta_{it}$ , který zahrnuje chyby v měření či specifikaci. Model tak nabývá podoby

$$Y_{it} = f(X_{it}^*; \beta) \tau_i \zeta_{it}. \quad (7.32)$$

Zahrnutí tohoto chybového členu (chyb měření) vytváří stochastickou hranici, a odtud tedy máme název modelu *stochastických hranic*. Pokud je hranice výrobních možností,  $f(\cdot)$ , v log-lineární podobě (např. Cobb-Douglasova produkční

funkce nebo produkční funkce TRANSLOG), můžeme využít logaritmování a přepsat vztah (7.32) jako

$$y_{it} = X_{it}\beta + \epsilon_{it} - z_i, \quad (7.33)$$

kde  $\beta = (\beta_1, \dots, \beta_k)'$ ,  $y_{it} = \ln(Y_{it})$ ,  $\epsilon_{it} = \ln(\zeta_{it})$ ,  $z_i = -\ln(\tau_i)$  a  $X_{it}$  je protějškem  $X_{it}^*$ , kdy vstupy jsou transformovány za pomoci logaritmů. Veličina  $z_i$  je označován jako neefektivita, a protože  $0 < \tau_i \leq 1$ , jedná se o nezápornou náhodnou veličinu. Předpokládáme, že  $X_{it}$  obsahuje úrovnovou konstantu a koeficient  $\beta_1$ . Všimněme si, že tento model má podobu modelu individuálních vlivů. Tzn., že výraz  $\beta_1 - z_i$  hraje tutéž roli jako  $\alpha_i$  v sekci 7.4.2. V modelu stochastických hranic nám však sama ekonomická teorie dává vodítko k výběru hierarchické apriorní hustoty.

Je potřeba poznamenat, že v případě, kdy není produkční funkce log-lineární (např. CES produkční funkce, tedy produkční funkce s konstantní elasticitou substituace), je potřeba bayesovskou analýzu provádět ve vzájemné kombinaci technik z kapitoly 5 a technik uváděnými na tomto místě.

Rovnice (7.33) může být zapsána jako

$$y_{it} = X_{it}\beta + \epsilon_i - z_i \nu_T, \quad (7.34)$$

pokud si setřídíme všechny proměnné do matic stejným způsobem, jak bylo popsáno v úvodu této kapitoly (část 7.1). Připomeňme si, že  $\nu_T$  je v našem značení  $T$ -rozměrný vektor jedniček.

### 7.7.2 Věrohodnostní funkce

Podoba věrohodnostní funkce závisí na předpokladech kladených na náhodnou složku. Kromě standardních předpokladů vyjádřených v úvodu této kapitoly (viz část 7.1) budeme předpokládat, že  $z_i$  a  $\epsilon_j$  jsou vzájemně nezávisle pro všechna  $i$  a  $j$ . Výsledná věrohodnostní funkce tak má podobu

$$p(y|\beta, h, z) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i\beta + z_i \nu_T)' (y_i - X_i\beta + z_i \nu_T) \right] \right\}, \quad (7.35)$$

kde  $z = (z_1, \dots, z_N)'$ .

V rámci této specifikace chápeme  $z$  jako vektor neznámých parametrů, které vstupují do věrohodnostní funkce. V přístupu „klasické“ ekonometrie by věrohodnostní funkce byla definována jako  $p(y|\beta, h, \theta) = \int p(y|\beta, h, z)p(z|\theta)dz$ , kde  $p(z|\theta)$  odpovídá našemu předpokladu o rozdělení neefektivity, které závisí na vektoru neznámých parametrů  $\theta$ . Tento postup je matematicky ekvivalentní bayesovskému přístupu využívajícímu  $p(z|\theta)$  jako hierarchickou apriorní hustotu. Jinými slovy, v modelech jako tento je volba toho, co označíme jako „věrohodnostní funkce“ a co jako „hierarchickou apriorní hustotu“, čistě sémantickou záležitostí, která nemá vliv na statistické postupy další analýzy.

### 7.7.3 Hierarchická apriorní hustota pro model stochastických hranic

Pro koeficienty hranice výrobních možností a přesnost chyby využijeme nám dobře známou nezávislou normální-gama apriorní hustotu:

$$\beta \sim N(\underline{\beta}, \underline{V}) \quad (7.36)$$

a

$$h \sim G(\underline{s}^{-2}, \underline{\nu}). \quad (7.37)$$

Pro míru neefektivity použijeme hierarchickou apriorní hustotu. Protože víme, že  $z_i > 0$ , není vhodné použít hierarchická hustota odpovídající normální hustotě pravděpodobnosti z části 7.4.2. Volba apriorní hustoty obvykle v literatuře odpovídá omezenému normálnímu rozdělení nebo některému rozdělení z rodiny gama rozdělení. Zde si ukážeme bayesovskou analýzu modelů stochastických hranic s využitím exponenciálního rozdělení, což je gama rozdělení s dvěma stupni volnosti (viz Příloha B). Předpokládejme tedy, že  $z_i$  a  $z_j$  jsou a priori nezávislé pro  $i \neq j$ , kdy platí

$$z_i \sim G(\mu_z, 2). \quad (7.38)$$

Hierarchická podstata apriorní hustoty znamená, že chápeme střední hodnotu rozdělení neefektivity jako parametr, který vyžaduje svůj vlastní prior. Protože je  $z_i > 0$ , vyplývá z toho, že rovněž  $\mu_z > 0$ . Je snadnější pracovat s  $\mu_z^{-1}$  než přímo s  $\mu_z$ , podobně jako pracujeme s přesností chyby ( $h$ ) místo rozptylu chyby ( $\sigma^2$ ), což nám umožňuje setrvat v nám známých třídách rozdělení chybových členů. Budeme tedy pracovat a apriorní hustotou v podobě

$$\mu_z^{-1} \sim G(\underline{\mu}_z^{-1}, \underline{\nu}_z). \quad (7.39)$$

Apriorní hyperparametry pro  $\underline{\mu}_z^{-1}$  a  $\underline{\nu}_z$  mohou být stanoveny na základě předpokladů o rozdělení efektivity. Můžeme tedy často mít k dispozici apriorní informaci o tom, kde je rozdělení efektivity umístěno. Necht  $\tau^*$  označuje apriorní medián tohoto rozdělení. Pokud očekáváme, že firmy v našem vzorku jsou spíše efektivní, můžeme nastavit hodnotu  $\tau^*$  na vysokou hodnotu (např. 0.95). Pokud očekáváme, že řada firem je neefektivních, nastavíme ji na hodnotu nižší. V literatuře je ukázáno, že nastavení  $\underline{\nu}_z = 2$  implikuje relativně neinformativní apriorní hustotu a z nastavení  $\underline{\mu}_z = -\ln(\tau^*)$  vyplývá, že medián apriorního rozdělení efektivity je  $\tau^*$ .

To nám ilustruje obvyklou strategii nastavení prioru. Apriorní hustotu stanovujeme prostřednictvím hyperparametrů, které jsou snadno interpretovatelné v kontextu výchozí ekonomické teorie (v tomto případě  $\tau^*$ ). Následně provádí zpětnou transformaci pro nalezení hyperparametrů použitých v modelu (v našem případě  $\underline{\mu}_z$  a  $\underline{\nu}_z$ ).

Užitečné je rovněž upozornit na skutečnost, že je to často ekonomická teorie, která nám nabízí různá omezení, která mohou být zavedena skrze náš prior.



Například bychom chtěli implementovat restrikcí, že hranice výrobních možností je monotónně rostoucí ve vstupech. V jiných variantách modelu stochastických hranic je žádoucí zahrnout restrikcí, že nákladová funkce je konkávní nebo, že je možný technologický úpadek (jako opak technologického pokroku). Toto vše znamená omezení parametrů ve tvaru nerovností, což lze provést metodami z kapitoly 4, části 4.2.

#### 7.7.4 Bayesovský výpočet

Posteriorní analýza může být provedena stejně jako v modelech individuálních vlivů pomocí Gibbsova vzorkovače. Gibbsův vzorkovač potřebuje pouze znalost plně podmíněných posteriorních hustot. Ukážeme si tedy pouze příslušná podmíněná rozdělení (ne celou posteriorní hustotu), která jsou stejná jako v případě modelu individuálních vlivů s hierarchickou apriorní hustotou, tedy až na hustoty odpovídající  $z$  a  $\mu_z$ .

Pro parametry hranice výrobních možností získáváme

$$\beta|y, h, z, \mu_z \sim N(\bar{\beta}, \bar{V}), \quad (7.40)$$

kde

$$\bar{V} = \left( \underline{V}^{-1} h \sum_{i=1}^N X_i' X_i \right)^{-1},$$

$$\bar{\beta} = \bar{V} \left( \underline{V}^{-1} \underline{\beta} + h \sum_{i=1}^N X_i' [y_i + z_i \iota_T] \right).$$

Pro přesnost chyby máme standardní výsledky

$$h|y, \beta, z, \mu_z \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (7.41)$$

kde

$$\bar{\nu} = TN + \nu,$$

$$\bar{s}^{-2} = \frac{\sum_{i=1}^N (y_i z_i \iota_T - X_i \beta)' (y_i z_i \iota_T - X_i \beta) + \nu s^2}{\bar{\nu}}.$$

Podmíněné posteriorní hustoty pro neefektivitu jsou na sobě nezávislé (tj.  $z_i$  a  $z_j$  jsou nezávislé pro  $i \neq j$ ) a odpovídají normálnímu rozdělení omezenému na kladné hodnoty s funkcí hustoty pravděpodobnosti danou jako

$$p(z_i | y_i, X_i, \beta, h, \mu_z) \propto f_N \left( z_i | \bar{X}_i \beta - \bar{y}_i - (Th \mu_z)^{-1}, (Th)^{-1} \right) 1(z_i \geq 0), \quad (7.42)$$

kde  $\bar{y}_i = \frac{\sum_{t=1}^T y_{it}}{T}$  a  $\bar{X}_i$  je matice rozměru  $(1 \times k)$  obsahující průměrné hodnoty každé vysvětlující proměnné pro každého jednotlivce  $i$ . Opět,  $1(z_i \geq 0)$  je indikační funkce rovna jedničce pokud  $z_i \geq 0$  a nule v ostatních případech.

Podmíněná posteriorní hustota pro  $\mu_z^{-1}$  je dána jako

$$\mu_z^{-1} | y, \beta, z \sim G(\bar{\mu}_z, \bar{\nu}_z), \quad (7.43)$$

kde

$$\begin{aligned} \bar{\nu}_z &= 2N + \nu_z, \\ \bar{\mu}_z &= \frac{N + \frac{\nu_z}{2}}{\sum_{i=1}^N z_i \mu_z}. \end{aligned}$$

Bayesiánská analýza modelu stochastické hranice může být provedena pomocí Gibbsova vzorkovače zahrnujícího sekvenční výběry z podmíněných hustot (7.40) až (7.43). Výběry z omezeného normálního rozdělení (viz (7.42)) můžeme provést tak, že provedeme výběry z odpovídajícího neomezeného normálního rozdělení a zahodíme výběry, kdy  $z_i < 0$ . Alternativně lze využít algoritmy pro výběry z omezeného normálního rozdělení, snadno dostupné např. pro MATLAB. Tradičním způsobem provádíme predikční analýzu a standardními MCMC diagnostikami ověřujeme konvergenci Gibbsova vzorkovače. Porovnání modelů lze využít za pomoci technik této kapitoly, např. pomocí Chibovy metody můžeme vypočítat marginální věrohodnost.

Užitečné je rovněž vědět, že tyto metody se využívají i pro čistě průřezovou verzi tohoto modelu (tzn.  $T = 1$ ). V případě  $T = 1$  však není přijatelné použití určitých nepravých priorů, neboť vedou k nepravým posteriorům. Intuitivně, pokud je  $T = 1$ , je počet parametrů v celém modelu větší než velikost vzorku (parametry  $z$ ,  $\mu_z$ ,  $\beta$ ,  $h$  dohromady dávají  $N + K + 2$  parametrů a my máme k dispozici jen  $N$  pozorování), což poněkud omezuje rozumnou posteriorní analýzu při absenci apriorní informace. Otázky použití neinformativních priorů v modelech stochastických hranic jsou podrobně diskutovány v odborné literatuře.

### 7.7.5 Empirická ilustrace

BUDE DOPLNĚNO!

## 7.8 Rozšíření

BUDE DOPLNĚNO!

## 7.9 Shrnutí

BUDE DOPLNĚNO!

# Kapitola 8

## Úvod do časových řad

### 8.1 Úvod

K problematice časových řad se váže rozsáhlá ekonometrická literatura. V této kapitole se budeme zabývat jednou třídou modelů zvanou *stavové modely* (*state space models*), které jsou běžně využívány právě za použití veličin a dat v podobě časových řad.

Pro zavedení stavových modelů hovoří tři důvody:

1. Stavové modely jsou ve své podstatě hierarchické. Právě bayesovské metody s hierarchickou apriorní hustotou jsou v praxi velmi atraktivní.
2. Bayesiánská analýza hlavního alternativního přístupu<sup>18</sup> k ekonometrii časových řad je detailně pokryta v knize Bauwens, Lubrano a Richard [1]. Tato kapitola nabízí jiný pohled na problematiku časových řad.
3. Stavové modely nejsou až tak odlišnou třídou modelů než používají Bauwens, Lubrano a Richard [1]. Nabízejí však jiný způsob zápisu téhož modelu.<sup>19</sup>

Využitím stavových modelů je možno řešit stejné problémy, které jsou obsahem publikace autorů Bauwens, Lubrano a Richard [1], ovšem zůstáváme v hierarchickém kontextu, který je důvěrně známý a výpočetně pohodlný.

Některé koncepty zahrnující časové řady byl řešeny v kapitole 6 v rámci analýzy lineárního regresního modelu s autokorelovanými náhodnými složkami. Při práci s časovými řadami tedy používáme  $t$  a  $T$  místo  $i$  a  $N$ , tedy  $y_t$  pro  $i = 1, \dots, T$  značí pozorování závisle proměnné v období 1 až  $T$ . Před tím, než se dostaneme ke stavovým modelům, je dobré připomenout, že techniky analyzované v kapitole 6 lze v praxi dobře uplatnit. Například, lineární regresní model s autokorelovanými chybami je modelem časových řad, který může být v

<sup>18</sup>Tento alternativní přístup v sobě zahrnuje ARIMA modely a dynamické regresní modely, které v sobě zahrnují problematiku jednotkových kořenů a kointegrace.

<sup>19</sup>Například, pro jakýkoliv ARIMA model existuje i stavová reprezentace.

řadě případů více než vhodný. Tento model obsahuje chybovou složku  $\epsilon_t$ , generovanou  $AR(p)$  procesem. Běžný model časových řad pro jednu vysvětlovanou proměnnou (tj. model analyzující chování jedné časové řady  $y$ ) pro  $y_t$  odpovídá  $AR(p)$  procesu:

$$(1 - \rho_1 L - \dots - \rho_p L^p)y_t = u_t \quad (8.1)$$

Výpočetní metody bayesovské analýzy tohoto modelu jsou zjednodušením metod diskutovaných v předchozích kapitolách. Ve skutečnosti je (8.1) lineární regresní model, kde vysvětlující proměnné jsou zpožděné hodnoty proměnné vysvětlované

$$y_t = \rho_1 y_{t-1} + \dots + \rho_p y_{t-p} + u_t \quad (8.2)$$

Všechny základní regresní techniky diskutované v předchozích kapitolách jsou relevantní. Rovnice (8.2) může být rozšířena o další vysvětlující proměnné a jejich zpoždění, přičemž se stále budeme nacházet v regresním konceptu:

$$y_t = \rho_1 y_{t-1} + \dots + \rho_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + u_t \quad (8.3)$$

V tomto na regresi založeném přístupu však vzniká několik komplikací. Volně řečeno, největší část literatury pojednávající o časových řadách se týká restrikcí kladených na koeficienty (8.3), případně jejich transformaci. Rovněž je zde řada důležitých problémů týkajících se definování apriorní hustoty, které nevznikají v kontextu průřezových dat.<sup>20</sup>

V rámci úvodu do stavových modelů začneme s nejjednodušším případem zvaným *local level model*. Základní otázky týkající se formulace apriorní hustoty a výpočtu mohou být diskutovány v kontextu tohoto modelu. Potom přistoupíme obecnějšímu stavovému modelu. Detailnější výklad nabízí West a Harrison [29].<sup>21</sup> Publikace Kima a Nelsona [18] je další bayesiánskou knihou, která je úvodem i rozšířením problematiky stavových modelů.

V této kapitole využijeme stavové modely k zavedení empirických bayesovských metod. Tyto metody jsou velmi populární v hierarchických modelech všeho druhu. Poskytují na datech založené metody k získání apriorních hyperparametrů. Empirické bayesovské metody nabízejí atraktivní alternativu<sup>22</sup> pro ty, kdo nechtějí využívat subjektivního definování informativních apriorních hustot nebo nechtějí používat neinformativních priorů.

## 8.2 Local level model

Local level model je dán jako

$$y_t = \alpha_t + \epsilon_t \quad (8.4)$$

<sup>20</sup>Kromě knihy Bauwense, Lubrana a Richarda [1], další podrobnosti nabízejí příspěvky z tematických vydání časopisů *Econometric Theory* (ročník 10, srpen-říjen 1994) a *Journal of Applied Econometrics* (ročník 6, říjen-prosinec 1991)

<sup>21</sup>Další články pojednávající o bayesiánské analýze stavových modelů jsou Carlin, Polson a Stoffer [3], Carter a Kohn [4], de Jong a Shephard [7], Fruhwirth-Schnatter [11], Koop and van Dijk [20] a Shively a Kohn [27]. Durbin a Koopman [9] je dobrým zdrojem s bayesovským obsahem.

<sup>22</sup>Statistický úvod do empirických bayesovských metod nabízí Carlin a Louis [2].

kde  $\epsilon_t$  je i.i.d.  $N(0, h^{-1})$ . Jedinečným aspektem tohoto modelu je člen  $\alpha_t$ , který není pozorován a o kterém předpokládáme, že je generován jako náhodná procházka (*random walk*)

$$\alpha_{t+1} = \alpha_t + u_t \quad (8.5)$$

kde  $u_t$  je i.i.d.  $N(0, \eta h^{-1})$  a  $\epsilon_t$  a  $u_s$  jsou vzájemně nezávislé pro všechna  $t$  a  $s$ . Ve vztahu (8.4)  $t$  probíhá od 1 do  $T$ , ve vztahu (8.5) od 1 do  $T - 1$ . Rovnice (8.5) neposkytuje explicitně vyjádření pro  $\alpha_1$ , označované jako *počáteční podmínka*. Rovnice (8.4) je tzv. *rovnice pozorování* či *měření* (*observation equation*, *measurement equation*), rovnice (8.5) je pak nazývána *stavovou rovnicí* (*state equation*).

V kapitole 6 byl diskutován  $AR(1)$  model a bylo zdůrazněno, že pro koeficient u zpožděné závisle proměnné  $\rho = 1$  je tato řada nestacionární. Implikací tohoto je, že  $\alpha_t$  má *stochastický trend*. Pojem stochastický trend vychází ze skutečnosti, že pro modely jako (8.5) se příslušná řada může v čase dosti volně vyvíjet (tj. má trend), nicméně do tohoto trendového chování vstupuje prvek náhodnosti. V protikladu k *deterministickému trendu* jako

$$\alpha_t = \alpha + \beta t$$

kde proměnná je právě funkcí času, stochastický trend v sobě obsahuje náhodnou složku  $u_t$ . Skutečnost, že (8.5) vyjadřuje trendové chování pro  $\alpha_t$ , je možno vyvodit ze zápisu

$$\alpha_t = \alpha_1 + \sum_{j=1}^{t-1} u_j \quad (8.6)$$

a tedy (při zanedbání počátečních podmínek)  $var(\alpha_t) = (t - 1)\eta h^{-1}$ . Navíc  $\alpha_t$  a  $\alpha_{t-1}$  mají tendenci ležet blízko u sebe (tj.  $E(\alpha_t | \alpha_{t-1}) = \alpha_{t-1}$ ). Stochastický trend má tedy variabilitu, která je rostoucí funkcí času (a řada tedy může procházet stále se rozšiřující oblastí), ovšem  $\alpha_t$  se mění v čase jen pozvolna. To je konzistentní s intuitivním konceptem trendu, který je něco co postupně roste nebo klesá v čase.

Rovnice (8.4) dekomponuje pozorovanou řadu  $y_t$  na trendovou komponentu a chybovou či nepravidelnou komponentu  $\epsilon_t$ .<sup>23</sup> Obecně řečeno, stavové modely lze interpretovat jakožto modely dekomponující časovou řadu na různé části. V local level modelu jsou to komponenty: trend a náhodná složka. V složitějších modelech mohou přibýt i další složky, jako např. sezónní složka.

Je dobré poznamenat, že local level model se používá pro měření relativní velikosti trendu a nepravidelné složky. To je zároveň i motivací, proč byly zapsány rozptyly předchozích náhodných složek ( $h^{-1}$  a  $\eta h^{-1}$ ). Tento způsob přímo implikuje smysl  $\eta$  jakožto míru velikosti náhodné procházky relativně k rozptylu chyby v rovnici měření. Je tak možno vidět, že pro  $\eta \rightarrow 0$  náhodná složka v (8.5) vypadává a  $\alpha_t = \alpha_1$  pro všechna  $t$  a (8.4) nabyde tvaru  $y_t = \alpha_1 + \epsilon_t$ .

<sup>23</sup>Z makroekonomického pohledu by tak trend mohl pokrývat dlouhodobou růstovou tendenci ekonomiky (např. díky růstu pracovní síly, růstu kapitálové zásoby nebo technickému pokroku), přičemž nepravidelná komponenta reflektuje náhodné krátkodobé šoky postihující ekonomiku (např. efekty hospodářského cyklu).

V tomto případě vykazuje  $y_t$  náhodné fluktuace kolem konstantní úrovně  $\alpha_1$  a nemá už tedy trendový charakter. Ovšem pro rostoucí  $\eta$  roste rovněž i rozptyl  $u_t$  a tudíž narůstá role stochastického trendu. Analýza  $\eta$  je tedy dobrý způsob pro měření významu trendového chování v časové řadě. Jen na okraj, test toho, zda-li  $\eta = 0$ , je jedním ze způsobů testování *jednotkového kořene* (*unit root*). Tento test hraje svou důležitou roli v moderní empirické makroekonomii, a stavové modely umožňují toto testování vcelku intuitivním a jasným způsobem.

Jiný způsob interpretace (8.4) a (8.5) vychází z toho, že  $\alpha_t$  je střední hodnota (či úroveň, tedy *level*) pro  $y_t$ . Protože tato střední hodnota se mění v čase, používá se terminologie *local level model*. Interpretovat  $\alpha_t$  tímto způsobem, jako parametr, je v bayesovském chápání vcelku přirozené. Rovnici (8.4) je tak možno chápat jako velmi jednoduchý případ lineárního regresního modelu zahrnujícího jen úrovnovou konstantu. Významnou inovací je to, že se tento parametr může měnit v čase. *Local level model* je tedy jednoduchým příkladem *modelu v čase proměnných parametrů* (*time varying parameters model*). Sofistikovanější stavové modely v sobě mohou zahrnovat v čase proměnné parametry (regresní koeficienty) nebo v čase proměnné rozptyly náhodných složek. Pokud je  $\alpha = (\alpha_1, \dots, \alpha_T)'$  interpretováno jako vektor parametrů, potom v bayesiánském kontextu je třeba definovat příslušnou apriorní hustotu. Ovšem (8.5) tuto apriorní hustotu přímo nabízí, neboť definuje hierarchickou apriorní hustotu pro  $\alpha$ . S touto interpretací je *local level model* velmi podobný modelu individuálních vlivů v rámci analýzy panelových dat v kapitole 7 s  $T = 1$ . Samozřejmě model individuálních vlivů má úrovnovou konstantu, která se mění v rámci jednotlivců, přičemž *local level model* má úrovnovou konstantu proměnnou v čase. Základní struktura těchto dvou modelů je však stejná. Nástroje kapitoly 7 s využitím nezávislé normální-gama apriorní hustoty zde lze s určitými modifikacemi využít. Z tohoto důvodu zde použijeme přirozeně konjugovanou apriorní hustotu a zavedeme novou techniku pro definování apriorní hustoty.

Bayesovské metody využívající nezávislou normální-gama apriorní hustotu jsou podobné těm z kapitoly 7, nebudeme je zde tedy opakovat. Lze zde odvodit Gibbsův vzorkovač. Tento algoritmus bude odvozen v kontextu obecnějšího stavového modelu (vzhledem k této obecnosti je půjde využít i pro *local level model*). V následující části bude využita přirozeně konjugovaná apriorní hustota k zavedení empirických bayesiánských metod.

### 8.2.1 Věrohodnostní funkce a apriorní hustota

Pokud budeme definovat  $y = (y_1, \dots, y_T)'$  a  $\epsilon = (\epsilon_1, \dots, \epsilon_T)'$ , lze zapsat *local level model* v maticovém zápisu jako

$$y = I_T \alpha + \epsilon \quad (8.7)$$

Budeme-li předpokládat standardní požadavky kladené na  $\epsilon$ , tedy že  $\epsilon$  má vícerozměrné normální rozdělení se střední hodnotou  $0_T$  a kovarianční maticí  $h^{-1} I_T$ , potom se bude jednat o normální lineární regresní model, kde matice vysvětlujících proměnných bude jednotková matice ( $X = I_T$ ) a  $\alpha$  bude  $T$ -rozměrný vektor

regresních koeficientů. Věrohodnostní funkce tak bude mít standardní podobu normálního lineárního regresního modelu (viz kapitola 3).

Samozřejmě v jakémkoliv bayesovském příkladu můžeme využít apriorní hustotu dle našich potřeb a přání. Stavová rovnice v (8.5) však sama nabízí hierarchickou apriorní hustotu. Využijeme tedy tuto hustotu v přirozeně konjugované podobě. Pro porovnání s výsledky kapitoly 3 pro normální lineární regresní model s přirozeně konjugovanou apriorní hustotou je dobré zapsat tento model poněkud odlišným způsobem. Začneme definováním *matice prvních diferencí* rozměru  $(T-1) \times T$ :

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & 0 & -1 & 1 \end{bmatrix} \quad (8.8)$$

Pro představu souvislostí se stavovým modelem poznamenejme, že

$$D\alpha = \begin{pmatrix} \alpha_2 - \alpha_1 \\ \vdots \\ \alpha_T - \alpha_{T-1} \end{pmatrix}$$

a stavová rovnice (8.5) tak může být zapsána v podobě

$$D\alpha = u$$

kde  $u = (u_1, \dots, u_{T-1})'$ . Předpoklad, že  $u$  je normální lze tedy interpretovat tak, že stavová rovnice definuje normální hierarchickou apriorní hustotu pro  $D\alpha$ .

Ke specifikaci kompletních apriorních hustot pro všechny parametry modelu je třeba specifikovat apriorní hustotu pro  $h$  a  $\alpha_1$ . Abychom tak učinili, zapíšeme (8.7) v podobě

$$y = W\theta + \epsilon \quad (8.9)$$

kde

$$\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \vdots \\ \alpha_T - \alpha_{T-1} \end{pmatrix}$$

a

$$W = \begin{pmatrix} 1 & 0_{t-1}' \\ \iota_{T-1} & C \end{pmatrix},$$

kde  $\iota_{T-1}$  je  $(T-1)$ -rozměrný vektor jedniček. Maticovým násobením lze ukázat, že (8.9) je ekvivalentní k (8.7). Přímá maticová inverze může být využita k tomu, aby se ukázalo, že  $C$  je dolní trojúhelníková matice rozměru  $(T-1) \times (T-1)$  se všemi nenulovými prvky rovnými jedné (jedná se o inverzi matice  $D$  s vynechaným prvním sloupcem). Tedy  $C$  má všechny prvky na a pod hlavní diagonálou rovny jedné a všechny prvky nad hlavní diagonálou rovny nule.

Nyní si definujme přirozeně konjugovanou apriorní hustotu pro  $\theta$  a  $h$ :

$$\theta, h \sim NG(\underline{\theta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}) \quad (8.10)$$

Budeme předpokládat specifickou strukturu pro  $\underline{\theta}$  a  $\underline{V}$  která bude zahrnovat apriorní informaci obsaženou ve stavové rovnici:

$$\underline{\theta} = \begin{pmatrix} \theta_1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad (8.11)$$

$$\underline{V} = \begin{pmatrix} \underline{V}_{11} & 0_{T-1}' \\ 0_{T-1} & \eta I_{T-1} \end{pmatrix} \quad (8.12)$$

Poznamenejme, že tato apriorní hustota implikuje, že  $\alpha_{t+1} - \alpha_t$  odpovídá normální hustotě,  $N(0, \eta h^{-1})$ , což je přesně to, co jsme předpokládali na začátku této sekce. Skutečnost, že apriorní hustota závisí na parametru  $\eta$  ji činí hierarchickou. Navíc jsme poskytly apriorní hustotu pro počáteční podmínku  $\alpha_1 \sim N(\underline{\theta}_1, h^{-1} \underline{V}_{11})$ .

Na tomto místě je vhodně shrnout, co jsme doposud udělali. Zapsali jsme local level model jako náš dobře známý normální lineární regresní model s přirozeně konjugovanou apriorní hustotou. Skutečnost, že se jedná o problematiku časových řad zahrnující stavový model, se projevuje výhradně díky výběru apriorní hustoty. V rámci bayesovského paradigmatu je interpretace stavové rovnice jakožto prioru zcela přirozená a atraktivní. Je však třeba dodat, že nebayesovská ekonometrie by tuto hierarchickou apriorní hustotu interpretovala jako část věrohodnostní funkce. V řadě modelů je jen otázkou rozhodnutí, kterou část modelu označíme jako "věrohodnostní funkci" a kterou jako "prior".

## 8.2.2 Posteriorní hustota

Standarní výsledky pro NLRM s přirozeně konjugovanou apriorní hustotou (viz kapitola 3) nám říkají, že posteriorní hustota pro parametry  $\theta$  a  $h$ , označována jako  $p(\theta, h|y)$ , odpovídá  $NG(\bar{\theta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$  kde

$$\bar{\theta} = \bar{V}(\underline{V}^{-1}\underline{\theta} + W'y) \quad (8.13)$$

$$\bar{V} = (\underline{V}^{-1} + W'W)^{-1} \quad (8.14)$$

$$\bar{\nu} = \underline{\nu} + T \quad (8.15)$$

a

$$\bar{\nu} \bar{s}^2 = \underline{\nu} \underline{s}^2 + (y - W\bar{\theta})'(y - W\bar{\theta}) + (\bar{\theta} - \underline{\theta})'\underline{V}^{-1}(\bar{\theta} - \underline{\theta}) \quad (8.16)$$

Vlastnosti normálního-gama rozdělení implikují, že je snadné zpětně transformovat parametrizaci v (8.9) na původní parametrizaci danou v (8.7). Protože



$p(\theta|h, y)$  je normální a platí, že lineární kombinace veličiny z normálního rozdělení je opět veličina z normálního rozdělení, potom za podmínky, že posteriorní hustota  $p(\theta, h)$  odpovídá  $NG(\bar{\theta}, \bar{V}, \bar{s}^{-2}, \bar{\nu})$ , je posteriorní rozdělení  $(\alpha, h)$  analogické rozdělení  $NG(\bar{\alpha}, \bar{V}_\alpha, \bar{s}^{-2}, \bar{\nu})$  kde

$$\bar{\alpha} = W\bar{\theta} \quad (8.17)$$

a

$$\bar{V}_\alpha = W\bar{V}W' \quad (8.18)$$

Protože jsme použili přirozeně konjugovanou apriorní hustotu, máme k dispozici analytické výsledky, a nepotřebujeme tak posteriorní simulátor. Je dobré poznamenat, že local level model je regresní model, kde počet regresních koeficientů je roven počtu pozorování. V regresní analýze je obvyklejší případ, kdy počet regresních koeficientů je výrazně nižší než počet pozorování (tj. v notaci předchozích kapitol  $k \ll N$ ). Local level model tak ukazuje, že apriorní informace může být v mnoha případech využita k získání hodnotné posteriorní analýzy i v modelech s velkým počtem parametrů. V tomto kontextu vzniká jiná otázka, a to, proč nezískáváme degenerované posteriorní rozdělení v bodě  $y = \alpha$ . Pokud bychom nastavili  $\alpha_t = y_t$  pro všechna  $t$ , získali bychom perfektně padnoucí model ve smyslu, že  $\epsilon_t = 0$  pro všechna  $t$ . Lze ověřit, že věrohodnostní funkce nabývá v tomto bodě nekonečně velké hodnoty. Bayesiánská posteriorní hustota však není do tohoto bodu nekonečné věrohodnosti umístěna, a to díky apriorní informaci. Stavová rovnice nám říká, že  $\alpha_{t+1}$  a  $\alpha_t$  leží velmi blízko u sebe, což odsunuje posteriorní hustotu dále od bodu perfektní shody modelu s daty. V literatuře zabývající se stavovými modely je tato skutečnost či jev nazývána *vyhlazením (smoothing)* stavového vektoru.

Jelikož se jednalo v tomto oddíle o NLRM s přirozeně konjugovanou apriorní hustotou, porovnání modelů a predikce je prováděna metodami z kapitoly 3.

### 8.2.3 Empirické bayesiánské metody

V předchozích kapitolách byla apriorní hustota formulována buď subjektivně nebo se využívalo neinformativních apriorních hustot. V kontextu této kapitoly to znamená volbu  $\underline{\theta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}$  nebo nastavení neinformativních hodnot  $\underline{\nu} = 0$  a  $\underline{V}^{-1} = 0_{T \times T}$ .<sup>24</sup> Oba tyto přístupy však mají své nedostatky. Subjektivní volbu apriorních priorů je někdy obtížné udělat, případně může být zdrojem kritiky jiných výzkumníků majících apriorní hyperparametry (či hustoty) jiné. Neinformativní apriorní hustoty obvykle znemožňují bayesovské porovnání modelů z důvodů nedefinovanosti marginálních věrohodností. Mnozí bayesiánci proto využívají tzv. empirických bayesovských metod, které se snaží překlenout výše zmiňované problémy. Local level model je vhodným místem pro jejich zavedení. Samozřejmě tyto metody lze využít s jakýmkoli modelem, přičemž jsou obzvláště populární v modelech s hierarchickými apriorními hustotami (jako jsou modely této kapitoly a kapitoly 7). Je však třeba poznamenat, že empirické bayesovské metody jsou kritizovány pro jejich implicitní "dvojpočet" s daty, tj. data jsou

<sup>24</sup>Připomeňme si, že při této volbě jsou hodnoty  $\underline{\theta}$  a  $\underline{s}^{-2}$  irelevantní.

nejprve využita pro volbu apriorních hyperparametrů a po jejich vobě jsou tatáž data využita po druhé v rámci standardní bayesiánské analýzy.

Empirické bayesovské metody v sobě zahrnují odhad apriorních hyperparametrů z dat. Ideálním nástrojem pro tento postup tak bývá marginální věrohodnost. Pro jakoukoliv volbu apriorních hyperparametrů jsme schopni spočítat marginální věrohodnost. Hodnoty apriorních hyperparametrů, kterými dostáváme nejvyšší apriorní věrohodnost, pak využijeme v empirické bayesovské analýze. Hledání skrze všechny apriorní hyperparametry však může být velmi náročné. Empirické bayesovské metody se tak využívají nejčastěji pro jeden nebo dva klíčové hyperparametry. Zde si ukážeme postup této analýzy v rámci local level modelu.

Apriorní hustota pro local level model specifikována v (8.10), (8.11) a (8.12) závisí na pěti hyperparametrech  $\eta$ ,  $\underline{\theta}_1$ ,  $\underline{V}_{11}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$ . Z nich je nejvýznamnějším parametrem  $\eta$  a zdá se tak být největším kandidátem na použití empirického bayesovského přístupu. Tento parametr lze interpretovat jako parametr vztahující se k velikosti komponenty náhodné procházky ve stavovém modelu a lze tak pro ni obtížně definovat subjektivně nějakou hodnotu. Navíc zjevná "neinformativní" limitní volba  $\eta \rightarrow \infty$  nedává velký smysl, protože to impikuje fakt, že stochastický trend převáží nad nesystematickou (nepravidelnou) komponentou. To je dosti „informativní“ tvrzení. Zaměříme se tedy na  $\eta$ . Budeme předpokládat, že jsme schopni subjektivně definovat hodnoty pro  $\underline{\theta}_1$ ,  $\underline{V}_{11}$ ,  $\underline{s}^{-2}$  a  $\underline{\nu}$ .

Výsledky kapitoly 3 (viz (3.34)) implikují, že marginální věrohodnost pro analyzovaný model má podobu:

$$p(y|\eta) = c \left( \frac{|\bar{V}|}{|V|} \right)^{\frac{1}{2}} (\underline{\nu s^2})^{-\frac{\nu}{2}} \quad (8.19)$$

kde

$$c = \frac{\Gamma(\frac{\nu}{2}) (\underline{\nu s^2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2}) \pi^{\frac{\nu}{2}}} \quad (8.20)$$

Značení v (8.19) ukazuje, že bereme marginální věrohodnost jako funkci  $\eta$  (v předchozích kapitolách jsme používali značení  $p(y)$  nebo  $p(y|M_j)$ ), ale tady chceme explicitně vyjádřit její závislost na  $\eta$ . Standardní způsob provedení empirické bayesovské analýzy by byl zvolit takové  $\eta = \hat{\eta}$ , pro které bude maximalizována  $p(y|\eta)$ . Toto  $\hat{\eta}$  by se pak vložilo do (8.12) a posteriorní analýza by proběhla standardním způsobem využitím (8.13)-(8.18). V local level modelu je možno  $\hat{\eta}$  nalézt síťovou vyhledávací metodou (grid search method), kdy se zkusí dát každá potenciální hodnota  $\eta$  do odpovídající sítě či mřížky a zvolí se  $\hat{\eta}$  maximalizující  $p(y|\eta)$ .

Formálnější způsob je pak možný explicitním chápáním  $\eta$  jakožto parametru a využitím zákonů podmíněné pravděpodobnosti pro bayesovskou analýzu. Pokud je  $\eta$  bráno jako neznámý parametr, potom Bayesův teorém implikuje  $p(\eta|y) \propto p(y|\eta)p(\eta)$ , kde  $p(\eta)$  je apriorní hustota, a můžeme tak psát

$$p(\eta|y) = c \left( \frac{|\bar{V}|}{|V|} \right)^{\frac{1}{2}} (\underline{\nu s^2})^{-\frac{\nu}{2}} p(\eta) \quad (8.21)$$

Tato posteriorní hustota může být využita pro analýzu  $\eta$ . Pokud se náš zájem zaměřuje i na ostatní parametry modelu, potom lze využít skutečnosti, že

$$p(\theta, h, \eta|y) = p(\theta, h|y, \eta)p(\eta|y)$$

Jelikož  $p(\theta, h|y, \eta)$  je normální-gama hustota (podmíněná specifickou hodnotou  $\eta$ , čímž platí posteriorní výsledky v (8.13)-(8.18)) a  $p(\eta|y)$  je jednorozměrná hustota, můžeme využít Monte Carlo integraci pro posteriorní analýzu tohoto modelu. Provedeme tak náhodný výběr z  $p(\eta|y) \propto p(y|\eta)p(\eta)$  a podmíněno tímto výsledkem provedeme náhodný výběr z  $p(\theta, h|y, \eta)$ , čímž získáme výběr ze sdružené posteriorní hustoty. To, jak dostaneme výběr z  $p(\eta|y)$ , bude záviset na přesné podobě  $p(\eta)$ . Jednoduchý způsob výběru z jednorozměrného rozdělení zahrnuje jeho aproximaci diskrétní alternativou. Tedy, vyhodnocení  $p(\eta|y)$  v  $B$  různých bodech mřížky  $\eta_1, \dots, \eta_B$  vede k získání  $p(\eta_1|y), \dots, p(\eta_B|y)$ . Výběry  $\eta$  brané z tohoto diskrétního rozdělení (tj. rozdělení definovaného pravděpodobnostmi  $p(\eta = \eta_i) = p(\eta_i|y)$  pro  $i = 1, \dots, B$ ) budou aproximativně odpovídat výběrům z  $p(\eta|y)$ . Pro rostoucí  $B$  se bude zvyšovat i kvalita aproximace.

Popsané empirické bayesovské metody pro local level model vyžadují výběr  $\underline{\theta}_1, \underline{V}_{11}, \underline{s}^{-2}$  a  $\underline{\nu}$  (případně  $p(\eta)$  při využití druhého přístupu zmiňovaného v předchozím odstavci). Obvyklá je volba neinformativních apriorních hyperparametrů a ve většině modelů s hierarchickou apriorní hustotou tato volba funguje dobře (např. model pro panelová data z kapitoly 7). Ovšem pro případ local level modelu tento postup nefunguje. Podrobněji si tedy řekneme, proč tomu tak je, což zároveň bude ilustrovat problém, který může vzniknout při bayesovské analýze modelů s velkým počtem parametrů.

Uvažujme, co se stane, když nastavíme  $\underline{\nu}$  a  $\underline{V}_{11}^{-1}$  na jejich limitní hodnoty  $\underline{\nu} = \underline{V}_{11}^{-1} = 0$ . V takovémto případě budou hodnoty  $\underline{s}^{-2}$  a  $\underline{\theta}_1$  irelevantní. Pro tuto neinformativní volbu lze přímo ověřit, že  $p(\theta, \sigma^{-2}|y, \eta)$  je dobře definovaná posteriorní hustota. S ohledem na marginální věrohodnost však nastávají dva problémy.

První problém je, že integrační konstanta v (8.20) je nedeterminovatelná. To je standardní problém diskutovaný v kapitole 2 v oddíle zabývajícím se porovnáním modelů. Pokud se však zaměříme jen na  $\eta$ , popřípadě je marginální věrohodnost využita pouze pro porovnání současného modelu s jiným s touž neinformativní apriorní hustotou, není tento první problém nijak vážný. Konstanta  $c$  buď do vztahů nevstoupí nebo se vykrátí v rámci např. Bayesova faktoru a může být tedy ignorována.

Druhý problém spočívá v tom, že člen  $\overline{\nu s^2}$  se blíží nule pro  $\eta \rightarrow \infty$ . Abychom to viděli, poznamenejme, že pokud jsou všechny hyperparametry nastaveny na své neinformativní hodnoty,  $\bar{\theta} = (W'W)^{-1}W'y$  a  $y - W\bar{\theta} = 0_T$ . Nebudeme uvádět formální důkaz, nicméně se jedná o případ, kdy tato degenerativnost je dostatečná k tomu, aby marginální věrohodnost byla nekonečno pro  $\eta \rightarrow \infty$ . Empirická bayesovská analýza tak nastaví  $\hat{\eta} \rightarrow \infty$  pro jakoukoliv datovou sadu. Lze ukázat, že tato volba povede k tomu, že  $E(\alpha|y) = y$  a nenastává tak žádné vyhlazení (smoothing) stavového vektoru. Empirické bayesovské metody selhávají v rámci local level model při nastavení  $\underline{\nu}$  a  $\underline{V}_{11}^{-1}$  na neinformativní

hodnoty. Tento problém (který nevzniká ve většině jiných modelů) nastává z toho důvodu, že počet vysvětlujících proměnných daných v (8.7) je roven počtu pozorování a regresní přímka (či křivka) tak může přesně proložit pozorovaná data. Při práci s nepravými neinformativními apriorními hustotami tak je třeba být velmi obezřetný, jestliže pracujeme s modely s velkým počtem parametrů.

Lze ověřit, že pokud nastavíme buď  $\underline{\nu} > 0$  nebo  $\underline{V}_{-11}^{-1} > 0$  (a zvolíme  $\underline{s}^2$  nebo  $\underline{\theta}_1$  adekvátním způsobem), potom empirické bayesovské metody použít můžeme. Obě volby zamezí tomu, aby  $\underline{\nu s}^2$  v (8.19) se blížilo nule pro  $\eta \rightarrow \infty$ . Není tedy nutné mít informativní apriorní hustotu jako pro  $h$ , tak pro  $\theta_1$ .

V rámci alternativního přístupu, který chápe  $\eta$  jako parametr (viz (8.21)), nastává podobně nepříjemná situace pro volbu  $\underline{\nu} = \underline{V}_{-11}^{-1} = 0$  a pro nastavení nepravého prioru pro  $\eta$ . Pokud například zvolíme  $\underline{\nu} = \underline{V}_{-11}^{-1} = 0$  a ještě budeme brát  $p(\eta)$  jako nepravou apriorní hustotu odpovídající uniformnímu rozdělení na intervalu  $(0, \infty)$ , potom bude platit, že  $p(\eta|y)$  není platná hustota pravděpodobnosti (je tedy nepravá). Pokud však zvolíme buď  $\underline{\nu} > 0$  nebo  $\underline{V}_{-11}^{-1} > 0$  anebo  $p(\eta)$  jako platnou p.d.f., potom i  $p(\eta|y)$  je platná posteriorní hustota. Jestliže tedy budeme chápat  $\eta$  jako neznámý parametr, lze provést bayesiánskou analýzu v tom případě, kdy bude k dispozici informativní apriorní hustota jednoho z parametrů  $\eta$ ,  $h$  nebo  $\theta_1$ .

### 8.2.4 Empirická ilustrace: Local level model

BUDE DOPLNĚNO!

## 8.3 Obecný stavový model

V tomto oddíle se zaměříme na obecnější stavový model ve tvaru

$$y_t = X_t \beta + Z_t \alpha_t + \epsilon_t \quad (8.22)$$

kde

$$\alpha_{t+1} = T_t \alpha_t + u_t \quad (8.23)$$

Tento model využívá trochu odlišný zápis než v případě local level modelu, kdy předpokládá, že  $\alpha_t$  je  $p \times 1$  rozměrný vektor obsahující  $p$  stavových rovnic. Předpokládáme, že  $\epsilon_t$  je i.i.d.  $N(0, h^{-1})$ , ovšem  $u_t$  je nyní  $p \times 1$  rozměrný vektor, který je i.i.d.  $N(0, H^{-1})$  a  $\epsilon_t$  a  $u_s$  jsou vzájemně nezávislé pro všechna  $s$  a  $t$ .  $X_t$  a  $Z_t$  jsou postupně vektory rozměru  $1 \times k$  a  $1 \times p$  obsahující vysvětlující proměnné a (nebo) známé konstanty. Matice  $T_t$  je matice známých konstant rozměru  $p \times p$ . Případ, kdy  $T_t$  obsahuje neznámé parametry lze řešit způsobem uvedeným níže.

Zde uváděný model není tím zcela nejobecnějším, nicméně i tak je užitečné ilustrovat si v tomto kontextu několik speciálních případů. Local level model je speciálním případem při volbě  $p = 1$ ,  $k = 0$ ,  $T_t = 1$  a  $Z_t = 1$ , tudíž je možné i tento stavový model využít pro dekompozici časové řady na stochastický trend a nepravidelnou (nesystematickou) složku. Dokonce i normální lineární regresní

model můžeme snadno získat volbou  $Z_t = 0$ . Normální lineární regresní model s v čase proměnnými parametry pak získáme tehdy, pokud  $Z_t$  obsahuje některé nebo všechny vysvětlující proměnné. Existují i tzv. *strukturální modely časových řad*, které lze převést do podoby (8.22) a (8.23). Durbin a Koopman [9] analyzují tento typ modelů, a to včetně problematiky sezónnosti a jak lze běžně používané ARIMA modely (autoregresivní integrované modely klouzavých součtů) převést do podoby stavových modelů. Na tomto místě si ukážeme, jak jeden z běžných strukturálních modelů časových řad zvaný *local linear trend model* můžeme přepsat do stavového tvaru. Tento model je podobný local level modelu, ovšem zahrnuje navíc v sobě lineární trend. Tedy

$$\begin{aligned}y_t &= \mu_t + \epsilon_t \\ \mu_{t+1} &= \mu_t + \nu_t + \xi_t\end{aligned}$$

a

$$\nu_{t+1} = \nu_t + \zeta_t$$

kde  $\xi_t$  je i.i.d.  $N(0, \sigma_\xi^2)$ ,  $\zeta_t$  je i.i.d.  $N(0, \sigma_\zeta^2)$  a všechny náhodné chyby jsou vzájemně nezávislé. Je možno vidět, že tento local linear trend model lze přepsat do podoby stavového modelu volbou

$$\begin{aligned}\alpha_t &= \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} \\ u_t &= \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix} \\ T_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ Z_t &= (1 \quad 0) \\ H^{-1} &= \begin{pmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}\end{aligned}$$

a  $\beta = 0$ . Krátce a stručně shrnuto: řada užitečných regresních modelů i modelů časových řad lze zapsat jako stavové modely.

### 8.3.1 Bayesovský výpočet pro stavový model

Jednou z výhod bayesovského přístupu je to, že metody posteriorní analýzy v řadě komplikovaných modelů lze odvodit jednoduchou kombinací výsledků jednodušších modelů. Stavový model je dobrým příkladem tohoto postupu. Aniž bychom si postupně odvozovali věrohodnostní funkci, apriorní hustoty a posteriorní hustoty, rovnou přeskočíme na problém bayesovského výpočtu a ukážeme si, jak využít výsledky předchozích kapitol pro tento stavový model. Jak je možno vidět, pro posteriorní simulaci nastává komplikace toho rázu, že posteriorní hustoty pro  $\alpha$  (analogické podobě jako v kapitole 7), (8.23) nebudou v čase nezávislé. To implikuje, že  $\alpha_t$  a  $\alpha_{t-1}$  nebudou vzájemně nezávislé. Nejsme tak

schopni najednou generovat výběry pro  $\alpha_t$  a případná přímá implementace Gibbsova vzorkovače by zahrnovala výběry z  $T$ -rozměrného normálního rozdělení. To by mohlo být poněkud zdlouhavé, nicméně De Jong a Shephard [7] popisují efektivní metodu Gibbsova vzorkovače pro tuto třídu modelů.

Bližší pohled na (8.23) naznačuje, že pokud by  $\alpha_t$  pro  $t = 1, \dots, T$  bylo známé (jakožto protiklad tomu, že je nepozorováno), potom by se stavový model redukoval na normální lineární regresní model:

$$y_t^* = X_t\beta + \epsilon_t$$

kde  $y_t^* = y_t - Z_t\alpha_t$ . Všechny výsledky pro normální lineární regresní model předchozích kapitol by tak byly využitelné, pouze vysvětlovaná proměnná by byla  $y_t^*$  místo  $y$ . To naznačuje, že pro stavový model by mohl být využit Gibbsův vzorkovač s obohacenými daty (*data augmentation*)<sup>25</sup>. V závislosti na zvolené apriorní hustotě,  $p(\beta, h|y, \alpha_1, \dots, \alpha_T)$  bude mít jednu z jednoduchých podob daných v kapitole 3 nebo 4. Pokud by podobně byly známé  $\alpha_t$  pro  $i = 1, \dots, T$ , potom stavové rovnice dané v (8.23) jsou jednoduchou variantou modelu zdánlivě nesusouvisejících regresí (SUR) diskutovaných v kapitole 6 a  $p(H|y, \alpha_1, \dots, \alpha_T)$  má známou formu.<sup>26</sup> Pokud tedy jsme schopni odvodit metodu pro generování náhodných výběrů z podmíněné hustoty  $p(\alpha_1, \dots, \alpha_T|y, \beta, h, H)$ , budeme mít plně specifikovaný Gibbsův vzorkovač s obohacenými daty, jenž nám umožňuje bayesovskou analýzu stavového modelu. V následující části bude ukázán takovýto Gibbsův vzorkovač pro specifickou volbu apriorní hustoty, nicméně lze využít i jiných apriorních hustot, a to s minimálními dopady a změnami.

V našem případě využijeme pro parametry  $\beta$  a  $h$  nezávislou normální-gama apriorní hustotu, pro matici  $H$  Wishartovu apriorní hustotu a pro  $\alpha_1, \dots, \alpha_T$  apriorní hustotu implikovanou stavovou rovnicí. Konkrétně tedy předpokládáme apriorní hustotu v podobě

$$p(\beta, h, H, \alpha_1, \dots, \alpha_T) = p(\beta)p(h)p(H)p(\alpha_1, \dots, \alpha_T|H),$$

kde

$$p(\beta) = f_N(\beta|\underline{\beta}, \underline{V}), \quad (8.24)$$

$$p(h) = f_G(h|\underline{s}^{-2}, \underline{\nu}) \quad (8.25)$$

a

$$p(H) = f_W(H|\underline{\nu}_H, \underline{H}) \quad (8.26)$$

<sup>25</sup>Pod pojmem *data augmentation* je obvykle chápána metoda pro konstrukci iterativních algoritmů (jakým je rozhodně Gibbsův vzorkovač) skrze zavedení nepozorovaných dat či latentních proměnných (resp. latentních dat). Jako český ekvivalent metodám s *data augmentation* budu používat pojmu metody s „obohacenými daty“ či „pomocnými proměnnými“ (*auxiliary variables*).

<sup>26</sup>Případ, kdy  $T_t$  obsahuje neznámé parametry, by v sobě zahrnoval výběry z podmíněné hustoty  $p(H, T_1, \dots, T_t|y, \alpha_1, \dots, \alpha_T)$ , což lze snadno provést. V obvyklém "v čase neměnném" případě, kde  $T_1 = \dots = T_t$ ,  $p(H, T_1, \dots, T_t|y, \alpha_1, \dots, \alpha_T)$  bude mít právě podobu SUR modelu.

Pro prvky stavového vektoru zacházíme s (8.23) jako s hierarchickou apriorní hustotou. Bereme-li v tomto případě časový index jakožto index začínající od nuly (tzn.  $t = 0, 1, \dots, T$ ) a předpokládáme-li, že  $\alpha_0 = 0$ , potom nám stavová rovnice dokonce poskytuje apriorní hustotu pro počáteční podmínku. Formálně můžeme tuto část apriorní hustoty psát jako

$$p(\alpha_1, \dots, \alpha_T | H) = p(\alpha_1 | H) p(\alpha_2 | \alpha_1, H) \dots p(\alpha_T | \alpha_{T-1}, H),$$

kde pro  $t = 1, \dots, T - 1$

$$p(\alpha_{t+1} | \alpha_t, H) = f_N(\alpha_{t+1} | T_t \alpha_t, H) \quad (8.27)$$

a

$$p(\alpha_1) = f_N(\alpha_1 | 0, H) \quad (8.28)$$

Poznamenejme, že  $H$  hraje podobnou roli jako  $\eta$  v local level modelu. V tomto případě je ovšem  $H$  matice rozměru  $p \times p$ , tudíž využití empirických bayesovských metod by bylo v případě tohoto modelu obtížné. Navíc nevyužíváme přirozeně konjugované apriorní hustoty, což má za následek neexistenci analytického řešení.

Výše uvedené úvahy napovídají, že směřujeme ke Gibbsově vzorkovači s obohacenými daty, který bude využívat sekvenční výběry z  $p(\beta | y, \alpha_1, \dots, \alpha_T)$ ,  $p(h | y, \alpha_1, \dots, \alpha_T)$ ,  $p(H | y, \alpha_1, \dots, \alpha_T)$  a  $p(\alpha_1, \dots, \alpha_T | y, \beta, h, H)$ . První tři posteriorní hustoty lze odvodit z předchozích kapitol. Z kapitoly 4 (oddíl 4.1) víme, že

$$\beta | y, h, \alpha_1, \dots, \alpha_T \sim N(\bar{\beta}, \bar{V}) \quad (8.29)$$

a

$$h | y, \beta, \alpha_1, \dots, \alpha_T \sim G(\bar{s}^{-2}, \bar{\nu}), \quad (8.30)$$

kde

$$\bar{V} = \left( \underline{V}^{-1} + h \sum_{t=1}^T X_t' X_t \right)^{-1}, \quad (8.31)$$

$$\bar{\mathbf{1}} = \bar{V} \left( \underline{V}^{-1} \underline{\beta} + h \sum_{t=1}^T X_t' (y_t - Z_t \alpha_t) \right), \quad (8.32)$$

$$\bar{\nu} = T + \underline{\nu} \quad (8.33)$$

a

$$\bar{s}^2 = \frac{\sum_{t=1}^T (y_t - X_t \beta - Z_t \alpha_t)^2 + \underline{\nu} \underline{s}^2}{\bar{\nu}}. \quad (8.34)$$

Za využití výsledků pro SUR model (bez vysvětlujících proměnných) z kapitoly 6 získáváme

$$H | y, \alpha_1, \dots, \alpha_T \sim W(\bar{\nu}_H, \bar{H}), \quad (8.35)$$

kde

$$\bar{\nu}_H = T + \underline{\nu}_H \quad (8.36)$$

a

$$\bar{H} = \left[ \underline{H}^{-1} + \sum_{t=0}^{T-1} (\alpha_{t+1} - T_t \alpha_t)(\alpha_{t+1} - T_t \alpha_t)' \right]^{-1} \quad (8.37)$$

Pro úplnost Gibbsova vzorkovače potřebujeme odvodit podmíněnou posteriorní hustotu  $p(\alpha_1, \dots, \alpha_T | y, \beta, h, H)$  a způsob jak z této hustoty generovat náhodné výběry. Ačkoliv není těžké zapsat tuto hustotu jako vícerozměrné normální rozdělení, je obtížné získat z ní prakticky náhodné výběry, neboť se jedná o  $T$ -rozměrné rozdělení a její prvky jsou vysoce korelovány. V této souvislosti existuje řada příspěvků snažících se nalézt efektivní způsob generování náhodných výběrů z tohoto rozdělení (významnými příspěvky jsou Carter a Kohn [4] a DeJong a Shephard [7]). Zde budeme prezentovat metodu DeJonga a Shepharda, která se ukázala v řadě praktických aplikací jako vysoce efektivní. Důkazy a další odvození jsou obsaženy v příslušném příspěvku. Oba pánové pracují s trochu odlišnou obecnou verzí stavového modelu v podobě

$$y_t = X_t \beta + Z_t \alpha_t + G_t \nu_t \quad (8.38)$$

a

$$\alpha_{t+1} = T_t \alpha_t + J_t \nu_t \quad (8.39)$$

pro  $t = 1, \dots, T$  v (8.38) a  $t = 0, \dots, T$  v (8.39) a  $\alpha_0 = 0$ . Náhodná složka  $\nu_t$  je i.i.d.  $N(0, h^{-1} I_{p+1})$ . Ostatní proměnné a parametry jsou definovány stejně jako v původní definici stavového modelu. Lze ukázat, že tento model je ekvivalentní s původní formulací pokud zvolíme

$$\nu_t = \begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix}$$

$G_t$  bude řádkový vektor rozměru  $(p+1)$  daný jako

$$G_t = (1 \quad 0 \quad \dots \quad 0)$$

a  $J_t$  je matice rozměru  $p \times (p+1)$  daná jako

$$J_t = [0_p \quad A],$$

kde  $A$  je matice rozměru  $p \times p$  implicitně definována vztahem

$$H^{-1} = \frac{1}{h} A A'$$

Jelikož Gibbsův vzorkovač v sobě zahrnuje výběry z podmíněné posteriorní hustoty  $p(\alpha_1, \dots, \alpha_T | y, \beta, h, H)$ , vše v (8.38) a (8.39) může být bráno jako známé hodnoty, s výjimkou  $\alpha_t$  a  $\nu_t$ . Příspěvkem DeJonga a Shepharda [7] je návrh efektivního algoritmu<sup>27</sup> pro výběry  $\eta_t = F_t \nu_t$  pro různé volby  $F_t$ . Výběry z  $\eta_t$  pak lze transformovat do výběrů z  $\alpha_t$ . Jejich algoritmus je nastaven pro arbitrární

<sup>27</sup>Existují i další výhody tohoto algoritmu, zahrnující požadavky na paměť počítače a vyhnutí se určitým případům degenerace.



$F_t$ , ovšem je třeba poznamenat, že obvyklou volnou je  $F_t = J_t$ , čímž získáváme výběry chyb stavové rovnice, které lze přímo transformovat do požadovaných výběrů z  $\alpha_t$ .

DeJong a Shephard nazývají svůj algoritmus jako simulační vyhlazovač (*simulation smoother*). Tento simulační smoother začíná nastavením  $a_1 = 0$ ,  $P_1 = J_0 J_0'$  a výpočtem následujících veličin pro  $t = 1, \dots, T$ :<sup>28</sup>

$$e_t = y_t - X_t \beta - Z_t \alpha_t \quad (8.40)$$

$$D_t = Z_t P_t Z_t' + G_t G_t' \quad (8.41)$$

$$K_t = (T_t P_t Z_t' + J_t G_t') D_t^{-1} \quad (8.42)$$

$$a_{t+1} = T_t a_t + K_t e_t \quad (8.43)$$

a

$$P_{t+1} = T_t P_t (T_t - K_t Z_t')' + J_t (J_t - K_t G_t)' \quad (8.44)$$

potom si uložíme získané veličiny  $e_t$ ,  $D_t$  a  $K_t$ . Následně je vypočítána nová sada veličin v obrácené časové posloupnosti (tj.  $t = T, T-1, \dots, 1$ ). Tento postup začíná nastavením  $r_T = 0$  a  $U_T = 0$  a výpočtem

$$C_t = F_t (I - G_t' D_t^{-1} G_t - [J_t - K_t G_t]' U_t [J_t - K_t G_t]) F_t' \quad (8.45)$$

$$\xi_t \sim N(0, h^{-1} C_t) \quad (8.46)$$

$$V_t = F_t (G_t' D_t^{-1} Z_t + [J_t - K_t G_t]' U_t [T_t - K_t Z_t]) \quad (8.47)$$

$$r_{t-1} = Z_t' D_t^{-1} e_t + (T_t - K_t Z_t)' r_t - V_t' C_t^{-1} \xi_t \quad (8.48)$$

$$U_{t-1} = Z_t' D_t^{-1} Z_t + (T_t - K_t Z_t)' U_t (T_t - K_t Z_t) + V_t' C_t^{-1} V_t \quad (8.49)$$

a

$$\eta_t = F_t (G_t' D_t^{-1} e_t + [J_t - K_t G_t]' r_t) + \xi_t \quad (8.50)$$

kde  $G_0 = 0$ . Tímto algoritmem získáme  $\eta = (\eta_0, \dots, \eta_T)'$  a lze dokázat, že se jedná o náhodný výběr z  $p(\eta|y, \beta, h, H)$ . V závislosti na podobě  $F_t$  lze tento výběr transformovat na požadovaný náhodný výběr  $\alpha_t$  pro  $t = 1, \dots, T$ . Při obvyklé volbě  $F_t = J_t$  tento algoritmus poskytuje výběry náhodných chyb ve stavové rovnici (tj.  $\eta_t = J_t \nu_t$ ), které lze transformovat na výběry z  $\alpha_t$  za využití (8.39) a skutečnosti, že  $\alpha_0 = 0$ .

Tyto vzorce vypadají složitě. Nicméně algoritmus je jednoduchou sérií výpočtů zahrnující matice nízkých rozměrů a náhodný výběr z normálního rozdělení pro získání  $\xi_t$ . To velmi zrychluje výpočet neboť práce s maticemi velkých rozměrů (tj.  $T \times T$ ) je dosti pomalá. Pro většinu aplikací budou mít matice  $F_t$ ,  $G_t$ ,  $J_t$  a  $T_t$  jednoduchou podobu, což předchází rovnice ještě více zjednodušuje. Naprogramovat si takovýto Gibbsův vzorkovač tak není nijak obtížné.

Byl tedy odvozen Gibbsův vzorkovač s obohacenými daty, který prováděl sekvenční výběry z  $p(\beta|y, \alpha_1, \dots, \alpha_T)$ ,  $p(h|y, \alpha_1, \dots, \alpha_T)$ ,  $p(H|y, \alpha_1, \dots, \alpha_T)$  a

<sup>28</sup>Tyto výpočty jsou označovány jako běh Kalmanova filtru.

$p(\alpha_1, \dots, \alpha_T | y, \beta, h, H)$ . Na základě výstupů takového posteriorního simulátoru, lze povést posteriorná analýzu obdobnou té z kapitoly 4. Stejně tak je možno provést predikční analýzu, spočítat prediční  $p$ -hodnoty nebo HPDI. Marginální věrohodnost pro stavový model lze vypočítat pomocí Chibovy metody, je obdobná jako její aplikace pro model individuálních vlivů, kdy  $\alpha_1, \dots, \alpha_T$  jsou brány jako latentní data (kapitola 7).

### 8.3.2 Empirická ilustrace

BUDE DOPLNĚNO!

## 8.4 Rozšíření

BUDE DOPLNĚNO!

## 8.5 Shrnutí

BUDE DOPLNĚNO!

## Kapitola 9

# Modely kvalitativních a omezených vysvětlovaných proměnných

### 9.1 Úvod

Normální lineární regresní model je mocný analytický nástroj využitelný pro analýzy založené na široké paletě dat. Má však jedno velké omezení, a to takové, že rozdělení vysvětlované proměnné  $y$  (podmíněné maticí vysvětlujících proměnných  $X$ ) odpovídá normálnímu rozdělení. Pro řadu aplikací je tento předpoklad nereálný. Tato kapitola je tedy věnována určitým typům dat (a na nich založeným modelům), pro které není použití normálního lineárního regresního modelu adekvátní. Ukážeme si však, že je možné koncept normálního lineárního regresního modelu rozšířit do takové podoby, že již bude přiměřená těmto „ne-standardním“ typům dat.

„Nestandardními“ typy je zde myšleno to, že vysvětlovaná proměnná je kvalitativní povahy případně je nějakým způsobem omezena. Jako malou ilustraci je možno na úvod zmínit příklad z ekonomie dopravy. Může nás zajímat, proč někteří lidé volí jako dopravní prostředek pro svou cestu do práce auto a proč jiní cestují za prací veřejnou dopravou. Data, na kterých je takováto analýza postavena obvykle pocházejí z nějakého druhu výběrového (dotazníkového) šetření, kdy jsou respondenti dotazováni na to, jestli do práce jezdí autem nebo veřejnou dopravou a současně s tím poskytují o sobě i další více či méně osobnější údaje (např. jak daleko to do práce z domova mají, jaká je jejich mzda atd.). Pokud bychom chtěli sestavit regresní model, byly by vysvětlujícími proměnnými právě tyto osobní charakteristiky. Vysvětlovaná proměnná by však měla kvalitativní podobu. Jednalo by se o umělou proměnnou, která by nabývala hodnoty 1, pokud respondent dojíždí do práce autem, a hodnotu 0, pokud volí hromadnou dopravu. Bylo by asi dosti naivní předpokládat, že takováto umělá proměnná

(podmíněná vysvětlujícími proměnnými) bude pocházet z normálního rozdělení.

Jako druhý příklad uvažujme teoretický model, který zkoumá závislost žádoucí úrovně investic nějaké firmy na jejich různých charakteristikách. V odpovídajícím empirickém modelu tak bude požadovaná úroveň investic firmy vysvětlovaná proměnná a vysvětlující proměnné budou ony různé charakteristiky firmy. V praxi jsou však data o požadovaných investicích zřídka kdy dostupná. Místo toho pozorujeme skutečně realizované investice firmy. Pokud nepřipadají v úvahu negativní investice, potom je aktuální úroveň investic rovna požadovaným investicím jen v případě, pokud jsou požadované investice kladné. Negativní hodnota požadovaných investic odpovídá nulovým hodnotám skutečných investic. Pokud bychom tak pro regresní analýzu využívali skutečné investice jako závisle proměnnou, dopouštěli bychom se vážné chyby (právě z toho důvodu že nulové skutečně pozorované investice neodpovídají nulovým požadovaným investicím). V tomto případě je totiž závisle proměnná omezená (*censored*). Jedná se o příklad modelu s omezenou vysvětlovanou proměnnou (*limited dependent variable*).

Oba tyto příklady implikují to, že model obsahuje latentní (nepozorovanou) závisle proměnnou, u které můžeme vcelku rozumně předpokládat její normalitu. V prvním případě odpovídají tato latentní data užítku, který svou volbou toho či onoho dopravního prostředku dosahuje každý z dotazovaných respondentů. V druhém případě je latentní proměnnou požadovaná úroveň investic. Bohužel však v žádném z těchto příkladů nejsme schopni latentní data zcela perfektně pozorovat. V prvním případě pozorujeme jen skutečnou volbu respondenta, ne užitek, který mu tato volba přináší. V druhém případě pozorujeme latentní proměnnou v omezené (cenzorované) podobě. To, že jsme schopni tyto příklady interpretovat ve vztahu k latentním datům, nám dává vodítko k tomu, jak přistupovat k bayesovské analýze tohoto problému. Podmíníme-li naše analýzy latentními daty, bude se v každém případě jednat o normální lineární regresní model a techniky z předchozích kapitol budou více než využitelné pro posteriorní simulaci. Budeme-li schopni odvodit podmíněnou posteriorní hustotu latentních dat podmíněnou skutečně pozorovanými daty a parametry modelu, jsme schopni v rámci bayesovské analýzy nasadit Gibbsův vzorkovač s obohacenými daty (*data augmentation*)<sup>29</sup>. To je přesně postup identifikace modelů v této kapitole.

V další části bude trochu formálněji vyjádřen postup naznačený v předchozích odstavcích. Bude ukázáno, jak lze tuto obecnou strategii implementovat do tří typových modelů známých jako *tobit*, *probit* a *uspořádaný (ordered) probit*. Dále budeme předpokládat případ, kdy je vysvětlovaná proměnná vícerozměrná a zaměříme se na model známý jako *multinomiální (multinomial) probit*. Bayesovská analýza modelů *tobit*, *probit* a *uspořádaný probit* zahrnuje kombinaci metod pro normální lineární regresní model s modelem spojujícím latentní a pozorovaná data. Bayesovská analýza multinomiálního *probit* modelu je po-

<sup>29</sup>Jen pro připomenutí, pod pojmem *data augmentation* je obvykle chápána metoda pro konstrukci iterativních algoritmů skrze zavedení nepozorovaných dat či latentních proměnných (resp. latentních dat). Jako český ekvivalent metodám s *data augmentation* budu používat metody s „obohacenými daty“ či „pomocnými proměnnými“ (*auxiliary variables*).

dobná s tou výjimkou, že normální lineární regresní model je nahrazen modelem zdánlivě nesouvisějících regresí (SUR).

Přestože klíčovým prvkem této kapitoly bude (normální) lineární regrese (a zaměříme se tak na třídy modelů tobit a probit), je třeba zdůraznit, že existuje celá řada dalších modelů využívaných pro práci s kvalitativními a omezenými vysvětlovanými proměnnými, přičemž tyto modely již nejsou úzce spjaty s normálním lineárním regresním modelem. K těmto modelům existuje velmi pestrá literatura. Jako příklad využití multinomiálních probit modelů s panelovými daty zmiňme oblast marketingu, pro kterou je tento typ modelů zcela běžný.

## 9.2 Jednorozměrné modely pro kvalitativní a omezené vysvětlované proměnné

Pro začátek vyjděme z nám dobře známého normálního regresního modelu. V rámci značení však učiníme drobnou změnu, a to takovou, že vysvětlovaná proměnná bude označována jako  $y^* = (y_1^*, \dots, y_N^*)'$ . Model tak zapíšeme jako

$$y_i^* = x_i' \beta + \epsilon_i, \quad (9.1)$$

kde  $x_i = (1, x_{i2}, \dots, x_{ik})'$ . V maticovém vyjádření je značení

$$y^* = X\beta + \epsilon. \quad (9.2)$$

Na vektor náhodných složek  $\epsilon$  budeme klást standardní předpoklady, tedy

1. vektor náhodných složek  $\epsilon$  je z vícerozměrného normálního rozdělení se střední hodnotou  $0_N$  a kovarianční maticí  $h^{-1}I_N$ ,
2. všechny prvky matice  $X$  jsou pevná čísla (tj. nenáhodné veličiny). V případě náhodných veličin předpokládáme, že tyto jsou nezávislé na všech prvcích vektoru  $\epsilon$  a jejich funkce hustoty pravděpodobnosti je  $p(X|\lambda)$ , kde  $\lambda$  je vektor parametrů, který neobsahuje  $\beta$  ani  $h$ .

Jestliže bychom byli schopni veličinu  $y^*$  pozorovat, analýza by probíhala v intencích předchozích kapitol. V této kapitole však budeme předpokládat, že  $y^*$  obsahuje nepozorovaná, latentní data, která jsou nějakým způsobem propojena se skutečně pozorovanými daty obsaženými ve vektoru  $y$ . Aby byla zaručena „funkčnost“ dále popisovaných metod, musíme předpokládat, že vztah mezi  $y^*$  a  $y$  je takový, že  $p(\beta, h|y^*, y) = p(\beta, h|y^*)$  (pokud pracujeme s přirozeně konjugovanou apriorní hustotou jako v kapitole 3) nebo  $p(\beta|y^*, y, h) = p(\beta|y^*, h)$  a  $p(h|y^*, y, \beta) = p(h|y^*, \beta)$  (pracujeme-li s nezávislou normální-gama apriorní hustotou jak je tomu v kapitole 4). Tyto podmínky nám neříkají nic jiného než to, že pokud bychom pozorovali  $y^*$ , nepřineslo by nám dodatečné pozorování  $y$  žádnou novou informaci. Tato podmínka je platná pro celou řadu užitečných modelů včetně těch diskutovaných v této kapitole.

Pokud jsou tedy výše uvedené podmínky pro vztah mezi  $y$  a  $y^*$  splněny, je možno provést bayesovskou analýzu s využitím Gibbsova vzorkovače s obohacenými daty. V případě neinformativní apriorní hustoty (popř. její neinformativní verze) provádí posteriorní simulátor postupné výběry z  $p(\beta, h|y^*)$  a  $p(y^*|y, \beta, h)$ . V případě nezávislé normální-gama apriorní hustoty je posteriorní simulace založena na postupných výběrech z  $p(\beta|y^*, h)$ ,  $p(h|y^*, \beta)$  a  $p(y^*|y, \beta, h)$ . V obou případech je jedinou novou podmíněnou hustotou, kterou musíme odvodit (ostatní odpovídají těm z kapitol 3 a 4), hustota  $p(y^*|y, \beta, h)$ . Z tohoto důvodu se zaměříme pouze na tuto podmíněnou hustotu.

### 9.3 Tobit model

Tobit model je jednoduchým příkladem modelu, ve kterém jsou data nějakým způsobem omezena. Příkladem může být v úvodu kapitoly naznačený problém skutečných investic, které jsou omezeným pozorováním požadované úrovně investic. Tento případ tedy uvažuje vztah mezi  $y$  a  $y^*$  v podobě

$$\begin{aligned} y_i &= y_i^* & \text{pokud } y_i^* > 0 \\ y_i &= 0 & \text{pokud } y_i^* \leq 0 \end{aligned} \tag{9.3}$$

Asi není překvapením, že pokud bychom znali  $y^*$ , potom bychom znali rovněž i  $y$ . Z tohoto důvodu by  $p(\beta, h|y^*) = p(\beta, h|y, y^*)$  a mohli bychom pro získání výběrů z podmíněné posteriorní hustoty parametrů (podmíněné  $y^*$ ) snadno využít výsledky z kapitol 3 a 4. Nicméně většinou tato optimistická varianta nenastává a musíme tudíž odvodit  $p(y^*|y, \beta, h)$  pro využití Gibbsova vzorkovače.

Posteriorní hustotu latentních dat podmíněnou parametry modelu lze odvodit vcelku jasným způsobem. Poznamenejme nejprve, že jsme předpokládali, že náhodné chyby jsou vzájemně nezávislé a tudíž i latentní data budou mít stejnou vlastnost. Můžeme tak psát

$$p(y^*|y, \beta, h) = \prod_{i=1}^N p(y_i^*|y_i, \beta, h)$$

a zaměřit se tím na  $p(y_i^*|y_i, \beta, h)$ . Musíme vzít do úvahy dva případ, a to případ kdy  $y_i > 0$  a případ, kdy  $y_i = 0$ . První z nich je jednoduchý: pokud je  $y_i > 0$ , potom máme  $y_i^* = y_i$ . Formálně řečeno, jestliže je  $y_i > 0$ , potom je podmíněná posteriorní hustota pro  $y_i^*$  degenerovaná hustota s veškerou pravděpodobností směřovanou do bodu  $y_i^* = y_i$ . Druhý případ lze řešit kombinací (9.1) (tzn. skutečností, že nepodmíněná hustota  $y_i^*$  je normální) s faktem, že  $y_i = 0$  implikuje nerovnost  $y_i^* \leq 0$ . Proměnná  $y_i^*$  má omezené normální rozdělení v případě, kdy  $y_i = 0$ . Formálně tak můžeme zapsat  $p(y_i^*|y_i, \beta, h)$  jako

$$y_i^*|y_i, \beta, h \sim \begin{cases} y_i^* = y_i & \text{pokud } y_i > 0 \\ N(x_i'\beta, h^{-1})1(y_i^* < 0) & \text{pokud } y_i = 0 \end{cases} \tag{9.4}$$

přičemž  $1(y_i^* < 0)$  je indikační funkce, která je rovna jedné pro  $y_i^* < 0$  a nule jinak.

Posteriorní analýza tobit modelu může být úspěšně provedena za pomoci Gibbsova vzorkovače, který kombinuje výsledky předchozích kapitol s (9.4). Lze zde aplikovat všechny doposud probírané nástroje pro porovnání modelů a predikci využívajících MCMC algoritmus. Savage-Dickey density ratio využijeme k výpočtu Bayesova faktoru porovnávajícího různé modely, případně jako alternativu lze pro výpočet matginální věrohodnosti tobit modelu použít Chibovu metodu. Predikční analýza je opět zřejmou aplikací výsledků popsaných v části 4.1.6. Nesmíme samozřejmě zapomenout na ověření konvergence pomocí konvergenčních diagnostik.

V této části jsme předpokládali případ, kdy omezení (cenzoring) závisle proměnné nastává při nulové hodnotě. Stačí však vcelku triviální rozšíření výše uvedeného modelu, abychom umožnili omezení vysvětlované proměnné při nějaké známé hodnotě  $c$ . Stejně tak je možné omezit závisle proměnnou ve známých hodnotách zhora i zdola. Všechna taková rozšíření znamenají toliko změnu bodu(-ů) omezení v rovnici (9.4). Příklad, kdy k omezení dochází při nějaké hodnotě  $c$ , která je ovšem neznámým parametrem je trochu složitější rozšíření výše uvedeného modelu. Nicméně i tento problém lze řešit pomocí dodatečného bloku v rámci Gibbsova vzorkovače. Příkladem aplikace takového rozšíření tobit modelu ukazuje Li [23].

### 9.3.1 Empirická ilustrace: tobit model

BUDE ČASEM DOPLNĚNO!

## 9.4 Probit model

Probit model se obvykle používá v situacích, kdy je vysvětlovaná proměnná kvalitativního ražení indikující výsledek v rámci jedné či dvou kategorií (např. jednotlivec cestuje do práce vlastním automobilem nebo veřejnou dopravou). Motivace tohoto typu modelu je založena na tom, že jednotlivec provádí v rámci svého rozhodování nějaký druh volby. Samozřejmě probit model je možno využít i v kontextu, kdy je závisle proměnná umělou proměnnou nabývající hodnoty nula nebo jedna.

Předpokládejme nicméně, že se jednotlivec rozhoduje mezi dvěma alternativami. Ekonomickou formalizací takovéto situace by byla specifikace užitkové funkce. Nechť tedy  $U_{ij}$  je užitek, který jednotlivec  $i$  (pro  $i = 1, \dots, N$ ) přiřazuje volbě  $j$  (pro  $j = 0, 1$ ). Jednotlivec tak provede volbu 1 tehdy, pokud  $U_{1i} \geq U_{0i}$  a volbu 0 jinak. Výběr tak závisí na rozdílu v užitech vycházejících z volby jednotlivých alternativ. Tento rozdíl můžeme definovat jako

$$y_i^* = U_{1i} - U_{0i}$$

Probit model předpokládá, že tento rozdíl v užitech lze vysvětlit normálním lineárním regresním modelem definovaným rovnicemi (9.1) nebo (9.2). To tedy znamená, že rozdíl v užitech jednotlivců závisí na pozorovaných charakteristikách v  $x_i$  (např. vzdálenost do práce, úroveň platu, atd.) a chybovém členu,

## 162 Modely kvalitativních a omezených vysvětlovaných proměnných

o kterém předpokládáme, že má normální rozdělení. Díky této náhodné chybové složce jsou probit model a další podobné modely nazývány jako *modely náhodného užítka* (*random utility models*).

Problém je samozřejmě to, že nejsme schopni  $y_i^*$  pozorovat přímo. Jsme schopni pouze pozorovat konkrétní volbu, kterou jednotlivec  $i$  provede. Stejně jako v tobit modelu můžeme  $y^*$  chápat jako proměnnou obsahující latentní data. Na tomto základě samozřejmě budeme schopni k bayesovské analýze použít Gibbsův vzorkovač s obohacenými daty. Z důvodů diskutovaných v části (9.2) potřebujeme odvodit pouze podmíněnou hustotu pro  $y^*$ , tedy  $p(y^*|y, \beta, h)$ .

V rámci probit modelu má vztah mezi  $y$  a  $y^*$  podobu

$$\begin{aligned} lcy_i &= 1 & \text{pokud } y_i^* &\geq 0 \\ y_i &= 0 & \text{pokud } y_i^* &< 0 \end{aligned} \quad (9.5)$$

Opět lze okamžitě vidět, že bychom při znalosti  $y^*$  znali rovněž i  $y$ , a z tohoto důvodu by bylo  $p(\beta, h|y^*) = p(\beta, h|y, y^*)$ . Mohli bychom tak pro získání výběrů z podmíněné posteriorní hustoty parametrů (podmíněné  $y^*$  a závislé na zvolené apriorní hustotě) snadno využít výsledky z kapitol 3 a 4.

Podobu podmíněné hustoty  $p(y^*|y, \beta, h)$  lze odvodit podobným způsobem jako v rámci tobit modelu. Nezávislost pozorování mezi jednotlivci implikuje

$$p(y^*|y, \beta, h) = \prod_{i=1}^N p(y_i^*|y_i, \beta, h)$$

a stačí se tak zaměřit na  $p(y_i^*|y_i, \beta, h)$ . Předpoklady vyřčené v rámci normálního lineárního regresního modelu implikují, že i hustota  $p(y_i^*|\beta, h)$  odpovídá normálnímu rozdělení. K získání  $p(y_i^*|y_i, \beta, h)$  stačí zkombinovat tento výsledek s informací obsaženou v  $y_i$ . Pokud je  $y_i = 1$ , potom získáme normální rozdělení omezené zleva v hodnotě 0. Pokud je naopak  $y_i = 0$ , potom získáme normální rozdělení omezené zprava ve stejné nulové hodnotě. Precizněji to lze vyjádřit následovně:

$$\begin{aligned} y_i^*|y_i, \beta, h &\sim N(x_i'\beta, h^{-1})1(y_i^* \geq 0) & \text{pokud } y_i &= 1 \\ y_i^*|y_i, \beta, h &\sim N(x_i'\beta, h^{-1})1(y_i^* < 0) & \text{pokud } y_i &= 0 \end{aligned} \quad (9.6)$$

Posteriorní analýzu tak lze provést pomocí Gibbsova vzorkovače využívajícím obohacená data, kterým postupně generujeme výběry z (9.6) a  $p(\beta, h|y^*)$  (opět zde využijeme poznatky z kapitol 3 a 4). Porovnání modelů a predikce může být snadno implementována s využitím dříve uváděných nástrojů.

K samotným odhadům parametrů je často užitečné uvádět i informaci o pravděpodobnostech jednotlivých voleb. Tyto pravděpodobnosti lze odvodit z posteriorního rozdělení parametrů, pokud vezmeme v úvahu skutečnost, že pro určité hodnoty parametrů platí

$$\begin{aligned} \Pr(y_i = 1|\beta, h) &= \Pr(y_i^* \geq 0|\beta, h) \\ &= \Pr(x_i'\beta + \epsilon_i \geq 0|\beta, h) = \Pr(\sqrt{h}\epsilon_i \geq -\sqrt{h}x_i'\beta|\beta, h) \end{aligned} \quad (9.7)$$



Protože jsou chybové členy normálně rozdělené, je poslední člen v rovnici (9.7) roven rozdílu jedničky a kumulativní distribuční funkce standardizovaného normálního rozdělení (tj.  $\sqrt{h}\epsilon_i$  odpovídá  $N(0,1)$ ). Pokud budeme definovat kumulativní distribuční funkci standardizovaného normálního rozdělení jako  $\Phi(a)$ , potom pravděpodobnost volby alternativy 1 bude rovna výrazu  $1 - \Phi(-\sqrt{h}x_i'\beta)$ .

Členy ve vztahu (9.7) jsou funkcemi parametrů modelu, čímž lze jejich posteriorní charakteristiky vypočítat standardně s využitím výstupu Gibbsova vzorkovače. Použitím značení z kapitoly 4, části 4.1, odpovídají členy rovnice (9.7) výrazu  $g(\theta)$  pro specifickou volbu funkce  $g(\cdot)$ .

Rovnice (9.7) ilustruje *identifikační problém*, se kterým jsme se nesetkali u tobit modelu. O identifikačním problému hovoříme tehdy, pokud více hodnot parametrů modelu způsobí stejnou hodnotu věrohodnostní funkce. V rámci probit modelu existuje nekonečný počet hodnot parametrů  $\beta$  a  $h$ , které vedou přesně k totožnému modelu. To si můžeme lehce vybavit, pokud si uvědomíme, že  $\Pr(x_i'\beta + \epsilon_i \geq 0 | \beta, h) = \Pr(x_i'c\beta + c\epsilon_i \geq 0 | \beta, h)$  pro jakoukoli kladnou konstantu  $c$ . Protože transformovaná náhodná veličina  $c\epsilon_i$  má rozdělení  $N(0, c^2h^{-1})$ , jedná se o totožné probit modely, lišící se jen různými koeficienty a přesností chyb. Ekvivalentním způsobem důkazu tohoto tvrzení by bylo vyjádření a porovnání příslušných věrohodnostních funkcí. Lze tak ukázat, že hodnota věrohodnostní funkce je totožná pro hodnoty  $(\beta = \beta_0, h = h_0)$  a  $(\beta = c\beta_0, h = \frac{h_0}{c^2})$ , přičemž  $\beta_0$  a  $h_0$  jsou arbitrárně zvolené hodnoty pro  $\beta$  a  $h$ . Jinými slovy, probit model není schopen rozlišit odděleně  $\beta$  a  $h$ , je schopen pouze identifikovat součin  $\beta\sqrt{h}$ . Z ekonomického úhlu pohledu není tento závěr překvapující, protože s podobnou vlastností se lze setkat u užitkové funkce. Pokud je např.  $U(x)$  užitková funkce definovaná na množině kombinací zboží  $x$ , potom i  $cU(x)$  vyjádří naprosto identické spotřebitelské preference týkající se rozhodování mezi jednotlivými kombinacemi statků.

Řešením tohoto problému v rámci probit modelu je nastavení  $h = 1$  (jak tomu je v empirické ilustraci). Alternativním řešením je nastavení jednoho z prvků vektoru  $\beta$  na pevně danou hodnotu (např. nastavení jednoho z koeficientů na hodnotu 1). Ovšem takovéto řešení vyžaduje to, abychom znali znaménko příslušného koeficientu. Pokud totiž nastavíme jeden z koeficientů na hodnotu 1, implikuje nám to, že tato vysvětlující proměnná bude mít pozitivní efekt na úroveň užitku (tzn. vysoké hodnoty této proměnné zvýší pravděpodobnost provedení volby 1). V praxi je tento druh informace (pokud jde o znaménko) zřídka dostupný, a tak je obvykle preferováno normalizování parametru  $h$  na hodnotu jedna, tedy  $h = 1$ .

### 9.4.1 Empirická ilustrace: probit model

BUDE ČASEM DOPLNĚNO!

## 9.5 Uspořádaný probit model

Probit model dovoluje dvě alternativy (např. člověk jezdící do práce volí mezi autem nebo veřejnou dopravou). Řada empirických aplikací však zahrnuje tři nebo více alternativ (cestovat do práce se dá třeba i na kole). V další části bude řeč o multinomiálním probit modelu, který obecně dovoluje zahrnutí více alternativ. Ještě před tím si však zavedeme pojem uspořádaný probit model (*ordered probit model*), který je jednoduchým rozšířením probit modelu z předchozí kapitoly. Uspořádaný probit model dovoluje volbu mezi několika alternativami, ovšem tyto alternativy mají specifickou podobu. Ta spočívá v tom, že tyto alternativy jsme schopni jednoznačně seřadit v ordinálním slova smyslu (což samozřejmě v řadě aplikací nemusí být rozumné). Ovšem v mnoha situacích existuje jejich logické seřazení, z čehož vyplývá, že uspořádaný probit model může být vcelku kvalitním řešením problému. Představme si takový marketingový průzkum, ve kterém jsou spotřebitelé dotazováni na své názory ohledně nějakého konkrétního produktu, přičemž volí mezi alternativami “velmi špatný”, “špatný”, “neutrální”, “dobrý”, “velmi dobrý”. V tomto případě má těchto pět možností vcelku logické pořadí od “velmi špatný” po “velmi dobrý”. V ekonomii práce se uspořádaný probit model využívá k analýze pracovních úrazů, kdy je zraněný pracovník ohodnocen dle vážnosti úrazu (datová množina tak obsahuje řadu kategorií od lehkých zranění, až po zranění velmi vážná).

K popisu uspořádaného probit modelu je nutné zobecnění dosavadního značení. Model lze interpretovat jako normální lineární regresní model, kdy je vysvětlovaná proměnná latentní (jako v (9.2)). Stejně jako v předchozích částech je klíčový vztah mezi  $y^*$  a  $y$ . V uspořádaném probit modelu nabývá  $y_i$  hodnoty  $j = 1, \dots, J$ , kde  $J$  je počet uspořádaných alternativ a získáváme tak vztah

$$y_i = j \quad \text{pokud} \quad \gamma_{j-1} < y_i^* \leq \gamma_j \quad (9.8)$$

kde  $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_J)'$  je vektor parametrů, přičemž  $\gamma_0 \leq \dots \leq \gamma_J$ .

Stejným způsobem jako ve vztahu (9.7) můžeme využít předpokladů normality regresního modelu pro latentní data k vyjádření pravděpodobnosti volby jednotlivých alternativ. Podobně jako v probit modelu je nutno zavést omezení pro identifikovatelnost parametru. Budeme tak předpokládat omezení  $h = 1$ . Dostáváme tak

$$\begin{aligned} \Pr(y_i = j | \beta, \gamma) &= \Pr(\gamma_{j-1} < y_i^* \leq \gamma_j | \beta, \gamma) = & (9.9) \\ \Pr(\gamma_{j-1} < x_i' \beta + \epsilon_i \leq \gamma_j | \beta, \gamma) &= \Pr(\gamma_{j-1} - x_i' \beta < \epsilon_i \leq \gamma_j - x_i' \beta | \beta, \gamma) \end{aligned}$$

Protože chybové členy  $\epsilon_i$  pocházejí z  $N(0, 1)$ , odpovídají pravděpodobnosti volby kumulativní distribuční funkci standardizovaného normálního rozdělení. Konkrétně, při značení definovaném v rámci vztahu (9.7) dostáváme

$$\Pr(y_i = j | \beta, \gamma) = \Phi(\gamma_j - x_i' \beta) - \Phi(\gamma_{j-1} - x_i' \beta) \quad (9.10)$$

$$(9.11)$$

Uspořádaný probit počítá pravděpodobnosti volby pro každého jednotlivce tak, že vyjdeme z normálního rozdělení (jehož integrál přes všechny meze je

roven jedné) a zvolíme  $\gamma_0, \dots, \gamma_J$  takovým způsobem, abychom rozdělili pravděpodobnost mezi všechny možnosti volby. S využitím této intuice je zřejmé, že bude potřeba mít více restrikcí pro identifikaci parametrů. Předpokládejme případ, kdy  $J = 3$  a pravděpodobnost tak musí být rozdělena mezi tři alternativy. Představme si normální rozdělení, v rámci kterého si můžeme volně zvolit jeho střední hodnotu (tj.  $x_i'\beta$ ) a čtyři body v oblasti normálního rozdělení (tj.  $\gamma_0, \gamma_1, \gamma_2$  a  $\gamma_3$ ). Existuje celá řada různých způsobů, jak zvolit tyto parametry abychom dostali dané rozdělení pravděpodobnosti mezi jednotlivé alternativy. Předpokládejme, že  $x_i$  obsahuje pouze úroňovou konstantu a budeme chtít  $\Pr(y_i = 1|\beta, \gamma) = 0.025$ ,  $\Pr(y_i = 2|\beta, \gamma) = 0.95$  a  $\Pr(y_i = 3|\beta, \gamma) = 0.025$ . Toho můžeme dosáhnout tak, že  $\beta = 0$ ,  $\gamma_0 = -\infty$ ,  $\gamma_1 = -1.96$ ,  $\gamma_2 = 1.96$  a  $\gamma_3 = \infty$ . Ovšem totožné pravděpodobnosti volby získáme pro  $\beta = 1$ ,  $\gamma_0 = -\infty$ ,  $\gamma_1 = -1.96$ ,  $\gamma_2 = 1.96$  a  $\gamma_3 = \infty$  a řadu dalších voleb hodnot parametrů. A tak tu máme samozřejmě identifikační problém. Standardní způsob řešení tohoto problému je nastavení  $\gamma_0 = -\infty$ ,  $\gamma_1 = 0$  a  $\gamma_J = \infty$ .

Pochopit nutnost omezení parametrů z identifikačních důvodů je možno i alternativně tak, že se vrátíme k předchozímu jednoduchému probit modelu. Tento model je ekvivalentní uspořádanému probit modelu s  $J = 2$ . Lze ověřit, že se vztahy (9.8) a (9.9) zredukuje na své "probitovské" ekvivalenty ve vztazích (9.5) a (9.7) v případě, kdy  $\gamma_0 = -\infty$ ,  $\gamma_1 = 0$  a  $\gamma_2 = \infty$ , což jsou přesně ona identifikační omezení zavedená v předchozím odstavci.

Stejně jako v probit modelu si můžeme  $y^*$  představit jako úroveň užítku. Protože jsou jednotlivé alternativy jednoznačně uspořádané, je celkem rozumné modelovat pravděpodobnosti volby založené na svých latentních užitečích způsobem, jakoby se jednalo o integrály postupně na sebe navazujících oblastí normálního rozdělení. Představme si tak náš marketingový případ, kdy je spotřebitel dotazován na svůj užitek, který získává spotřebou nějakého zboží a musí volit z alternativ „velmi špatný“, „špatný“, „neutrální“, „dobrý“, „velmi dobrý“. Předpokládejme  $y_i^*$ , což je takový užitek  $i$ -tého spotřebitele, který ho vede k vyjádření alternativy, že produkt je „špatný“. Pokud se užitek spotřebitele mírně zvýší, potom je uspořádáním kategorií myšleno to, že spotřebitel řekne, že produkt je pro něj nyní „neutrální“ (nebo zůstane u stávající volby, a to té, že se stále jedná o produkt „špatný“). S uspořádaným probit modelem neexistuje možnost, že se mírná změna v užítku náhle promítne do změny názoru na produkt ve smyslu volby „velmi dobrý“. Je nutné skutečně zdůraznit, že omezení užítku tímto způsobem dává smyslu jen tehdy jestliže se bude jednat o uspořádané kategorie příslušných alternativ. Pokud spotřebitel volí mezi několika neuspořádanými alternativami, je adekvátní modelovou volbou multinomiální probit model.

Bayesovská analýza uspořádaného modelu využívá Gibbsův vzorkovač s obhacenyými daty, v rámci něhož jsou postupně generovány výběry z  $p(\beta|y^*, \gamma)$ ,  $p(\gamma|y^*, y, \beta)$  a  $p(y^*|y, \beta, \gamma)$ . Stejně jako u probit modelu odpovídá podmíněná hustota  $p(\beta|y_i^*, \gamma)$  normálnímu rozdělení, samozřejmě za podmínky, že jsme jako apriorní hustotu pro vektor parametrů  $\beta$  využili normální rozdělení popřípadě neinformativní apriorní hustotu (viz kapitola 3, kdy ovšem předpokládáme, že  $h = 1$ ). Podmíněná hustota latentních dat  $p(y_i^*|y_i, \beta, \gamma)$  odpovídá ohraničené

normální hustotě pravděpodobnosti, která je jednoduchým rozšířením vztahu (9.6):

$$y_i^* | y_i = j, \beta, \gamma \sim N(x_i' \beta, 1) 1(\gamma_{j-1} < y_i^* \leq \gamma_j) \quad (9.12)$$

To co je u uspořádaného probit modelu nové, je podmíněná hustota pro  $\gamma$ , tedy  $p(\gamma | y_i^*, y_i, \beta)$ . Pro každý z těchto parametrů využijeme nepravou apriorní hustotu mající charakter nekonečně širokého rovnoměrného rozdělení, tj.  $p(\gamma_j) \propto c$ . S mírnými modifikacemi je možné využít i jiné typy apriorních hustot. Tento druh prioru umožňuje snadnější současné výběry každého prvku vektoru  $\gamma$ . Připomeňme si, že identifikovatelnost parametrů vyžaduje volbu  $\gamma_0 = -\infty$ ,  $\gamma_1 = 0$  a  $\gamma_J = \infty$  a je tak obvyklejší generovat náhodné výběry z  $p(\gamma_j | y_i^*, y, \beta, \gamma_{(-j)})$  pro  $j = 2, \dots, J-1$ . Označení  $\gamma_{(-j)}$  vyjadřuje vektor  $\gamma$  bez prvku  $\gamma_j$ , tzn.  $\gamma_{(-j)} = (\gamma_0, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_J)'$ . Podmíněnou hustotu  $p(\gamma_j | y_i^*, y, \beta, \gamma_{(-j)})$  lze snadno odvodit, vezmeme-li v úvahu několik jednoduchých skutečností. Za prvé, tato hustota je podmíněna vektorem  $\gamma_{(-j)}$  a víme tedy, že  $\gamma_j$  musí ležet v intervalu  $[\gamma_{j-1}, \gamma_{j+1}]$ . Za druhé, hustota je podmíněna jak vektorem  $y$ , tak i  $y^*$  a můžeme si tak ujasnit, jaké hodnoty latentních dat odpovídají příslušným hodnotám skutečných dat. Za třetí, v argumentech podmíněné hustoty není přítomna žádná další informace o  $\gamma_j$ . Tyto skutečnosti nám impikují rovnoměrné rozdělení:

$$\gamma_j | y_i^*, y, \beta, \gamma_{(-j)} \sim U(\bar{\gamma}_{j-1}, \bar{\gamma}_{j+1}) \quad (9.13)$$

pro  $j = 2, \dots, J-1$ , přičemž

$$\bar{\gamma}_{j-1} = \max \{ \max \{ y_i^* : y_i = j \}, \gamma_{j-1} \}$$

a

$$\bar{\gamma}_{j+1} = \min \{ \min \{ y_i^* : y_i = j+1 \}, \gamma_{j+1} \}$$

Označení  $\max \{ y_i^* : y_i = j \}$  označuje maximální hodnotu latentních dat mezi všemi jednotlivci, kteří si zvolili alternativu  $j$ . Výraz  $\min \{ y_i^* : y_i = j+1 \}$  je definován analogickým způsobem.

Shrneme-li si dosavadní poznatky, posteriorní analýzu uspořádaného probit modelu lze provést pomocí Gibbsova vzorkovače s obohacenými daty, kterým provádíme sekvenční výběry z (9.12), (9.13) a  $p(\beta | y^*)$ . Poslední hustota odpovídá normálnímu rozdělení, pokud budeme předpokládat normální či neinformativní apriorní hustotu a přesnost chyby  $h = 1$ . Porovnání modelů a predikce může být implementována pomocí standardních, v předcházejících kapitolách rozebíraných nástrojů. V závislosti na porovnávaných modelech lze využít buď Savage-Dickey density ratio nebo Chibovu metodu. Predikční analýza je implementována podobně jako je tomu v kapitole 4. Konvergenčními diagnostikami je třeba ověřit konvergenci Gibbsova vzorkovače.

## 9.6 Multinomiální probit model

Existují případy, kdy jednotlivci mohou volit mezi několika alternativami, nicméně zde nebude existovat jejich logické uspořádání. V takovýchto případech

si s uspořádaným probit modelem nevystačíme. Zavedem si tedy multinomiální probit model, který je asi nejvíce využívaným modelem v případech, kdy uvažujeme několik neuspořádaných alternativ.

Na úvod si trochu upravíme značení předchozích částí kapitoly. Budeme předpokládat, že  $y_i$  může nabývat hodnot  $\{j = 0, \dots, J\}$ . Máme tedy celkem  $J + 1$  alternativ indexovaných jako  $\{j = 0, \dots, J\}$ , kdy  $J > 1$ . Pro motivaci zavádění multinomiálního probit modelu si můžeme rozšířit předchozí případ konceptu „náhodných užiteků“. Výběr jednotlivce nezáleží na absolutní hodnotě užitku spojené s danou alternativou, ale na relativním užitku vzhledem k ostatním alternativám. Nechť  $U_{ji}$  je užitek  $i$ -tého jednotlivce volícího alternativu  $j$  (pro  $i = 1, \dots, N$  a  $j = 0, \dots, J$ ). Veškerá informace o volbě, která byla ve skutečnosti učiněna, je obsažena v rozdílech užiteků vzhledem k nějaké základní alternativě. Zvolíme si alternativu 0 jako základní volbu a definujeme latentní proměnnou diferencí užiteků jako

$$y_{ji}^* = U_{ji} - U_{0i}$$

pro  $j = 1, \dots, J$ . Multinomiální probit model předpokládá, že tento rozdíl užiteků lze popsat normálním lineárním regresním modelem:

$$y_{ji}^* = x'_{ji}\beta_j + \epsilon_{ji} \quad (9.14)$$

kde  $x_{ji}$  je  $k_j$ -rozměrný vektor obsahující vysvětlující proměnné, které ovlivňují užitek spojený s volbou  $j$  (relativně vzhledem k volbě 0),  $\beta_j$  je odpovídající vektor regresních koeficientů a  $\epsilon_{ji}$  je chybový člen regrese.

Protože rovnice (9.14) zahrnuje  $J$  rovnic, bude posteriorní simulátor kombinovat výsledky pro model zdánlivě nesouvisejících regresí (SUR) s metodami poskytujícími výběry pro latentní rozdíly užiteků. Bude tedy užitečné přepsat vztah (9.14) do podoby SUR modelu. Všechny rovnice tak dáme do společných vektorů a matic, konkrétně  $y_i^* = (y_{1i}^*, \dots, y_{Ji}^*)'$ ,  $\epsilon_i = (\epsilon_{1i}, \dots, \epsilon_{Ji})'$ ,

$$\beta = \begin{pmatrix} \beta_1 \\ \cdot \\ \beta_J \end{pmatrix} \quad X_i = \begin{pmatrix} x'_{1i} & 0 & \cdot & \cdot & 0 \\ 0 & x'_{2i} & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & x'_{Ji} \end{pmatrix}$$

Definujme  $k = \sum_{j=1}^J k_j$  a

$$y_i^* = X_i\beta + \epsilon_i \quad (9.15)$$

Pokud dále budeme definovat

$$y^* = \begin{pmatrix} y_1^* \\ \cdot \\ y_N^* \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \cdot \\ \epsilon_N \end{pmatrix} \quad X = \begin{pmatrix} X_1 \\ \cdot \\ X_N \end{pmatrix}$$

můžeme zapsat multinomiální probit model (pomocí latentních rozdílů užiteků) jako

$$y^* = X\beta + \epsilon \quad (9.16)$$

## 168 Modely kvalitativních a omezených vysvětlovaných proměnných

Rovnice (9.16) je v podobě SUR modelu a můžeme tak využít standardní předpoklady kladené na náhodné složky SUR modelu. Předpokládáme tedy, že  $\epsilon_i$  jsou nezávislé a stejnoměrně rozdělena, přičemž jejich rozdělení odpovídá  $N(0, H^{-1})$  pro  $i = 1, \dots, N$ , kdy  $H$  je matice přesností chyb rozměrů  $J \times J$ . Alternativní možností vyjádření těchto předpokladů je tvrzení o tom, že rozdělení vektoru  $\epsilon$  odpovídá  $N(0, \Omega)$ , kde  $\Omega$  je blokově diagonální matice rozměru  $NJ \times NJ$ .

$$\Omega = \begin{pmatrix} H^{-1} & 0 & \cdot & \cdot & 0 \\ 0 & H^{-1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & H^{-1} \end{pmatrix} \quad (9.17)$$

Jakožto ekonometři nepozorujeme  $y_{ji}^*$  přímo, nicméně pozorujeme  $y_i$ , kdy platí

$$\begin{aligned} y_i &= 0 & \text{pokud} & \max(y_i^*) < 0 \\ y_i &= j & \text{pokud} & \max(y_i^*) = y_{ji}^* \geq 0 \end{aligned} \quad (9.18)$$

přičemž  $\max(y_i^*)$  je maximum  $J$ -rozměrného vektoru  $y_i^*$ . Jednotlivce tak volí tu či onu alternativu, která bude maximalizovat jeho užitek, ale my pozorujeme pouze jeho volbu.

Připomeňme si, že jsem v rámci jednorozměrného probit modelu kombinovali normální lineární regresní model se specifikací pro latentní data  $y^*$ . Pro bayesovskou analýzu jsme následně využili Gibbsův vzorkovač s obohacenými daty, v rámci něhož jsme generovali výběry z podmíněných hustot  $p(\beta|y^*)$  a  $p(y^*|y, \beta)$ . V kontextu multinomiálního probit modelu využijeme podobnou strategii, kdy budeme kombinovat výsledky identifikace SUR modelu spolu se specifikací latentních dat. Odvodíme si tak Gibbsův vzorkovač, který využívá výsledků z kapitoly 6, části 6.6, pro získání výběrů z  $p(\beta|y^*, H)$  a  $p(H|y^*, \beta)$ , a jistou podobu vícerozměrného ohraničeného normálního rozdělení pro podmíněnou hustotu  $p(y^*|y, \beta, H)$ .

Metody generování výběrů z  $p(y^*|y, \beta, H)$  lze odvodit v podobném duchu jako v případě tobit nebo probit modelu. Nezávislost chování mezi jednotlivci implikuje

$$p(y^*|y, \beta, H) = \prod_{i=1}^N p(y_i^*|y_i, \beta, H)$$

díky čemuž se můžeme zaměřit jen na  $p(y_i^*|y_i, \beta, H)$ . Rovnice (9.15) nám napovídá, že  $p(y_i^*|\beta, H)$  odpovídá normální hustotě pravděpodobnosti. Pokud toto tvrzení zkombinujeme s informací obsaženou v pozorovaných datech  $y_i$ , získáme ohraničenou normální hustotu pravděpodobnosti, konkrétně pak

$$\begin{aligned} lcy_i^*|y_i, \beta, H &\sim N(X_i'\beta, H^{-1})1(\max(y_i^*) < 0) & \text{pokud} & y_i = 0 \\ y_i^*|y_i, \beta, H &\sim N(X_i'\beta, H^{-1})1(\max(y_i^*) = y_{ji}^* \geq 0) & \text{pokud} & y_i = j \end{aligned} \quad (9.19)$$

Ekonometrická analýza multinomiálního probit modelu byla mnoho let mimo oblast hlavního zájmu (jak z hlediska bayesovského, tak i klasického přístupu),

a to z důvodu výpočetních obtíží vztahujících se k ohraničenému normálnímu rozdělení. Bayesovský přístup vyžaduje generování vzorků z ohraničené vícerozměrné normální hustoty pravděpodobnosti, klasický přístup by pak vyžadoval výpočet integrálů nad různými oblastmi parametrického prostoru ohraničené vícerozměrné normální hustoty pravděpodobnosti. Problém nastával zejména v případě velkého počtu alternativ. Současné pokroky ve výpočetní technice však výrazně zjednodušují bayesovskou (i klasickou) analýzu multinomiálního probit modelu. Existují volně dostupné programové kódy pro generování výběrů z ohraničeného vícerozměrného normálního rozdělení zahrnující omezení v podobě lineárních nerovností.

Jestliže využijeme pro vektory a matice parametrů,  $\beta$  a  $H$ , nezávislou normální-Wishartovu apriorní hustotu, můžeme provést bayesovskou analýzu pomocí posteriorní simulace zahrnující postupné výběry z podmíněných hustot  $p(y^*|y, \beta, H)$  (vztah (9.19)),  $p(\beta|y^*, H)$  (což je normální rozdělení odpovídající rovnicím (6.49)-(6.51)) a  $p(H|y^*, \beta)$  (což je Wishartovo rozdělení vycházející ze vztahů (6.52)-(6.54)). Vyvstává nám zde však problém týkající se identifikovatelnosti modelu. U jednorozměrného probit modelu byl tento problém řešen volbou  $h = 1$ . Tato podmínka je jednoduchá a významně zjednodušuje výpočet. V rámci multinomiálního probit modelu je však zahrnutí identifikačních omezení mnohem složitější a poněkud se tím komplikují i výpočetní postupy. Důvody, proč je multinomiální probit model neidentifikovatelný, jsou podobné jako u probit modelu.

Pokud budeme definovat kovarianční matici chyb jako  $\Sigma = H^{-1}$  a označíme-li  $\sigma_{ij}$  jako  $ij$ -tý prvek matice  $\Sigma$ , potom je standardním způsobem řešení identifikovatelnosti volba  $\sigma_{ij} = 1$ . Za těchto podmínek však podmíněná hustota  $p(H|y^*, \beta)$  nebude odpovídat Wishartovu rozdělení a nelze tak využít výsledky analýzy SUR modelu. Řadu způsobů řešení toto problému lze nalézt v literatuře. Například McCulloch a Rossi [25] jednoduše ignorují problém identifikovatelnosti a nepracují tak s předpokladem  $\sigma_{11} = 1$ . V tomto případě odpovídá podmíněná hustota  $p(H|y^*, \beta)$  Wishartovu rozdělení a následné výpočty jsou tak jasné. Místo prezentace empirických výsledků pro parametry  $\beta$  prezentují autoři své výsledky pro identifikovanou kombinaci parametrů  $\frac{\beta}{\sigma_{11}}$ . Někteří autoři však práci s neidentifikovanými modely nepreferují. Dostí nebezpečná je práce s neinformativními apriorními hustotami, kde nastávají výpočetní problémy. Velmi obvyklá je však bayesovská analýza multinomiálního probit modelu s využitím informativní apriorní hustoty ovšem s ignorováním identifikačních omezení.

Pro ty, kdo chtějí pracovat pouze s identifikovanými modely nabízejí užitečný přístup McCulloch, Polson a Rossi [24]. Jejich regresní model předpokládá v (9.15), že  $\epsilon_i$  odpovídá  $N(0, \Sigma)$ . Připomeňme si, že jakoukoli sdruženou hustotu pravděpodobnosti lze zapsat pomocí marginálních a podmíněných rozdělení. Na tomto základě je tedy možné rozdělit vektor  $\epsilon_i$  do podoby

$$\epsilon_i = \begin{bmatrix} \epsilon_{1i} \\ v_i \end{bmatrix}$$

## 170 Modely kvalitativních a omezených vysvětlovaných proměnných

kde  $v_i = (\epsilon_{2i}, \dots, \epsilon_{Ji})'$ . Matice  $\Sigma$  je rozdělena podobným způsobem, konkrétně

$$\Sigma = \begin{bmatrix} \sigma_{11} & \delta' \\ \delta & \Sigma_v \end{bmatrix} \quad (9.20)$$

Zákony pravděpodobnosti nám říkají, že  $p(\epsilon_i) = p(\epsilon_{1i})p(v_i|\epsilon_{1i})$ . Z vlastností vícerozměrného normálního rozdělení lze vyvodit, že hustoty pravděpodobnosti  $p(\epsilon_{1i})$  a  $p(v_i|\epsilon_{1i})$  odpovídají normálním rozdělením. Přesněji,

$$\epsilon_{1i} \sim N(0, \sigma_{11}) \quad (9.21)$$

a

$$v_i|\epsilon_{1i} \sim N\left(\frac{\delta}{\sigma_{11}}\epsilon_{1i}, \Phi\right), \quad (9.22)$$

kde  $\Phi = \Sigma_v - \frac{\delta\delta'}{\sigma_{11}}$ . Místo přímé práce s kovarianční maticí chyb  $\Sigma$  rozměru  $J \times J$  můžeme pracovat s parametry  $\sigma_{11}$ ,  $\delta$  a  $\Phi$ . Jednoduše tak můžeme nastavit  $\sigma_{11} = 1$ , zvolit apriorní hustoty pro  $\delta$  a  $\Phi$  a odvodit Gibbsův vzorkovač.

Je obvyklé předpokládat normální apriorní hustotu pro  $\delta$  a Wishartovu apriorní hustotu pro  $\Phi^{-1}$ . Volíme tedy

$$p(\delta, \Phi^{-1}) = p(\delta)p(\Phi^{-1})$$

, kde

$$p(\delta) = f_N(\delta|\underline{\delta}, \underline{V}_\delta) \quad (9.23)$$

a

$$p(\Phi^{-1}) = f_W(\Phi^{-1}|\underline{\nu}_\Phi, \underline{\Phi}^{-1}) \quad (9.24)$$

S těmito apriorními hustotami McCulloch, Polson a Rossi [24] ukazují, že potřebné podmíněné hustoty pro Gibbsův vzorkovač jsou

$$p(\delta|y^*, \Phi, \beta) = f_N(\delta|\bar{\delta}, \bar{V}_\delta) \quad (9.25)$$

a

$$p(\Phi^{-1}|y^*, \delta, \beta) = f_W(\Phi^{-1}|\bar{\nu}_\Phi, \bar{\Phi}^{-1}) \quad (9.26)$$

Jednotlivé parametry těchto podmíněných hustot jsou dány následovně:

$$\begin{aligned} \bar{V}_\delta &= \left( \underline{V}_\delta^{-1} + \Phi^{-1} \sum_{i=1}^N \epsilon_{1i}^2 \right)^{-1} \\ \bar{\delta} &= \bar{V}_\delta \left( \underline{V}_\delta^{-1} \underline{\delta} + \Phi^{-1} \sum_{i=1}^N v_i \epsilon_{1i} \right) \\ \bar{\Phi}^{-1} &= \left[ \underline{\Phi} + \sum_{i=1}^N (v_i - \epsilon_{1i} \delta)(v_i - \epsilon_{1i} \delta)' \right]^{-1} \\ \bar{\nu}_\Phi &= \underline{\nu}_\Phi + N \end{aligned}$$



Je nutné si uvědomit, že se jedná o podmíněné hustoty, kdy podmíněnost je dána vektory  $y^*$  a  $\beta$ , tudíž lze využít vztah (9.15) a brát vektor  $\epsilon_i = (\epsilon_{1i}, v_i')$  jako známý vektor.

Shrneme-li si naše dosavadní poznatky, můžeme říct, že bayesiánskou analýzu multinomiálního probit modelu lze snadno provést s využitím Gibbsova vzorkovače s obohacenými daty. Pokud ignorujeme identifikační omezení, zahrnuje Gibbsovův vzorkovač postupné výběry z podmíněných hustot (9.16), (6.49) a (6.52). Budeme-li uvažovat identifikační omezení  $\sigma_{11} = 1$ , potom Gibbsovův vzorkovač obnáší výběry z (9.16), (9.25), (9.26) a (6.49). Ve všech případech lze pro porovnání modelů a predikci využít nástroje z předcházejících kapitol.

Je třeba upozornit, že v některých aplikacích byla zjištěna pomalá konvergence Gibbsova vzorkovače. V literatuře tak lze nalést práce zaměřené na efektivnější výpočetní algoritmy.

V této podkapitole byly pro kovarianční matici chyb specifikovány dvoje rozdílné apriorní hustoty. Existuje však i řada jiných specifikací apriorních hustot, které lze s úspěchem využít. Multinomiální probit model je někdy kritizován pro svou přeparametrizovanost. To znamená, že pokud je počet alternativ velký, obsahuje kovarianční matice  $\Sigma$  příliš hodně parametrů, což může vést k nepřesným odhadům. Mnozí výzkumníci tak byli vedeni snahou o specifikaci omezených variant multinomiálního probit modelu nebo o rozvoj informativních priorů zahrnujících dodatečnou strukturalizaci modelu. Možností omezení je například omezení matice  $\Sigma$  na diagonální matici. Pokud je nějaké takové omezení rozumné, může významně snížit riziko přeparametrizování a zjednodušit praktické výpočty.

### 9.6.1 Empirická ilustrace: Multinomiální probit model

BUDE ČASEM DOPLNĚNO!

## 9.7 Rozšíření probit modelů

Asi není překvapivé, že existuje řada rozšíření modelů typu probit, uspořádaný probit a multinomiální probit. Jedním z nejvýznamnějších rozšíření zahrnuje využití panelových dat. Tento typ dat je obvykle dostupný v marketingu. Díky čtečkám čárových kódů v supermarketech je možné sledovat chování a výběry mnoha jednotlivců během jejich různých návštěv. Je navíc zřejmé, že se jednotlivci mohou lišit ve svých užítkových funkcích. To nám naznačuje možné využití modelu náhodných koeficientů z kapitoly 7 (část 7.4) pro latentní rozdíly v užítcích, a to v kontextu probit modelu. Například v případě, kdy existují dvě alternativy (tj. nějaký produkt je buď koupen nebo nekoupen) a k dispozici máme panelová data, lze rovnici pro normální lineární regresní model s latentní závisle proměnnou (viz (9.1)) zapsat v podobě

$$y_{it}^* = x_{it}'\beta_i + \epsilon_{it}. \quad (9.27)$$

K odvození podmíněných hustot  $p(\beta_i|y^*)$  pro  $i = 1, \dots, N$  lze využít metody z kapitoly 7 zahrnující hierarchickou apriorní hustotu pro vektory parametrů  $\beta_i$ , kterými je vyjádřena heterogenita spotřebitelů. Výraz pro  $p(y_{it}^*|y_{it}, \beta)$  je jednoduchým rozšířením vztahu (9.6). Snadno tak lze opět využít Gibbsův vzorkovač s obohacenými daty. Bayesovská analýza *panelového multinomiálního probitu náhodných koeficientů* kombinuje odvození v rámci multinomiálního probit modelu, modelu náhodných koeficientů a SUR modelu. Další rozšíření může vést k ošetření problémů či předpokladů spojených s časovými řadami, kam můžeme zařadit problém autokorelace náhodných chyb. Příkladem je tzv. multinomiální vícečasový probit model (*multinomial multiperiod probit model*).

Důležitým poznatkem bayesiánství jako takového je skutečnost, že bayesovská analýza panelového probit modelu a jeho multinomiální varianty zahrnuje kombinaci bloků Gibbsova vzorkovače pocházejících z jednodušších modelů předchozích kapitol. Bayesovská analýza je obvykle implementována s využitím posteriorních simulátorů, které obvykle pracují s bloky parametrů podmíněné ostatními bloky. Takováto modulární povaha posteriorních simulátorů vede k snadné a intuitivní kombinaci částí (výsledků analýzy) různých modelů. Jakýkoliv probit (tobit) model je možno kombinovat s modelem či modely předchozích kapitol. Probit (tobit) model s nelineární regresní funkcí, heteroskedasticitou náhodných složek nebo jejich autokorelací lze vyvinout s využitím výsledků této kapitoly a příslušných kapitol předchozích. Všechna taková rozšíření potvrzují modulární charakter Gibbsova vzorkovače.

## 9.8 Další rozšíření

V rámci celé této kapitoly jsme se zaměřili na modely, ve kterých vhodně definovaná latentní data (a jejich chování) mohou být vysvětlena pomocí normálního lineárního regresního modelu. Existuje ovšem celá plejáda jiných populárních modelů, které nelze zakomponovat do tohoto schématu. Většinu z nich lze zařadit do skupiny modelů, které mají charakter lineárních regresních modelů, nicméně náhodné chyby mají ne-normální rozdělení. Pokud je například vysvětlována proměnná počet něčeho (počet patentů registrovaných nějakou firmou, počet zemřelých doprovázející nějaký druh lékařského zákroku), potom předpoklad normálních chyb je zcela nevhodný. V takových případech je obvykle využíváno Poissonovo rozdělení. V případech, kdy je vysvětlovanou proměnnou doba trvání nějaké události (např. počet týdnů trvání nezaměstnanosti), je obvyklé využít v rámci regresních modelů exponenciální nebo Weibullovo rozdělení.

S ohledem na kvalitativní podobu dat jsou populární konkurencí probit modelů třída tzv. logit modelů. Ve skutečnosti existuje ke každé podobě probit modelu jeho „logitovský“ protějšek (uspořádaný logit model, multinomiální logit model apod.). Tyto modely jsou založeny na podobné intuici ohledně existence náhodných užitek jako v případě probit modelů a jejich jediným rozdílem je předpoklad logistického rozdělení náhodných chyb. Klíčovou vlastností tohoto rozdělení je to, že existuje analytické vyjádření příslušné kumulativní distribuční

funkce. Tuto vlastnost nemá normální rozdělení. Až do doby rozvoje posteriorních simulátorů se odhad multinomiálního probit modelu s velkým počtem alternativ potýkal s výpočetními obtížemi. Multinomiální logit model s analytickým vyjádřením kumulativní distribuční funkce, která je klíčovou částí věrohodnostní funkce, tento problém samozřejmě nemá. Problematiku logit modelů ve svých různých variantách dobře popisuje odborná literatura.

Třída logit modelů nabízí atraktivní alternativu k probit modelům a je samozřejmě, že lze využít bayesovských nástrojů k rozhodování mezi oběma typy modelů (posteriorní podíl šancí, posteriorní predikční p-hodnota). Pokud však existují více než dvě alternativy, má multinomiální logit model vlastnost, která není v řadě aplikací příliš vítána. Pravděpodobnosti voby implikované multinomiálním logit modelem musí splňovat vlastnost tzv. nezávislosti irelevantních alternativ (*Independence of Irrelevant Alternatives - IIA*). Tím je myšleno, že poměr pravděpodobností jakýchkoli dvou voleb bude stejný bez ohledu na to, jaké budou ostatní alternativy. Předpokládejme, že si cestující do práce může vybrat mezi autobusem ( $y = 1$ ) a autem ( $y = 0$ ) a pravděpodobnosti těchto voleb jsou stejně pravděpodobné (tj.  $\frac{p(y=0)}{p(y=1)} = 1$ ). Dále předpokládejme, že ve městě postaví stezku pro cyklisty, čímž se rozšiřují možnosti cestování do práce na prostředek kolo. Vlastnost IIA říká, že dodání této alternativy neovlivní skutečnost, že  $\frac{p(y=0)}{p(y=1)} = 1$ . V tomto příkladu je možné (i když spíše nepravděpodobné), že vlastnost IIA může být rozumně splněna. Původně jsme předpokládali, že  $p(y = 0) = p(y = 1) = 0.5$ . Předpokládejme, že bude existovat 20% šance, že si náš pracující zvolí pro cestu do práce kolo (když už byla postavena ona nová stezka pro cyklisty). To je konzistentní s faktem, že  $p(y = 0) = p(y = 1) = 0.4$ , kdy je v důsledku stále zachován poměr  $\frac{p(y=0)}{p(y=1)} = 1$ . Pro ilustraci nepřijatelnosti podmínky IIA si představme obměnu výše nastoleného příkladu do podoby problému tzv. červeného a modrého autobusu (Red Bus-Blue Bus problem). V rámci něho předpokládáme možnost volby mezi červeným autobusem ( $y = 1$ ) a autem ( $y = 0$ ). Pokud začne v oblasti operovat nová společnost s modrými autobusy ( $y = 2$ ), pokrývající i linku do práce, potom není moc rozumný předpoklad podmínky IIA. Vyjděme z toho, že původně platilo  $p(y = 0) = p(y = 1) = 0.5$  (a tedy i  $\frac{p(y=0)}{p(y=1)} = 1$ ). Pokud je linka modrých autobusů identická s červenými, má zavedení této alternativy pramalý vliv na volbu cestování automobilem, a tedy platí  $p(y = 0) = 0.5$  a  $p(y = 1) = p(y = 2) = 0.25$ . Zavedení nové alternativy tak vede k poměru  $\frac{p(y=0)}{p(y=1)} = 2$ , což porušuje vlastnost IIA, a takováto změna není v rámci multinomiálního logit modelu povolena.

Pro překonání této omezující vlastnosti byly vyvinuty různé varianty logit modelů. Jednou z nich je tzv. vnořený logit model (*nested logit model*), který předpokládá vnořenou strukturu rozhodovacího procesu. V příkladu červeného a modrého autobusu by ekonometr využil logit model pro cestujícího volícího mezi autem a veřejnou dopravou. V případě volby veřejné dopravy by pak použil druhý logit model pro volbu mezi červeným a modrým autobusem. Jeden logit model je tak vnořen do druhého logit modelu. Analýza vnořených logit modelů je v literatuře opět dobře popsána.

## **9.9 Shrnutí**

BUDE ČASEM DOPLNĚNO!

## Kapitola 10

# Flexibilní modely

Modely z předchozích kapitol zahrnovaly nějaký předpoklad o funkčním tvaru závislosti mezi proměnnými, rozdělení náhodných složek apod. V normálním lineárním regresním modelu jsme předpokládali, že náhodná složka  $\epsilon$  pochází u normálního rozdělení a vztah mezi závisle proměnnou a vysvětlujícími proměnnými je lineární. Tyto předpoklady jsou nezbytné pro konstrukci věrohodnostní funkce, která tvoří základ bayesovské analýzy. Ekonomická teorie nám však málokdy dokáže přesně říct, jakou funkční podobu a typ rozdělení máme uvažovat. Např. v oblasti teorie produkce nám ekonomická teorie často říká, že výstup firmy je rostoucí ve vstupu a případně zde existují klesající výnosy z rozsahu. Ekonomická teorie nám neřekne, že máme použít „produkční funkci s konstantní elasticitou substituce“. V praxi je tak třeba dbát na pečlivé využívání porovnávání modelů a kvality souladu modelu s daty (např. pomocí posteriorní predikční  $p$ -hodnotě nebo posteriorního podílu šancí), abychom se ujistili, že naše předpoklady o věrohodnostní funkci jsou korektní. Ve světle skutečností, že předpoklady o věrohodnostní funkci nemusí být vhodné a silně ovlivňují empirické výsledky, dochází k rozvoji nebayesiánské literatury věnované *neparametrickým* a *semiparametrickým metodám*. Proč se tyto metody nazývají právě takto je velice intuitivní. Věrohodnostní funkce závisí na parametrech a specifikace konkrétních předpokladů o rozdělení a funkční podobě nám dává *parametrickou věrohodnostní funkci*. Myšlenka stojící v pozadí neparametrického přístupu je založena na snaze zbavit se takovýchto parametrických předpokladů buď zcela (v případě neparametrických metod) nebo z části (v případě semiparametrických metod).<sup>30</sup>

---

<sup>30</sup>Cíl „nechat zcela promluvit data“ je často v praxi těžko dosažitelný, protože je vždy nutné vytvořit pro daný problém určitou strukturu, abychom získali rozumné empirické výsledky. Neparametrické metody zahrnují rovněž tvorbu určitých předpokladů, takže není správné argumentovat, že přístup založený na věrohodnosti vychází z nějakých předpokladů, přičemž neparametrický přístup zcela „nechá promluvit data“. To co odlišuje neparametrické a na věrohodnostním přístupu založené metody je právě typ vytvářených předpokladů. Například nelineární regresní model vychází z předpokladů, že „vztah mezi  $y$  a  $x$  má specifickou funkční podobu“. Neparametrický regresní model vychází z předpokladů o určité hladkosti regresní křivky. Otázku rozumnosti toho či onoho druhu předpokladů lze zodpovědět pouze v kontextu

Bayesiánská analýza je vždy založena na parametrické věrohodnostní funkci a tedy bychom neměli hovořit o bayesovských „nearametrických“ nebo „semiparametrických“ metodách. To je důvod, proč je tato kapitola pojmenována „Flexibilní modely“. Nicméně, existuje řada bayesiánských modelů, které jsou myšlenkou velmi podobné nebayesiánským neparametrickým metodám, což dává příčinu pro vznik a neustálý růst literatury používající označení „Bayesiánská neparametrická“. Tato oblast je velmi široká, tudíž je tato kapitola zaměřena na dva druhy bayesovských neparametrických přístupů, které jsou vcelku jednoduché a vystačíme si v nich s metodami předchozích kapitol.

Pro motivaci nemusíme chodit daleko a představme si předpoklady kladené v rámci normálního lineárního regresního modelu. Mohli bychom chtít uvolnit předpoklad o lineární závislosti (tj. uvolnit předpoklad o funkční formě) nebo uvolnit předpoklad o normálně rozložených chybových členech (tj. uvolnit předpoklad o rozdělení). Těmito dvěma aspekty se budeme zabývat. Část *Bayesiánská neparametrická a parametrická regrese* uvolňuje předpoklad o funkční formě, a část věnovaná modelování s pomocí směšování či mixu (*mixture*) normálních rozdělení uvolňuje předpoklady o rozdělení. Jak uvidíme, bayesiánskou semiparametrickou regresi jsme schopni provést pomocí technik z kapitoly 3 věnované normálnímu lineárnímu regresnímu modelu s přirozeně konjugovanou apriorní hustotou. Modelování směšováním normálních hustot můžeme provést za pomoci Gibbsova vzorkovače, který je rozšířením toho, co bylo obsahem kapitoly 6 (část 6.4) pro model s náhodnou složkou mající Studentovo  $t$ -rozdělení.

## 10.1 Bayesovská neparametrická a semiparametrická regrese

### 10.1.1 Přehled

V kapitole 5 jsme diskutovali nelineární regresní model v podobě

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

kde  $X_i$  je  $i$ -tý řádek matice vysvětlujících proměnných  $X$ ,  $f(\cdot)$  je známá funkce závislejší na  $X_i$  a vektoru parametrů  $\gamma$ . V této části začneme podobně a zapíšeme so neparametrický regresní model jako

$$y_i = f(X_i) + \epsilon_i. \quad (10.1)$$

V tomto případě je  $f(\cdot)$  *neznámá funkce*. V rámci celé této části si zadefinujeme standardní předpoklady:

1.  $\epsilon$  pochází z  $N(0_N, h^{-1}I_N)$ -
2. Všechny prvky matice  $X$  jsou buď pevná čísla (nenáhodné vleičiny), nebo pokud jsou náhodné vleičiny, jsou nezávislé na všech prvcích vektoru náhodných složek  $\epsilon$  a jejich funkce hustoty pravděpodobnosti je  $p(X|\lambda)$ , kde  $\lambda$  je vektor parametrů, který nezahrnuje žádný z jiných parametrů modelu.

konkrétní empirické aplikace.

Než se dostaneme k diskuzi nad samotnou neparametrickou regresí, je dobré zdůraznit, že pomocí nelineárních regresních metod, které předpokládají extrémně flexibilní volbu pro  $f(X_i, \gamma)$ , dokážeme dosáhnout podobné výsledky jako v případě neparametrické ekonometrie, a to s použitím již nám dobře známých metod. Můžeme použít některou z obvyklých metod rozvoje funkce (Taylorův, Fourierův nebo Mutzův-Szatzův rozvoj) a dostat tak parametrickou podobu pro  $f(X_i, \gamma)$ , které je dostatečně flexibilní k aproximaci jakékoli neznámé funkce. Výběr bodu omezení rozvoje řady nám umožňuje řídit přesnost aproximace.

Neparametrické metody vycházejí z myšlenky, že  $f(\cdot)$  je hladká funkce. To znamená, že když  $X_i$  a  $X_j$  leží blízko sebe, potom by i  $f(X_i)$  a  $f(X_j)$  měly ležet ve své blízkosti. Neparametrické regresní metody tedy odhadují neparametrickou regresní křivku na základě lokálních průměrů sousedních pozorování. Řada neparametrických regresních estimátorů pro  $f(X_i)$  má podobu

$$\hat{f}(X_i) \sum_{j \in N_i} w_j y_j, \quad (10.2)$$

kde  $w_j$  je váha příslušná  $j$ -tému pozorování a  $N_i$  označuje okolí  $X_i$ . Přístupy se liší v tom, jak jsou definovány jednotlivé váhy a okolí. Bohužel, pokud je příliš mnoho vysvětlujících proměnných, trpí neparametrické metody tzv. *prokletím dimenzionality* (*curse of dimensionality*). Neparametrické metody tak v podstatě „průměrují“ okolní pozorování k aproximaci regresního vztahu. Pro pevně danou velikost vzorku (poorování) s růstem dimenze  $X_i$  jsou od sebe okolní pozorování čím dál vzdálenější a neparametrické metody se stávají méně a méně přijatelnými. Málodky tedy vidíme, že by se neparametrické regresní modely (10.1) přímo používaly v aplikacích zahrnujících hodně vysvětlujících proměnných. Místo toho jsou používány modely, které se tomuto problému dimenzionality vyhýbají. Ukážeme si dva takové modely, přičemž prvním z nich je *Parciální lineární model*.

### 10.1.2 Parciální lineární model

Parciální lineární model rozděluje vysvětlující proměnné do dvou skupin. V jedné skupině jsou proměnné chápány parametricky ( $z$ ), ve druhé jsou proměnné brány neparametricky ( $x$ ). Pokud má  $x$  nízkou dimenzi, problém dimenzionality je vyřešen. Výběr toho, které vysvětlující proměnné chápat neparametricky závisí na konkrétní aplikaci. Obvykle jsou v  $x$  obsaženy nejdůležitější proměnné (nebo proměnná), která je klíčová z hlediska korektního zjištění jejich (jejího) marginálního vlivu. Na tomto místě budeme předpokládat, že  $x$  je skalár a krátce pohovoříme o tom, jak lze řešit případ neskalárního  $x$ .

Formálně je parciální lineární model dán jako

$$y_i = z_i \beta + f(x_i) + \epsilon_i, \quad (10.3)$$

kde  $y_i$  je závisle proměnná,  $z_i$  je vektor  $k$  vysvětlujících proměnných,  $x_i$  je skalární vysvětlující proměnná a  $f(\cdot)$  je neznámá funkce. Poznamenejme, že  $z_i$

neobsahuje úrovnovou konstantu, protože tuto roli převítá  $f(x_i)$ . Funkci  $f()$  označujeme jako *neparametrickou regresní křivku*.

Základní myšlenka bayesovského odhadu tohoto modelu je založena na tom, že  $f(x_i)$  pro  $i = 1, \dots, N$  mohou být chápány jako neznámé parametry. Pokud tomu tak je, potom 10.3 je normální lineární regresní model (i když s více vysvětlujícími proměnnými než pozorováními). Bayesiánská analýza s přirozeně konjugovanou apriorní hustotou může být provedena stejně jako v kapitole 3. Bayesiánská analýza je tudíž jasná a jednoduchá v případě parciálního lineárního modelu.

Začneme tím, že si seřadíme pozorování tak, že  $x_1 \leq x_2 \leq \dots \leq x_N$ . Protože jsou jednotlivá data na sobě nezávislá, je jejich přesně řazení nepodstatné a volba seřazení vzestupně činí smysl „okolních“ či „blízkých“ pozorování jasnější. Uspořádáme si všechny proměnné do matic, a to obvyklým způsobem, tedy  $y = (y_1, \dots, y_N)'$ ,  $Z = (z_1', \dots, z_N')'$  a  $\epsilon = (\epsilon_1, \dots, \epsilon_N)'$ . Pokud definujeme  $\gamma = (f(x_1), \dots, f(x_N))'$ ,

$$W = [Z \quad I_N]$$

a  $\delta = (\beta', \gamma')'$ , potom můžeme (10.3) psát jako

$$y = W\delta + \epsilon. \quad (10.4)$$

Poznamenejme, že  $\gamma$  je  $N$ -rozměrný vektor obsahující každý bod neparametrické regresní křivky. V této fázi jsme nezavedli žádné restriktce na prvky vektoru  $\gamma$ . Aplikujeme tedy neparametrický přístup v tom smyslu, že  $f(x_i)$  může být cokoli a  $f()$  je zcela neomezená, neznámá funkce. Dále, (10.4) je zřejmě regresní model s vysvětlujícími proměnnými v matici  $W$  rozměru  $N \times (N+k)$ . Ovšem, (10.4) je neobvyklý regresní model, protože počet neznámých prvků ve vektoru  $\delta$  je větší než počet pozorování, tzn.  $N+k \geq N$ . TO nám implikuje, že můžeme provést perfektní proložení, kdy součet čtverců chyb bude nulový. Pokud budeme mít odhad vektoru  $\delta$  v podobě

$$\hat{\delta} = \begin{pmatrix} 0_k \\ y \end{pmatrix},$$

potom výsledné chyby budou nulové. Vektor  $\hat{\delta}$  nám implikuje, že body na neparametrické regresní křivce jsou odhadnuty jako  $\hat{f}(x_i) = y_i$ . Takovýto odhad nám tedy nenabízí žádné vyhlazení neparametrické regresní křivky. Ve smyslu výrazu (10.2), odpovídající váhy jsou  $w_i = 1$  a  $w_j = 0$  pro  $j \neq i$ . Takovýto estimátor je nedostačující. Tuto anomálii však lze překonat pomocí apriorní informace.

V literatuře věnované neparametrické regresi jsou estimátory založeny na myšlence, že  $f()$  je hladká funkce. To znamená, že pokud  $x_i$  a  $x_{i-1}$  leží velmi blízko sebe, měly by i jejich funkční hodnoty,  $f(x_i)$  a  $f(x_{i-1})$ , ležet blízko sebe. V bayesiánské analýze lze tuto informaci zahrnout do apriorní hustoty. Existuje celá řada způsobů, jak toho docílit. Zde si ukážeme jeden z nich. Předpokládejme přirozeně konjugovanou normální-gama apriorní hustotu pro  $\beta$ ,  $\gamma$  a  $h$ . Díky tomu získáme jednoduché analytické výsledky bez nutnosti použití posteriorních simulátorů. Abychom se zaměřili na neparametrickou část parciálního lineárního



modeli, budeme předpokládat standardní neinformativní apriorní hustotu pro  $h$  a  $\beta$ :

$$p(\beta, h) \propto h. \quad (10.5)$$

Pro koeficienty neparametrické části modelu využijeme částečně informativní apriorní hustotu prvních diferencí  $\gamma$ :

$$R\delta \sim N(0_{N-1}, h^{-1}V(\eta)), \quad (10.6)$$

kde  $V(\eta)$  je pozitivně definitní matice závisající na hyperparametru  $\eta$  (který bude vysvětlen později) a  $R = [0_{(N-1) \times k} \quad D]$ , kde  $D$  je matice prvních diferencí rozměru  $(N-1) \times N$ :

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & 0 & -1 & 1 \end{bmatrix} \quad (10.7)$$

Poznamenejme, že takto specifikovaná struktura nám implikuje, že máme apriorní informaci pouze o rozdílu  $f(x_i) - f(x_{i-1})$ . Skutečnost, že očekáváme že sousedící body neparametrické regrese si budou co do hodnoty podobné, je obsažena v (10.6) díky předpokladu, že  $E[f(x_i) - f(x_{i-1})] = 0$ . Matice  $V(\eta)$  nám umožňuje řídit očekávanou velikost odchylky  $f(x_i) - f(x_{i-1})$  a tedy stupeň hladkosti neparametrické regresní křivky.

Je třeba na tomto místě rovněž zdůraznit, že v rámci apriorní informace můžeme chtít pro neznámou funkci, popisující regresní křivku, zavést i různé druhy omezení ve tvaru nerovnosti. Příkladem může být požadavek, že  $f(\cdot)$  je monotónně rostoucí funkce. To lze snadno provést s využitím technik z kapitoly 4, část 4.2.

Než e pustíme do vyjádření posteriorní hustoty tohoto modelu, je dobré zdůraznit dvě věci. Za prvé, můžeme si všimnout, že struktura parciálního lineárního modelu je téměř identická s local level modelem z kapitoly 8. Ve skutečnosti, pokud zanedbáme parametrický člen ( $Z$ ) a změním index  $i$  na index  $t$ , potom je neparametrický regresní model identický se stavovým modelem. To není překvapivé, pokud si všimneme, že oba modely mají zeřazená data a struktura stavové rovnice (8.5) je identická s apriorní hustotou definovanou v rovnici (10.6). Skutečnost, že metody pro stavové modely lze využít k neparametrické regresi je zmiňována v literatuře. Vše v kapitole 8, část 8.2, lze tedy využít. Využity mohou být i empirické bayesovské metody z části 8.2.3, pokud nechceme rozhodovat o apriorním hyperparametru jakým je parametr  $\eta$ . Za druhé, můžeme se částečně podívat i nad tím, že jsme se ve vztahu (10.6) zmiňovali možnost řízení „stupně hladkosti“ neparametrické regresní křivky skrze apriorní informaci ohledně prvních diferencí. Obvykle je míra hladkosti funkce měřena pomocí druhých diferencí funkce, což by nám napovídalo, že bychom měli použít apriorní informaci ohledně druhých diferencí (tzn.  $[f(x_{i+1}) - f(x_i)] - [f(x_i) - f(x_{i-1})]$ ). Apriorní informaci o druhých derivacích můžeme zavést jednoduchým způsobem pomocí předdefinování matice  $D$  v (10.7) na matici druhých diferencí.

Není obtížné dokázat, že posteriorní hustota pro normální lineární regresní model s částečně neinformativní normální-gamma apriorní hustotou je

$$\delta, h|y \sim NG\left(\tilde{\delta}, \tilde{V}, \tilde{s}^{-2}, \tilde{\nu}\right), \quad (10.8)$$

kde

$$\tilde{V} = (R'V(\eta)^{-1}R + W'W)^{-1}, \quad (10.9)$$

$$\tilde{\delta} = \tilde{V}(W'y), \quad (10.10)$$

$$\tilde{\nu} = N, \quad (10.11)$$

$$\tilde{\nu}\tilde{s}^2 = (y - W\tilde{\delta})'(y - W\tilde{\delta}) + (R\tilde{\delta})'V(\eta)^{-1}(R\tilde{\delta}). \quad (10.12)$$

Posteriorní hustota je korektní a platná funkce hustoty pravděpodobnosti a to navzdory faktu, že počet vysvětlujících proměnných v regresním modelu je větší než počet pozorování. Intuitivně, apriorní informace o stupni hladkosti v neparametrické regresní funkci je dostačující ke korekci nepřijemné možnosti perfektního proložení dat.

V empirických studiích se zaměřujeme zejména na neparametrickou část modelu. S využitím (10.8) a z vlastností vícerozměrného rozdělení plyne

$$E(\gamma|y) = [M_z + D'V(\eta)^{-1}D]^{-1}M_z y, \quad (10.13)$$

kde  $M_z = I_N - Z(Z'Z)^{-1}Z'$ . Rovnici (10.13) lze využít k odhadu  $f()$  a nazveme ji jako *vyrovnanou neparametrickou regresní přímku*. Matice  $M_z$  je zajímavá matice, která se často vyskytuje v klasické ekonometrii. Součin  $M_z y$  odpovídá reziduíům získaným metodou nejmenších čtverců v rámci regrese  $y$  na  $Z$ . Rovnici (10.13) můžeme interpretovat tak, že nejdříve odstraníme vliv  $Z$  na  $y$  (právě proto, že  $M_z y$  jsou rezidua) a výsledek pak vyhladíme použitím matice  $[M_z + D'V(\eta)^{-1}D]^{-1}$ . Poznamenejme, že v čistě neparametrickém případě (kdy tedy  $Z$  nevstupuje do modelu) se v případě, kdy se apriorní hustota v (10.6) stane neinformativní (tzn.  $V(\eta)^{-1} \rightarrow 0_{N-1, N-1}$ ), je  $E(\gamma|y) = y$  a neparametrická část modelu zcela vyrovná pozorovaná data (není zde tedy žádné vyhlazení).

Doposud jsme si neřekli nic o  $V(\eta)$ , přičemž zde existuje řada možností volby. Jednoduchá volba, které odpovídá pouze předpokladu o vyhlazení (tzn.  $f(x_i) - f(x_{i-1})$  je malé), by odpovídala volbě  $V(\eta) = \eta I_{N-1}$ .<sup>31</sup> Tato apriorní hustota závisí pouze na skalárním hyperparametru  $\eta$ , který si můžeme zvolit sami podle toho, jak chceme řídit míru vyhlazení. Abychom si přiblížili způsob, jak bayesovská posteriorní hustota řeší průměrování sousedních pozorování. Napomůže nám pohled na podmíněnou střední hodnotu  $E(\gamma_i|y, \gamma^{(i)})$ , kde  $\gamma^{(i)} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_N)$ . Pro čistě neparametrickou regresi (kdy do regrese nevstupuje  $Z$ ) lze uázat, že

$$E(\gamma_i|y, \gamma^{(i)}) = \frac{1}{2 + \eta}(\gamma_{i-1} + \gamma_{i+1}) + \frac{\eta}{2 + \eta}y_i,$$

<sup>31</sup>V malých datových vzorcích může být rozdíl mezi  $x_i$  a  $x_{i-1}$  velký a je tak žádoucí tuto skutečnost zakomponovat do apriorní hustoty. Jednoduchý způsob by bylo využití prioru zahrnujícího  $V(\eta)$  jakožto diagonální matici s prvky na diagonále rovnými  $\nu_i = \eta(x_i - x_{i-1})$ .

pro  $i = 2, \dots, N - 1$ .  $E(\gamma_i|y, \gamma^{(i)})$  je vážený průměr  $y_i$  a nejbližších bodů neparametrické regresní křivky pod a nad  $i$  (tzn.  $\gamma_{i-1}$  a  $\gamma_{i+1}$ ). Protože parametr  $\eta$  kontroluje stupeň hladkosti, který chceme přiřadit funkci  $f(\cdot)$ , je smysluplné, že pro  $\eta \rightarrow \infty$  dostaneme  $E(\gamma_i|y, \gamma^{(i)}) = \frac{1}{2}(\gamma_{i-1} + \gamma_{i+1})$ . Dále lze ukázat, že  $\text{var}(\gamma_i|y, \gamma^{(i)}) = \frac{\sigma^2 \eta}{2 + \eta}$ , která se blíží nule pro  $\eta \rightarrow 0$ . V limitním případě  $\eta \rightarrow 0$  získáme  $\gamma_i = \frac{1}{2}(\gamma_{i-1} + \gamma_{i+1})$  a neparametrická část regrese odpovídá přímce.

Shrnuto, bayesiánská analýza parciálního lineárního regresního modelu může být provedena s využitím normálního lineárního regresního modelu s přirozeně konjugovanou apriorní hustotou, pokud budeme chápat neznámé body neparametrické regresní přímky jako parametry. Přes skutečnost, že počet vysvětlujících proměnných je větší než počet pozorování, získáváme řádnou posteriorní hustotu. Porovnání modelů lze provést stejně jako v kapitole 3.

V řadě případů chceme přiřadit parametru  $\eta$  konkrétní hodnotu. Stejně tak lze ale pro volbu hodnoty  $\eta$  využít i empirické bayesovské metody (viz kapitola 8, část 8.2.3). Vyplatí se ale zmínit i další metodu výběru tohoto parametru, s využitím dat. Tato nová metoda, která je hojně využívaná neparametrickými statistiky je označována jako *krosvalidace* (*cross-validation*). Základní myšlenka je ta, že některá data jsou „zatajena“. Model je tak odhadnut s využitím zbylých dat a využít k predikci těchto „zatajených“ dat. Modely jsou porovnávány na základě toho, jak dobře jsou schopni tato data predikovat.<sup>32</sup> V tomto kontextu můžeme definovat krosvalidační funkci jako

$$CV(\eta) = \frac{1}{N} \sum_{i=1}^N \left( y_i - E(\gamma_i|y^{(i)}) \right)^2,$$

kde  $y^{(i)} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_N)'$ . To znamená, že vždy vymažeme jedno pozorování v čase a spočítáme neparametrickou regresní přímku s využitím zbylých dat. Následně použijeme  $(y_i - E(\gamma_i|y^{(i)}, \eta))^2$  jako metriku toho, jak dobře nám výsledná neparametrická regresní přímka vysvětluje zbylá data. Parametr  $\eta$  volíme tak, aby byla minimalizována krosvalidační funkce.

<sup>32</sup>Postup krosvalidace lze použít pro porovnání či hodnocení modelů v případě jakéhokoli modelu, nejen neparametrického.

Empirická ilustrace: Parciální lineární model

Rozšíření: Semiparametrický probit a tobit

### 10.1.3 Aditivní verze parciálního lineárního modelu

Ilustrace: Aditivní model

## 10.2 Kompozice (mixture) normálních modelů

### 10.2.1 Přehled

### 10.2.2 Věrohodnostní funkce

### 10.2.3 Apriorní hustota

### 10.2.4 Bayesovský výpočet

### 10.2.5 Porovnání modelů: Informační kritéria

### 10.2.6 Empirická ilustrace: Kompozice normálních modelů

## 10.3 Rozšíření a alternativní přístupy

## 10.4 Souhrn

## Kapitola 11

# Bayesiánské průměrování modelů



## Kapitola 12

# Další modely, metody a otázky bayesiánské ekonometrie

### 12.1 Úvod

Je logické, že jediná kniha nemůže pojmout všechny různé modely a metody využívané v rámci bayesovské ekonometrie. Logiku bayesiánské analýzy však lze využít v rámci jakéhokoliv modelu a nástroje prezentované v předchozích kapitolách (např. z oblasti posteriorní simulace) mají velmi širokou uplatnitelnost. V předchozích kapitolách tedy byly představeny nezbytné nástroje bayesovské ekonometrie pro výzkum v kontextu jakéhokoliv modelu, který by připadal v úvahu. Pokud se pouštíme do analýzy dat, měli bychom se vždy zamyslet nad tím, jaký vhodný model by připadal v úvahu, což v podstatě obnáší stanovení věrohodnostních funkcí a apriorních hustot. Vhodný model může být jak ten, který již dříve vyvinutý výzkumníky v dané oblasti problémů, a stal se tak jakýmsi standardem, tak i takový, který je zcela nový a spíše na doposad využívané dosavadní modely navazuje (pokud vůbec). Pokud si zvolíme množinu modelů, měli bychom využít pravidla pravděpodobnosti k nalezení posteriorních hustot, marginálních věrohodnostní a predikčních hustot. V poslední fázi je pak třeba vyvinout metody a techniky pro práci s těmito charakteristikami. Tyto metody budou obvykle zahrnovat posteriorní simulaci, kdy lze obvykle s úspěchem využít Gibbsův vzorkovač nebo Metropolis-Hastings algoritmy. V této kapitole tak nebude od věci, popsat si v krátkosti některé další obvyklé modely, metody a otázky bayesovské analýzy, pro něž existuje bohatá odborná literatura.

## 12.2 Další metody

Pro řadu modelů je využíván Gibbsův vzorkovač s obohacenými daty. Zvláště při práci s latentními daty však můžeme narazit na problém konvergence. To znamená, že nezbytný počet odstraněných replikací, jakož i replikací „přeživších“ může být natolik vysoký, že výpočetná náročnost bude příliš vysoká. Tento problém nastává zejména v případech, kdy je v modelu velký počet parametrů a korelace mezi nimi je vysoká. Protože i latentní data mohou být interpretována jako vektor parametrů, je počet parametrů v řadě modelů předchozích kapitol vcelku vysoký. Bayesiánská literatura se tak pokouší o vytvoření mnohem efektivnějších algoritmů posteriorní simulace (pro některé třídy modelů).

Algoritmy jako Metropolis-Hastings nebo Metropolis-within-Gibbs využívají náhodných výběrů z kandidátské hustoty a ty pak akceptují s určitou pravděpodobností. Pokud není kandidát přijat, zůstává řetězec na místě. Jestliže je kandidátské rozdělení nevhodně zvoleno, může být akceptační pravděpodobnost stále nízká a řetězec výběrů zůstává na jednom místě po dosti dlouhou dobu. V takových případech může řetězci trvat velmi dlouhou dobu, než začne konvergovat, čímž se zvyšují výpočetní nároky. Jednoduchou radou je v tomto případě volba lepší kandidátské hustoty. Ovšem toto není vždy možné, zejména v případě, dy je počet parametrů velký.

V řadě modelů nejsou problémy s konvergencí Gibbsova vzorkovače či s volbou kandidátské hustoty příliš patrné, zejména pokud jde o regresní modely. Pokud však nastanou, existuje několik přístupů, o kterých není na škodu zmínit pár slov.

Posteriorní simulátory z předchozích kapitol v sobě zahrnovaly výběry z obvyklých hustot pravděpodobnosti, pro které běžně existují počítačové algoritmy (např. pro normální, gama či Wishartovo rozdělení). Existuje však několik metod, které jsou mnohem obecnější a umožňují výběry z široké palety nestandardních hustot. Pokud některou z těchto metod je možno přímo využít (pro generování výběrů z posteriorní či podmíněné posteriorní hustoty), není potřeba zapojovat Metropolis-Hastings algoritmus nebo Metropolis-within-Gibbs, které mohou mít onen neduh nízkých akceptačních pravděpodobností.

Třemi populárními metodami jsou *acceptance sampling*, *Griddyho-Gibbsův vzorkovač* a *adaptivní rejection sampling*. „Acceptance sampling“ je založeno na následujícím teorému

**Teorém 12.1** (Acceptance Sampling). *Nechť  $p^*(\theta|y)$  je jádro posteriorní hustoty (tj.  $p(\theta|y) = cp^*(\theta|y)$ ) a  $q(\theta)$  je jádro hustoty pravděpodobnosti, ze které není problém generovat náhodné výběry. Hustota  $q(\theta)$  se nazývá zdrojovou hustotou (source density) a musí být definována nad definičním oborem hustoty  $p(\theta|y)$ . Předpokládejme, že jádra zdrojové a posteriorní hustoty splňují podmínku*

$$0 \leq \frac{p^*(\theta|y)}{q(\theta)} \leq a, \quad (12.1)$$

kde  $a$  je konečná konstanta. Potom algoritmus, který bude generovat vzorky z



$q(\theta)$  a akceptovat je s pravděpodobností

$$\frac{p^*(\theta|y)}{q(\theta)a}, \quad (12.2)$$

bude tímto generovat náhodné výběry z  $p(\theta|y)$ .

Úspěch tohoto simulátoru závisí na nalezení zdrojové hustoty, která bude splňovat omezení z (12.1). Ne vždy je možné takovou to hustotu nalézt. Zdrojová hustota musí být navíc taková, že akceptační pravděpodobnost v (12.2) nebude mít tendenci být příliš malá. Není obtížné nalézt příklady, kdy je akceptační pravděpodobnost tak nepatrná, že acceptance sampling není výpočetně zvládnutelné. V mnoha případech se však jedná o účinný a hojně využívaný nástroj. Například generátory náhodných čísel z normálního a gama rozdělení obsažené v řadě výpočetních balíčků typicky využívají acceptance sampling.

Griddy-Gibbsův vzorkovač je aproximativní metoda, jejíž hlavní ideu lze vcelku snadno popsat. Předpokládejme, že máme nestandardní posteriorní hustotu (nebo podmíněnou posteriorní hustotu v rámci Gibbsova vzorkovače), ze které bychom rádi generovaly náhodné výběry. Tato hustota může být aproximována tak, že vezmeme mřížku (grid - proto v názvu Griddy) bodů a vyhodnotíme posteriorní hustotu v každém bodě takovéto mřížky či možná snad sítě. Ty pak mohou být využity k multinomiální aproximaci posteriorní hustoty. Protože z multinomiálního rozdělení je snadné generovat náhodné výběry, lze aproximativní výběry z nestandardních rozdělení získat vcelku jasným způsobem. Jinými slovy, Griddy-Gibbsův vzorkovač v sobě zahrnuje diskrétní aproximaci nestandardní posteriorní hustoty, která nás zajímá.

Adaptivní rejection sampling je populární metoda pro výběry z log-konkávních hustot. Pro hlubší proniknutí do dalších metod generujících výběry z nestandardních posteriorních hustot lze nalézt v bohaté literatuře. Obecně tyto metody zvyšují výkonnost a efektivitu MCMC algoritmů.

Existuje i řada aproximací, které dovolují abstrahovat od samotné posteriorní simulace. V nejjednodušším případě totiž v řadě situací konverguje posteriorní rozdělení k normálnímu rozdělení pro  $N \rightarrow \infty$ . Typický teorem *Bayesiánské asymptotické* literatury má následující podobu.

**Teorem 12.2** (Bayesovský centální limitní teorem). *Za určitých podmínek regularity lze posteriorní rozdělení pro  $N \rightarrow \infty$  aproximovat jako*

$$\theta|y \sim N(\tilde{\theta}, [I(\tilde{\theta})]^{-1}), \quad (12.3)$$

kde  $\tilde{\theta}$  je posteriorní modus<sup>33</sup> a  $I(\tilde{\theta})$  je pozorovaná informační matice vyhodnocená v bodě  $\tilde{\theta}$ . Pozorovaná informační matice je definována jako záporná hodnota Hessiánu logaritmu posteriorní hustoty (tj. matice druhých parciálních derivací logaritmu posteriorní hustoty vzhledem k jednotlivým prvkům vektoru  $\theta$ ).

<sup>33</sup>Vektor  $\tilde{\theta}$  může být i odhad metodou maximální věrohodnosti a bude platit totožný teorem. Intuitivně to lze vyjádřit tak, že v některých případech, kdy se zvyšuje velikost vzorku, se datová informace stává čím dál více dominantnější nad apriorní informací. Asymptoticky je tak apriorní hustota irrelevantní.

Tento teorém lze dokázat při řadě různých podmínek regularity. Jinými slovy, tyto podmínky se využijí pro vyřazení potenciálně nevhodné apriorní hustoty (např. apriorní hustota nesmí přiřadit nulové váhy oblastem parametrického prostoru v blízkosti odhadů metodou maximální věrohodnosti) a pro vyloučení určitých případů, kdy se dimenzionalita parametrického prostoru zvyšuje s velikostí vzorku. Výše uvedený teorém tak lze využít pro aproximaci bayesiánské analýzy ve většině modelů. Z výpočetního hlediska je tím jediným, co potřebujeme, program k maximalizaci posteriorní hustoty (či věrohodnostní funkce) a vyhodnocení Hessiánu. Většina programů má funkce k maximalizaci či minimalizaci arbitrárně zvolené funkce a k výpočtu Hessiánu (ačkoliv je obvykle lepší využívat v programu analyticky získaný Hessián).

## **12.3 Další otázky**

### **12.3.1 Identifikace**

## **12.4 Další modely**

### **12.4.1 Modely časových řad**

### **12.4.2 Endogenita, výběr vzorku a další otázky**

### **12.4.3 Modely s nestandardními vysvětlovanými proměnnými**

### **12.4.4 Strukturální modely**

### **12.4.5 Bayesovské neparametrické metody**

# Příloha A

## Úvod do maticové algebry

Protože je v průběhu celého textu hojně využíván zápis pomocí matic, nebude od věci připomenout si základy maticové algebry.

**Definice A.1** (Matice a vektor). Matice  $A$  rozměru  $N \times K$  je uspořádání  $N \cdot K$  prvků (např. náhodných veličin) do  $N$  řádků a  $K$  sloupců:

$$X = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1K} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{N1} & a_{N2} & \cdot & \cdot & a_{NK} \end{bmatrix},$$

kde člen  $a_{nk}$  je prvek  $n$ -tého řádku a  $k$ -tého sloupce. Pokud je  $K = 1$ , potom  $A$  je sloupcový vektor a pokud  $N = 1$  pak  $A$  je vektor řádkový. Matice s  $N = 1$  a  $K = 1$  se nazývá skalár.

**Definice A.2** (Maticové sčítání a odčítání). Pokud  $A$  a  $B$  jsou matice rozměru  $N \times K$ , potom  $A + B$  je matice rozměru  $N \times K$  s prvky v  $n$ -tém řádku a  $k$ -tém sloupci danými jako  $a_{nk} + b_{nk}$ . Rozdíl matic  $A - B$  je matice rozměru  $N \times K$  s prvky v  $n$ -tém řádku a  $k$ -tém sloupci danými jako  $a_{nk} - b_{nk}$ . Sčítat a odčítat můžeme jen matice stejných rozměrů.

**Teorém A.1** (Vlastnosti maticového součtu). Pro maticový součet platí pravidla komutativní a asociativní (jako v základní aritmetice). Pokud tedy matice  $A$ ,  $B$  a  $C$  mají rozměr  $N \times K$ , platí:

$$A + B = B + A$$

$$A + (B + C) = (A + B) + C.$$

**Definice A.3** (Skalární součin). Pokud  $c$  je skalár a  $A$  je matice rozměru  $N \times K$ , pak  $cA$  je matice rozměru  $N \times K$ , kde prvky  $n$ -tého řádku a  $k$ -tého sloupce jsou dány jako  $ca_{nk}$ .

**Teorém A.2** (Vlastnosti skalárního součinu). Pokud  $c$  a  $d$  jsou skaláry a  $A$  a  $B$  jsou matice rozměru  $N \times K$ , potom platí:

$$(c + d)A = cA + dA$$

$$c(A + B) = cA + cB.$$

**Definice A.4** (Maticový součin). Nechť  $A$  je matice rozměru  $N \times K$  a  $B$  je matice rozměru  $K \times J$ . Potom  $C = AB$  je matice rozměru  $N \times J$  s prvky  $n$ -tého řádku a  $j$ -tého sloupce  $c_{nj}$ , přičemž

$$c_{nj} = \sum_{k=1}^K a_{nk}b_{kj}.$$

**Teorém A.3** (Vlastnosti maticového součinu). S ohledem na předchozí definici obecně pro maticový součin neplatí komutativní zákon, nicméně tento součin podléhá asociativnímu a distribučnímu zákonu. Obecně tedy platí  $AB \neq BA$  ( $BA$  je definován jen pokud  $N = J$ ). Opačně však platí  $A(BC) = (AB)C$  a  $A(B + C) = AB + AC$ , a to za podmínky, že matice  $A$ ,  $B$ ,  $C$  mají takové rozměry, že tyto operace jsou definovány.

**Definice A.5** (Transpozice matic). Nechť  $A$  je matice rozměru  $N \times K$  s prvky  $n$ -tého řádku a  $k$ -tého sloupce  $a_{nk}$ . Pak transponovaná matice  $A$ , označovaná jako  $A'$ , je matice rozměru  $K \times N$  s prvky  $k$ -tého řádku a  $n$ -tého sloupce  $a_{nk}$ .

*Poznámka:* transpozice zaměňuje řádky a sloupce, tedy  $n$ -tý řádek matice  $A$  odpovídá  $n$ -tému sloupci matice  $A'$ .

**Teorém A.4** (Vlastnosti transpozice). Nechť  $A$  je matice rozměru  $N \times K$  a  $B$  je matice rozměru  $K \times J$ . Potom  $(AB)' = B'A'$ .

**Definice A.6** (Speciální matice). *Čtvercová matice* je matice se stejným počtem řádků jako sloupců. *Diagonální matice* je čtvercová matice, kde všechny prvky mimo diagonálu jsou rovny nule (tj.  $a_{nk} = 0$  pro  $n \neq k$ ). *Horní trojúhelníková matice* má všechny prvky pod hlavní diagonálou rovny nule (tj.  $a_{nk} = 0$  pro  $n > k$ ). *Dolní trojúhelníková matice* má všechny prvky nad hlavní diagonálou rovny nule (tj.  $a_{nk} = 0$  pro  $n < k$ ). *Symetrická matice* je čtvercová matice, kde  $a_{nk} = a_{kn}$ .

**Definice A.7** (Užitečné matice). *Nulová matice* je matice rozměru  $N \times K$  se všemi prvky rovnými nule a je označována jako  $0_{N \times K}$ . Pokud  $K = 1$ , označuje se  $0_N$ . *Jedničková matice* je matice rozměru  $N \times K$  se všemi prvky rovnými jedné a je označována jako  $1_{N \times K}$ . Pokud je  $K = 1$ , označuje se  $1_N$ . *Jednotková matice* je diagonální matice rozměru  $N \times N$  s prvky na hlavní diagonále rovnými jedné (tj.  $a_{nk} = 1$  pro  $n = k$  a  $a_{nk} = 0$  pro  $n \neq k$ ). Je označována jako  $I_N$ . Je-li z kontextu znám rozměr těchto matic, označují se bez uvedení indexu.

**Teorém A.5** (Některé vlastnosti užitečných matic). Nechť  $A$  je matice rozměru  $N \times K$ , potom

$$A + 0_{N \times K} = 0_{N \times K} + A = A$$

$$A \cdot 0_{K \times J} = 0_{N \times J}, \quad 0_{J \times N} \cdot A = 0_{J \times K}$$

$$A \cdot I_N = I_K A = A.$$

**Definice A.8** (Lineární nezávislost). Nechť  $A_1, \dots, A_K$  označuje  $K$  sloupců matice  $A$  o rozměru  $N \times K$ . Tyto sloupce jsou *lineárně závislé*, pokud existují skaláry  $c_1, \dots, c_K$  (ne všechny nulové) takové, že

$$c_1 A_1 + c_2 A_2 + \dots + c_K A_K = 0_N.$$

Pokud takové skaláry neexistují, jsou sloupce matice  $A$  *lineárně nezávislé*.

**Definice A.9** (Hodnota matice). Hodnota matice  $A$  označována jako  $h(A)$  či  $rank(A)$ , je maximální počet lineárně nezávislých sloupců matice  $A$ .

**Definice A.10** (Determinant matice). Volně řečeno je determinant výhradně čtvercové matice  $A$  rozměru  $N \times N$  skalár a označuje se jako  $|A|$ . Lze jej intuitivně interpretovat jako číslo vyjadřující velikost matice.

**Definice A.11** (Stopa matice). Stopa matice  $A$  označována jako  $tr(A)$  je součet diagonálních prvků této matice.

**Teorém A.6** (Některé vlastnosti determinantu). Nechť matice  $A$  a  $B$  jsou čtvercové matice o rozměru  $N \times N$  a  $c$  je skalár. Potom platí:

$$|AB| = |A||B|,$$

$$|cA| = c^N |A|,$$

$$|A'| = |A|.$$

Obecně však  $|A + B| \neq |A| + |B|$ .

**Definice A.12** (Inverze matice). Inverze čtvercové matice  $A$  rozměru  $N \times N$  označována jako  $A^{-1}$  je matice rozměru  $N \times N$ , pro kterou platí  $AA^{-1} = I_N$ . Pokud  $A^{-1}$  existuje, pak se jedná o matici *regulární*, v opačném případě se jedná o matici *singulární*.

**Teorém A.7** (Singularita matice). Nechť  $A$  je matice rozměru  $N \times N$ , potom sloupce této matice jsou lineárně závislé tehdy a jen tehdy, pokud  $|A| = 0$ . Podobně, jestliže  $|A| = 0$ , potom  $h(A) < N$ . Matice  $A$  je singulární tehdy a jen tehdy, pokud  $|A| = 0$  (či ekvivalentně  $h(A) < N$ ).

**Teorém A.8** (Vlastnosti inverze). Nechť  $A$  a  $B$  jsou nesingulární matice rozměru  $N \times N$ , potom platí:

$$(A^{-1})^{-1} = A,$$

$$(AB)^{-1} = B^{-1}A^{-1},$$

$$(A')^{-1} = (A^{-1})',$$

$$|A^{-1}| = |A|^{-1}.$$

**Teorém A.9** (Inverze a determinant dělené matice). Nechť  $A$  je regulární matice rozměru  $N \times N$  rozdělená do následujících submatic:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

kde  $A_{11}$  a  $A_{22}$  jsou regulární matice rozměrů  $N_1 \times N_1$  a  $N_2 \times N_2$ , kdy  $N_1 + N_2 = N$ . Submatice  $A_{21}$  a  $A_{12}$  jsou matice rozměrů  $N_2 \times N_1$  resp.  $N_1 \times N_2$ . Inverzní matice  $A^{-1}$  je rozdělena obdobným způsobem:

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}.$$

Platí:

$$|A| = |A_{22}| |A_{11} - A_{12} A_{22}^{-1} A_{21}| = |A_{11}| |A_{22} - A_{21} A_{11}^{-1} A_{12}|.$$

Části inverzní matice  $A^{-1}$  lze spočítat následovně:

$$\begin{aligned} A^{11} &= (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}, \\ A^{22} &= (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}, \\ A^{12} &= -A_{11}^{-1} A_{12} A^{22}, \\ A^{21} &= -A_{22}^{-1} A_{21} A^{11}. \end{aligned}$$

**Definice A.13** (Kvadratická forma). Nechť  $x$  je vektor rozměru  $N \times 1$  a  $A$  je symetrická matice rozměru  $N \times N$ . Pak skalár  $x'Ax$  se nazývá *kvadratická forma*.

*Poznámka:* Nechť  $x_i$  označují prvky vektoru  $x$  a  $a_{ij}$  označují prvky matice  $A$ , která je symetrická, tedy  $a_{ij} = a_{ji}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, N$ . Potom  $x'Ax = \sum_{i=1}^N \sum_{j=1}^N a_{ij} x_i x_j$  je kvadratická funkce ve čtvercích a křížových součinech prvků  $x$ . Volně řečeno, kvadratická forma je maticové zobecnění součtů čtverců.

**Definice A.14** (Definitnost matice). Symetrická matice  $A$  rozměru  $N \times N$  je:

- pozitivně definitní tehdy a jen tehdy, jestliže  $x'Ax > 0$  pro všechny nenulové  $x$ ,
- negativně definitní tehdy a jen tehdy, jestliže  $-A$  je pozitivně definitní,
- pozitivně semidefinitní tehdy a jen tehdy, jestliže  $x'Ax \geq 0$  pro všechny  $x$  a  $x'Ax = 0$  pro některé nenulové  $x$ ,
- negativně semidefinitní tehdy a jen tehdy, jestliže  $-A$  je pozitivně semidefinitní.

*Poznámka:* Kovarianční matice náhodného vektoru je pozitivně definitní (nebo pozitivně semidefinitní). Podobnost se skalárním případem je v tom, že variance náhodné veličiny je kladná (nebo nezáporná). Užitečnou vlastností je i to, že pozitivně definitní matice je regulární.

**Teorém A.10** (Diagonalizace symetrické matice). Nechť  $A$  je čtvercová matice rozměru  $N \times N$ , potom existují matice  $X$  a  $D$  takové, že  $X'X = I_N$  a  $D$  je diagonální matice, přičemž platí  $X'AX = D$ . Jestliže je  $A$  pozitivně definitní, pak  $X$  a  $D$  jsou regulární.

*Poznámka:* Tento výsledek se používá obvykle pro transformaci modelu s obecnou kovarianční maticí do podoby modelu s kovarianční maticí rovnou  $cI$ , kde  $c$  je skalár.

**Teorém A.11** (Choleského dekompozice). Nechť  $A$  je čtvercová, pozitivně definitní matice rozměru  $N \times N$ . Pak existuje dolní trojúhelníková regulární matice  $X$  rozměru  $N \times N$ , pro kterou platí  $A = X'X$ .

*Poznámka:* Matice  $X$  není jednoznačně určena. Obvyklý způsob učinit ji jedinečnou je specifikovat požadavek, aby všechny její diagonální prvky byly kladné. Podobnost Choleského dekompozice ve skalárním případě je operátor druhé odmocniny. Pokud je  $A$  kovarianční matice, pak je analogií Choleského dekompozice pro skalární případ výpočet směrodatné odchylky. Pro Choleského dekompozici platí  $A^{-1} = (X')^{-1}X^{-1}$ . Tato vlastnost je často využívána v počítačových programech pro výpočet inverze pozitivně definitní matice. Choleského dekompozici je užitečné využít i v případě, kdy dokážeme vygenerovat pouze náhodný výběr ze standardizovaného normálního rozdělení. Konkrétně lze využít Choleského dekompozici kovarianční matice pro transformaci výběrů ze standardizovaného normálního rozdělení do podoby výběru z vícerozměrného normálního rozdělení, a to s arbitrárně určenou kovarianční maticí.





# Příloha B

## Úvod do pravděpodobnosti a matematické statistiky

V tomto dodatku jsou obsaženy potřebné základy pravděpodobnosti a matematické statistiky. V některých případech se jedná spíše o intuitivní či neformální definice.

### B.1 Základy pravděpodobnosti

**Definice B.1** (Experiment a náhodné jevy). Experiment je proces, jehož výsledek není předem znám. Možné výsledky experimentu se nazývají náhodné jevy. Množina všech možných výsledků je výběrový prostor.

**Definice B.2** (Diskrétní a spojitě proměnné). Proměnnou (či veličinu) nazýváme diskrétní, jestliže existuje konečný či spočítatelný počet hodnot, jichž může nabývat. Proměnná je spojitá, pokud může nabývat jakékoliv hodnoty na přímce reálných hodnot nebo na určitém intervalu reálných hodnot.

**Definice B.3** (Náhodné proměnné a pravděpodobnost). Obvykle bývají otázky vztahující se k pravděpodobnosti, experimentu a náhodným jevům reprezentovány proměnnou (veličinou), ať již diskrétní nebo spojitou. Jelikož není výsledek experimentu předem znám, je tato proměnná nazývána náhodnou veličinou. Pravděpodobnost lze intuitivně chápat jako reflexi věrohodnosti toho, že každý z náhodných jevů nastane. Pravděpodobnost realizace jevu  $A$  je označována  $\Pr(A)$ . Je důležité rozlišovat mezi náhodnou veličinou  $X$  označovanou velkými písmeny a její realizací  $x$  obvykle značenou písmenem malým. Jako příklad lze uvažovat experiment házení kostkou. Výběrový prostor je v tomto případě  $[1, 2, 3, 4, 5, 6]$  a diskrétní náhodná veličina  $X$  nabývá hodnot  $1, 2, 3, 4, 5, 6$  s pravděpodobnostmi danými jako  $\Pr(X = 1) = \Pr(X = 2) = \dots = \Pr(X = 6) = \frac{1}{6}$ . Alternativně je náhodná veličina  $X$  funkcí definovanou v bodech  $1, 2, 3, 4, 5, 6$ . Funkce je implicitně definována pravděpodobnostmi  $\Pr(X = 1) = \Pr(X = 2) = \dots = \Pr(X = 6) = \frac{1}{6}$ .

**Definice B.4** (Nezávislost). Dva jevy  $A$  a  $B$  jsou nezávislé, jestliže platí  $\Pr(A, B) = \Pr(A)\Pr(B)$ , kde  $\Pr(A, B)$  je pravděpodobnost současné realizace jevů  $A$  a  $B$ .

**Definice B.5** (Podmíněná pravděpodobnost). Podmíněná pravděpodobnost jevu  $A$  jevem  $B$ , označována  $\Pr(A|B)$ , je pravděpodobnost realizace jevu  $A$  za podmínky realizace  $B$ .

**Teorém B.1** (Pravidla podmíněné pravděpodobnosti). Nechť  $A$  a  $B$  označují dva jevy, potom platí:

$$\Pr(A|B) = \frac{\Pr(A, B)}{\Pr(B)},$$

$$\Pr(B|A) = \frac{\Pr(A, B)}{\Pr(A)},$$

a kombinací pak získáme Bayesův teorém

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$

**Definice B.6** (Pravděpodobnost a distribuční funkce). Nechť diskrétní náhodná veličina  $X$  je definovaná na výběrovém prostoru  $x_1, \dots, x_n$ . Pravděpodobnostní funkce této veličiny se označuje  $p(x)$ , přičemž platí:

$$p(x) = \begin{cases} \Pr(X = x_i) & \text{pro } x = x_i, \\ 0 & \text{jinak,} \end{cases}$$

pro  $i = 1, 2, \dots, N$ . Distribuční funkce (DF) označována  $P(x)$  je definována jako:

$$P(x) = \Pr(X \leq x) = \sum_{j \in J} \Pr(x_j),$$

kde  $J$  je množina indexů  $j$  pro které platí  $x_j \leq x$ .

Pravděpodobnostní a distribuční funkce mají následující vlastnosti:

$$p(x_i) > 0 \quad i = 1, 2, \dots, N$$

$$\sum_{i=1}^N p(x_i) = P(x_N) = 1$$

**Definice B.7** (Hustota pravděpodobnosti a DF). DF příslušná spojité náhodné veličině  $X$  je  $P(x) = \Pr(X \leq x) = \int_{-\infty}^x p(t)dt$ , kde  $p(\cdot)$  je hustota pravděpodobnosti (probability density function - p.d.f.). Pro tyto funkce platí:

$$p(x) \geq 0 \quad \forall x,$$

$$\int_{-\infty}^{\infty} p(t)dt = P(\infty) = 1,$$

$$p(x) = \frac{dP(x)}{dx},$$

$$\Pr(a \leq x \leq b) = P(b) - P(a) = \int_a^b p(x)dx.$$

**Definice B.8** (Očekávaná hodnota). Nechť  $g(\cdot)$  je funkce, pak očekávaná hodnota  $g(X)$ , označována jako  $E[g(X)]$  je definována jako:

$$E[g(X)] = \sum_{i=1}^N g(x_i)p(x_i),$$

pokud  $X$  je diskrétní náhodná veličina na výběrovém prostoru  $x_1, \dots, x_N$ , a

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)p(x)dx,$$

pokud  $X$  je spojitá náhodná veličina (platí-li  $E[g(X)] < \infty$ ). Speciálními případy této obecné definice zahrnují:

- střední hodnota  $\mu \equiv E(X)$ ,
- rozptyl  $\sigma^2 \equiv \text{var}(X) = E[X - \mu]^2 = E(X^2) - \mu^2$ ,
- $r$ -tý moment  $E(X^r)$ ,
- $r$ -tý centrovaný moment  $E(X - \mu)^r$ .

Třetí a čtvrtý centrovaný moment jsou obvyklé míry šikmosti a špičatosti náhodné veličiny, což reflektuje tloušťku krajů (chvostů) funkce hustoty pravděpodobnosti.

**Teorém B.2** (Vlastnosti očekávané hodnoty). Nechť jsou dány náhodné veličiny  $X$  a  $Y$ , funkce  $g(\cdot)$  a  $h(\cdot)$  a konstanty  $a$  a  $b$ , potom platí:

- $E[ag(X) + bh(Y)] = aE[g(X)] + bE[h(Y)]$ ,
- $\text{var}[ag(X) + bh(Y)] = a^2\text{var}[g(X)] + b^2\text{var}[h(Y)]$ , pokud  $X$  a  $Y$  jsou nezávislé.

**Definice B.9** (Modus, medián a mezikvartilové rozpětí). Střední hodnota je obvyklým měřítkem polohy funkce hustoty pravděpodobnosti nebo pravděpodobnostní funkce. Dalšími alternativami je modus a medián. Medián  $x_{med}$  je charakterizován jako  $P(x_{med}) = \frac{1}{2}$ . Modus  $x_{mod}$  je charakterizován jako  $x_{mod} = \arg \max[p(x)]$ .

**Definice B.10** (Sdružená pravděpodobnost a DF). Nechť je dán vektor  $N$  diskrétních náhodných veličin  $X = (X_1, \dots, X_N)'$  na výběrovém prostoru  $X_i$  daným  $x_{i1}, \dots, x_{iN}$ . Potom pravděpodobnostní funkce  $X$  označována jako  $p(x)$  je dána výrazem:

$$p(x) = \Pr(X = x_1, \dots, X_N = x_N),$$

kde  $x$  je  $N$ -rozměrný vektor  $x = (x_1, \dots, x_N)'$ . Pokud  $x$  nenáleží do výběrového prostoru  $X$ , potom  $p(x) = 0$ . Distribuční funkce  $P(X)$  je definována jako:

$$P(x) = \Pr(X_1 \leq x_1, \dots, X_N \leq x_N)$$

**Definice B.11** (Sdružená hustota a DF). Distribuční funkce definovaná pro náhodný vektor  $X = (X_1, \dots, X_N)'$  označována jako  $P(x)$ , kde  $x = (x_1, \dots, x_N)'$ , je

$$P(x) = \Pr(X_1 \leq x_1, \dots, X_N \leq x_N) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_N} p(t) dt_1 \dots dt_N,$$

přičemž  $p(x)$  je sdružená hustota pravděpodobnosti.

**Definice B.12** (Marginální pravděpodobnost a DF). Mějme vektor  $N$  náhodných veličin  $X = (X_1, \dots, X_N)'$  (diskrétních či spojitých) a  $X^* = (X_1, \dots, X_J)'$ , kde  $J < N$ . Potom sdružená marginální distribuční funkce  $X^*$  je svázána s distribuční funkcí  $X$  následovně:

$$P(x^*) = P(x_1, \dots, x_J) = \lim_{x_i \rightarrow \infty} P(x_1, \dots, x_J, \infty, \dots, \infty).$$

Pokud je  $X$  spojitý náhodný vektor, potom je sdružená funkce marginální hustoty pravděpodobnosti  $p(x)$  definována jako

$$p(x^*) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x) dx_{J+1} \dots dx_N.$$

Pokud  $J = 1$ , užívá se v terminologii výrazů funkce marginální hustoty pravděpodobnosti, marginální distribuční funkce a funkce marginální pravděpodobnosti.

**Definice B.13** (Podmíněná hustota a DF). Nechť  $X = (X_1, \dots, X_N)'$  je vektor  $N$  spojitých náhodných veličin. Definujme  $X^* = (X_1, \dots, X_J)'$ ,  $X^{**} = (X_{J+1}, \dots, X_N)'$  a nechť  $x$ ,  $x^*$  a  $x^{**}$  jsou příslušné realizace. Potom funkce podmíněná hustota pravděpodobnosti  $X^*$  za podmínky realizace  $X^{**}$  je definována jako

$$p(x_1, \dots, x_J | x_{J+1}, \dots, x_N) = p(x^* | x^{**}) = \frac{p(x)}{p(x^{**})} = \frac{p(x^*, x^{**})}{p(x^{**})}.$$

Podmíněná distribuční funkce  $P(x^* | x^{**})$  je definována jako

$$\begin{aligned} P(x^* | x^{**}) &= \Pr(X_1 \leq x_1, \dots, X_J \leq x_j | X_{J+1} = x_{J+1}, \dots, X_N = x_N) \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_J} p(x_1, \dots, x_J | x_{J+1}, \dots, x_N) dx_1 \dots dx_J \end{aligned}$$

Definice pro případ diskrétního náhodného vektoru je zobecněním výše uvedeného.

**Definice B.14** ( Vícerozměrná očekávaná hodnota). Nechť je dán vektor  $N$  spojitých náhodných veličin  $X = (X_1, \dots, X_N)'$  s hustotou pravděpodobnosti  $p(x)$ , kde  $x = (x_1, \dots, x_N)'$ . Potom očekávaná hodnota skalární funkce  $g(X)$  značená  $E[g(X)]$  je dána jako:

$$E[g(X)] \equiv \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x) p(x) dx_1 \dots dx_N.$$

Je to zobecnění definice střední hodnoty pro vícerozměrný případ. Rozšíření na podmíněnou očekávanou hodnotu  $E[g(X^*)|x^{**}]$  lze provést nahrazením  $p(x)$  funkcí  $p(x^*|x^{**})$ . Snadno lze tuto definici zobecnit pro diskretní náhodný vektor.

**Definice B.15** (Kovariance a korelace). Nechť  $X_1$  a  $X_2$  jsou náhodné veličiny, kdy  $E(X_1) = \mu_1$  a  $E(X_2) = \mu_2$ . Potom kovariance mezi  $X_1$  a  $X_2$  se označuje jako  $cov(X_1, X_2)$  a je definována jako:

$$\begin{aligned} cov(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E(X_1 X_2) - \mu_1 \mu_2. \end{aligned}$$

Korelace mezi  $X_1$  a  $X_2$  se označuje jako  $corr(X_1, X_2)$  a definuje se jako:

$$corr(X_1, X_2) = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)var(X_2)}}.$$

Pokud  $X_1 = X_2$ , je kovariance rovna rozptylu. Korelaci je možno interpretovat jako stupeň míry propojení mezi dvěma náhodnými veličinami.

**Definice B.16** (Kovarianční matice). Nechť  $X = (X_1, \dots, X_N)'$  je vektor  $N$  náhodných veličin a definujme  $N$ -rozměrný vektor  $\mu \equiv E(X) \equiv [E(X_1), \dots, E(X_N)]' \equiv [\mu_1, \dots, \mu_N]$ . Potom je definována kovarianční matice jako  $var(X)$ . Jedná se o matici rozměru  $N \times N$  obsahující variance a kovariance všech prvků  $X$ :

$$var(X) = E[(X - \mu)(X - \mu)']$$

$$= \begin{bmatrix} var(X_1) & cov(X_1, X_2) & \cdot & \cdot & cov(X_1, X_N) \\ cov(X_1, X_2) & var(X_2) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & cov(X_{N-1}, X_N) \\ cov(X_1, X_N) & \cdot & \cdot & cov(X_{N-1}, X_N) & var(X_N) \end{bmatrix}.$$

**Teorém B.3** (Vlastnosti očekávání a kovariance). Nechť je definována nestochastická matice  $A$  rozměru  $M \times N$  a  $Y = AX$ , přičemž výrazy jsou definovány stejně jako v předchozí definici. Potom platí:

$$E(Y) = AE(X) = A\mu,$$

$$var(Y) = Avar(X)A'.$$

Pokud  $a$  a  $b$  jsou skaláry a  $X_1$  a  $X_2$  náhodné vektory, pak platí:

$$var(aX_1 + bX_2) = a^2 var(X_1) + b^2 var(X_2) + 2abcov(X_1, X_2).$$

## B.2 Běžná rozdělení pravděpodobnosti

**Definice B.17** (Binomiální rozdělení).

**Definice B.18** (Poissonovo rozdělení).

**Definice B.19** (Uniformní rozdělení).

**Definice B.20** (Gama rozdělení). Spojitá náhodná veličina  $Y$  má *gama* rozdělení se střední hodnotou  $\mu > 0$  a stupni volnosti  $\nu > 0$ , označované jako  $Y \sim G(\mu, \nu)$ , pokud je její funkce hustoty pravděpodobnosti definována jako:

$$f_G(y|\mu, \nu) \equiv \begin{cases} c_G^{-1} y^{\nu-2} \exp\left(-\frac{y\mu}{2\nu}\right) & \text{pokud } 0 < y < \infty \\ 0 & \text{jinak} \end{cases}$$

kde integrační konstant je dána jako  $c_G = \left(\frac{2\mu}{\nu}\right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)$ , kde  $\Gamma(a)$  je gama funkce.

**Teorém B.4** (Střední hodnota a rozptyl gama rozdělení). Pokud  $Y \sim G(\mu, \nu)$ , potom  $E(Y) = \mu$  a  $\text{var}(Y) = \frac{2\mu^2}{\nu}$ .

**Poznámky:** gama rozdělení je v bayesiánské ekonometrii důležité rozdělení a obvykle je vztaheno k přesnosti chyby. Další rozdělení, která jsou úzce spojena s gama rozdělením, jsou *chí-kvadrát* rozdělení, což je gama rozdělení s  $\nu = \mu$ . Označuje se jako  $Y \sim \chi^2(\nu)$ . *Exponenciální* rozdělení je gama rozdělení s  $\nu = 2$ . *Invertované gama* rozdělení má tu vlastnost, že pokud je  $Y$  z gama rozdělení, potom  $\frac{1}{Y}$  pochází z gama rozdělení. V některých bayesiánských učebnicích a textech pracují autoři přímo s rozptylem náhodných složek (místo přesností chyb) a právě invertované gama rozdělení je v tomto případě s úspěchem využito.

**Definice B.21** (Multinomiální rozdělení).

**Definice B.22** (Vícerozměrné normální rozdělení). Spojitý  $k$ -rozměrný vektor  $Y = (Y_1, \dots, Y_k)'$  má *normální rozdělení* se střední hodnotou  $\mu$  ( $k$ -rozměrný vektor) a s kovarianční maticí  $\Sigma$  (pozitivně definitní matice rozměru  $k \times k$ ) a označuje se jako  $Y \sim N(\mu, \Sigma)$ , pokud je jeho funkce hustoty pravděpodobnosti dána jako

$$f_N(y|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right].$$

**Poznámka:** speciální případ, kdy  $k = 1$ ,  $\mu = 0$  a  $\Sigma = 1$ , je nazýván *standardizované normální rozdělení*. Většina statistických a ekonometrických učebnic obsahuje tabelované hodnoty percentilů standardizovaného normálního rozdělení.

**Definice B.23** (Vícerozměrné Studentovo  $t$ -rozdělení).

**Definice B.24** (Normální-gama rozdělení).

**Definice B.25** (Wishartovo rozdělení).

**Definice B.26** (Dirichletovo a beta rozdělení).

## B.3 Úvod do teorie výběru

**Definice B.27** (Náhodný výběr).

**Teorém B.5** (Vlastnosti náhodného výběru).

**Definice B.28** (Konvergence v pravděpodobnosti).

**Definice B.29** (Slabý zákon velkých čísel).

**Teorém B.6** (Slabý zákon velkých čísel pro náhodný výběr).

**Definice B.30** (Konvergence (limita) v distribuci). Necht  $\{Y_T\}$  je řada náhodných veličin,  $\{P_T(\cdot)\}$  je odpovídající řada distribučních funkcí, a necht  $Y$  je náhodná veličina s distribuční funkcí  $P(y)$ . Potom  $\{Y_T\}$  konverguje v distribuci k náhodné veličině  $Y$ , což je označováno jako  $Y_T \xrightarrow{d} Y$ , jestliže

$$\lim_{T \rightarrow \infty} P_T(y) = P(y).$$

$P(y)$  se nazývá limitní rozdělení. Konvergence v distribuci má tu vlastnost, že pokud  $Y_T \xrightarrow{d} Y$ , potom  $g(Y_T) \xrightarrow{d} g(Y)$ , kde  $g(\cdot)$  je spojitá funkce.

**Definice B.31** (Centrální limitní věta). Necht  $\{Y_T\}$  je řada náhodných veličin,  $Y$  je náhodná veličina a označme výběrový průměr závisující na délce řady  $T$  jako

$$\bar{Y}_T = \frac{\sum_{t=1}^T Y_t}{T}.$$

Potom  $\bar{Y}_T$  vyhovuje centrální limitní větě (CLV), jestliže  $\bar{Y}_T \xrightarrow{d} Y$ .

*Poznámka:* V ekonometrických situacích a problémech je limitní rozdělení vždy normální. V bayesovské analýze je CLV využívána k výpočtu numerických standardních chyb pro odhady získané z řady náhodných výběrů z posteriorní simulace. Stejně jako v případě slabého zákona velkých čísel zde existuje celá řada různých centrálních limitních vět, na které se lze odkazovat v závislosti na vlastnostech řady (např. jestli řada obsahuje nezávislé náhodné veličiny nebo zda-li jsou veličiny závislé, jestli je řada vybírána s téhož rozdělení nebo naopak z různých rozdělení apod.). Jako příklad je zde následující CLV, která je relevantní pro případ, kdy je řada náhodným výběrem. Ačkoliv mnohé posteriorní simulátory (např. Gibbsův vzorkovač) poskytují řadu, která není náhodným výběrem, platí, že v textu zmiňované simulátory splňují centrální limitní věty nutné pro zajištění konvergence v distribuci.

**Teorém B.7** (Centrální limitní věta pro náhodný výběr). Necht máme řadu  $\{Y_T\}$  nezávislých a stejně rozdělených náhodných veličin ze stejného rozdělení se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ . Definujme novou sekvenci nezávislých a stejně rozdělených náhodných veličin  $\{Z_T\}$ , kde

$$Z_T = \frac{\sqrt{T}(\bar{Y}_T - \mu)}{\sigma}.$$

Potom  $Z_T \xrightarrow{d} Z$ , kde  $Z \sim N(0, 1)$ .

*Poznámka:* Tento teorém je často označován jako Lindberg-Levyho centrální limitní věta. Existuje řada rozšíření zde uváděných základních centrálních limitních vět a slabých zákonů velkých čísel např. vícerozměrné rozšíření výše uváděných jednorozměrných teorémů.

## B.4 Další užitečné teorémy

**Teorém B.8** (Teorém změny proměnné - jednorozměrný).

**Teorém B.9** (Teorém změny proměnné - vícerozměrný).



## Příloha C

# Užitečné funkce v Matlabu

Na tomto místě nebude od věci prezentovat užitečné funkce a příkazy programového prostředí Matlab, které lze využít při bayesiánské analýze. Jedná se zejména o funkce, které jsou součástí Statistického toolboxu. Veškeré podrobnější informace o dané funkci lze získat příkazem `help <název funkce>` (napsaným do příkazového řádku Matlabu), případně v dokumentaci (nápovědě). Součástí této přílohy jsou i funkce LeSageho Ekonometrického toolboxu [22], který je volně dostupný na [www.spatial-econometrics.com](http://www.spatial-econometrics.com). Tyto funkce jsou rovněž dobře popsány v manuálu, který tento toolbox doprovází.

### C.1 Základní příkazy Matlabu

### C.2 Generátory náhodných čísel

Spíše než o generátory náhodných čísel se jedná o pseudo generátory. Základními funkcemi jsou `rand` (pro generování pseudo náhodných čísel z uniformního rozdělení  $U(0, 1)$ ) a `randn` (generátor ze standardizovaného normálního rozdělení). Obě funkce jsou využity pro generování náhodných čísel z dalších rozdělení, které jsou součástí Statistického toolboxu.

- `betarnd` – náhodná čísla z beta rozdělení
- `chi2rnd` – náhodná čísla z chí-kvadrát rozdělení
- `exprnd` – náhodná čísla z exponenciálního rozdělení
- `frnd` – náhodná čísla z  $F$ -rozdělení
- `gamrnd` – náhodná čísla z gama rozdělení
- `lognrnd` – náhodná čísla z log-normálního rozdělení
- `normrnd` – náhodná čísla z normálního rozdělení

- `trnd` – náhodná čísla z  $t$ -rozdělení
- `wishrnd` – náhodná čísla z Wishartova rozdělení
- `mvnrnd` – náhodná čísla z vícerozměrného normálního rozdělení
- `mvtrnd` – náhodná čísla z vícerozměrného  $t$ -rozdělení

### C.3 Ekonometrický toolbox

# Literatura

- [1] BAUWENS, L., LUBRANO, M., AND RICHARD, J. F. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, 1999.
- [2] CARLIN, B., AND LOUIS, T. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall, 2000.
- [3] CARLIN, B., POLSON, N., AND STOFFER, D. Monte carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association* 87 (1992), 493–500.
- [4] CARTER, C., AND KOHN, R. On gibbs sampling for state space models. *Biometrika* 81 (1994), 541–553.
- [5] CHIB, S. Marginal likelihood from the gibbs sampler. *Journal of the American Statistical Association* 90 (1995), 1313–1321.
- [6] CHIB, S., AND GREENBERG, E. Understanding the metropolis-hastings algorithm. *The American Statistician* 49 (1995), 327–335.
- [7] DEJONG, P., AND SHEPHARD, N. The simulation smoother for time series models. *Biometrika* 82, 2 (1995), 339–350.
- [8] DEVROYE, L. *Non-Uniform Random Number Generation*. Springer-Verlag, 1986.
- [9] DURBIN, J., AND KOOPMAN, S. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
- [10] ENDERS, W. *Applied Econometric Time Series*. John Wiley & Sons, 1995.
- [11] FRUHWIRTH-SCHNATTER, S. Model discrimination and bayes factors for linear gaussian state space models. *Journal of the Royal Statistical Society, Series B* 56 (1995), 237–246.
- [12] GELFAND, A., AND DEY, D. Bayesian model choice: Asymptotic and exact calculations. *Journal of the Royal Statistical Society Series B* 56 (1994), 501–514.

- 
- [13] GEWEKE, J. Bayesian inference in econometrics models using monte carlo integration. *Econometrica* 57 (1989), 1317–1340.
- [14] GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4* (1992), 641–649.
- [15] GEWEKE, J. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8 (1993), 19–40.
- [16] GEWEKE, J. Using simulation methods for bayesian econometric models: Inference, development, and communication. *Econometric Reviews* 18 (1999), 1–126.
- [17] JUILLARD, M. Dynare toolbox for matlab. ver. 4.
- [18] KIM, C., AND NELSON, C. *State Space Models with Regime Switching*. MIT Press, 1999.
- [19] KOOP, G. *Bayesian Econometrics*. Wiley, 2003.
- [20] KOOP, G., AND VAN DIJK, H. Testing for integration using evolving trend and seasonals models: A bayesian approach. *Journal of Econometrics* 97 (2000), 261–291.
- [21] LANCASTER, T. *An introduction to Modern Bayesian Econometrics*. Blackwell Publishing, 2004.
- [22] LESAGE, J. P. Econometrics toolbox. version 7.
- [23] LI, K. Exchange rate target zone models: A bayesian evaluation. *Journal of Applied Econometrics* 14 (1999), 461–490.
- [24] MCCULLOCH, R., POLSON, N., AND ROSSI, P. A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* 99 (2000), 173–193.
- [25] MCCULLOCH, R., AND ROSSI, P. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64 (1994), 207–240.
- [26] POIRIER, D. *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press, 1995.
- [27] SHIVELY, T., AND KOHN, R. A bayesian approach to model selection in stochastic coefficient regression models and structural time series models. *Journal of Econometrics* 76 (1997), 39–52.
- [28] VERDINELLI, I., AND WASSERMAN, L. Computing bayes factor using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association* 90 (1995), 614–618.
- [29] WEST, M., AND HARRISON, P. *Bayesian Forecasting and Dynamic Models*. Springer, 1997.